IntechOpen

Statistics

Growing Data Sets and Growing Demand for Statistics

Edited by Türkmen Göksel

STATISTICS - GROWING DATA SETS AND GROWING DEMAND FOR STATISTICS

Edited by **Türkmen Göksel**

Statistics - Growing Data Sets and Growing Demand for Statistics
http://dx.doi.org/10.5772/intechopen.72139
Edited by Türkmen Göksel

Contributors

Elena Zarova, Tatiana Burtseva, Wilmar Hernandez, Alfredo Mendez, Laura Ibarra Garcia, Gordana Djurovic, Milica Muhadinovic, Vasilije Djurovic, Martin Bojaj, Thomas Rose, Trevor Sweeting, Foued Aloulou, Türkmen Göksel

Notice

First published in London, United Kingdom, 2018 by IntechOpen
eBook (PDF) Published by IntechOpen, 2019
IntechOpen is the global imprint of INTECHOPEN LIMITED, registered in England and Wales, registration number: 11086078, The Shard, 25th floor, 32 London Bridge Street
London, SE19SG – United Kingdom
Printed in Croatia

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Additional hard and PDF copies can be obtained from orders@intechopen.com

Statistics - Growing Data Sets and Growing Demand for Statistics
Edited by Türkmen Göksel

p. cm.

Print ISBN 978-1-78984-395-8

Online ISBN 978-1-78984-396-5

eBook (PDF) ISBN 978-1-83881-354-3

Meet the editor

Türkmen Göksel holds a PhD degree in Economics (2009) from the University of Minnesota, USA. He received both his BSc and MSc degrees in Economics from Ankara University in 2002 and 2004, respectively. In 2013, he became an associate professor of economics. Currently, he is employed at the Department of Economics, Ankara University, Turkey. His main research interests are macroeconomics, international economics, mathematical modeling, programming languages, and electoral systems. He has published many papers in local and international peer-reviewed journals in these fields, and has also published books about electoral systems and applications of MATLAB in social sciences. In addition to Ankara University, he has taught courses at the University of Minnesota, USA, and Boğaziçi University, Turkey.

Contents

Preface

Since data grows faster than ever, the role of statistics becomes more and more crucial nowadays, and there is no doubt that statistics will be even more critical in the future.

As we all know, we are living in the "Information Age" in which information has become a commodity that is quickly and widely disseminated and easily available. However, data has no meaning by itself. It should be organized and interpreted by human beings before it becomes meaningful. Therefore, the "Information Age" requires us to become fluent in understanding data, including the ways it is collected, analyzed, and interpreted. In a nutshell, it is essential to understand the role and significance of statistics both theoretically and in our daily lives.

The application of statistics is extensive, and in our daily lives there is almost no human activity where the use of statistics is not needed. In this limited volume, we try to cover as many different and multidisciplinary fields in statistics as possible.

Motivated by the above facts, *Statistics: Growing Data Sets and Growing Demand for Statistics* aims to present recent developments and applications of statistical analysis. Hopefully, this book will serve as a supplementary resource for students, researchers, and practitioners in the area of statistics. The book consists of three sections and eight chapters, each focusing on a particular aspect of theory or application. Within every episode, the reader will be given an overview of background information and in many cases a description of the authors' novel approach.

In this book, we try to collect texts that emphasize the crucial role of statistics. Furthermore, we also want to highlight the applications of statistics in economics, development, and other various topics. Therefore, this book is organized into three sections.

The first section "The Role of Statistics on Quantification" consists of two chapters. The first (introductory) chapter in this section discusses the significant role of statistics both theoretically and in our daily lives. This chapter also highlights both the benefits and misuses of statistics.

The second chapter in this section looks at the relationship between the data-gathering, statistically oriented empirical research and theoretical formulation. In particular, this chapter presents the criteria that sociological explanations and theories must cover and their role in empirical study based on quantitative methods.

The next section "Applications of Statistics on Economics and Development" provides examples of statistical analysis in the field of economics and development. The first chapter of this section is devoted to the modeling of macroeconomic business cycles with statistical methods. This chapter proposes and tests a novel approach, and this novelty can serve as a basis for further research in both theoretical and applied aspects.

The second chapter of this section analyzes the investment appeal of Russian regions with various statistical methods. Proposed methodological developments in this chapter can be applied to a large variety of tasks related to the monitoring of development strategies of areas and other objects of strategic planning.

The third chapter of this section measures the progress in Montenegro's national strategy for sustainable development. This chapter compares Montenegro's 2030 and EU's Agenda and makes a number of policy recommendations.

The final chapter of this section highlights the application of discrete choice models on the transport economy by specifying their contribution to the estimation of transport demand. In particular, this chapter focuses on unordered multinomial logit modeling to analyze real-life problems in Tunisia.

The final section "Applications of Statistics on Various Topics" analyzes statistical data in various other fields. The first chapter covers the application of statistical methods to image compression. This chapter also offers a novel construction of principal components to reduce the computational cost of their calculation, although decreasing the accuracy.

The second chapter discusses nuclear accidents with core melting. This chapter analyzes the probability of severe nuclear accidents related to power generation. Moreover, the chapter investigates the learning effects of reactor operators using generalized linear models with a frequentist and Bayesian approach. This chapter also contains a discussion of core melt accidents and predictions for future events.

The editor and authors of the book would like to thank InTech publishing company for the opportunity to present selected views on modern issues of statistical analysis. Of course, this limited volume cannot cover all the topics of statistics. However, we believe and wish that this book will open new perspectives to interested readers.

Assoc. Prof. Türkmen Göksel
(PhD University of Minnesota)
Ankara University
Faculty of Political Science
Ankara, Turkey

Section 1

The Role of Statistics on Quantification

Introductory Chapter: Statistics

Türkmen Göksel

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.79773

1. Introduction

"Statistics is the grammar of Science," a famous quote by Karl Pearson who was the British statistician and leading founder of the modern field of statistics. Pearson highlights the importance of statistics and particularly emphasizes the significance of quantification for various fields of scientific study in his publication, The Grammar of Science [1].

Statistics is defined as the study of the collection, analysis, interpretation, presentation, and organization of data by the Oxford Dictionary of Statistical Terms [2]. Since the data grow faster than ever and information is increasing tremendously nowadays, the role of statistics becomes more crucial.

In general, we can group statistical analysis into two parts: (i) descriptive statistics and (ii) inferential (analytical) statistics. Descriptive statistics is used to summarize and/or describe a collection of data. Therefore, descriptive statistics provides a powerful way to summarize what already exists in data. However, inferential statistics focuses on the patterns in the data and then draws inferences from these patterns. In other words, by analyzing data gathered from samples (smaller subsets of the entire population), statistical methods infer about populations.

The field of statistics is the science of learning from data. In other words, statistics is the tool we use to convert data into information. Decisions based on data and information will provide better outcomes than those just based on intuition or gut feelings. In our daily life, there is almost no human activity where the application of statistics is not needed. Therefore, application of statistics plays a very significant role in almost every field such as Mathematics, Physics, Chemistry, Biology, Botany, Medicine, Economics, Education, Public Policy, Psychology, Astronomy, Zoology, Bio-Technology, Information Technology, Manufacturing, Service Industry, Business, and Commerce, among many other fields.

Since the application of statistics is very wide, different and multidisciplinary fields have evolved over time. These are some examples of application of statistics to other disciplines: Astrostatistics, Biostatistics, Econometrics, Business Analytics, Environmental Statistics, Statistical Mechanics, Statistical Physics, Actuarial Science, and so on. For example, Astrostatistics is the field which applies statistics to astronomical data which indicate that astrostatistics is a combination of astrophysics and statistical analysis. Biostatistics is the application of statistics to a wide range of topics in biology. Econometrics is the field where statistical tools are used to explain economic theories, and business analytics is the branch in which the statistical analysis applied to understanding of business performance and opportunities. Statistical physics is the branch which uses statistical methods to answer physical problems, and actuarial science is the field that uses statistical methods to analyze the risk insurance and some other financial issues.

To underline the importance of statistics in our daily life, we can look at the following examples: (i) *weather forecasting*: most of the people watch weather news and make decisions according to this news. Ineluctably, there are statistical models behind these forecasts that predict the weather conditions. (ii) *Insurance*: most of the people have some kind of insurance, such as medical, home, car, etc. Most of the insurance companies use statistical models to calculate the risk of giving insurance. (iii) *Medical field*: before any drugs prescribed, scientist should demonstrate a statistically valid rate of effectiveness. (iv) *Financial markets*: traders and businessmen use data and statistics to invest money. Of course one can easily extend the number of examples in which we use statistics in our daily life.

Among many others, some benefits of statistical analysis can be summarized as follows. First of all, it helps to present and compare the facts from data in a definite form. In other words, expressing results and/or conclusions in numbers develops a necessary and common form of communication for scientists, policy makers, and many others. Secondly, it helps us to formalize our thinking. In particular, statistical methods are used in formulating/testing the hypotheses or a new theory. By using these methods, we can determine the likelihood that a hypothesis should be either rejected or not. Thirdly, statistical methods help us to draw conclusions about populations based only on sample results. Last but not least, statistics is very important especially when it comes to the conclusion of the research, and in this sense statistical methods allow us for forecasting and policy making.

Among many benefits, of course, there are also some misuses of statistics. Main examples of misuses among others are overgeneralization, biased samples, insufficient sample size, and spurious correlations. All these mistakes may give us misleading conclusions. Let me just give some intuition about these misuses. (i) *Overgeneralization*: the results from one population may not be valid to another population. For example, a result/conclusion which is valid for one age group may not be true for other age groups. Hence, one should be careful about his conclusions when it comes to generalization. (ii) *Biased samples*: a random sample should be used, since a non-random sample can bias the results from the beginning. Making an argument or claim about an entire population based on a sample that is not representative of the whole is an important example of misuse of statistics. If one is analyzing whether or not the school lunch program provided enough food for students and conducts a survey only with the basketball team then this would be a very good example of biased sample. The reason for this is that most likely a player on a basketball team burns more calories and eats more than an average student. In this sense, a basketball player is not representing the population very well which indicates that we have a biased sample. (iii) *Insufficient sample size*: when measuring a population, it may not be possible collect information from every member of that population. However, good news is that a "sample" can perform the job just as well. However, the important issue is to determine the right size for a sample to get accurate results. Using various methods, one can determine the right sample size which makes the data collection statistically significant. On the other hand, if the sample size is not sufficiently big enough, than the data may give us misleading conclusions. (iv) *Spurious correlations*: it does not necessarily mean that because two factors are correlated, one of these factors caused the variations in the other. Many statistical methods help us to analyze if the apparent relationships are meaningful and not simply chance occurrences.

Author details

Türkmen Göksel

Address all correspondence to: turkmen.goksel@ankara.edu.tr

Ankara University, Ankara, Turkey

References

[1] Pearson K. The Grammar of Science. London: Walter Scott/Dover Publications; 1892, 2004. ISBN: 0-486-49581-7

[2] Dodge Y. The Oxford Dictionary of Statistical Terms. New York, USA: Oxford University Press Inc.; 2006

Chapter 2

The Relationship between Theory, Scientific Explanation and Statistics in the Social Sciences

Laura Ibarra García

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.75117

Abstract

This chapter presents the criteria that sociological explanations and theories must cover and their role in empirical research based on quantitative methods. These criteria are: (1) must regain the phenomenon's intrinsic causality; (2) must reconstruct the conditions in which a social phenomenon surges and develops, although a historical methodology is insufficient in many cases; (3) must develop a second-order observation. Empirical sociological research that resorts to the quantitative method is among first-order observations. The second-order research are those that makes the first-order observations their subject of observation, which is why sociology, unlike other sciences, is a discipline that has itself as object of study (the sociology of sociology); and (4) must separate the intentions of the actors' actions from their effects. The first-order observation based solely on establishing statistical relationships, such as those performed by sociobiology, can offer the appearance of high scientific validity. In its last part, this chapter offers an example of the need to complete or reject data with second-order theoretical observations.

Keywords: theory, explanation, statistics, modern thinking, first-and second-order observation

1. Introduction

This chapter answers the following question: what relationship is there between the data-gathering and statistically oriented empirical research and the theoretical formulation and explanations of the social sciences? Even though it might seem obsolete after the discussions promoted by the rise of positivism, this question has obtained renewed relevance, due, among other reasons, to the predominance of the comparative-descriptive method in sociological and psychological research and that it has become one of the main analysis methodologies for society. Away from other debates, the comparative-descriptive analysis's only justification is its empirical reference; its goal doesn't offer anything beyond a picture of the similarities and differences observed in national societies. For example, there are investigations about the differences in moral values or unemployment, birth, marriage, divorce, mortality rates, and so on. This stems from the conviction that the comparison between social groups or between regions within a nation facilitates getting to know a society. Indeed, its results make it clear that there are differences within societies, for example, between north and south, east and west, and so on. However, these differences between one region and another, between the north and south of Italy, for example, tell very little about what makes the Italian society a society. This type of research appeals to empirical researchers or sociologists, although it barely helps to explain the emergence of society as a rising phenomenon. Society does not result from it. Many of its parts are classified and assessed, but their sum does not make a society [1].

Another reason why it is important to ask the question about the relationship between the statistical data elaboration method and theory is the current conditions in social science research. Subjected to the logic of project financing, many observers of society are forced to present "useful" results, inspired by logical positivism. At the end of their work, they invariably maintain that their conclusions have been rigorously tested, that they are based on first-hand experiences, and that they have resulted from logical analysis. The result has been a colossal amount of empirical knowledge, which is hardly the object of a second-order observation, with explanatory aims, or rebuttals that lead to social theories.

In terms of Piaget's theory, it could be said that a formal-operational thinking exercise, which considers the thought and not only reality itself, is needed [2]. Data alone say nothing. To understand it, a reasoning is necessary that thoughtfully relates it to ideas that give it a meaning. Concisely, it can be said that there is too much sociological research and a growing theoretical deficit. One of the reasons that have led to this is easily identified: in terms of projects financing, the institutional preference to support quantitative research that provides immediate political usefulness instead of the thoughtful sociological observation.

Another reason why we are inclined to consider the relationship between empirical research and the scientific and theoretical explanation is the recent emergence of a global society and regional societies of considerable dimensions, for example, the European society, which represents a challenge for quantification and theorization [3]. So far, the comparative-descriptive method and its statistical resources in sociological research had application within the limits of national borders [4]. When we look at the reports provided by the institutes devoted to statistical information, as well as the publications resulting from empirical research in social sciences, it is possible to argue we are before a "nationalist" methodology [4]. The nation state constitutes the basic unit for the calculation of every social research [5]. However, comparative-descriptive research has broadened its application area. Studies, like those performed by the OECD in its member states or those done by the European Union through academic agencies or institutions, gather valuable information about nation states´ populations and analyze their similarities and differences. However, as indicated before, these comparisons help to understand that there are significant differences between countries, for example, they help us to understand that the economies of Mexico and Germany are very different, but their contribution to understand the world order's reorganization or the globalization processes is not very significant. The challenge for quantification does not derive from the comparison between countries with similar social structures but between large regions. Undoubtedly, the European national societies can be compared to each other since they are units with multiple similar internal factors in a homogeneous social space and can be seen as elements of an institutionally defined space that encompasses them. However, what can Europe, for example, be compared with? Obviously, only with something comparable; with the United States, or China, or Latin America, probably [1]. However, due to the large dimensions of these units, their complex internal structures of subunits, which interact in a very particular manner, and their lack of a common space, it is difficult to establish comparison points. In addition, the way to compare regions in a way that contributes to understand the formation process of a global society is an open question.

To identify the contribution of quantitative strategies to the formulation of theory, we must pay attention to the conditions social sciences thought theories must fulfill. For this, we must first explain what is meant by "modern thinking." From this follows a guide for the construction of any social theory. Once the logical cognitive scheme every scientific explanation should be based upon is presented, the chapter describes the requirements that should be covered by any explanatory or theoretical reflection. First, it points out the need to resort to the conditions in which a phenomenon emerges. Any object or event results from a formation process that must be causally observed and reconstructed to achieve an explanation. This is also true in natural sciences, indispensable to understand diseases or the characteristics of a territory, for example.

Subsequently, the chapter addresses the distinction between first- and second-order observations, which allows us to distinguish between the empirical research activity and a reflection that relates the results of this activity to the wealth of conceptualizations and theories, which constitute a solid knowledge and have achieved the understanding of other phenomena, or, in any case, constructs it.

Separating the intentions of the members from their effects has resulted essential for the study of societies. Society is no longer seen as the result of a collective idea inspired on the achievement of the common good. Therefore, in the next point, this distinction is addressed, which allows for the entry of categories such as power, competence, or conflict, to explain the dynamics of social order and places morality as a relevant means only in a community where the neighborly, kinship, or face-to-face relationship ties prevail.

Then, we identify three requirements that may be considered the starting point for any explanation and theorization. In each of them, the role played by quantification is indicated. Given the results of statistical instruments that are often linked to assumptions that escape empirical verification, I present a sociobiology study at the end as example, which is supported by a rigorous statistical exercise but whose premises and conclusions lack an analytical foundation. Absent this, most thesis, no matter how crazy, can find a statistical basis to give it some credibility.

Surely, at this point, specialists in empirical research await an explanation of the observations′ methodology followed in this chapter, but methodological strategy is a procedure proper of the first-order observations, as we will see hereunder.

2. Requirements for scientific explanations

To answer the question about the relationship between the statistics producing research and the scientific explanation, it is necessary to first analyze how explanations are formulated in social sciences. The emergence of natural and social sciences is closely linked to a fundamental change in the historical development of thought that occurred at the beginning of the modern era. This change is particularly characterized by a break in the logic applied to understand and explain the phenomena of the world, or, in a given case, the world itself. Therefore, we must clarify what "explaining" means in the modern sense. From this surges every approach for society's theory construction as second-order observation. This radical transformation in the scheme or logic of thought, which may well be understood as a cognitive paradigm, should be understood as the basic condition for the development of any explanation or theory that claims scientific validity.

2.1. The desistance of absolutes as explanation

"The ancients" thought every phenomenon through until they recognized "one clear terminus" [6]. Nowadays, scientists agree that any hypothesis must desist from claiming an absolute origin that provides the ultimate argument. This is the result of the world's process of secularization that accompanied the three modernity revolutions: the revolutions in the natural sciences, in politics, and in economics.

From the epistemological point of view, this means that it is not possible to understand what exists as the product of an absolute origin and that all questions regarding its cause are answered by resorting to a single ultimate origin. Günter Dux explains this way of proceeding: "The absolute in premodern thinking, where spirit predominates, was absolute because it contained what was to come out of it as substance. The way of explaining consisted of attributing the explanation to the absolute to make it rise from there as an emanation." [7]. The explanatory potential of resorting to the absolute resided in the impossibility to question it. It was impossible to explain something within this logic, since the absolute origin subtracted itself from the question about its own cause. In this regard, Dux points out: "In a world that turned radically secular, in which… nothing that subtracts itself from a set of conditions can be found, the constructively achieved worlds – and their logics – must be explained through the conditions that made it be" [7]. Then, it's not that a certain explanation lost its convincing power along the history of thought, but the substitution of an explaining logic for another, of a paradigm for another. According to the old logic, explaining meant to part from the concrete object or phenomenon—natural or social—and to assign it an ultimate origin to claim it as its cause. This way of proceeding forced us to think of the cause as an origin analogous to what would emerge from it. Cause and effect were (partially) identical.

The attempt to find explanations through this process lost strength for the same question about the origin that could not be answered by resorting to an absolute origin anymore, not without falling into an endless return.

Accordingly, every form of explanation where the explainer is present in the initial explanation from which it emerges is unsatisfactory, for example, in the explanation of a social phenomenon where the said phenomenon is already embedded in the actors' wills, or in the explanation of specific social phenomena that are already in the own qualities of that society [4].

Modern thinking imposes then a first requirement for sociological research: it is not feasible to opt for a research strategy where the concept is defined beforehand to apply it then to the phenomenon being researched; we would rather have to work the other way, we must perform an observation and then specify how it developed.

2.1.1. The survival of the old explanatory scheme

Nevertheless, in many modern sociological and psychological theoretical approaches survive remnants of the absolutist and substantialist logic, which are especially visible in sociology's classical theories and authors because during the historical transition from one logic to another—and despite their attempt to abandon the old explanatory scheme—they retain the idea of an absolute as explanation, but under a different tag. In Durkheim's sociology of religion, for example, "society" is attributed to all forms of religious thought, but this concept in his theory does not provide any information about the creation process of mental constructions. Society is the only origin of the social. With regard to religious phenomena, Durkheim reaffirms the formula that inspired his sociological theory and is based on the identity logic that resort to the Absolute. In it, "society" appears as the origin and cause of all phenomena that require explanation.

The general result of the book (The elementary forms of religious life) is that religion is a prominently social issue, religious ideas are collective ideas that express collective realities: rites are behaviors that can only arise within groups and must be useful to maintain or restore certain psychic states of these groups [8]. In a different place: "just like the concept of religious power and divinity, the concept of soul is not devoid of reality…since society – the only source of everything sacred – is not content with moving us from outside temporarily; it is permanently installed in us." [8].

As shown, Durkheim's concept of society is an ontological construction placed in an all-explaining position by theory, and society precedes everything which sabotages any other possible explanation. The best proof that we are dealing with a secularized form of the metaphysical structure of thought is provided by Durkheim himself, when he declares that it will be necessary to choose between god and society as explanation [9].

A similar structure is behind several philosophical approaches. For example, in moral philosophy, any attempt to discover the origin of morality already involves morality itself. In philosophy, during the fundamentation process, what is intended to be justified is already given. In its explanatory reflection, what is to be explained is already in the explanation: language is in the origin of language, communication is in the origin of communication, and morality, of course, is derived from morality. In the transcendental theory of knowledge, morality as a part of reason comes, of course, from reason, but not as a slowly forming construct along the human societies natural history—having as a condition the brain in which the reason is not contained beforehand—but in its substantial form, which holds what will emerge from it [10].

Likewise, a fall in premodern absolutist logic also occurs in current theoretical approaches, which attribute certain characteristics to the concept of society, and then the phenomena in this society are seen as consequences of those characteristics. This circularity is found, for example, when it is affirmed that the risk society contains several risks, the society of options offers many options, or that in the society of experience many things can be experienced. All the previous statements are late absolutist historical manifestations, in which finding an explanation means to derive effects from a cause established a priori, and in which they are already contained. This same logic underlies the systems theory when it says that the explanation of the differences in society requires as starting point the unit of the system of society that creates these differences. With the development of thinking in the modern era, this construction of the concept of society results obsolete [4].

Hence, it is only possible to scientifically accept a concept of society when it allows to explain what is to be understood by resorting to empirically verifiable actions in society and asking about the existing conceptions of society in praxis. For example, when society is conceptualized as reciprocal forms [11] or as "a network of connections of practical forms, in which men lead their lives" [12].

The concept of society must be pointed toward the relationships between elements and must neither admit absolutist nor substantial connotations. Vobruba even recommends desisting from a pre-formulated definition of society. He points out that this does not mean the end of a theory of society; rather, it opens the possibility to articulate a theory with empirical content. A theory that can clarify the existing phenomena in practice and can question its causes and its effects [4].

Evidently, in this sense, empirical research, including that which uses quantification and variable relations, can be a very valuable instrument for the process of articulation of a concept of society. If sociology must deal with social and political events subject to a network of actions and social relationships in which these events occur and produce effects, the analysis that empirically verifies them, classifies them, and evaluates them represents a first and valuable step for a latter theoretical exercise that must retake them and explain them. This way of proceeding ensures that the explanation will not be included in the explainer, instead, the explanation will resort to the elaboration of a complex relationship between causes and effects based on data that portrays reality, that is, that ensure an empirical content.

2.2. What does explaining mean?

If scientific explanations must desist from resorting to the absolute as argument, what must they appeal to then?

Günter Dux very precisely points out what the change in the logic of explanation consists of: the absolute is discarded to recover the real causes of the phenomenon itself, in other words, a causality that thought imposes on phenomena is abandoned, to reflexively reconstruct the real causes inherent to the observed process, that is, the objective relations between things [10]. Therefore, the historical decline of absolutist logic as explanation for the phenomena was accompanied by the emergence of the diversity of science. At the beginning of the modern era, it was no longer possible to resort to an ultimate cause, but each phenomenon showed specific determinants that gave rise to the different sciences. A lengthy list of systems replaces the limited relationships of the previous logic. According to Piaget:

> *"There are logical operations such as those resulting from a class system (meeting of individuals) or relations, arithmetic operations (addition, multiplication, etc., as well as their investments), geometric operations (intersection, deviations, etc.), temporary operations (ordering sequences of events and insertion of intervals), mechanical, physical operations, etc."* [13].

In social sciences, the study of the conditions in which cultural forms of life arise made it visible that a methodological strategy that assumed the incorporation of history as explanation was required. To understand social phenomena, it was inescapable to resort to their formation process, their development. Although the theories of sociology's classic theorists, like Durkheim or Marx, have lost their explanatory power for distinct reasons, they have at least the merit of introducing history's explanatory role. The importance of *The Capital* for the development of knowledge consists of following: the formation of economic development from its genesis, first in the commercial capital and then in the industrial capital of the eighteenth and nineteenth centuries.

In the case of psychology, regardless of the current value attributed to psychoanalysis, identifying mental pathologies as the result of traumatic experiences during the history of the individual was the fundamental contribution of Freudian theory. Albeit, a single experience and experience type—sexuality—were given too much importance, psychic disturbances found their explanation in the subject's own past. Nowadays, everyone accepts the close relationship, discovered by the Freudianism, between an individual's affectivity and his past, especially his childhood [13]. Everything looks different after a revolution.

However, it must be emphasized that scientific work consists of apprehending the intrinsic causality in the dynamics of the process. Every sociocultural form of life can only be explained if one resorts to the conditions in which it started and under which it has continued developing. This is what Marx means when he says that there is only one science: history [14].

As we have seen, explaining is a concept that assumes a specific meaning in social theories: it is about reconstructing the conditions in which the sociocultural forms of organization emerged and then evolved.

A scientific explanation can set spatial and temporal limits to the observation period of the conditions in which the social phenomenon of interest arises, while a theory must contemplate the core aspect of all its development. It must be capable of generalizing in a way that its application covers broader periods of time or a wider range of phenomena.

In a social science theory, the starting point is the set of constellations formed by systematically organized conditions that influence each other and that only in their joint action give rise to the result. In this sense, the methodological logic that is followed is a relational and systemic logic. However, it is also a procedural logic. By this, we mean that the result, the thought, language, or morality organization forms are not found within the constellation of conditions that constitute the starting point. The result is formed in the process. The difference with the methodology followed by philosophy is evident: in philosophy's logic of derivations, there never is a new phenomenon. Within the procedural logic, there is room for a new phenomenon to arise: life can surge from lifeless matter, from life devoid of spirituality, cognition, language, or morality rise.

However, the reconstruction of the conditions in which the sociocultural forms of life were formed represents only the first step. If we intend to objectively and scientifically analyze the sociocultural forms of life, including current ones, it is necessary to explain why the sociocultural forms of life developed and grew in a specific way. Of course, this approach involves the task of reflecting on the development of history in its entirety. However, a sociological theory is not about advancing the knowledge of the succession of events in their outward appearance, but to follow the footsteps of something like a development logic that sustains the process. Marx already tried this regarding production systems when he found out that the development of the productive forces in human history follows a line that started with hunting and gathering, continues with the agricultural forms of production and all the way till reaching industrial production. And, who would dare today to deny that, in this sense, social evolution is moving in this direction and hence follows a logic? In the same way, Piaget endeavored to demonstrate that the ontogenetic cognitive development parts from a sensory-motor intelligence go through the preoperational stage until reaching the operational and formal-operational stages. And although in the 1980s, a discussion arose regarding the universality of the sequence of stages, epistemologists do not hesitate to admit that cognition is subject to development and that it registers a logic, as Piaget formulated it.

The postmodernist idea that history follows the irrationality tenet undermines the knowledge that countless empirical studies have achieved, and we can consider reliable information. At this point, the postmodern stand is only based on the reluctance to place oneself in the field of knowledge that empirical studies have made available. History is not simply the sequence of innumerable events in which each of them differs from the previous state, the sequence of life forms with which man becomes accessible and reality follows a logic of development. This is precisely what a theory seeks to prove. History, as a man-determined history, in which he interprets himself, is susceptible to explanation. In a rigorous analysis of history, it is not possible to argue that the development has lacked logic, neither in its beginning, nor in its direction. If history is understood as a series of life forms in which man becomes accessible to the world and interprets himself in it, then it becomes evident that development possesses a logic. However, the historical method is not enough for everything [13].

Although one might think that the social sciences should resort to history, this does not necessarily mean a remote past that goes to the very process of formation of the constitution of the anthropological species, the conditions in which a phenomenon arises and develops to include conditions in the present or recent history, which require empirical verification, especially in their causality and effectiveness. Quantitative studies provide valuable information that can be considered the solid foundation for the explanations and theory formulation. To this condition is added a third that distinguishes sociology from other disciplines of the social sciences.

2.3. Society under two observation perspectives

To answer the question: what should we consider a scientific approach in a science of society? We first must bear in mind that there are two orders of observation and interpretation [15]. An observation of the first order is the people's perspective, who observe and interpret social relations and, according to this, act. In the second-order perspective, people's actions, along with their observations and interpretations, are sociologically observed.

The main difference between the first-order observations level and the second-order observations level lies in the people, which act in relation with a reality on the foundation of their first-order observations, while second-order sociological observations are not linked to the action. Any sociological definition of society must include as reference the actions of the members of society and the knowledge of their actions in which "society" can be identified. Since society is built even without sociology's participation, sociological observation must adhere to the way in which people observe and interpret social relationships and act in society accordingly. As sociology observes and interprets the observations, interpretations, and actions of the people, it assumes the reality with which people act and thus escape the danger of constructing arbitrary conceptualizations, since people's interpretations cannot be of any kind if you do not want actions to cause any damage.

Andreas Balog points out that the solid foundation on which the formulation of a concept is based, what provides certainty over the identity of the social phenomena and sets limitations to sociological categorization, is common knowledge, which constitutes the basis for the orientation of the actors in their daily world and is manifested in their actions and in the language of daily life [16]. In this sense, sociology is always a second-order observation.

Thus, since sociology is a second-order observation, its subject of study is society in the sense that there is a group that understands itself as society. It is totally irrelevant if the sociologist thinks there are no reasons to consider that Paraguay and Uruguay are two different societies, what matters is that for the people living in Paraguay and Uruguay—that is, in the first-order observation—there are enough differences to consider them as two different societies. In the second-order perspective, the concept of society is applied to recover what in society (as empirical group) is perceived and interpreted as society. The concept of society in the second order of observation is that which exists and acts in society.

Given that everything in society can be observed in the second-order sociology, even sociology itself, the observes can become themselves observable, or first-order observers. And, since everything observed from the second-order perspective turns into first-order observation, empirical sociological research becomes first-order observation and sociologists themselves become "the people" [4]. When this occurs, it is observed how sociologists observe, interpret, and act in the sociological practice. According to Vobruba, sociology can have itself as object of study, since it can become its own object of observation (id.).

In the sociological perspective of first-order observations with the purpose of conducting research, the concept of society can only be required to cover the field of study. Hence, the concepts of society cannot be considered as true or false, but simply as practical or impractical. Every sociologist, or first-order observer, can formulate a concept of society and deduce the consequences that result from it. Sociology's scientific means do not allow to establish "the right concept of society." Therefore, each author can formulate their own concept, and it can also be expected for other observers to present their own different concepts, which can be documented empirically with the number of concepts of society available [4].

This differentiates sociology from other disciplines. There cannot be a medicine of medicine, nor a legal science of social science, nor an economics of economics. However, there is a sociology of medicine, a sociology of law, a sociology of economics, and, also, a sociology of sociology in the sense that a sociological analysis of the cognitive requirements sociology must present.

Statistics is a first-order observation. However, this is a better knowledge than which comes straight from the senses and which has a subjective character. In this context, better means a more direct picture of reality. Hard data are an attempt to understand objective reality in a more precise and correct way. However, data alone do not say anything. Second-order conceptual research is necessary to place it in a context that gives it meaning.

2.4. Separating intentions and consequences in social integration

In the past, the concept of society used to designate social groups that were intentionally established and were aimed for specific objectives. At the beginning of the modern era, society was understood as an intentional institution of equals (bourgeois) for economic purposes. Because of this, the concept was used in a more plural often. In the seventeenth and eighteenth centuries, Hobbes, Locke, and Rousseau thought of society as the product of a constitutive act at some point in history, the result of their intention to ensure peace, protect their property, or a willingness to ensure equality and the freedom. Society was understood as the product of a social pact that ensured its future integration. Later, the concept of society was transferred to problems of social order, which maintained the intentional structure of the understanding of society. It was believed that, due to the importance of politics as a guiding system for society, it was possible to translate the intentions into any desired political model, society was thought of as a malleable political entity.

Durkheim strove to overcome the vision of a society built by a social agreement or to discard the idea of an "invisible hand" that regulated social relations. Because of the division of labor, men in modern society find themselves increasingly in dependency relationships, which moderates their selfishness and converts them into moral beings. "In one word: since the division of labor becomes the main source of social solidarity, it simultaneously becomes the basis of the moral order" [17]. Society's integration is the result then of a single principle: the moral order. However, modern society, unlike a community in which relationships occur face to face, is not based on morality. The term "community" is applied to premodern relationships, in which ties of kinship and friendship play a key role in social integration, but in society this is not the case. In society, a shared morality, outside of a minimum, that allows interaction, is not a condition for its constitution. From the epistemological perspective, Durkheim's thinking is halfway between the break between absolutist thinking and modern thought, between a substantialist conception of society and a conception that understands it as a rising process.

In the second third of the twentieth century, other contributions with important repercussions in the way of understanding society emerged. Among them, a new observation approach for social relationships linked to the growing inclusiveness of the concept of society was introduced. In this approach, the effects on society are seen and systematized as separate from the intentions; the individual intents can have collective consequences that distinguish themselves qualitatively from these intents. That was the beginning of the end of the idea that society is composed of individual efforts or intentions that seek social unity. The creation of a level of observation in which it is possible to separate the intentions from the effects is the requirement to deprive the social from the moralization and cognitivization. Society thus becomes an object of observation and the materialization of the effects that result from social relations.

As soon as the actions´ intentions and effects are disengaged, that is, as soon as it is possible to think of society as a result of unintended effects, "the good purposes" lose their significance. With this, categories such as inequalities, power struggle, conflict, or antagonisms regain relevance in the conceptualizations of society, and solidarity and morality are reduced to a motivation for action among many others. In modern sociological theories, social phenomena can be attributed to actions, but not to the intentions that guide them.

Due to the introduction of the difference between intentions and effects and the disassociation of the concept of society from the intentions, a semantic space emerges to separate the concepts of society (unintended) and community (intended). However, there still are sociological approaches that try to explain society and all the events that happen in it by the hand of its actors´ intentions, and because of this they acquire a futuristic character that starts from the good intentions of the social actors.

3. The "short circuit" between empirical results and second-degree observation: The case of sociobiology

Empirical research has gathered a considerable amount of valuable descriptions with its statistical-descriptive method. The main problem of these investigations is that they are often linked to conclusions, which are based on assumptions that are subtracted from empirical verification. This is the case of sociobiology when, before formulating the assumptions that will be verified, parts from the idea that social capabilities are contained in the genome, for example: studies that conclude a relationship between "racial ancestry" and social phenomena subject to other forms of determination. The following is stated in one of them [18], for example:

> *We find that European ancestry is consistently and usually strongly positively correlated with cognitive ability and socioeconomic outcomes (mean r for cognitive ability = .708; for socioeconomic well-being = .643). And even further:*
>
> *12 zero order correlational analyses found a substantial positive relationship between European ancestry and both cognitive ability and general socioeconomic well-being…*
>
> *While the association between racial ancestry and outcomes is consistent with a genetic hypothesis, to obtain decisive evidence in support of such a position, one would need to identify specific alleles that vary between ancestral groups which are directly* [19] *or plausibly indirectly* [20] *associated with cognitive and/or socioeconomic outcomes on the individual level…*[21].[1]

In the conception of the individual that this study parts from, remnants of absolutist thought are still present: the sociocultural forms of life are presented as if they could be explained from a genetic a priori. In a second-degree observation, one may ask: What does this mean? How important is, for the behavior of an individual, what past generations, in the natural history of the species, have contributed as genetic material? How exactly does this material lead to certain behaviors?

This study shares the false supposition that genetics determines the cognitive level with the rest of the sociobiological investigations. The great wealth of psychogenetic and educational studies, initiated by Jean Piaget, has shown that the determinants of any individual's behavior do not reside in the genes but in learning, the actions of a subject are the result of its life experiences. The theoretical attempts to re-naturalize mental capacity have failed. To effectively evaluate the influence of the genetic background in cognitive abilities—which the study intends to show—first we need to remember the role played by the genes and the development of learning in the natural evolutionary history of the species. This way, it must be clear what biological equipment can and cannot do. Let's take a better look at this.

3.1. Theoretical excursus

Undoubtedly, the key to understand the cultural forms of life, including cognition, morality, love, and so on, should be sought in the natural history of the species first. The development of what characterizes the anthropological species was linked to the natural evolutionary history. Therefore, sciences try to reconstruct the development of these forms in the long transition process from animal to man. Some of the key questions sciences seek to answer are: What

[1] A more detailed analysis of this investigation as well as the reply of its authors can be found in [22].

were the conditions that made it possible for the anthropological species to develop the capacity to learn? How was the human being formation process as such possible?

The key process to understand the evolutionary development of the anthropological species is the ontogenesis. For millions of years, our species development was characterized by two parallel processes: the dissolution of instinctive and inflexible genetically determined forms of behavior and enormous learning in the first phases of ontogenesis. Through its relationship with the person responsible for its care, the member of the species in initial stages was able to acquire experiences that allowed it to increase its actions competence and arrange reality into a comprehensible world, which led to a gradual radical change of the genetic material, so instinctive structures lost their strength as determinants of the actions. In the transition from animal to man, the lengthening of the primary relationship in ontogenetic development played a fundamental role. Everything seems to indicate the lengthening of the close mother-son relationship during the first years of life, and the huge amount of learning that takes place here made the instinctive mechanisms ineffective, in a process that lasted thousands of years. In the determination of human actions, the genetic basis plays a limited role, it is only present in fields closely linked to survival, such as feeding, sexuality, and defense. The evolutionary consequence of this process is the extreme inability of the anthropological species to survive by itself at birth.

The evolution of the *Homo sapiens* would not have been possible if functional mechanisms for the construction behavior pattern linked to learning had not developed along with the dissolution of instinct. Brain development was key for this. Even though we do not know much about the brain's evolutionary development and the formation of differentiated neuronal zones, these made the acculturation process of men possible, and, with it, the formation of cognitive and normative structures. There is no doubt that without the brain's constructive capacity, these structures wouldn't have formed. However, it must be clear that these are only brain capacities and not cognitive competences, the latter must be acquired in a constructive way by every member of the species in their experiences in the outside world.

Since the anthropological constitution hardly has any survival instincts, the member of the species in initial stages is forced to develop mechanisms that allow it to lead an independent life. In its interaction with the adult, the new member of the species coordinates its motor skills, achieves greater competence in its actions, and develops structures that it uses to organize the world. The requirement for this to happen is precisely the social relationship with the adult in charge of its care, usually the mother. The newborn finds in her not only a figure that keeps his body alive but also a representative of the outside world from whom he learns a huge amount of behaviors and knowledge. The mother–child dyad is the stage during early ontogenesis in which the decisive learning process of survival takes place. Here begins every way of understanding reality, all the knowledge about our surrounding environment, every thought. For this reason, mother–child interaction and the mental capacities acquired in it are of paramount importance in the construction of cultural forms of life, especially cognition.

If we consider the null cultural situation of the newborn and the interactions established in the first stage of the biography, there can be no doubt that the conditions at the beginning of the formative process at any age and in any society are the same. In every age and in all societies, for 40,000 years, the member of the species in the early stage, with nothing but a few instincts to ensure its survival, is forced to develop mechanisms that allow it to lead an independent life. It is exactly this situation which repeats everywhere what creates equality in the early cognitive structures. The same conditions in which the process takes place determine the same results. Therefore, there is no known society, present, or past, in which its members hadn't started to coordinate their motor skills or build the schemes of object/subject, space, time, and causality in the early stage. The same thing that happened to the phylogenetic development continues to occur at the beginning of every biography. At this level of knowledge, ethnic and cultural differences play no role.

As Piaget pointed out, almost every competence has a biological basis from which they part, but human competences are constructive processes, neither human conscience nor subjectivity is determined by nature, which only offers the opportunity to develop them. The newborn doesn't count with a world in order; for him, there is no space nor time, objects lack shape and permanence, and there are no causal relationships between them. Nor does he have the capacity to organize his toward an end and doesn't even perceive himself as being independent of the outside world. He will have to learn all of this through a long constructive process.

3.2. Reconsideration of ethnic determination in cognition

After briefly reviewing the natural history development of the anthropological species, we turn our attention to sociobiological studies.

It should be clear by now that human beings do not count with a competence for action nor a structure of the self provided beforehand by nature. He doesn't bring with him the principles of social organization, nor does it have enough knowledge of the outer world he is born into. Unlike other species, he doesn't possess an array of the natural elements that, fixed in the genetic code, reduce them to the elements relevant for action, as it happens with instinct.

If we bear in mind that at birth every member of the species is in this null cultural situation and undertakes a constructive process of the world, and that all of this happens in every society and in every age, one must ask, under the precepts of sociobiology: what is the point of trying to statistically demonstrate that belonging to a race, or genetic constitution, has any relationship with cognitive abilities?

If European, indigenous, or African children are born with a biological equipment that does not ensure their survival in any way, and all of them have the same task of undertaking the lengthy process of constructing their world and survival competences of action before them, how is it possible then to search the reason for the cognitive abilities in genetic differences?

Sociobiology parts from the assumption that cognition forms are found in the genome (more recently, in the brain). It assumes that cognition forms are already given in the genetic material, and while stored there only wait for an opportunity to manifest. However, the construction of the cognition forms happen on a stage different from the biochemical evolution.

http://dx.doi.org/10.5772/intechopen.75117

Scientific review of what happens in evolution allows us to confirm that sociocultural life forms, including cognition, do not take shape in the biological sphere itself, the genome, or the brain. The biological sphere does not know of mental abilities, it only provides the conditions to build them. A "naturalization of the mind" on the understanding of human forms of existence as links to a species evolution is only admissible in the sense that cognitive, linguistic, normative, and esthetic competences are constructed over natural capacities without being stored in them. If we consider the defining role of learning in the acquisition of cognitive competences, belonging to a race lacks explanatory validity when trying to find the reasons that truly allow to understand the differences between individuals in the field of cognition.

3.3. The detachment of the "ethnicity" variable from the historical and social context

A second argument in sociobiological studies to be challenged is that the "European ancestry" variable appears detached from its historical and social context; as a consequence, statistical relationships formulated with it do not contribute to consider the true problem posed by the fact that a white minority, those with a European ancestry, in some Latin-American countries, show better cognitive abilities and significantly higher income rates than the rest.

If we want to comprehend the relationships between these variables, it is necessary to resort to history to understand what this relationship truly means and contemplate the changes introduced by the democratization in the composition of power in democratic Latin-American societies and which have resulted insufficient to put an end to social inequalities. This is not the place to explain in detail what the links with European ancestors mean sociologically, because this would require a broad and deep reflection, but the situation with the minority of "European ancestors" can at least be mentioned and referred to what is truly significant in current sociological studies: its status in the composition of power.

In Latin America, the European arrival at the beginning of the sixteenth century occurred through a cruel war of conquest. The Spaniards destroyed the old indigenous world and imposed a regime of domination, in which they occupied the positions of civil, ecclesiastical, and military power. The indigenous people were not only plundered but were also subjected and enslaved. After 300 years, the Latin-American criollos started to drive the Spaniards out through the independence war, but this didn't mean a profound change in the conditions of the poor peasants and indigenous people. Independence made way for the development of the middle class and brought a certain improvement in the situation of the mestizos. In Mexico, it wasn't until after the 1910 Revolution that the old rural regime, based on the concentration of land in large private states, disappeared to finally make way in the 1917 Constitution for some of the demands of peasants and workers.

Despite industrial progress, the expansion of the middle class, the considerable increase in miscegenation, and even the democratic advances in Latin-American countries, regional elites have maintained certain ethnic purity since colonial times. However, traits such as skin color and race purity are unimportant in sociological analysis; what truly matters is the permanence of the same elite at the top of the composition of power and the mechanisms that have kept them in power.

The statistical analysis of European ancestry in a network of variables is irrelevant considering the dramatic fact that, despite the arrival of democracy, those who have European ancestry continue to hold privileged positions in the composition of power, and indigenous people (10% of the Mexican population) keep occupying the population segments in extreme poverty.

4. Conclusions

The need to rethink the relationship between empirical research that uses quantitative methods and the articulation of explanations and theories results from three circumstances:

1. The valuable material provided by the comparative-descriptive analysis is insufficient to formulate an explanation of social phenomena.
2. There is a growing theoretical deficit in sociological research due to the impressive amount of first-order observations.
3. Globalization represents a challenge for empirical research, but also for its explanation and theorization.

The aim of this chapter is to state the criteria that sociological reflection must fulfill to articulate an explanation and a theory.

The first one comes from the fundamental change in the history of thought at the beginnings of modern thought, which replaces a logic that resorts to an ultimate absolute origin as explanation for a relational-procedural logic. However, remains of the old logic survive today in numerous explanations, approaches, and theories.

In the modern sense, explaining means to resort to the conditions a phenomenon arises from, in other words, to cognitively recover its intrinsic causality. However, a method that resorts to history is not enough. It requires a second-order observation that analyzes the first-order empirical observations and associates them to other knowledge. Statistical data from empirical research must be subject of a professional sociological observation that correlates it with the knowledge that has been recognized as objective and valid by the scientific community.

The reflections made by a second-order observation of the statistical material from sociobiology, especially those studies that statistically link "European ancestry" with income and cognitive levels, show that in quantitative methodology, variables that in social reality are immersed in a network of historical, social, and political relationships are taken out of context to form part of a two-dimensional statistical model. Without a reflection that contributes to give meaning to the data, the mere enunciation of correlations does not offer possibilities to explain any of the phenomena it deals with.

In the case of the existing relationships between race and cognitive abilities analyzed by the sociobiology, they follow the already historically overcome early logic. This logic parts from what is before it and refers to an origin in which is already contained what is to result from it.

In this sense, whether in the race or genes, what is to emerge from them is already contained in them. In this reasoning, what truly determines cognitive abilities, the learning development, is left out.

Author details

Laura Ibarra García

Address all correspondence to: 95nubi@megared.net.mx

University of Guadalajara, Mexico

References

[1] Eder K. Die EU als entstehender Kommunikationsraum. Zum Theoriedefizit der soziologischen Europaforschung und ein Vorschlag dieses zu verringern. In: Eigmüller M, Mau S, editors. Gesellschaftstheorie und Europapolitik. Sozialwissenschaftliche Ansätze zur Europaforschung. Neue Bibliothek der Sozialwissenschaften. Wiesbaden: Springer VS Verlag für Sozialwissenschaften; 2010. pp. 80-108

[2] Piaget J. Psicología de la inteligencia. Buenos Aires: Editorial Psique; 1972

[3] Bach M. Die Konstitution von Räumen und Grenzbildung in Europa. Von verhandlungsresistenten zu verhandlunsabhängigen Grenzen. In: Eigmüller M, Mau S, editors. Gesellschaftstheorie und Europapolitik. Sozialwissenschaftliche Ansätze zur Europaforschung. Neue Bibliothek der Sozialwissenschaften. Wiesbaden: Springer VS Verlag für Sozialwissenschaften; 2010. pp. 153-178

[4] Vobruba G. Geseellschaftstheoretische Grundlagen der Europasoziologie. Die soziologische Beobachtung der Gesellscahft under Europäischen Integration. In: Eigmüller M, Mau S, editors. Gesellschaftstheorie und Europapolitik. Sozialwissenschaftliche Ansätze zur Europaforschung. Neue Bibliothek der Sozialwissenschafte. Wiesbaden: Springer VS Verlag für Sozialwissenschaften; 2010. pp. 431-470

[5] Bach M. Die Europäisierung der nationalen Gesellschaft Problemstellungen und Perspektiven einer Soziologie der europäischen Integration. In: Bach M, editor. Die Europäisierung nationaler Gesellschaften. Sonderheft der Kölner Zeitschrift für Soziologie und Sozialpsychologie. Vol. 40. 2000. pp. 11-38

[6] Wittgenstein L. Tractatus lógico-philosophicus. Logisch-philosophische Abhandlung. Frankfurt a. M: Suhrkamp; 1971/1918/1921

[7] Dux G. Historisch-genetische Theorie der Kultur. Instabile Welten. Zur prozesualen Logik im kulturellen Wandel. Weilerswist: Velbrück; 2000

[8] Durkheim E. Las formas elementales de la vida religiosa. España: Ediciones Akal; 1982

[9] Durkheim E. Sociología y Filosofía. Madrid: Miño y Avila Editores; 2000

[10] Dux G. Historisch-genetische Theorie der Kultur. Instabile Welten. Zur prozesualen Logik im kulturellen Wandel. Weilerswist: Velbrück; 2002

[11] Simmel G. Soziologie Untersuchungen über die Formen der Vergesellschaftung. Frank furt a. M: Suhrkamp; 1908/1992

[12] Dux G. Das Subjekt in der Grenze der gesellschaft. In: Nikos Psarros P, Stekeler-Weithofer GV, editors. Die Entwicklung sozialer Wirklichkeit. Auseinandersetzungen mit der historisch-genetischen Theorie der Gesellschaft. Weilerswist: Velbrück; 2003

[13] Piaget J. Theorien und Methoden der modernen Erziehung. Wien: Verlag Fritz Molden; 1972/1964

[14] Marx K, Engels F. Die deutsche Ideologie. MEW. Vol. 3. Berlin: Dietz Verlag; 1969. pp. 9-530

[15] Luhmann N. Die Wissenschaft der Gesellschaft. Frankfurt a. M: Suhrkamp; 1991

[16] Balog A. Der Begrif Gesellschaft. Österreichische Zeitschrift für Soziologie. 1999;**24**(2):66-93

[17] Durkheim E. La división del trabajo social. Madrid: Akal Ediciones; 1995

[18] Fuerst J, Kirkegaard E. The Genealogy of Differences in the Americas. Mankind Quarterly. 2016;**56**(3):256-374/252-253

[19] Piffer D. A rewiew of intelligence GWAS hits: Their relationship to country IQ and the issue of spatial autocorrelation. Intelligence. 2015;**53**:43-50. Available from: https://www.sciencedirect.com/science/article/pii/S0160289615001087

[20] Fedderke J, Klitgaard R, Mac Murray J, Napolini V. Diagnosing deep roots of development: Genetic, desease and environmental factors. Economic Research Southern Africa. 2014. Available from: https://econrsa.org/system/files/publications/working_papers/working_paper_465.pdf

[21] Rindermann H. Evolution vs. Culture as background factors for international intelligence differences. Presented at the London Conference of Intelligence. 2015. Retrived from: https://mail.google.com/mail/u/0/#inbox/14eee641d9da3248?compose=14ef3e99e65c886e&orojector=1

[22] Fuerst J, Kirkegaard E. The genealogy of differences in the Americas. Mankind Quarterly. 2016;**56**(3):256-374-252-253

Section 2

Applications of Statistics on Economics and Development

Chapter 3

Statistical Methodology for Evaluating Business Cycles with the Conditions of Their Synchronization and Harmonization

Elena Zarova

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.75580

Abstract

The importance of the topic of business cycle research and their interaction is due to the fact that the cyclical nature of development is a universal feature of the market economy (regardless of the level of development of the country's economy and the principles of its organization). In all cases, cyclical ups and downs depend not only on internal system cyclical processes and their factors in countries but also on the consequences of intercountry interaction. The ability to measure and predict business cycles, taking into account their mutual influence, is a prerequisite for the development of an adequate business policy of countries and their associations.

Keywords: business cycle, indicator, synchronization, harmonization, principle component, modeling

1. Introduction

This chapter is devoted to the substantiation of methods of statistical assessment and modeling of macroeconomic business cycles on the basis of their understanding as an integrated effect of changing business phases in different sectors, as well as the impact of synchronization and harmonization of business cycles in both the economy of one country and the intercountry levels.

The main directions of quantitative research of business cycles based on the econometric approach, which are widely presented in the literature, fall into two main groups. The first of these is the identification of stable cyclic components in the dynamics of macroeconomic indicators. In most cases, the authors of scientific publications use the real GDP (gross domestic product), as an indicator for investigation of macroeconomic business cycle. The development of this direction is the study of the interplay of the "total" business cycles identified in the GDP dynamics by countries. The second direction is the definition and quantitative description of the cyclic components in the dynamics of indicators characterizing the processes in individual sectors or spheres of the economy. The "specific cycles" thus estimated are the basis for the identification and quantification of the so-called common cycle using cluster analysis methods. In this case, the "common" (or as it is also called the "reference") cycle is represented as a multivariate value of the actual reaction of individual industries or economic subsystems, e.g., financial, investment, labor market, etc., observed with the help of statistical indicators. This chapter proposes a method for identifying and quantifying the common business cycle as a directly unmeasurable phenomenon, which manifests itself in fluctuations in the dynamics of the specific indicators of industries and economic systems, but has an objective and independent economic nature. In the author's opinion, specific cycles, even if they have a leading character with respect to the general cycle, are its economic consequence. The chapter suggests methods and their applications for identifying and quantifying the macroeconomic business cycle as a latent system-wide phenomenon, as well as methods for estimating intercountry synchronization and harmonization of common cycles. At the same time, the author gives her own definition of these forms of the interaction of business cycles. This approach is different to the view expressed in many publications about the identity of the concepts of synchronization, harmonization, concordance, and correlation of business cycles.

The chapter in addition to this introduction includes two main sections and a conclusion. The first section is devoted to the review of scientific publications that disclose the concept of the "business cycle," and also this section presents the author's systematization of the methods outlined in a number of publications for identifying cyclic components in the dynamics of macroeconomic indicators. In the second section of the chapter, an algorithm for quantifying the overall business cycle based on the principal component method is proposed, and methods for estimating synchronization and harmonizing business cycles are substantiated. The conclusion of the chapter contains a concentrated expression of scientific novelty of the author's methodological proposals of the business cycle quantification and the features of the algorithms for evaluating their synchronization and harmonization at the macroeconomic level.

The proof of the concept proposed in this chapter and the approbation of the corresponding algorithm were realized on the basis of statistics of Eurostat and Rosstat. Most of the examples in this chapter are compiled from the results of calculations for Germany and Russia, which allows a comparative analysis provided that there is a significant difference in the duration and stages of the history of the market economy that have a significant impact on the formation of sustainable multi-year business cycles.

Approbation of the proposed algorithm can be based on R-packages or the "STATISTICA" program.

http://dx.doi.org/10.5772/intechopen.75580

2. A literature review of the definition of a business cycle and methods for extracting cyclic components

The number of scientific papers devoted to the theory and methods of quantifying business cycles is measured in hundreds. Nevertheless, there is a basis for their distribution into two large parts.

This basis is the definition of the concept of fluctuations in economic activity. This determines the choice of mathematical and statistical methods used to determine business cycles (Kijek [20]).

The first part includes the papers which are based on Burns and Mitchell's definition of a cycle [3]. According to them, a business cycle represents the four distinct phases of "aggregate economic activity" development that evolve from one into another: expansion, recession, depression, and revival.

The second part is the works based on a different view of business cycles, which was presented by Lucas [23], who does not interpret cycles as inevitable transitions between different phases of the cycle. He sees the business cycle as a process of oscillation of GNP around a long-term trend.

Some authors define the business cycle in accordance with Burns and Mitchell and then continue to measure the fluctuations of the macroeconomic aggregate values (GNP, GDP, industrial production, investment, and so on) relatively to the long-term trend.

This approach cannot be adopted without a preliminary study confirming that the observed cyclicality of the national economy is a violation of macroeconomic equilibrium rather than synchronous periodic fluctuations of various economic activities without changing of their balance.

The two basic approaches to the definition of the concept of a business cycle, mentioned above, were the basis for developing analytical tools for recognizing cycles.

Harding and Pagan [17] define three directions ("traditions") in the deviation of approaches presented in the literature on the development of cycle indicators from information available in a time-continuous random variable (yt).

The main idea of the first direction is changing of the initial time series (yt) by a series of binary random variable St, which includes turning points. In this case, peaks (troughs) are considered as local maxima (minima) in the series yt, and they are taking the value unity and zero otherwise:

$$\begin{aligned} \vee \mathrm{t} &= 1\left(y_t < y_{t\pm1} \leq \mathrm{j} \leq \mathrm{k}\right); \\ \wedge \mathrm{t} &= 1\left(y_t < y_{t\pm1} \leq \mathrm{j} \leq \mathrm{k}\right). \end{aligned} \tag{1}$$

where ∨t (∧t) are binary variables and k is the length of period for local maxima (minima) estimation.

Harding and Pagan [18] show that for Burns and Mitchell's specific cycle dating procedures it is necessary to set k = 5 for monthly data or k = 2 for quarterly data, but it is not the single variant. Mathematics and IT providing of extracting and estimation of turning points in structure of economic time series have been actively developing.

Other two directions, or "traditions," by Harding and Pagan [17] are combined because they both are based on the prior transformation of yt so as to remove a permanent component, leaving only a transitory one (zt). The first direction of them provides for the isolation of the cycle on the basis of an analysis of the presence or absence of peaks in the spectral density. The second direction combines the methods of extracting a cycle by the results of the analysis of serial correlations in zt series.

The abovementioned approach to the systematization of methods for extracting and quantifying cycles in long-term economic dynamics, set forth in papers of Harding and Pagan, is accepted by many scientists, but is not the only one. An example of a somewhat different classification of methods for isolating stable fluctuations reflecting business cycles is presented in the paper of Kijek [20].

According to this author's view, three dominant approaches are used to distinguish cycles. Two of them coincide with the groups of methods singled out by Harding and Pagan (the last two "traditions"). They are (1) the presentation of the time series as a difference-stationary process and the application of autoregressive integrated moving average (ARIMA) models and (2) the presentation of economic dynamics as a trend-stationary process and treat it as a sum (but also it can be multiple interconnection) of polynomial as deterministic trend and stochastic deviation around it. The last one is considered as a residual cyclical component, which includes business cycle's pattern and random deviation.

The third concept of time series decomposition according to Kijek [20] is the use of frequency filters, methodology, and software which are widely represented in the statistical literature.

Some authors include in the classification of methods for extracting from the time series of macroeconomic indicators of cyclic components adequate to real business cycles and groups of methods based on the evaluation of regression model parameters (such as the Fourier series models and multifactor models of the dependence of the cyclicality of GDP on the dynamics of factor variables).

We consider that it is not the methods of extraction of business cycles, but the methods for the next step of business cycle analysis—the quantification of cycle's characteristics (such as its phases, amplitude, duration of the period, and others).

The results of grouping of methods for extraction of cyclic components from the time series of macroeconomic indicators in conditions of its primly seasonal adjusting are presented in **Figure 1**. They include the group of "turning point" methods: methods for spectral density analysis and their realization with band-pass filtering (the statistical tools that *pass frequencies within a certain range and reject frequencies outside that range*) and methods for determining and subsequently evaluating cyclic components. We will not dwell on detrending methods in detail, since they are detailed in the articles, including in connection with the algorithms for isolating and quantifying the cycles (e.g., Canova [4]).

http://dx.doi.org/10.5772/intechopen.75580

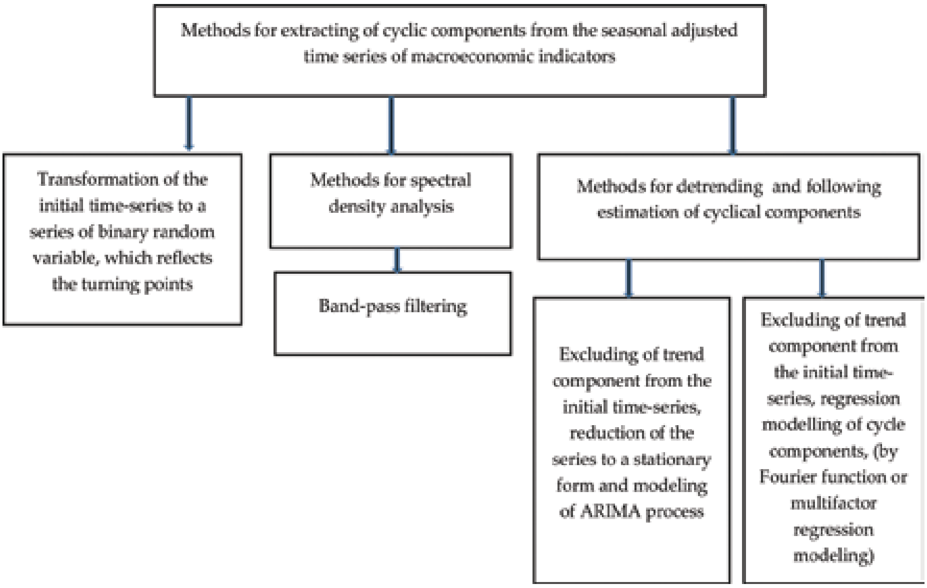

Figure 1. System of methods for extraction of cyclic components from the seasonal adjusted time series of macroeconomic indicators.

The above methods are detailed and combined in different documents, but we must take into account that "The crucial question is not which method is more appropriate but whether different concepts of cycle are likely to produce alternative information which can be used to get a better perspective into economic phenomena and to validate theories" (Canova [4]).

The purpose of this paper is to create and demonstrate the results of the application of the algorithm of business cycle synchronization and harmonization proceed from economic content definition of business cycle by Burns and Mitchell and methods of estimation of economic fluctuations presented in publications. The main idea of it is based on the two key features in Burns and Mitchell's definition of business cycles. They are determined in many articles in the same variant. For example, in the paper of Diebold and Rudebusch ([8], p. 1), we can read "The first (key feature) is the comovement among individual economic variables…In their analysis, Burns and Mitchell considered the historical concordance of hundreds of series, including those measuring commodity output, income, prices, interest rates, banking transactions, and transportation services…The second prominent element of Burns and Mitchell's definition of business cycles is their division of business cycles into separate phases or regimes."

We believe that the first point should be based on another focus of "aggregate economic activity" in determining the business cycles of Burns and Mitchell: not only as a combination of the abovementioned indicators characterizing aggregate economic activity but above all the multiple effect of the entry of various economic activities into one and the same phase of the business cycle (the second key features in Burns and Mitchell's definition of business cycles).

Consequently, at the macroeconomic level, the economic cycle is an integral result of business cycles of different economic activities that are at different or equal (to a greater or lesser extent) phases of the "common"[1] cycle.

According to this idea, we included following two points in the practical part of this article:

1. Explanation of the applied algorithm for measuring the business cycle in relation to different types of economic activities by countries and conducting cross-country analysis.

2. Interpretation of synchronization and harmonization of business cycles as economic definitions and their evaluation at the macroeconomic level.

3. Algorithm for the quantification of business cycles, their synchronization and harmonization, and its application

Measuring business cycles is necessary and usually the starting point of their research. Measurement in this context means quantifying the following characteristics of a business cycle (2002):

- The duration of the cycle and its phases
- The amplitude of the cycle and its phases
- Any asymmetric behavior of the phases
- Cumulative movements within phases

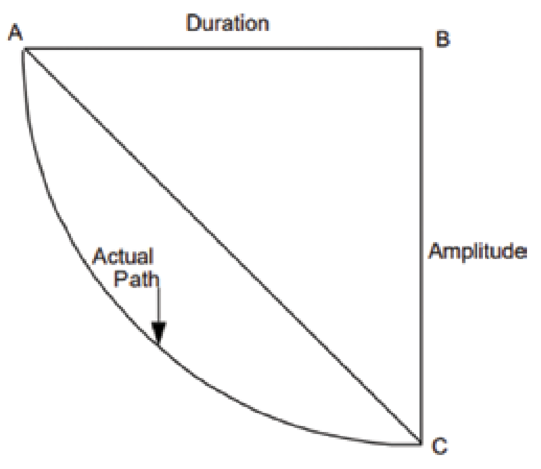

Figure 2. Stylized recession phase (Harding and Pagan [16]).

[1]The idea of the "common cycles" is presented by A. S. Blinder and S. Fischer [2].

http://dx.doi.org/10.5772/intechopen.75580

Harding and Pagan [16] graphically showed and explained the ratio of the peak (A) and the trough (C) as an example of a stylized recession (**Figure 2**). The height of the triangle in this graph is the amplitude, and the base is the duration. In this article, the authors also rightly notice that the knowledge of these two elements for any cycle makes it possible to calculate the area of the triangle and thereby estimate (say) the total loss in the output from the peak to the trough.

For the purposes of quantification of the synchronization and harmonization of business cycles on the macroeconomic level and further cross-country comparison, we suggest the following algorithm, presented in **Figure 3**.

At the first stage of our algorithm, we extract trend, cyclical, and irregular components of the initial seasonal and calendar adjusted time series of quarterly indicators of gross value added by the types of activity (sectors) by country. Examples of implementation of the proposed algorithm of band-pass filtering and the following analysis are based on the seasonal and calendar adjusted time series of quarterly indicators of gross value added by the types of activity (sector) represented by Eurostat[2] and Rosstat[3]. The indicators used are presented in fixed prices and converted to the natural logarithm, so they reflect the relative growth of the value added, forming the output of gross domestic product (GDP).

According to the approach of Baxter and King [1], the ideal band-pass filter should satisfy the six requirements:

1. The filter must extract the specified range of periodicity, which means that it passes through time series components with periodic oscillations between low and high frequencies, defining a specific cycle. Baxter and King recommend a filter that approximates periodic oscillations between 6 and 32 quarters (according to the definition of the Mitchell cycle).
2. An ideal band-pass filter should not introduce a phase shift, i.e., do not change the time stamps of the turning points at any frequency.
3. The expression of the discrepancy between the exact and approximate filters should be expressed mathematically.
4. The application of the band-pass filter must extract of a deterministic trend from a time series and result in a stationary time series, even when applied to trending data.
5. The filter should allocate the same components of the business cycle, regardless of the length of the observation period. Baxter and King notes that "Technically, this means that the moving averages we construct."
6. Method of filtering must be operational.

[2]Eurostat: http://ec.europa.eu/eurostat
[3]Russian Federal State Statistics Service (Rosstat): http://www.gks.ru

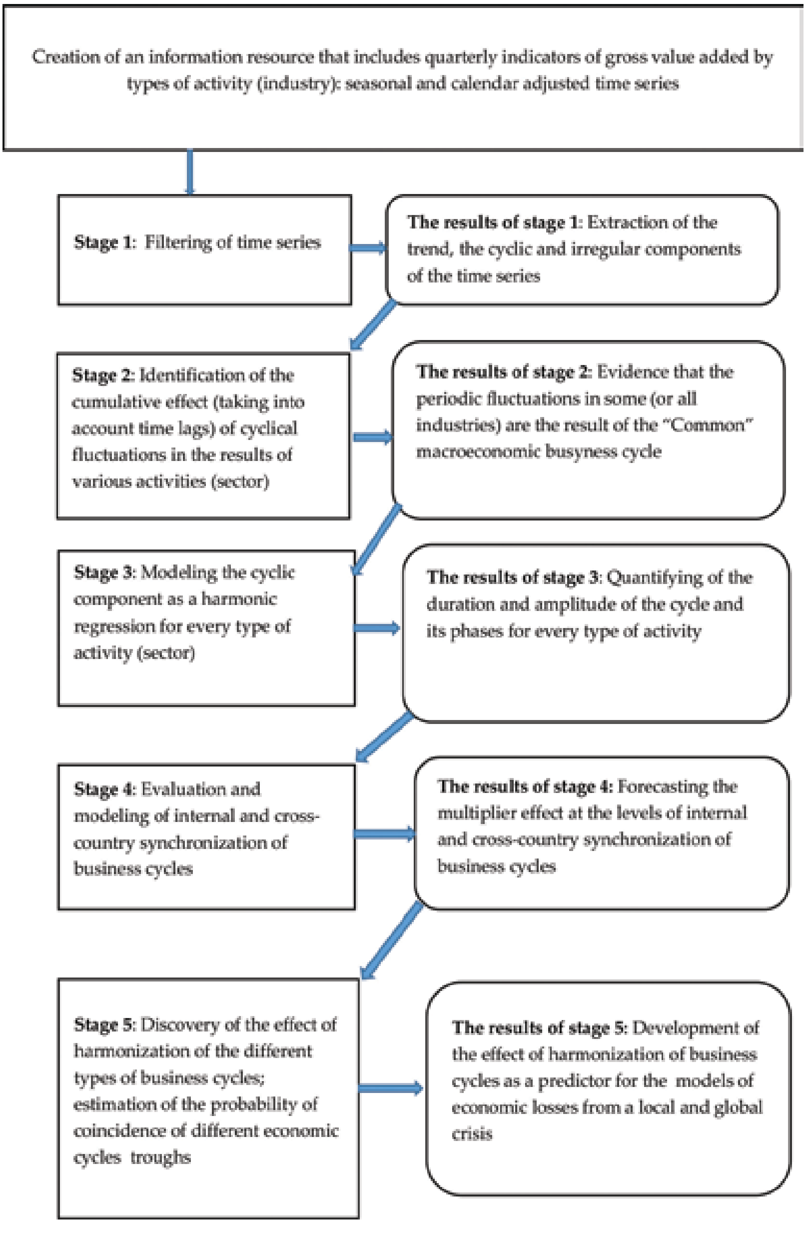

Figure 3. An algorithm for quantifying business cycles and their synchronization and harmonization at the macroeconomic level.

We applied two filters for calculations, which are widely represented in the economic literature: the Hodrick-Prescott (HP) filter and the Baxter-King (BK) filter. These filters, calculated in the "mFilter" package of R-CRAN [26], largely meet the above requirements.

Figure 4 shows that HP and BK filters on quarterly data give very close results.

http://dx.doi.org/10.5772/intechopen.75580

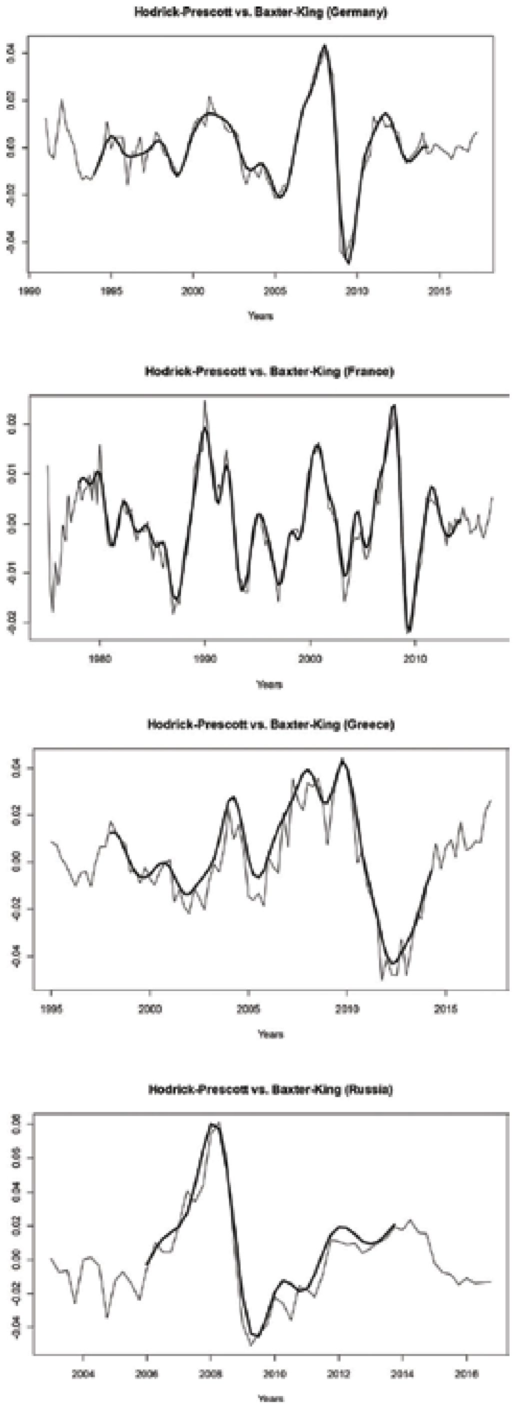

Figure 4. Comparison of business cycles extracted by filters HP (gray line) and BK (black line): Germany, France, Grease, and Russia.

The difference of the application of HP and BK filters is opened by Baxter and King [1]. As they note, the BK filter is easier in design and gives more exact results for data sampled at other-than-quarterly frequencies. However, the HP filter provides a longer cycle curve, because it is not based on moving average calculation (we can see it on **Figure 3**: the black line is shorter than the gray line).

Because we use the quarterly data by country, the HP filter is more suitable for business cycle extraction as a basis of further analysis.

At **stage 2** of our algorithm, we will evaluate the delays and correlations between the "total" business cycle reflected by the GDP curve and the business cycles of various economic activities (sectors) presented by added value curves, and on this basis, we will estimate the "common" cycle as the result of aggregating specific cycles by sectors. This approach corresponds to Mitchell's definition of business cycle presented above in this paper.

According to our approach, the difference between "total" and "common" cycle consists in the following:

"Total" cycle is a cyclical curve of any macroeconomic statistical indicator chosen for the purposes of measurement and analysis of business cycle (usually GDP or in our case the gross value added as a sum of added value by sectors (analogue of GDP). "Common" business cycle is a result of aggregation of different phases of sector's business cycles. The modern scientific problem is measurement of leading or lagging sector's business cycles relatively "common" cycle.

Coincidence of "total" and "common" cycles confirms the correctness of choosing the statistical indicator for measurement of business cycle on the macroeconomic level.

The results are presented by two countries (Germany and Russia) because of the limitation of an article volume.

The results presented in **Table 1** make it possible to conclude that in Germany the sectors enter the general cycle not simultaneously. Analyzed types of activities can be divided into three groups:

1. Industry production, information, and communications are the "leading" activities. Their phases are two quarters earlier as the same phases of the "total" cycle, reflected by fluctuations of the gross value added indicator. We can name them as "pro-cyclical," because they have the same (parallel) phases and shift peaks, relatively the peaks of "total" cycle in time (also as a shift of trough).

2. Agriculture, forestry and fishing, real estate activities, public administration, defense, education, health and social work are the economic activities which have lagging and "counter-cyclical" phases relatively the "total" cycle.

3. Construction, wholesale and retail, transport, accommodation and food service activities, financial and insurance activities, professional, scientific and technical activities and arts, entertainment and recreation, and other types of activities show simultaneous entry into the same phases with a "total" cycle. As the first group, these activities are "pro-cyclical."

Similar analysis for Russia (**Table 2**) led to the conclusion that the majority of sectors (D, F, G, H, I, J, K, N, O) show simultaneous and "pro-cyclical" oscillations with the "total" cycle. Such

http://dx.doi.org/10.5772/intechopen.75580

Types of activity (sectors) according to NACE2	Code	Lags of correlation with the "total" cycle in quarters				
		Lag = 0	Lag = −2	Lag = −4	Lag = −6	Lag = −8
Agriculture, forestry, and fishing	A	–	–	−0.343	−0.221	–
Industry (except construction)	B–E	0.840	0.855	0.320	–	−0.310
Manufacturing	C	0.857	0.863	0.311	–	−0.338
Construction	F	0.492	0.487	–	–	–
Wholesale and retail, transport, accommodation, and food service activities	G–I	0.803	0.802	–	−0.265	−0.419
Information and communication	J	0.568	0.573	–	–	−0.281
Financial and insurance activities	K	0.545	0.386	0.231	–	−0.293
Real estate activities	L	–	–	−0.266	−0.316	−0.307
Professional, scientific, and technical activities	M–N	0.680	0.680	–	–	−0.418
Public administration, defense, education, human health, and social work	O–Q	–	–	−0.321	−0.307	–
Arts, entertainment, and recreation: other service activities	R–U	0.479	0.471	–	–	–

Note: Dash indicates a statistically insignificant correlation

Table 1. Spearmen correlation coefficients of "total" business cycle estimated by gross value added and business cycles on different types of activity (sectors) with different lags of influence (in quarters of year) (Germany).

actions as "mining" and "public administration and defense and social security," demonstrate a statistically significant lagging "countercyclical" effect.

On the basis of the results of the previous stage of analysis, we can create a "common" cycle model as an aggregate of business cycles of various types of activity, taking into account their correlation and lagging effect interconnection with "total" cycle, which in this case plays the role of "connection bridge" between the business cycles of different types of activity and "common" cycle as a latent variable.

This conceptual approach determines the possibility of solving the problem of "Common" cycle estimation with the methodology of principle component analysis.

The factor loadings to "common" cycle are presented in **Table 3**. This and the following tables are the copies of the reports on the performed calculations in the software package of statistical analysis "STATISTICA," which gives an opportunity to imagine the practical side of the research (**Tables 3–5**).

The general cycle corresponds to the first principle component, since it is characterized by significant statistical relationships[4] with most types of activity, i.e., manifests its effect of the aggregated cycle by Mitchell's definition of business cycle. In both cases the first principle component will explain about 40% of the variance of the initial feature space.

[4]The statistically significant correlations are marked in red.

Types of activity (sectors) according to NACE2	Code	Lags of correlation with the "total" cycle in quarters				
		Lag = 0	Lag = −2	Lag = −4	Lag = −6	Lag = −8
Agriculture, hunting, and forestry	A	–	–	−0.381	−0.378	−0.295
Fishing and fish farming	B	–	0.281	–	–	–
Mining	C	–	–	–	0.437	0.598
Manufacturing	D	0.791	0.754	0.421	–	–
Production and distribution of electricity, gas, and water	E	–	–	–	–	0.313
Development	F	0.853	0.589	–	–	−0.423
Wholesale and retail trade, repair of motor vehicles, motorcycles, household goods, and personal items	G	0.901	0.670	0.231	–	−0.293
Hotels and restaurants	H	0.864	0.598	–	–	-0.493
Transport and communications	I	0.732	0.644	0.391	–	–
Finance	J	0.683	0.415	–	–0.337	–0.656
Real estate, renting, and business activities	K	0.531	–	–	–0.457	–0.744
Public administration and defense and social security	L	–	−0.309	−0.508	−0.563	−0.551
Education	M	–	–	−0.307	–0.336	–
Health and social services	N	0.663	0.349	–	–	–0.304
Other community and social and personal services	O	0.635	0.382	–	–	–0.512

Table 2. Spearmen correlation coefficients of "total" business cycle estimated by gross value added and business cycles on different types of activity (sectors) with different lags of influence (in quarters of year) (Russia).

In **Figures 5** and **6**, we can see the difference in the adequacy of the reflection of the business cycle in Germany and in Russia with the statistical indicator "gross value added" (accounting for almost 100% of GDP produced) (but in both cases, there are high correlation coefficients: 0.954 and 0.837).

Modeling the cyclic component as a harmonic regression on **stage 3** of our algorithm is necessary for quantifying the duration and amplitude both the "common" and specific cycles of different types of activity (sectors). For these purposes, it is expedient to use a Fourier series —the periodic function with a finite number of elements.

The business cycles extracted with the HP filter in accordance with the expansion of the Fourier series can be represented as a periodic function of time ($\widehat{y}_t$) in order to decrease the number of allocated harmonics, (i):

$$\widehat{y}_t = a_0 + \sum_1^i \left\{a_{ij}\sin\left(b_{ij}t + k_{ij}\right) + a_{ij}\cos\left(b_{ij}t + k_{ij}\right)\right\}, \quad (2)$$

where $\widehat{y}$ is the harmonic model of the business cycle

a_0 , a_{ij}, b_{ij}, and k_{ij} are the parameters of the harmonic model of the business cycle

j is the number of parameters

Factor loadings (Germany)

Variable	Factor 1	Factor 2	Factor 3	Factor 4
A_HP_cycle	0,198073	0,785048	0,129719	-0,108655
B-E_HP_cycle	-0,863405	-0,334347	0,075096	-0,175139
C_HP_cycle	-0,877166	-0,340340	0,058312	-0,160360
F_HP_cycle	-0,501404	-0,328351	0,375376	0,463128
G-I_HP_cycle	-0,801096	0,048955	-0,075760	-0,327621
J_HP_cycle	-0,581831	0,627588	-0,213228	0,073480
K_HP_cycle	-0,478678	0,384939	-0,435818	0,335469
L_HP_cycle	0,083806	0,161735	0,298852	-0,783388
M_N_HP_cycle	-0,856272	0,211469	0,024162	-0,022431
O-Q_HP_cycle	0,126418	0,207386	0,804045	0,209514
R-U_HP_cycle	-0,508612	0,348817	0,434247	0,151990
Expl.Var	4,029837	1,731736	1,332873	1,189153
Prp.Totl	0,366349	0,157431	0,121170	0,108105

Factor loadings (Russia)

Variable	Factor 1	Factor 2	Factor 3	Factor 4
A_HP_cycle	0,094444	0,080303	-0,210250	0,508336
B_HP_cycle	0,070069	-0,345123	0,437569	0,650009
C_HP_cycle	0,136829	-0,796738	0,222348	-0,217219
D_HP_cycle	0,876800	-0,273916	0,130628	-0,069699
E_HP_cycle	0,269795	-0,829359	-0,119065	-0,051618
F_HP_cycle	0,891412	0,095851	-0,064286	0,230713
G_HP_cycle	0,881254	0,111121	0,265041	0,096117
H_HP_cycle	0,922355	0,185863	0,052287	0,108052
I_HP_cycle	0,895478	-0,041845	-0,074617	-0,124972
J_HP_cycle	0,618859	0,466097	0,364537	0,339480
K_HP_cycle	0,581893	0,675720	-0,096186	0,231458
L_HP_cycle	-0,008477	0,251555	0,110716	0,740297
M_HP_cycle	-0,014588	0,115452	-0,892138	0,064979
N_HP_cycle	0,484983	0,273436	0,582791	0,126600
O_HP_cycle	0,838983	0,139996	-0,161603	0,178450
T_HP_cycle	0,970514	-0,156671	0,052845	0,015257
S_HP_cycle	0,106961	0,280366	-0,147597	0,877965
NT_HP_cycle	0,959853	-0,178622	0,127146	-0,032557
Expl.Var	7,633587	2,560923	1,755843	2,366452
Prp.Totl	0,424088	0,142274	0,097547	0,131470

Table 3. Factor loadings of specific cycles of the types of activity to "common" cycle latent variable (with lagging effect of influence).

A. Parameters of the harmonic model of the "Common" business cycle

Dep. Var. : Com
Level of confidence: 95.0% (alpha=0.050)

	Estimate	Standard error	t-value df = 93	p-value	Lo. Conf Limit	Up. Conf Limit
a0	-0,0555	0,048280	-1,1504	0,252934	-0,1514	0,0403
a1	0,5742	0,070292	8,1691	0,000000	0,4346	0,7138
b1	0,4807	0,004161	115,5311	0,000000	0,4724	0,4889
k1	-18,5253	0,252296	-73,4268	0,000000	-19,0263	-18,0243
a2	-0,5185	0,071880	-7,2128	0,000000	-0,6612	-0,3757
b2	0,3876	0,005044	76,8433	0,000000	0,3776	0,3976
k2	-4,3764	0,301399	-14,5204	0,000000	-4,9750	-3,7779
a3	0,9591	0,070957	13,5167	0,000000	0,8182	1,1000
b3	0,2391	0,002912	82,1234	0,000000	0,2333	0,2449
k3	-1,7077	0,173945	-9,8172	0,000000	-2,0531	-1,3622
a4	0,7546	0,069935	10,7905	0,000000	0,6158	0,8935
b4	0,3217	0,004306	74,7113	0,000000	0,3132	0,3303
k4	-14,5931	0,241599	-60,4019	0,000000	-15,0728	-14,1133

Multiple correlation coefficient: R=0.892

B. Parameters of the harmonic model of the Manufacturing type of activity, (sector "C"), business cycle

Dep. Var. : C
Level of confidence: 95.0% (alpha=0.050)

	Estimate	Standard error	t-value df = 93	p-value	Lo. Conf Limit	Up. Conf Limit
a0	-0,008	0,003387	-0,2231	0,823970	-0,0075	0,0060
a1	-0,315	0,004729	-6,6675	0,000000	-0,0409	-0,0221
b1	3,091	0,006163	50,1595	0,000000	0,2969	0,3214
k1	-88,83	0,391287	-22,7020	0,000000	-9,6600	-8,1060
a2	0,062	0,004673	1,3285	0,187255	-0,0031	0,0155
b2	0,876	0,027448	3,1903	0,001938	0,0331	0,1421
k2	24,191	1,777769	1,3607	0,176888	-1,1112	5,9494
a3	-0,265	0,004596	-5,7744	0,000000	-0,0357	-0,0174
b3	3,896	0,006292	61,9178	0,000000	0,3771	0,4021
k3	-148,036	0,392890	-37,6786	0,000000	-15,5838	-14,0234
a4	0,263	0,004761	5,5150	0,000000	0,0168	0,0357
b4	2,352	0,008097	29,0474	0,000000	0,2191	0,2513
k4	-88,399	0,480367	-18,4024	0,000000	-9,7938	-7,8860

Multiple correlation coefficient: R=0.730

Table 4. Parameters of the harmonic models: (**A**) of the "common" business cycle and (**B**) of the manufacturing type of activity (sector "C") business cycle with estimations of their significance level *(p-value)* and confident intervals; Germany.

A. Parameters of the harmonic model of the "Common" business cycle

Dep. Var. : Com
Level of confidence: 95.0% (alpha=0.050)

	Estimate	Standard error	t-value df = 49	p-value	Lo. Conf Limit	Up. Conf Limit
a0	-0,00379	0,060613	-0,0625	0,950428	-0,12559	0,11802
a1	0,61965	0,129919	4,7695	0,000017	0,35857	0,88074
b1	-0,19593	0,017425	-11,2443	0,000000	-0,23095	-0,16092
k1	9,01859	0,563702	15,9989	0,000000	7,88579	10,15139
a2	0,59223	0,129472	4,5742	0,000033	0,33205	0,85242
b2	0,26375	0,016224	16,2565	0,000000	0,23114	0,29635
k2	-5,00926	0,502540	-9,9679	0,000000	-6,01916	-3,99937

Multiple correlation coefficient: R=0.830

A. Parameters of the harmonic model of the Manufacturing type of activity, (sector "D"), business cycle

Model is: D=a0+a1*sin(b1*t+k1)+a2* cos(b2*t+k2) (RU_циклы_HP — копия
Dep. Var. : D
Level of confidence: 95.0% (alpha=0.050)

	Estimate	Standard error	t-value df = 49	p-value	Lo. Conf Limit	Up. Conf Limit
a0	-0,00047	0,004924	-0,0964	0,923559	-0,01037	0,00942
a1	-0,05082	0,006324	-8,0361	0,000000	-0,06353	-0,03811
b1	0,23708	0,010798	21,9558	0,000000	0,21538	0,25878
k1	-5,29867	0,335539	-15,7915	0,000000	-5,97296	-4,62438
a2	0,03601	0,006526	5,5179	0,000001	0,02290	0,04913
b2	0,39656	0,012630	31,3975	0,000000	0,37118	0,42194
k2	-7,88707	0,402672	-19,5868	0,000000	-8,69627	-7,07787

Multiple correlation coefficient: R=0.730

Table 5. Parameters of the harmonic models: (A) of the "common" business cycle and (B) of the manufacturing type of activity (sector "D") business cycle with estimations of their significance level (*p-value*) and confident interval; Russia.

The following tables (**Tables 4** and **5**) and figures (**Figures 7–10**) present the results of the assessment of harmonic models of "common" business cycles and the most closely related sectoral business cycles for Germany and Russia. We can see that the models as a whole and their parameters are significant.

The theory and methodology of measurement of the business cycle synchronization and harmonization can be divided into two groups: an approach based on the clustering of turning points and the evaluation of the concordance of business cycles.

For the implementation of **stages 3–5** of our algorithm (**Figure 3**), we suggest the next definitions of the synchronization and harmonization of business cycles in the general approach:

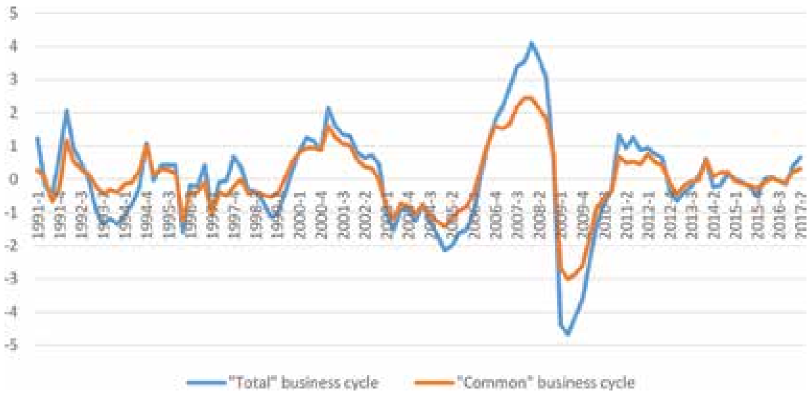

Figure 5. Business cycle graphs in Germany in "total" and "common" presentations.

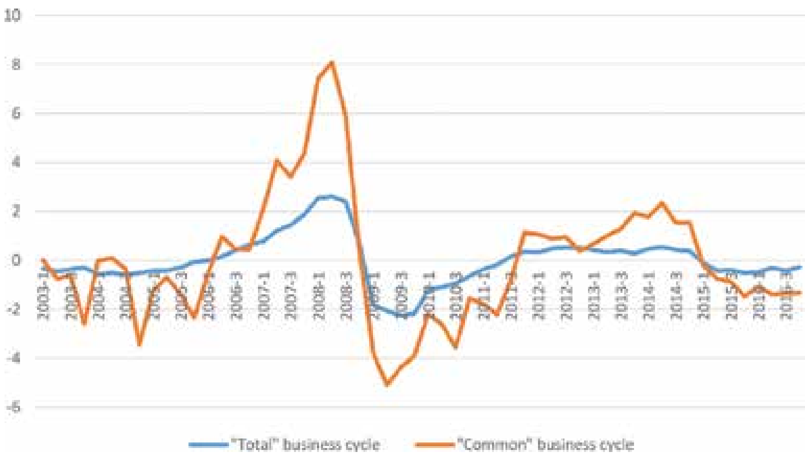

Figure 6. Business cycle graphs in Russia in "total" and "common" presentations.

- Business cycles are synchronous if their curves are parallel, with a possible shift in time.

For the comparison of two cycles, it can be described by the following mathematical conditions for any *i*-harmonics:

$$\begin{cases} \hat{y}_{it(1)} \cong \hat{y}_{it(2)} \\ T_{1,i} \approx T_{2,i}, \quad \text{for any } L_{2,i}\text{, where } L_{2,i} \text{ is any constant time shift } \hat{y}_{it(2)} \text{ relatively } \hat{y}_{it(1)}. \end{cases} \quad (3)$$

According to the theory of the Fourier analysis, for any business cycle harmonic model, the length of the period of the *i*-order harmonic is T(i) = 2π/ (i):

- Business cycles are harmonized if the peaks (or troughs) of these cycles fall on the same time point. The coincidence of the minima of the harmonic waves of such superstrong

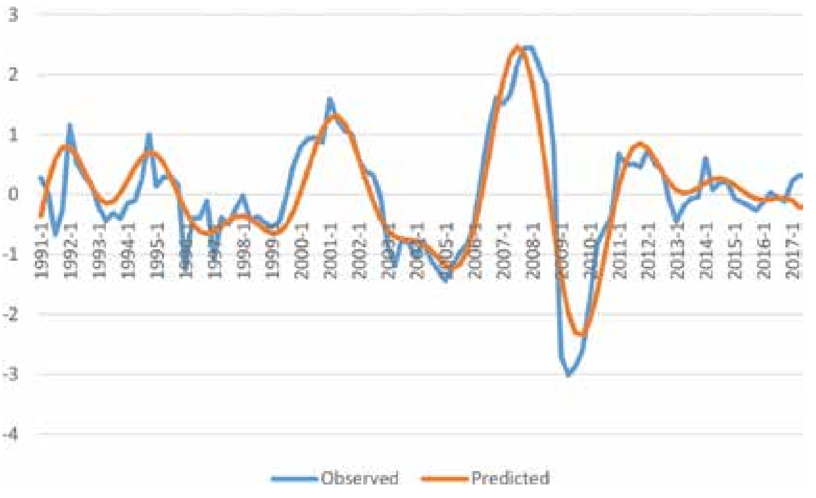

Figure 7. The "common" business cycle in Germany estimated above with principle components method (*observed*) and its harmonic model (*predicted*).

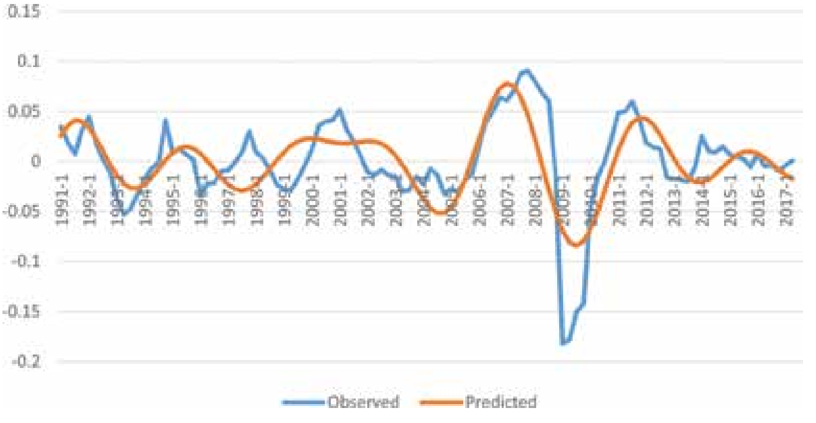

Figure 8. The manufacturing type of activity (sector "C") business cycle in Germany estimated above with principle components method (*observed*) and its harmonic model (*predicted*).

classical cycles as Kondratieff, Kitchin, Juglar, Kuznets, and others is the main cause of global economic crises, such as the 2007–2009 global financial and economic crisis or Great Recession [29]. Using the parameters of harmonic models of "common" and specific (by types of activity) cycles, we can estimate in-country and intercountry synchronization and harmonization of business cycles.

Using the parameters of harmonic models of "common" and specific (by types of activity) cycles, we can estimate in-country and intercountry synchronization and harmonization of business cycles.

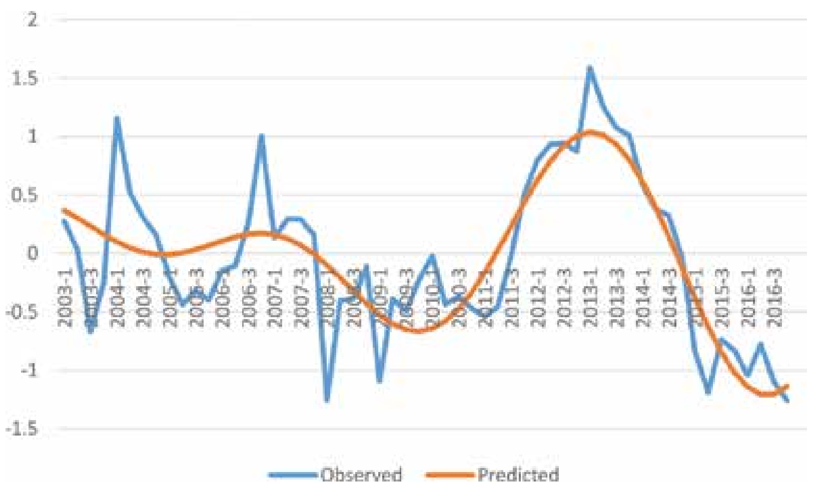

Figure 9. The "common" business cycle in Russia estimated above with principle components method (*observed*) and its harmonic model (*predicted*).

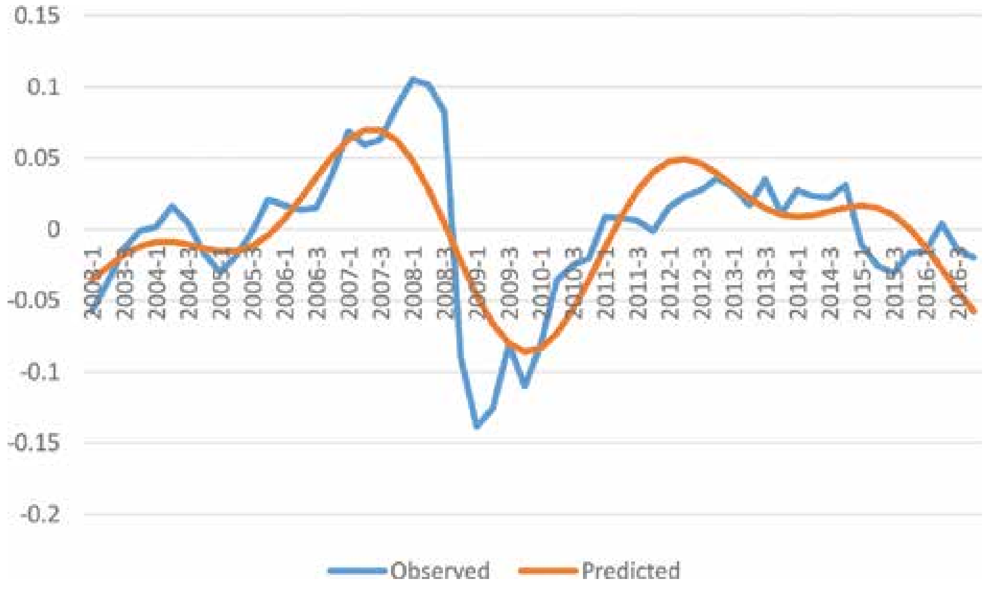

Figure 10. The manufacturing type of activity (sector "D") business cycle in Russia estimated above with principle components method (*observed*) and its harmonic model (*predicted*).

4. Conclusions

The approach proposed and tested in this chapter has a theoretical basis and originality. The theoretical basis of this approach is the classic definition of Mitchell's business cycle with an emphasis on its key position that the business cycle is an integrated multifactorial phenomenon. The effect of integration is understood as the interaction of business cycles of different types of activities (sectors), which are simultaneously at different phases of cyclicality.

The novelty of the approach is to assess the statistical relationship between the indicators of business cycles of sectors of the economy with a certain macroeconomic indicator characterizing the cyclical nature of the economy as a whole and representing a "total" cycle. There is an evaluation of some objective, but not directly measured, "common" cycle.

At the same time, the proposed algorithm takes into account the lagging effect of the mutual influence of business cycles, as well as their synchronization and harmonization.

Measurement of the effect of business cycle synchronization and harmonization is presented on the basis of the construction of harmonic models.

The above examples of calculations for Germany and Russia show the general and distinctive characteristics of the business cycles of these countries.

The results presented in the chapter can serve as a basis for further research in both theoretical and applied aspects. The main areas should be the development of methods for forecasting the entry of the economy into different phases of cyclicality and the expansion of groups of countries for analysis in-country and intercountry interaction of cycles.

Author details

Elena Zarova

Address all correspondence to: zarova.ru@gmail.com

Plekhanov Russian University of Economics, Moscow, Russia

References*

[1] Baxter M, King RG. Measuring Business Cycles Approximate Band-Pass Filters for Economic Time Series. Cambridge: National Bureau of Economic Research; 1995

[2] Blinder AS, Fischer S. Inventories, Rational Expectation, and the Business Cycle. Cambridge: National Bureau of Economic Research; 1979

[3] Burns AF, Mitchell WC. Measuring Business Cycles. New York, New York: National Bureau of Economic Research; 1946

[4] Canova F. Detrending and business cycle facts. Journal of Monetary Economics. 1998;**41**: 475-512

*Refs. 5, 6, 7, 9–15, 19, 21, 22, 24, 25, 27, 28 are recommended for reading for deep understanding of the topic of the article.

[5] Chin D, Geweke J. Predicting Turning Points Federal Reserve Bank of Minneapolis and Congressional Budget Office; 2000

[6] Christiano LJ, Fitzgerald TJ. The Band Pass Filter. Cambridge: National Bureau of Economic Research; 1999

[7] Cogley T. Data Filters. University of California; 2006

[8] Diebold FX, Rudebusch G. Measuring Business Cycles: A Modern Perspective. Review of Economics and Statistics. 1996;**78**:67-77

[9] Diebold FX, Rudebusch GD. Stochastic Properties of Revisions in Index of Leading Indicators. In: Proceeding of the Business and Economic Statistics Section. American Statistical Association; 1988. pp. 712-717

[10] Dudukovic S. Harmonic Analysis of the Real Business Cycles. Switzerland: Franklin College; 2007

[11] Dueker M, Wesche K. European Business Cycles: New Indices and Analysis of their Synchronicity. Federal Reserve Bank of St. Louis., Institute for International Economics, University of Bonn; 2001

[12] Extracting Business Cycles From Raw Data in R. https://econometricswithr.wordpress com/2015/10/27/extracting-business-cycles-from-raw-data-in-r/

[13] Fourier Transform: A R Tutorial. http://www.di.fc.ul.pt/~jpn/r/fourier/fourier.html

[14] Grigorą V, Stanciu IE. New Evidence on the Synchronization of Business Cycles: Is there an European Business Cycle? National Bank of Romania, Macroeconomic Modelling and Forecasting Department, Bucharest University of Economic Studies; 2015

[15] Günter GG, Lorenz HW. (1989) Business Cycle Theory: A Survey of Methods and Concepts. Springer-Verlag Berlin Heidelberg

[16] Harding D, Pagan A. Dissecting the cycle: A methodological investigation. Journal of Monetary Economics. 2002;**49**:365-381

[17] Harding D, Pagan A. Synchronization of Cycles. The University of Melbourne, Victoria 3010, Australia b The Australian National University and The University of New South Wales, Australia. Available online from: 5 March 2005

[18] Harding D, Pagan A. Measurement of Business Cycles. The University of Melbourne. Department of Economics; 2006

[19] Kedem B. Time Series Analysis by Higher Order Crossings. Department of Mathematics & ISR University of Maryland College Park, MD

[20] Kijek A. Spectral analysis of business cycles in the Visegrad group countries. Comparative Economic Research. 2017;**20**(2):53-71

[21] Kydland FE, Prescott EC. Time to build and aggregate fluctuations. Econometrica. 1982; **50**(6) (Nov., 1982):1345-1370. Published by: The Econometric Society

[22] Larsson G, Vasi T. Comparison of Detrending Methods. Department of Statistics, Uppsala University; 2012

[23] Lucas RE. Understanding Business Cycles. University of Chicago. http://icm.clsbe.lisboa.ucp.pt/docentes/url/jcn/mabes/LucasUnderstanding.pdf;1977

[24] Obradović S, Mihajlović V. Synchronization of Business Cycles in the Selected European Countries. Panoeconomicus. 2013;(6):759-773

[25] Package 'BC Dating'. https://cran.r-project.org/web/packages/BCDating/BCDating.pdf

[26] Package 'mFilter'. https://cran.r-project.org/web/packages/mFilter/mFilter.pdf

[27] Schirwitz B, Wälde K. Synchronization of Business Cycles in G7 and EU14 countries. University of Dresden and European Commission. Schirwitz@iwb-dresden.de, Klaus@waelde.com; 2004

[28] Stocka JH, Watsonb WM. Estimating turning points using large data sets. Journal of Econometrics. January 2014;**178**(2):368-381

[29] Zarova E. Spatial harmonization of economic cycles: Statistical confirmation of European-Russian interaction in real sectors of the economy. Statistical Journal of the IAOS. 2013; **29**(4):281-287

Chapter 4

Statistical Research of Investment Appeal of Russian Regions

Burtseva Tatiana Alexsandrovna

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.75465

Abstract

In this chapter, the methodological results directed on realization statistical research of investment appeal of Russian regions are offered. Methodological basis of research is the method of the dynamic standard, index and the coefficient analysis and the method of paired comparisons. The results of the study: (1) the method of the dynamic standard for creation of statistical model of region investment appeal is offered; (2) the normative model of region investment appeal to measure the productivity of the realization of regions investment policy in Russia is created; (3) new factors of region investment appeal are investigated and (4) statistically valid conclusions are drawn and practical recommendations are made. The results of the study are addressed to the Ministry of Economic Development of the Russian Federation in order to justify the amount of federal targeted investment programs financing in Russian regions.

Keywords: factors of region investment appeal, efficiency of regional investment policy, the system of region investment appeal, the normative model of region investment appeal, integrated statistical indicator

1. Introduction

Today, the regional policy becomes a key factor of the successful solution to some problems of economy modernization, investment, innovative activity and population life quality improvement. For the last decades, the Russian regions stored a vast experience of the economic development problem solution, which are worth to be studied and generalized. The last institutional changes in world politics constrain rethinking and keeping up to date approaches to managing social and economic development of regions. It brings to the forefront the questions of measurement and analysis of the productivity of the public and municipal administration in the new conditions. In this regard, the development of the methodology of statistical research of the strengthening investment attractiveness of the Russian region using modern regional investment policy mechanisms is deemed relevant. Attention is drawn to taking into account the regional space heterogeneity, a wide range of factors that are not always unequivocally affect investment attractiveness, occasionally forced necessity to develop dynamic models on the basis of data from short-time intervals. In this regard, methodology of statistical research of investment appeal development of the Russian region is actually represented. The purpose is a success assessment converting by means of modern investment policy mechanisms in the regions. Therefore, methodological development for realization investment appeal statistical research of the Russian regions, and its results should provide carrying out a statistical assessment complex of realization productivity of regional investment policy. This policy must be adequate to the Russian state development statistics.

This chapter reveals the author ideas on solving the abovementioned issues, which, in her opinion makes them relevant not only for the theory but also for the practice of statistical research of the regional investment situation, including the terms of its development prospects.

The main result for statistical science is the normative model of region investment appeal to measure the productivity of the realization of regions' investment policy in Russia.

The main results of a research of the author for economic science: (1) the method of the dynamic standard for creation of statistical model of region investment appeal is offered and (2) new factors of region investment appeal are investigated.

The results of the study are addressed to the Ministry of Economic Development of the Russian Federation in order to justify the amount of federal targeted investment programs financing in Russian regions.

2. Literature review

The role of investment as a factor of economic growth was justified by English scientist-economist Keynes [1]. He was the first who developed a macroeconomic model, which established the relationship between investment, employment, consumption and income, thereby justifying the leading role of the state in regulating the market economy in a period of instability and crises [2]. According to Keynes, the investment activity in the country is primarily determined by the expected return on investment. The growth of savings by itself has no effect on these expectations and does not automatically lead to an increase in investment. Therefore, Keynes detected the purpose of government in the impact on the change in the volume of public investment and level of the marginal profitability of capital investments. In the current context within the economic theory, it is customary to believe that a volume of budget investment has a significant impact on the growth of gross domestic product (the main gauge of economic growth at the global level). This thesis is based on the multiplier effect, which was also suggested by Keynes. This thesis is the only theoretical justification of the need of state investment in the economy [3].

At the moment, many states and regions do not have the opportunity to develop their economies through public investment. This makes it necessary to attract foreign investment. Therefore, state policy includes the goal of creating a favorable investment climate for foreign investors. The investment climate of the region is recognized by the international community as one of the main characteristics of the success of its economy development; therefore, the scientific interest in measuring the region investment appeal is constantly growing. Experts of the World Bank in the report "A Better Investment Climate for Everyone" for 2005 define investment climate as "the many location-specific factors that shape the opportunities and incentives for firms to invest productively, create jobs, and expand".

Porter has proposed to understand country's competitiveness and investment appeal as the productivity of using its resources [4, 5]. According to the author, this must be taken into account when assessing the investment climate of the country and the region.

So far, the large number works of domestic and foreign scientists are devoted to investment subject and investment regional appeal. Also, a lot of reports on the carried-out practical researches and recommendations about investment policy improvement at different management levels were prepared.

The most important achievements of basic Russian researches in the field of regional economics are connected with Schools of Sciences by the academician Lvov (research of management efficiency problems) and academician Granberg (research of structural regions distinctions on formation and distribution of investment resources) [6], [29]. Methodology of statistical research of economic development asymmetry of the Russian regions developed by Burtseva [7], Zubarevich [8] and other authors. Investment appeal of regions and branches of economy was considered in works by Frenkel et al. [9].

The analysis of the works devoted to the theory of investment market and investment modeling allows to allocate two conditional research directions. The first direction includes research developing general provisions of economic theory, namely, research of economic balance and economic growth. Within the second direction, actually economical and statistical research of investment is carried out. Generalization and ordering of scientific research on various aspects of statistical measurement of regional investment appeal and development reveal existence of a whole complex of insufficiently studied problems in assessing productivity of regional investment policy realization.

We should note actual problems of statistical research of the region investment appeal: regional space heterogeneity, wide range of factors, need to use dynamic models on small time spans. To solve these problems, it is proposed to develop a normative model that allows to obtain a quantitative level of balance in the indicators relative to each other and to identify the slowing or rising indicators, which will provide an integral statistical measure of the region investment appeal.

Solving these problems will allow providing new quality of an important problem of the state strategic planning—monitoring of investment strategy and medium-term programs of social and economic development of subjects for the Russian Federation.

Methodological development offered by the author allows to overcome successfully listed problems and also develops the methodology of statistical research of regional investment appeal according to the concept of the Russian state statistics development, which makes them relevant not only for theory but also for the practice of statistical research of the investment regional situation, including in perspective of its development.

The proposed methodological developments expand the methodology of Kaplan and Norton [10]. They allow realizing the monitoring of region investment policy on the basis of an integral statistical measure of the region investment appeal. The quantitative level of the integral statistical measure is the proportion of the number of completed relationships between the growth rates of the actual indicators characterizing a particular investigated object to the number of given ratios in the normative model.

3. The method of the dynamic standard

The essence of the method of the dynamic standard is the formation in accordance with some objective (e.g., maximization of the company's profit, increasing the cost of equity, increasing regional investment appeal), groups of indicators characterizing the purpose and, to the greatest extent, reflecting the real state of the object of research in dynamics. The quantitative composition of the indicators should be no lower than the established (no less than 6 and no more than 25). The method of the dynamic standard is the procedure of selection of economic indicators and their ordering.

The main idea of the method belongs to Syroezhin [11], it was further developed by his students [12]. It consists of the fact that not commensurable indicators in statics become commensurable in dynamics.

In modern scientific works and publications, there are examples of the application of the method of dynamic standard [13–17].

Syroezhin noticed that noncomparable static characteristics of the national economy are comparable in dynamics. The proposed dynamic standard is organized by pace (coefficients, indexes) growth (or base chain) set (system) of indicators, such that maintaining for a long-time interval specified in a dynamic normative order of indicators provides the maximization of integral evaluation. Form of expression the ordering of the indicators is the ranking of performance (assigning grades), if not all indicators are able to link strictly in order, the presentation days to serve the count of preferences and/or the corresponding matrix of preferences, in this case, the integral meter has the form of a normative model. The quantitative level of the integrated meter (integrated assessment) in this case is the ratio of the number of performed correlations between growth rates (indices) of growth of actual indicators characterizing the specific object under study, to the number of set relations in a normative model. Accordingly, the resulting quantitative levels vary in the range from 0 to 1, the closer the value is to 1, the more quantitative is the valuation level.

The algorithm for constructing normative models is disclosed in detail in the work of Pohostinsky [18]. The normative model differs from the matrix of preferences and its indicators are categorized by the transitive property, and this allows to obtain always a single measurement result.

Advantages of the method over other methods of building an integrated measure, for example, multidimensional average, are as follows:

1. it provides a complex convolution of directly incommensurable indicators, reflecting different aspects of the functioning of the economic system (presented in ordinal scale of measurement);
2. it makes possible to obtain integral evaluation, which is an integral meter, which characterize the system by studying the property as a whole taking into account the relationship of phenomena within it;
3. the normative model, each figure retains its own role and there is no effect of cancelation "positive" and "negative" changes captured by different indicators;
4. it has a high information capacity, which allows to obtain a dynamic integrated assessment based on a large number of indicators at a short time sample of observations, for example, only two periods to build an evaluation and three periods to sort the indicators and their groups (factors), its relative growth.

4. Research algorithm

4.1. Step 1: indicators of region investment appeal

The classical concept of "*investment appeal*" means the existence of certain investment conditions that affect the investor's goals and determine its choice when considering investment objects. The efficiency of the investment policy of the region is characterized by the degree of achievement of long-term goals of its development, the reproduction of the potential of the region and the growth of the quality of life of the population. Therefore, from our point of view, region investment appeal (RIA) is a complex of natural, geographical and socioeconomic factors that determine the effectiveness of the investment policy of the region and its socioeconomic development.

For the statistical research of investment appeal of Russian regions is proposed to use an integrated statistical indicator, which is based on measures that assess the "state of capacity development in the region" and the influence of the factors such as "the performance of business activities in the region" and "the performance of activities of public administration bodies in the region." **Figure 1** shows the structure of regional investment appeal. This interpretation is based on the scientific idea that the RIA is determined not only by the factors of the investment potential of the region, but also by the factors of the effectiveness of its investment policy.

Statistical estimation of the region potential is a traditional task of statistical measurement. The region potential includes components such as natural and geographical potential, property potential, financial potential, human potential and innovative potential. New for statistical measurement of its factors are "performance of government in the region" and "performance of business in the region."

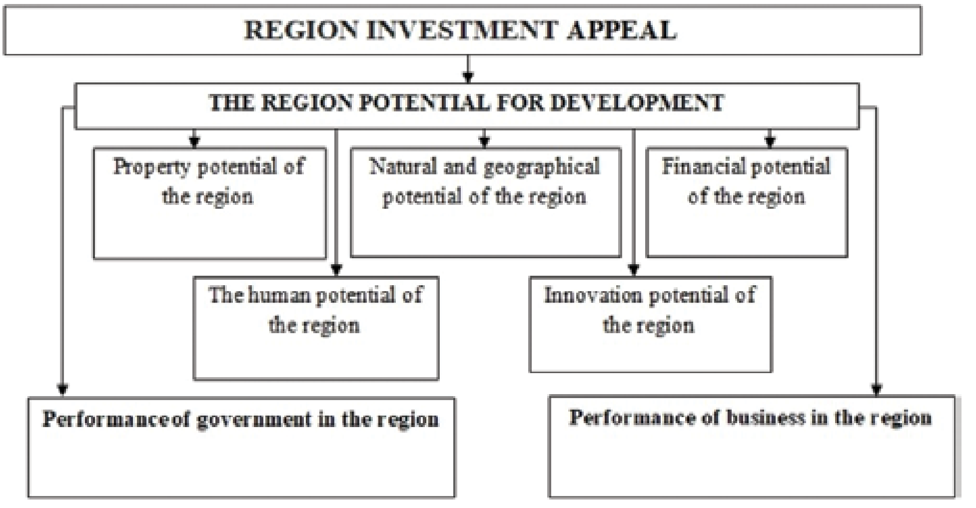

Figure 1. The structure of the region investment appeal as an object of statistical study.

Statistical assessment and monitoring of these factors are new tasks of statistical research. Their solution requires improvement and refinement of the methodology of statistical research of RIA in the search for adequate scientific methods of assessment, the principles of the formation of its information basis.

The study of existing methods for estimating RIA (the methodology of the agency "Universe," the agency "Expert-RA" and others) made it possible to find out that a large number of statistical indicators are used for its construction. The solution of such a problem became possible in the framework of the system approach. Therefore, the region was considered by us as a socio-economic system, and its investment attractiveness as its system-wide property, for the study of which a system of statistical indicators was formed. It consists of indicators that assess the potential and the results of the development of the region, which characterize the identified factors of the RIA. Absolute indicators selected in the system are presented in **Table 1**.

4.2. Step 2: the normative model of region investment appeal

The dynamic standard and the normative model of region investment appeal are presented in **Tables 2** and **3**. Formalization of the dynamic standard was tested using pairwise comparisons in accordance with targets of research (see **Table 2**). If, in accordance with the target installation rate in the row of the matrix needs to grow faster than the rate in the column below target was performed setting the "growth" that is put in matrix 1 at the intersection of row and column, while the symmetrical choice is −1. Otherwise, the −1, while the symmetric place puts 1. If relationship between the indicators is not set, then put a zero, the matrix diagonal has only zeros. Thus, in the matrix set 49 targets.

Formally, the dynamic standard of preferences is set by a matrix ($E = \{e_{ij}\}_{nxn}$), each element of which reflects the normative relation between performance (faster/slower) of the row and column of the matrix (see **Table 2**).

No.	Indicator name
Factor "Potential for development of the region"	
1	Population in the region
2	Value of fixed assets in the region
3	Residual value of fixed assets of the region
4	Internal costs for research and development of the region
5	Investments in fixed assets (Capex)
Factor "Performance of government in the region"	
6	*Number of employed in the economy of the region*
7	Incomes of the population of the region
8	Total number of unemployed in the region
9	Fund for remuneration of workers in the region
10	Gross regional product (GRP)
11	Tax revenues of the consolidated budget of the region
12	Revenues of the consolidated budget of the region
13	Expenditures of the consolidated budget of the region
Factor "Performance of business in the region"	
14	Balanced financial performance of organizations in the region
15	Number of employees employed in small enterprises in the region
16	Number of small enterprises in the region
17	Number of organizations in the region
18	The number of unprofitable organizations in the region

Table 1. Absolute indicators of RIA (Pi).

The matrix E is described as follows:

$$e_{ij}\begin{cases} 1, \text{if } GR(P_i) > GR(P_j)\ ; \\ -1, \text{if } GR(P_i) < GR(P_j)\ ; \\ 0, \text{if the reference ratio between } GR(P_i) \text{and } GR(P_j), \end{cases} \quad (1)$$

where i, j are the numbers of indicators in DS; Pi, Pj are indicators having the i-th and j-th numbers in DS, respectively; GR(Pi) > GR (Pj) and GR (Pi) < GR (Pj) are reference of ratio between rates (indices) of growth.

No and indicator name in DS, Pi	No indicator in DS																	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1. Population in the region	0	−1	−1	−1	0	0	−1	−1	−1	−1	−1	0	−1	−1	0	0	−1	1
2. Incomes of the population of the region	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
3. Number of employed in the economy of the region	1	0	0	−1	1	−1	−1	−1	−1	0	0	0	0	−1	0	0	−1	0
4. Number of employees employed in small enterprises in the region	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
5. Total number of unemployed in the region	0	0	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6. Fund for remuneration of workers in the region	0	−1	1	0	0	0	−1	0	0	0	−1	0	0	0	−1	0	0	0
7.GRP	1	0	1	0	0	1	0	1	1	0	0	0	−1	1	−1	1	1	1
8. Value of fixed assets in the region	1	0	1	0	0	0	−1	0	−1	0	1	0	0	−1	0	0	0	0
9. Residual value of fixed assets of the region	1	0	1	0	0	0	−1	1	0	0	1	0	0	−1	0	0	0	0
10. Number of small enterprises in the region	1	0	0	−1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
11. Number of organizations in the region	1	0	0	0	0	1	0	−1	−1	−1	0	1	−1	−1	−1	1	0	0
12. Number of unprofitable organizations in the region	0	0	0	0	0	0	0	0	0	0	−1	0	0	0	0	0	0	0
13. Internal costs for research and development of the region	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
14.Capex	1	0	1	0	0	0	−1	1	1	0	1	0	0	0	−1	0	0	0
15. Balanced financial performance of organizations in the region	0	0	0	0	0	1	1	0	0	0	1	0	0	1	0	0	0	0
16. Tax revenues of the consolidated budget of the region	0	−1	0	0	0	0	−1	0	0	0	−1	0	0	0	0	0	1	1
17. Revenues of the consolidated budget of the region	1	0	1	0	0	0	−1	0	0	0	0	0	0	0	0	−1	0	1
18. Expenditures of the consolidated budget of the region	−1	0	0	0	0	0	−1	0	0	0	0	0	0	0	0	−1	−1	0

Table 2. The dynamic standard (DS) of region investment attractiveness.

Thus, formed matrix E, after identifying additional relationships, is a normative model (**Table 3**).

Reflection of the results of the adopted and implemented managerial decisions is the actual relation matrix of indicators Pi. The closer the actual ordering of the indices to a given normative order in the model, the higher is the level of region investment attractiveness. The matrix of actual correlations of growth performance (F = {fij}nxn) is described as follows:

$$f_{ij}\begin{cases} \mathbf{1}, \boldsymbol{if}\ \boldsymbol{GR}(\boldsymbol{P}_i) > GR(\mathrm{P}_j) \ ; \\ -\mathbf{1}, \boldsymbol{if}\ \boldsymbol{GR}(\boldsymbol{P}_i) < GR(\boldsymbol{P}_j) \ ; \\ \mathbf{0}, \boldsymbol{if}\ \ \boldsymbol{GR}(\boldsymbol{P}_i) = \boldsymbol{GR}(\boldsymbol{P}_j), \end{cases} \tag{2}$$

where i, j are the numbers of indicators; Pi, Pj are indicators having the i-th and j-th numbers, respectively; GR(Pi), GR (Pj) are actual rate (index) of growth of the ith and jth indicators, respectively.

An integrated assessment of region investment attractiveness is the estimation of proximity of actual and normative models set in order the rates (indices) of growth indicators (Y).

$$\mathrm{y} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n} \mid e_{ij} \mid} \text{ where } b_{ij}\begin{cases} 1, if\ e_{ij} = 1\ simultaneously\ with\ f_{ij} \geq 0 \ ; \\ \qquad or \\ if\ e_{ij} = -1\ simultaneously\ with\ f_{ij} \leq 0 \ ; \\ \qquad 0, otherwise \end{cases} \tag{3}$$

n is the number of indicators in DD; i, j are the numbers of indicators in DD; b_{ij} is the element of the matrix of coincidence of actual and reference ratios of growth rates (B = $\{b_{ij}\}_{nxn}$); e_{ij} is the element of the matrix NM, f_{ij} is the element of the matrix F = $\{f_{ij}\}_{nxn}$.

Score Y varies from 0 to 1. Equal to 1, if all regulations set the ratio of the rate of improvement is actually implemented. Equal to 0, if the actual order of indices is opposite to the normative order of indicators in the model. The closer Y is to 1, the greater is the proportion of regulatory relationships between indicators implemented in reality.

The generated model can be considered as the factor system. The influence of each indicator on Y growth, which is the effective rate, determined by the formula:

$$\Delta Y(P_i) = \frac{\sum_{j=1}^{n} b_{ij}^{0} - \sum_{j=1}^{n} b_{ij}^{b}}{\sum_{i=1}^{n}\sum_{j=1}^{n} \mid e_{ij} \mid} \quad \text{and} \quad b_{ij}\begin{cases} 1, & if\left(e_{ij} = 1\ \ and\ \ f_{ij} \geq 0\right)\ or\ \left(e_{ij} = -1\ \ and\ \ f_{ij} \leq 0\right) \\ 0, & if\ \ others\ \ case \end{cases} \tag{4}$$

where $\Delta Y(P_i)$ is the increase in the assessment caused by the dynamics of the ratio of the growth rate of the ith indicator with others; n is the number of indicators; i, j are the numbers of indicators; b^0_{ij}, b^b_{ij} are the elements of the matrix of coincidence of actual and reference

No and indicator name in DS, Pi	No indicator in DS																	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1. Population in the region	0	−1	−1	−1	0	−1	−1	−1	−1	−1	−1	0	−1	−1	−1	−1	−1	1
2. Incomes of the population of the region	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0	1	1	1
3. Number of employed in the economy of the region	1	−1	0	−1	1	−1	−1	−1	−1	−1	−1	0	−1	−1	−1	−1	−1	1
4. Number of employees employed in small enterprises in the region	1	0	1	0	1	1	0	0	0	1	1	1	0	0	0	1	1	1
5. Total number of unemployed in the region	0	−1	−1	−1	0	−1	−1	−1	−1	−1	−1	0	−1	−1	−1	−1	−1	0
6. Fund for remuneration of workers in the region	1	−1	1	−1	1	0	−1	−1	−1	−1	−1	0	−1	−1	−1	0	0	1
7.GRP	1	0	1	0	1	1	0	1	1	0	1	1	−1	1	−1	1	1	1
8. Value of fixed assets in the region	1	0	1	0	1	1	−1	0	−1	0	1	1	−1	−1	−1	1	1	1
9. Residual value of fixed assets of the region	1	0	1	0	1	1	−1	1	0	0	1	1	−1	−1	−1	1	1	1
10. Number of small enterprises in the region	1	0	1	−1	1	1	0	0	0	0	1	1	0	0	0	1	1	1
11. Number of organizations in the region	1	0	1	−1	1	1	−1	−1	−1	−1	0	1	−1	−1	−1	1	1	1
12. Number of unprofitable organizations in the region	0	0	0	−1	0	0	−1	−1	−1	−1	−1	0	−1	−1	−1	0	0	0
13. Internal costs for research and development of the region	1	0	1	0	1	1	1	1	1	0	1	1	0	1	0	1	1	1
14.Capex	1	0	1	0	1	1	−1	1	1	0	1	1	−1	0	−1	1	1	1
15. Balanced financial performance of organizations in the region	1	0	1	0	1	1	1	1	1	0	1	1	0	1	0	1	1	1
16. Tax revenues of the consolidated budget of the region	1	−1	1	−1	1	0	−1	−1	−1	−1	−1	0	−1	−1	−1	0	1	1
17. Revenues of the consolidated budget of the region	1	−1	1	−1	1	0	−1	−1	−1	−1	−1	0	−1	−1	−1	−1	0	1
18. Expenditures of the consolidated budget of the region	−1	−1	−1	−1	0	−1	−1	−1	−1	−1	−1	0	−1	−1	−1	−1	−1	0

Table 3. The normative model of region investment appeal.

ratios of rates (indices) of growth in current and base periods, respectively; e_{ij} is the matrix element of the reference relationships between the growth indicators.

4.3. Step 3: Calculations

In this chapter, the presented results allow to claim that the proposed methodological developments can be applied to a large variety of tasks related to the monitoring of development strategies of regions and other objects of strategic planning. The decline in balance performance or growth with their help received a quantitative rating, which in turn allows you to implement monitoring of investment strategy in the tactical period. This methodology does not require serious mathematical tools; however, if growth of indicators is necessary to apply automated processing for calculation of the estimates, so the author used her own computer program and implemented the development of regulatory models for integrating quantitative evaluations of the investment attractiveness of the region, the city and municipal district.

5. Results

As the periods of research are selected: 2001–2007 (base period) and 2009–2015 (reporting period) as well as chain dynamics for 2009–2015. The choice of study periods for the baseline dynamics is due to the fact that in 2006–2007, Russia and its regions were given an investment rating by international agencies. The sense of the reporting period is that this was the period when the regions of Russia left the financial crisis and actively attracted foreign investments, and implemented a large number of investment projects.

Tables 4 and **5** show the quantitative levels of investment attractiveness of regions—leaders in Russia, calculated according to the author's algorithm.

Thus, the imposition of sanctions in 2014 significantly affected the investment climate of the Kaluga region, since its economy is more dependent on the activities of foreign investors. For the Tula, Voronezh and Moscow regions, Moscow and Russia in general, the sanctions played a stimulating role.

Regions	Y		ΔY	National rating of investment attractiveness of regions in 2015
	2001–2007	2009–2015		
Tula region	0.57	0.79	0.22	2
Russia	0.69	0.76	0.08	—
Voronezh region	0.52	0.76	0.24	4
Moscow region	0.61	0.66	0.05	5
Kaluga region	0.65	0.61	−0.04	1
Moscow	0.69	0.52	−0.17	3

Table 4. Quantitative levels of investment attractiveness of regions.

No.	Regions	Y						
		2009	2010	2011	2012	2013	2014	2015
1	Tula region	0.45	0.67	0.70	0.61	0.51	0.52	0.83
2	Voronezh region	0.47	0.42	0.74	0.71	0.51	0.61	0.69
3	Moscow	0.54	0.54	0.39	0.68	0.42	0.57	0.61
4	Russia	0.54	0.64	0.63	0.77	0.53	0.50	0.53
5	Moscow region	0.55	0.72	0.65	0.59	0.55	0.40	0.43
6	Kaluga region	0.48	0.55	0.67	0.71	0.53	0.61	0.40

Table 5. Dynamics of quantitative levels of investment attractiveness of regions.

6. Recommendations

The author has conducted a research of dynamics of region investment appeal of the Kaluga region—the leader in formation of favorable investment climate for foreign investors and to creation of industrial parks in Russia. It is as a result proved that in connection with imposition of economic sanctions the area has faced development problems, namely, deterioration in a solvency of the population, employment and effectiveness of activity of small business. All this has brought to rating downgrade of area to "National rating a state of investment climate in territorial subjects of the Russian Federation." For large businesses in the Kaluga region, the technology of creating industrial parks is effectively applied.

7. Conclusion

Author examines the effectiveness of region government policy on investment policy. Investment policy among her works plays a major role, through the human, innovation, and financial side and the development of strategies to attract investment. For realization of statistical research in Russia regions, she proposes normative model of region investment appeal, monitoring effectiveness of government and business for the development of the region. It is as a result proved that in connection with imposition of economic sanctions the leaders in formation of favorable investment climate in Russia regions have faced development problems. She proved that the imposition of sanctions in 2014 significantly affected the investment climate of the Kaluga region, since its economy is more dependent on the activities of foreign investors. For other leaders—the Tula, Voronezh, and Moscow regions, Moscow and Russia in general, the sanctions played a stimulating role.

Acknowledgements

The reported research was funded by Russian Foundation for Basic Research and the government of the Kaluga region of the Russian Federation, grant № 17-12-40003a(p) "Influence of

the results of financial and economic activity of enterprises of the commercial sector of the economy of an investment-attractive region on the achievement of the goals of its social and economic development (for the example of the Kaluga region)".

Author details

Burtseva Tatiana Alexsandrovna

Address all correspondence to: taburtseva@tksu.ru

Department of Management, Tsiolkovsky Kaluga State University, Kaluga, Russia

References

[1] Keynes JM. The Means to Prosperity. London: MacMillan and Co; 1933. 355 p

[2] Keynes JM. The General Theory of Employment, Interest and Money. 2011. Direct access: http://cas.umkc.edu/economics/people/facultypages/kregel/courses/econ645/winter2011/generalt heory.pdf

[3] McConnell KR, Brue SL. Economics in 2 volumes (Vol. 1. Economics: Principles, Problems, and Policies). Moscow: Infra-M; 2008

[4] Porter ME. The competitive advantage of nations. Harvard Business Review. 1990;**3**:73-79

[5] Russell M. The Russian economy: Will Russia ever catch up? European Parliamentary Research Service. 2015;**28**:36-42

[6] Burtseva TA. Methodology of statistical research of investment attractiveness of Russian regions. Statistics Problems. 2015;**1**:29-45

[7] Burtseva TA et al. Statistical research of investment development of Russian regions. International Journal of Environmental and Science Education. 2016;**11**:8182-8192. Direct access: http://www.ijese.net/makale/1073

[8] Zubarevich NV. Regions of Russia: Inequality, crisis, modernization independent. Moscow: Institute of Social Policy; 2010

[9] Frenkel AA et al. The rating of regions on an integrated indicator of investment attractiveness. 2009;**1**:56-60

[10] Kaplan RS, Norton DP. Using the balanced scorecard as a strategic management system. Harvard Business Review. 1996:75-85

[11] Syroezhin IM. Improving the System of Indicators of Efficiency and Quality. Moscow: Economics; 1980

[12] Dragisa S. A comparative analysis of the economic movement on the basis of growth matrix. Socio-Economic Planning Sciences. 1986;**20**(2):75-78

[13] Zakharchenko NN. Economic Measurement: Theory and Methods. Saint Petersburg: Saint Petersburg State University of Economics and Finance Press; 1993

[14] Zavgorodniaia AV. Diagnosis and Modelling of the Environment of the Firm. Saint Petersburg: Saint Petersburg State University of Economics and Finance Press; 1999

[15] Zhambekova RL. The methodology of the system of economic diagnostics of the enterprise: The Dissertation on competition of a scientific degree of the doctor of economic Sciences: 08.00.05. Saint Petersburg State University of Economics and Finance; 2000

[16] Burtseva TA. Indicative model of monitoring the investment attractiveness of the region. Statistics Problems. 2009;**6**:37-45

[17] Svetunkov SG, Zagranovskaia AV, Svetunkov IS. The Rating of Development of Regions and the Dynamics of their Socio-Economic Status. Complex-valued analysis and modeling of uneven socio-economic development of regions of Russia. 2012. www.sergey.svetunkov.ru/economics/complex/MD2012

[18] Pogostinskiy YA. A Normative Model of System Analysis of Economic Activity of the Enterprise: Monograph. Saint Petersburg: Saint Petersburg State University of Economics and Finance Press; 1999

Chapter 5

Agenda 2030: Measuring Progress in the Montenegro's National Strategy for Sustainable Development through SDG Indicators

Gordana Djurovic, Milica Muhadinovic,
Vasilije Djurovic and Martin M. Bojaj

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.75001

Abstract

The Government of Montenegro nationalized 2030 Agenda on Sustainable Development by adopting the National Strategy for Sustainable Development (NSSD) in 2016, together with a corresponding Action Plan for its implementation. The NSSD is umbrella, horizontal and long-term development strategy of Montenegro that relates not only to environment and economics, but also to human resources, valuable social capital that should ensure prosperous development, recommendations for establishing the framework of financing and governance for sustainable development. The NSSD represents strategic framework for the transposition of the UN sustainable development goals (SDGs) and its indicators into national context. The NSSD Action plan, divided into 6 thematic areas with 30 strategic goals of sustainable development of Montenegro and their 102 measures and 601 sub-measures, represents very complex mechanism for monitoring and reporting about achieving the UN sustainable development goals (SDGs) in Montenegro. Hence, measuring progress in the NSSD implementation is the focus of this research. Effective measuring of the progress in the NSSD implementation could be achieved under the following preconditions: developed coordination mechanism for reporting, methodology of designing indicators, IT support for data collection and reporting, and strengthening of inter-linkages between the EU agenda and UN 2030 Agenda.

Keywords: Montenegro, sustainable development strategy, SDGs, SDG indicators

1. Introduction

Montenegro is located in the South-Eastern Europe with 623.000 inhabitants which live on 13.912 km [1]. Following the results of the referendum withheld on May 21, 2006, Montenegro regained independence and became the 192nd member of the United Nations. The economy of Montenegro is mostly service-based: about two thirds of GDP and more than 80% of employment are in the service sector in the last few years. Economic growth model is driven mostly by foreign direct investments since average net FDI in the period 2006–2016 is 17.1% [2]. Country is in the late transition to a market economy and strategic development sectors are tourism, energy, industry and agriculture with rural development. Estimated GDP for 2017 is 4.202 mil € with real growth rate of 4% [3]. According to the Statistical office data GDP/pc in 2016 is 6.063 € [1]. Measuring GDP/pc in PPS, Montenegro has reached 42% of the EU average in 2016.

Diversity of geological base, landscape, climate and soil, as well as the very position of Montenegro on the Balkan peninsula and Adriatic sea, created conditions for formation of biological diversity with very high values, that puts Montenegro among biological "hot-spots" of Europe and world biodiversity [4].

Decisions to have Montenegro developed as an ecological state requires dedication, endurance, bold decisions and significant financial investments. Declaration of the Ecological state of Montenegro, adopted by Montenegrin Assembly on September 20, 1991, emphasizes that "by respecting our obligation to the nature which is the source of our health and inspiration of our freedom and culture, we turn ourselves to the protection of the mother nature, in the name of our own survival and the future of our successors." Following that decision, Constitution of Montenegro from 1992 contains definition of Montenegro as "an ecological state" and the same was confirmed in the Constitution from 2007, after regaining the independence. "Montenegro is a civil, democratic, ecological and the state of social justice, based on the rule of law" is prescribed in the article 1 of Montenegro's Constitution [5].

The Government of Montenegro adopted the first National Strategy for Sustainable Development (NSSD) in April 2007, together with a corresponding Action Plan for the period of 2007–2012. During its implementation period, the Government adopted five annual progress reports on the NSSD implementation. The monitoring and evaluation results indicated that the respectable progress has been achieved—environmental pillar score 47%, social development score 65% and economic development pillar score 49%, respectively, in average about 53% of 236 planned measures from the 2007 NSSD Action plan was realized by 2013 [6]. The final estimation about implementation level was about 60% of proposed strategic measures. These measures have been modified and transposed in the new NSSD 2030.

At the UN Sustainable Development Summit in September 2015, world leaders adopted the 2030 Agenda for Sustainable Development—including 17 Sustainable Development Goals (SDGs) and 169 targets—to end poverty, fight inequality and injustice, and tackle climate change [7]. The Agenda 2030 calls on governments to translate the global SDGs into national targets and policies. The task of national governments is to set a level of ambition and formulate a clear, long-term vision defining what they would like to achieve with the 2030 Agenda [8]. The SDGs, targets and SDG indicators, as well as their accompanying political declaration make it clear that countries are expected to make this national interpretation—setting their own goals, targets and priorities for implementing the SDGs, in response to national conditions and capabilities [9]. In March 2016, a global indicator framework for the SDGs—identified and proposed by the Inter-Agency Expert Group on SDG indicators (IEAG-SDGs)—was agreed by the 47th Session of the UN Statistical Commission [10]. This framework constitutes a key element of the 2030 Agenda for Sustainable Development's implementing architecture and is composed of 241 indicators for monitoring global progress towards the SDGs.

On July 7, 2016 Montenegro's *NSSD 2030* was adopted by the Montenegrin government. Montenegro became one of the first UN member countries to conduct a national review of its process to implement the global Agenda 2030 within the national policy context. The voluntary national reviews were conducted by 22 pilot countries and presented at the High-Level Political Forum (HLPF) 2016 meeting [6].

A periodic national monitoring report is needed to track progress and—depending on the political ambition—to promote accountability by explaining underlying developments or even to evaluate policy performance [8]. For these reasons, Montenegro's NSSD is accompanied with its Action plan which includes thematic areas, goals, measures, sub-measures, baseline indicators, targeted values and a proposed list of numerous indicators to measure progress in implementing the Strategy (SDG indicators, some national indicators, some indicators from the database of different international organizations and a few so-called composite indicators, such as Ecological footprint, Domestic material consumption, Lend consumption, Resource productivity, Environmental performance index, etc.).

The second part presents the research framework for measuring progress in the NSSD 2030 action plan through development of three research hypothesis. The third part presents Montenegro's NSSD revision through the prism of setting-up an integrated monitoring system for the NSSD, while the fourth part presents results of mapping and content analysis of all indicators. They encompass four groups: SDG indicators, national, international – from different databases, and proposed composite indicators. In the fifth part an analysis of needs necessities importance and main challenges of an integrated monitoring framework is presented. The focus of this part is on the analysis of complementarities between Montenegro's 2030 Agenda and the EU integration agenda. Finally, in the sixth part authors summarize the content of the chapter and main policy recommendations as conclusion.

2. Research framework for measuring progress in the NSSD 2030 action plan

Among the various steps, during the process of NSSD 2030 design, was the development of the NSSD monitoring and reporting framework. This paper is focused on analysis of the quality and capability of the developed monitoring and reporting framework in order to contribute to the preparation of regular reporting documents. At the same time, it is necessary to explain inter-linkages of the SDGs (UN agenda) and current European integration agenda of

Montenegro in order to create synergy in achieving reform goals and improve quality of life of Montenegro's citizens.

Proceeding from the above, the research hypothesis in this paper can be defined as follows:

1. The measurement of progress in the implementation of the NSSD 2030 Action Plan will be effective only if the following preconditions could be achieved: (a) the coordination mechanism for reporting is significantly strengthened and (b) a clear methodology for defining each SDG indicator is defined and (c) IT support for data collection and reporting is developed. It also involves respect of the planned dynamic of introducing SDG indicators in regular reporting process.

2. Measuring progress in the implementation of the NSSD Action plan may also be supported with introduction of other indicators proposed by the Strategy (national indicators, indicators from different databases of the selected international organizations and 10 complex indicators).

3. There are complementarities between the implementation process of NSSD 2030 and the process of the EU accession: fulfillment of the obligations of accession and later obligations of full membership in the Union, contribute strongly to the achievement of the Montenegro's NSSD 2030.

As such, this paper constitutes one of the first studies dealing with the national-level transposition of the global Agenda 2030 and its link with the process of integration to the EU. For this research we use, inter alia, a systemic approach, historical method, secondary data collection, comparative analysis and content analysis [11].

3. Structure of the NSSD 2030 and its action plan

Montenegro nationalized 2030 Agenda on Sustainable Development by adopting the Strategy for Sustainable Development (NSSD) in July 2016, together with a corresponding Action Plan. The NSSD is an umbrella, horizontal and long-term development strategy that relates not only to the environment and economics, but also to human resources, valuable social capital, recommendations for establishing the framework of financing and governance for sustainable development [12]. In other words, the NSSD and the NSSD Action plan are divided in the four thematic areas—four aspects of the national capital (human resources, social resources, natural resources and economic resources), governance and financing for sustainable development. The structure of thematic areas, goals, measures and sub-measures is presented in **Table 1**.

Mainstreaming the NSSD through other strategic documents of the Government is the key in successful achievement of the sustainable development in Montenegro, along with robust monitoring and reporting pertaining to results and its implementation.

The essential part of the NSSD 2030 is the Action Plan (NSSD AP). It is divided in 6 thematic areas with 30 strategic goals of sustainable development of Montenegro and their 102 measures

Goals	Thematic area I. Human resources: Improvement of the state of human resources and strengthening social inclusion	No. of measures	No. of sub-measures
1	1.1 Improve demographic trends and reduce demographic deficit	2	12
2	1.2 Improve health of citizens of all ages and reduce inequalities in health status	3	37
3	1.3 Ensure inclusive and quality education and promote lifelong learning opportunities for all	4	20
	Thematic area II. Social resources: Support values, norms and behavioral patterns of importance for the social sustainability		
4	2.1 Stimulate active relationship between key actors and development sustainability	7	42
5	2.2 Develop a system of values in accordance with the community sustainable development goals	3	13
6	2.3 Develop the state as an efficient rule of law	2	13
7	2.4. Overcome the managerial deficit problems and strengthen socially responsible business	1	5
8	2.5 Stimulate employability and social inclusion	3	26
9	2.6 Improve the importance of culture as a fundamental value of spiritual, social and economic development which significantly improve citizens' life quality	4	28
10	2.7 Establish an efficient and modern system of integrated protection, management and sustainable use of cultural heritage and landscape	3	31
11	2.8 Achieve equal socio-economic development in all local self-government units and regions based on competitiveness, innovations and employment, with specific emphasis on the Northern region	2	15
	Thematic area III. Natural resources: Conservation of natural capital		
12	3.1 Stop degradation of values of renewable energy sources: biodiversity, water, sea, air, soil	2	15
13	3.2 Improve the efficiency of renewable natural resource management	6	40
14	3.3 Enable symbiosis of effects of environmental performance improvement and protection of human health	2	9
15	3.4 Solve problems of unsustainable capacitating of space generated by unrealistic demand in terms of quantity and low quality of built-up areas	7	39
16	3.5 Enable resource-efficient use of metallic and non-metallic resources	1	6
17	3.6 Mitigate natural and anthropogenic hazard impacts	5	14
	Thematic area IV. Economic resources: Introduction of green economy		
18	4.1 Reduce level of greenhouse gasses emissions by 2030 by 30% compared to baseline 1990	4	25
19	4.2 Improve resource efficiency in key economic sectors	5	16
20	4.3 Improve waste management through circular economy approaches	5	36
21	4.4 Enable sustainable management of the coastal regional resources and encourage the blue economy	3	10
22	4.5 Support greening of economy through the development and implementation of operational instruments of sustainable consumption and production	3	16
23	4.6 Implement social responsibility in practice of all sectors in accordance with positive European and international experiences	3	12

Goals	Thematic area I. Human resources: Improvement of the state of human resources and strengthening social inclusion	No. of measures	No. of sub-measures
24	4.7 Increase Montenegrin economy competitiveness level for sustainable development & green jobs	4	18
	Thematic area V. Governance for sustainable development		
25	5.1 Strengthen governance system for sustainable development	2	9
26	5.2 Strengthen environmental management by improving implementation of environmental protection instruments	3	12
27	5.3 Implement reform for institutional organization of governance system for sustainable development	2	21
28	5.4 Establish a system for monitoring of national sustainable development, including monitoring of sustainable development goals implementation	2	15
	Thematic area VI. Financing for sustainable development		
29	6.1 Establish the system for sustainable environmental financing and preservation of natural capital as a component of financing for sustainable development	4	18
30	6.2 Enable introduction of green economy by mobilizing funds for sustainable development financing	5	28

Table 1. Action plan for Montenegro's NSSD 2030: 6 thematic areas, 30 strategic goals, 102 measures with 601 sub-measures to be achieved by 2030.

and 601 sub-measures. It is an important instrument for mainstreaming and monitoring the maturity level of the 2030 Agenda in Montenegro.

Out of 169 specific targets of sustainable development, structured in 17 SDGs, 167 were transposed into measures defined in the NSSD Action Plan, in accordance with national circumstances and future needs. Only two targets (9.a and 9.c) were not relevant for Montenegro, which confirms the universal nature and inter-sectorial and overarching character of the NSSD [12].

Gradually, all relevant institutions should determine a connection between certain SDGs and their sectoral strategies, plans and programs, measures and actions that they will be responsible for the implementation of the NSSD. Accordingly, it is necessary to determine relevance of sustainable development tasks and indicators in terms of their integration into sectoral policies which will enable harmonization of current relevant strategies, programs and plans with the NSSD 2030.

4. Statistical data producers and future monitoring mechanism

4.1. Statistical data producers

The Ministry for Sustainable Development and Tourism is the ministry responsible for coordination and reporting based on collected indicators. A numerous meetings and consultations

http://dx.doi.org/10.5772/intechopen.75001

were organized between the expert team and Montenegro's official and administrative data producers. Within 36 Montenegrin institutions consulted, 26 were found to be handling data for at least one SDG indicator. Institutions responsible for generating SDGs indicators (241 SDG indicators recognized in March 2016) are the following:

I. Official statistical data producers (independently analyze and process data, calculate indicators and prepare analytical reports), responsible for 121 SDG indicators:

- Statistical Office of Montenegro (Monstat)—responsible for 55 SDG indicators,
- Central Bank of Montenegro—responsible for 7 SDG indicators,
- Health Insurance Fund—responsible for 1 SDG indicator,
- Institute for Public Health—responsible for 20 indicators,
- Ministry of Finance—responsible for 30 SDG indicators,
- Ministry of Science—responsible for 4 SDG indicators,
- Customs Administration—responsible for 2 SDG indicators, and
- Department of Public Revenues—responsible for 2 SDG indicators.

II. Administrative statistical data producers, introduced by the NSSD, for the purpose of SDG indicator calculations (responsible for 82 indicators):

- Agency for Electronic Communications (2 SDG indicators),
- Ministry of Agriculture and Rural Development (14),
- Ministry of Justice (2),
- Ministry of Education (7),
- Ministry of Labor and Social Welfare (2),
- Ministry of Transport and Maritime Affairs (4),
- Ministry for Human and Minority Rights (8),
- Ministry of Economy (5),
- Ministry of Health (4),
- Institute for Marine Biology (3),
- Ministry of Foreign and European Affairs (7),
- Ministry of Interior Affairs (19)
- Ombudsman (1),
- Biotechnical Faculty (2),
- Public Procurement Administration (1) and
- Real Estate Administration (1).

III. Applicable institutions becoming official statistical data producers following the adoption of the NSSD (responsible for 38 SDG indicators): Ministry for Sustainable Development and Tourism (27 SDG indicators) and Environmental Protection Agency (11 SDG indicators).

For monitoring and reporting on the progress pertaining to the implementation of the Strategy, an ambitious plan is proposed. It comprises of a collection of 525 indicators to measure progress, as follows:

1. UN list of SDG indicators (in March 2016 there were 241 indicators in total: 121 + 82 + 38 SDG indicators); However, since nine indicators repeat under two (7.a.1 = 13.a.1; 8.4.1 = 12.2.1; 8.4.2 = 12.2.2; 10.3.1 = 16.b.1; 10.6.1. = 16.8.1; 15.7.1. = 15.c.1; 15.a.1 = 15.b.1;) or three different targets (1.5.1 = 11.5.1 = 13.1.2; 1.5.3 = 11.b.1 = 13.1.1;), the actual total number of individual indicators in the list was 230. According to data from March 2017, there are 232 indicators, since some changes were made [13].

2. 252 national indicators (Human resources—58 indicators, Social resources—65 indicators, Natural resources—52 indicators, Economic resources—68 indicators and Governance for sustainable development—9 indicators); the national indicators are important in relation to the specific national needs. These indicators are already being followed or it is easy to include them in the monitoring system in a short period. They do not represent an alternative to the content to the UN indicators, but complement them [13].

3. 20 international indicators produced by international organizations (WB, IMF, World Economic Forum, UNDP, etc.);

4. 10 recommended so-called complex indicators (Human Development Index, Gender Inequality Index, Social Progress Index, Ecological Footprint, Domestic Material Consumption, Land Consumption, Resource Productivity, Genuine Progress Indicator, Environmental Democracy Index and Environmental Performance Index).

Table 2 shows the framework plan for the development and monitoring of SDG indicators in Montenegro till 2024. In 2016, only 49 SDG indicators were followed in Montenegro by name and content (20%) and for 34 SDG indicators there is an alternative by content (14.1%).

By 2018, additional 54 indicators will be produced (22.4%), which is in total 137 SDG indicators or 56.8%. Regarding the dynamics of introduction of SDG indicators into the national

Year	Number of indicators measured/necessary to be additionally introduced	Total number of indicators that will be monitored in Montenegro	Share (%)
2016	83	49 (entirely, by name and content) +34 (there is an alternative by content)	34.4
2018	54	137	57
2020	32	169	70
2022	24	193	80
2024	27	220	90

Table 2. Overview of dynamics in development/monitoring SDGs indicators in Montenegro's NSSD till 2024.

http://dx.doi.org/10.5772/intechopen.75001

monitoring system through NSSD, 220 SDG indicators (90%), will be introduced by 2024. This ambitious plan is developed based on projected medium term plan for strengthening capacities of Montenegro's statistical data producers and expected budget for these activities [14].

In total, 137 SDG indicators will be monitored in Montenegro by 2018 (57% of the total number of indicators) and presented in the first report scheduled for 2019. The rest of the SDG indicators (10%) will be introduced by 2030, following regular modification of its number through the United Nations Statistical Commission work. This value is comparable to findings of a recent study by Colombia's National Administrative Department of Statistics, which found 91 out of 169 of the global SDG targets (54%) to be currently measurable through the almost 200 indicators existing in the National Development Plan of Colombia [8].

In addition, different UN Organizations are custodians of numerous SDG indicators and thus in a unique position to help the country to generate data/evidence. Having identified with the Government the main data gaps against the SDG indicators, the UN System in the country (led by UNICEF and UNDP in this regard) helps generate data through Multiple Indicators Cluster Survey and other surveys (for example, surveys that seek to measure social norm change and attitudes related to key topics such as violence, ethnic minorities, people and children with disabilities, gender, immunization and so on) and raising capacities for introducing five green economy indicators for which there is a UN methodology developed (Material footprint, Domestic material consumption, Tourism direct GDP as a proportion of total GDP and in growth rate, Number of jobs in tourism industries as a proportion of total jobs and growth rate of jobs, by sex, and Change in water-use efficiency over time) [15].

4.2. Responsible institutions for the coordination and reporting process

The initial information about implementation of the NSSD Action plan was prepared by the Ministry for Sustainable Development and Tourism in December 2017 [16] and submitted to the National Council for Sustainable Development as main advisory body and to the Government. NSSD AP will be evaluated through regular progress reports according to the dynamic plan presented in the **Table 3**. The first regular progress reports will be prepared in 2019. All phases of proposed evaluation process are presented in the table below.

The Ministry for sustainable development and tourism, as central coordinating body, is obliged to develop the detailed methodology for monitoring and reporting process. Development of a progress report methodology contains four basic elements: (a) preparation of detailed methodologies for reporting on all indicators, (b) producing reporting forms on each individual indicator being monitored ("indicator's passport"), (c) preparation of reporting methodology on the dynamics of the implementation of measures envisaged by the NSOR Action Plan, including mechanism of coordination, and (d) preparation of forms for reporting on individual measures and shares from the NSOR Action Plan. The reporting methodology is in the final phase of development and will be ready for the first evaluation process [17].

The process of implementation, monitoring and evaluation consists of few phases, and each of them is very important so the process itself represents a functional methodology. It necessitates setting up this process as to have precise defined obligations and responsibilities of all subjects

Phases	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030
Data collecting and processing	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
NSSD indicators introduction	x	x	x	x	x	x	x	x	x	x					
% of introduced SDG indicators			57		70		80		90						
Preparation of progress reports			x	x	x	x	x	x	x	x	x	x	x	x	
Publishing of progress reports				x		x		x		x		x		x	
Analysis of data in line with UN Statistical Commission regular reports				x	x	x	x	x	x	x	x	x	x	x	x
Development of recommendations for improving NSSD					x		x		x		x		x	x	x
Revision of the NSOR action plan					x					x					x
Revision of the NSOR															x

Table 3. Dynamics of evaluation process of the NSSD implementation by 2030.

involved in the NSSD implementation, with precisely defined procedures and process format and data exchange due to preparation of report of the NSSD [12]. Next scheme presents that coordination mechanism and its phases.

The UN Agenda 2030 [7] determined that "A robust, voluntary, effective, participatory, transparent and integrated follow-up and review framework will make a vital contribution to implementation and will help countries to maximize and track progress in implementing this Agenda in order to ensure that no one is left behind; As national ownership is key to achieving sustainable development, the outcome from national level processes will be the foundation for reviews at regional and global levels, given that the global review will be primarily based on national official data sources; at the same time follow-up and review in the national context is defined as voluntary instrument led by countries taking into account different national realities, capacities and levels of development and respecting national policies and priorities." Hence, it is necessary to enable continuous realization of the program for transposition of SDGs into national context and building of national capacities for the efficient implementation in upcoming monitoring and reporting process.

In addition, the establishment of a functional information system and data base represents the basis for successful measurement of progress in the 2030 NSSD implementation. Well organized functional information system should enable not only simultaneous collecting of data by subjects responsible for monitoring of certain sustainable development indicators, or the NSSD measures, but their archiving and further processing towards efficient national and international reporting.

http://dx.doi.org/10.5772/intechopen.75001

The central coordination institution is Ministry for sustainable development and tourism which covers three important phases of the monitoring and reporting process (**Figure 1**). According to the new systematization in the Ministry, the Department for Sustainable Development has been integrated into the General Directorate for Climate Changes and Mediterranean Affairs. As of December 2017, the Direction for Mediterranean affairs assumed the SDG related responsibilities of the previous "Department for Sustainable Development" [17]. New systematization and division of labor has introduced a lot of innovation into a system of sustainable development policies coordination. It will take time for the new internal organization scheme to be consolidated. Hence, there is potential risk of slowing down preparation of the first NSSD report scheduled for 2019.

The National council for sustainable development, climate change and integrated coastal zone management is currently chaired by the President of Montenegro and constitutes of 26 members from different stakeholder groups (national and local government authorities, the business sector, public institutions, civil society and independent experts) and four working groups (WGs): (1) WG for the revision and monitoring of the NSSD implementation; (2) WG on the Sustainable Management of Resources; (3) WG on mitigation and adaptation to Climate Changes; and (4) WG on integrated coastal zone management. The National council provides recommendations to the central government on the following: (1) implementation of sustainable development policies; (2) harmonization of sector policies under the principles and objectives of sustainable development, climate change and integrated coastal zone management;

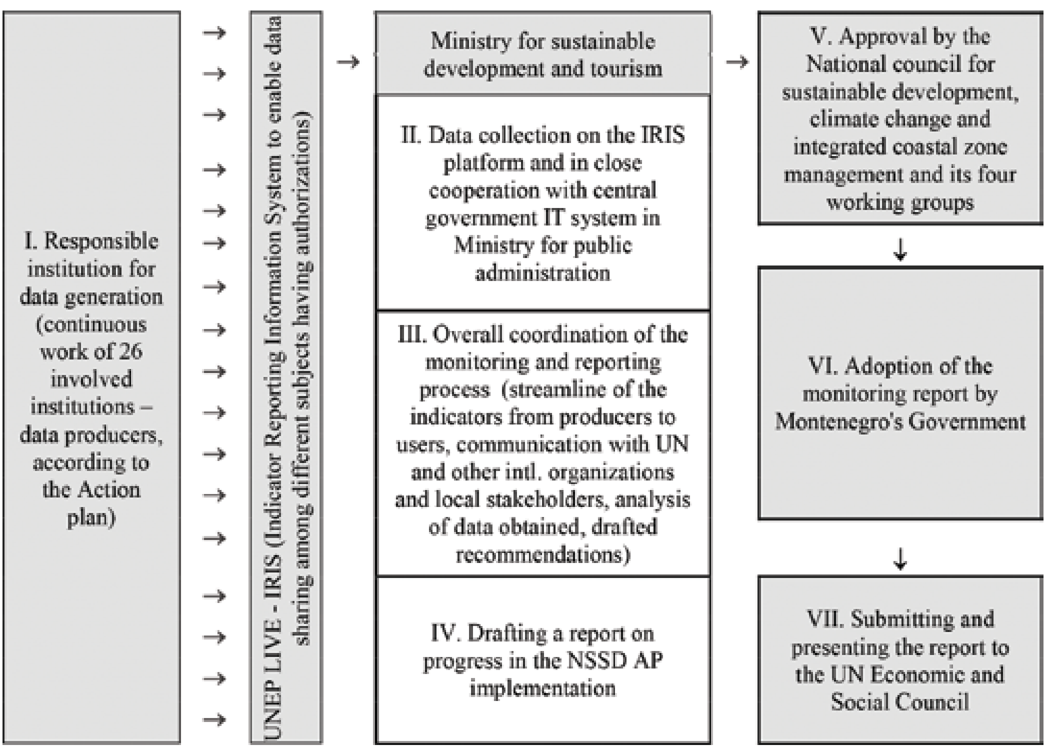

Figure 1. The NSSD 2030 coordination mechanism: Phases of the monitoring and reporting process.

(3) amendments to existing regulations; (4) adoption of new national regulations and (5) involvement in the process of monitoring the implementation of the NSSD 2030 [18] .

The final phases in the monitoring and reporting process are regular adoptions of the monitoring report by Montenegro's Government and its submission to the UN Economic and Social Council.

4.3. Establishment of a functional information system for data collection and reporting

Establishment of a functional information system and data base is the foundation for successful measurement of progress in the NSSD implementation. The established system should be simple, efficient and friendly oriented towards users since its primarily purpose is to facilitate implementation monitoring in order to have quality implementation of the NSSD, and not to be represent too massive and inefficient system that in practice will not work. Such organized functional information system should enable not only simultaneous collecting of data by subjects responsible for monitoring of certain indicators or measures, but their archiving and further processing towards efficient national and international reporting [12].

Existing UNEP's online reporting system (Indicator Reporting Information System—IRIS) is the solution recommended by the NSSD for organization and development of the information system. If adopted, after the ongoing test phase, IRIS will be managed by Montenegro. In this case the Ministry for Sustainable Development and Tourism, within a unit that will be in charge for the NSSD 2030 implementation [19].

5. Complementarities between Montenegro's 2030 agenda and EU agenda

Achieving NATO membership in June 2016, Montenegro's strategic foreign policy priority becomes EU membership [14]. Montenegro signed the Stabilization and Association Agreement (SAA) on October 15, 2007 and officially applied to join the EU on December 15, 2008. In November 2010, the Commission issued a favorable opinion on Montenegro's application and the Council granted it the candidate status [20]. The EU accession negotiation process with Montenegro was opened in June 2012 [21]. After five and a half years of negotiations, good dynamic has been achieved: 30 negotiating chapters have been opened (30 out of 33) and three have been temporarily closed (Science and research, Education and Culture, and Foreign affairs chapter) [22].

The programme of accession of Montenegro to the EU [23] is the roadmap for fulfillment of the accession criteria which comprises political, legal, economic and administrative criteria for full-fledge membership. Main methods for measuring progress in the EU accession talks are the following: level of harmonization of the legislation with the EU acquis and fulfillment of the closing benchmarks of the negotiation chapters.

At the same time, National strategy for sustainable development of Montenegro 2030 (NSSD 2030), adopted in 2016, represents comprehensive and long-term development strategy of

Montenegro that covers not only the framework period of accession process but also expected future period of Montenegro's development as the EU Member State. With adoption of the NSSD 2030 and a corresponding Action Plan for its implementation, the Government of Montenegro nationalized UN 2030 agenda for sustainable development, including its Sustainable Development Goals (SDGs) and SDG indicators.

The EU also implements the 2030 Agenda and the SDGs, together with its Member States, in line with the principle of subsidiarity. The EU's answer to the 2030 Agenda includes two work streams: the first is to mainstream the SDGs in the European policy framework and current Commission priorities; the second is to launch reflection on further developing of European longer term vision with the focus on sector policies after 2020—preparation for the long term implementation of the SDGs [24].

The sustainable development lies at the core of European values and constitutes an overarching objective of the Union. The EU has played a leading role in the process that led to the adoption of the 2030 Agenda for Sustainable Development and its 17 SDGs in September 2015. It is now determined to take the lead in its implementation. In June 2017, the European Council confirmed that "the EU and the Member States are strongly committed to implementing the 2030 Agenda in a full, coherent, comprehensive, integrated and effective manner and in close cooperation with partners and other stakeholders, including sub-national authorities, civil society, the private sector, social partners and academia." The European council underlined that the EU and its Member States will act as frontrunners in this respect [25].

Having in mind all of the above mentioned, Montenegro has committed to both the EU accession negotiations and the 2030 Agenda for Sustainable Development. This paper is also focused on analysis of relations between these processes and reporting characteristics. The main research questions are the following: (a) what is the level of similarities of those two reporting processes and (b) are there space for creating synergy and improve quality of reporting for both agendas.

Structural complementarities of these two agendas are presented in the **Table 4**. The Agenda 2030 is a long-term development framework focused on SDGs and its targets, while the EU agenda is comprised of the pre-accession period (accession negotiations process) and post-accession period and obligations arising from the full-fledged membership.

2030 Agenda as a plan of action for people, planet and prosperity, is based on achieving the sustainable development goals [12] while the EU integration process is based on achieving the accession criteria defined in negotiating chapters and related benchmarks [20]. Pre-accession assistance is provided by the EU through the Programme of the pre-accession assistance—IPA [26]. This programme is created in order to support Montenegro's integration process according to defined priorities in specific sectors. While IPA I (2007–2013) was implemented through its five components, IPA II (2014–2020) is created through sector-based approach for pre-accession support [27].

At the same time, United Nations Development Assistance Framework (UNDAF) 2017–2021 is a five-year strategic framework of cooperation between the Government of Montenegro and the UN for the period 2017–2021. Strategic document called Integrated UN Programme for

Montenegro		
2030 Agenda		EU Agenda
SDGs and targets		Accession criteria and benchmarks for the negotiation chapters
National Strategy for Sustainable Development NSSD 2030		Programme of the accession of Montenegro to the EU 2018–2021
Integrated UN Programme for Montenegro 2017–2021		Indicative strategy paper for Montenegro, IPA II (2014–2020)
SDG 1. No poverty	SDG 9. Industry, innovation and infrastructure	Sector 1—Democracy and Governance
SDG 2. Zero hunger	SDG 10. Reduced inequalities	Sector 2—Role of Law and Fundamental Rights
SDG 3. Good health and well-being	SDG 11. Sustainable cities and communities	Sector 3—Environment and Climate Action
SDG 4. Quality education	SDG 12. Responsible consumption and production	Sector 4—Transport
SDG 5. Gender equality	SDG 13. Climate action	Sector 5—Competitiveness and Innovation
6. Clean water and sanitation	SDG 14. Life below water	Sector 6—Education, Employment and Social Policies (including promotion of gender equality, and human resource development)
7. Affordable and clean energy	SDG 15. Life on land	
SDG 8. Decent work and economic growth	SDG 16. Peace, justice and strong institutions	
SDG 17. Partnerships for the goals		Sector 7—Agriculture and Rural Development
Reporting in 2019, 2021, 2023, 2025, 2027 and 2029;		Reporting semi-annually, until the conclusion of the accession negotiations; Later, continued through obligations of the EU Member State;

Table 4. Montenegro's 2030 agenda and EU agenda: goals, strategic documents, priority areas and reporting dynamic.

Montenegro 2017–2021 is designed to optimize the combined knowledge and resources of UN system organizations working under the Delivering as One modality. Strategic document was jointly developed by the UN Country Team in Montenegro and the Government of Montenegro, in collaboration with civil sector, academia and international stakeholders. The document, which articulates UN's assistance to Montenegro in addressing key national priorities, is aligned with the Agenda 2030 on Sustainable Development and country's aspiration to join European Union. The draft document provides a jointly agreed plan to support the country in the following key thematic areas: democratic governance, environmental sustainability, social inclusion and economic governance [28]. However, the UN 2030 Agenda does not bring an exact financial envelop that will provide financial support to the Government as the achieving SDG targets is responsibility of each of the UN member states [15]. Financial support, however, may come from external sources (vertical funds, bilateral, private donors, etc.), depending on economic/income status of the country.

http://dx.doi.org/10.5772/intechopen.75001

In mentioned context, 2030 Agenda as a universal, global development agenda is "nationalized" through adoption of the NSSD, while the European integration process is an instrument for "EU-isation" of Montenegro with strictly defined goals, i.e., achieving EU norms and standards through achieving closing benchmarks of negotiating chapters.

We can conclude that EU (integration) agenda is narrower than 2030 Agenda, having in mind that the scope of SDGs is significantly wider than the scope of accession criteria. But, implementing the EU agenda in the process of accession, Montenegro can significantly support implementation of the UN agenda for sustainable development.

The next level of analysis of inter-linkages between the EU accession criteria and 2030 Agenda for sustainable development in Montenegro-'s case is the analysis of EU negotiation chapters and SDGs and specific SDG targets. Based on above mentioned analysis, it is possible to make a link between negotiation chapters and sub-chapters with SDGs and SDG targets, with the following explanations:

- **SDG1** (No poverty) is linked with C19—Employment and social policy;
- **SDG2** (No hunger including sustainable agriculture) is linked with C11—Agriculture and rural development, but also with C12—Food safety, veterinary and phyto-sanitary policy and C25—Science and research;
- **SDG3** (Good health and well-being) is linked with C12—Food safety, veterinary and phyto-sanitary policy, C23—Justice and fundamental rights; C25—Science and research and C28—Consumer and health protection; For instance, EU requests are much smaller in C28 than scope of SDG 3 and its targets;
- **SDG4** (Quality education) is linked with C26 Education and culture, as well as with C25 Science and research; EU requests are much smaller than scope of SDG 4 and its targets;
- **SDG5** (Gender equality) is linked with C23—Justice and fundamental rights; It is important to mentioned that interim benchmarks in part "Fundamental rights" are very connected with SDGs 3, 5, 10 and 16;
- **SDG6** (Clean water and sanitation), **SDG12** (Responsible consumption and production), **SDG13** (Climate action) and **SDG15** (Life on land) are clearly linked to chapter 27—Environment;
- **SDG7** (Affordable and clean energy) is linked to C15—Energy;
- **SDG8** is linked to the following "economic" negotiation chapters: C1–C7, C9, C17, C19, C25 and C29;
- **SDG9** is linked to the following chapters: C10, C14, C20, C21 (Trans-European Networks and Regional Policy);
- **SDG10** (Reduced inequalities) is linked with C19 - Employment and social policy and C23 —Judiciary and fundamental rights;

- **SDG11** (Sustainable cities and communities) is linked with 27—Environment, but also with C26—Education and culture;
- **SDG12** (Sustainable consumption and production) as well as **SDG13** (Climate action) is linked with 27—Environment;
- **SDG14** (Life below water) is linked with C13—Fishery and C27—Environment;
- **SDG15** (Life on land) is also linked with 27—Environment;
- **SDG16** (Peace, justice and strong institutions) is linked with C8 (Competition and State Aid), C22 Regional policy, C23 & C23—Rule of Law chapters, C31—Foreign, security and defense policy and C33—Financial and budgetary provisions;
- **SDG17** (Partnership for the goals) is linked with C16—Taxation, C17—EMU, C18 –Statistics, C22 - Regional policy, C24—Justice, freedom and security, C30—Foreign policy, C32 (Financial control) and C33 (Financial and budgetary provisions);

Using the explained methodology, it is possible to "cover" all SDGs with areas of the negotiation, but not with the same intensity. The European integration process has better potential to 'accelerate' SDG achievement when they are linked to economic development, but less in the areas of social dimension of development.

Today, social realties within European countries (including both Member States and Candidate Countries) differ greatly, especially in the areas such as education and health, employment patterns, wages, incomes and social protection systems. Unemployment rates are falling but differ substantially across Europe, youth unemployment is falling but remains high, employment is recovering and growing steadily, but employment rates differ; social protection systems also differ across Europe [29]. Hence, special attention should be given in the forthcoming period to the **social** dimension of development and negative externalities of economic growth even during the process of European integration.

Through the pre-accession support, Berlin process and the Connectivity agenda (followed with the international financial institutions), the EU will continue to support building infrastructure in energy, transport, environment and social policy area. We can conclude that European pre-accession assistance and future EU structural fund support are excellent opportunities also in the areas of green economy and support for local sustainable economic development projects.

If we compare the closing benchmarks with the level of achievement of SDGs and SDG targets measured through SDG indicators—the conclusion could be the same. Based on the content analysis, 74 closing benchmarks and 83 interim benchmarks, defined for specific negotiation chapters within the process of EU accession talks of Montenegro [14], cover more than 60% of SDG related targets, which confirm high level of complementarities of those two processes. It is specifically important for potential improvement of their monitoring and reporting documents through close cooperation of their coordination structures for reporting and using of the same indicators in a lot of cases.

http://dx.doi.org/10.5772/intechopen.75001

6. Conclusions

The Government of Montenegro nationalized 2030 Agenda on Sustainable Development by adopting the National Strategy for Sustainable Development (NSSD) in 2016, together with a corresponding Action Plan for its implementation. The NSSD is umbrella, horizontal and long-term development strategy of Montenegro that relates not only to environment and economics, but also to human resources, valuable social capital that should ensure prosperous development, recommendations for establishing the framework of financing and governance for sustainable development. The NSSD represents strategic framework for the transposition of the UN sustainable development goals (SDGs) and its indicators into national context. The NSSD Action plan, divided in six thematic areas with 30 strategic goals of sustainable development of Montenegro and their 102 measures and 601 sub-measures, represents very complex mechanism for monitoring and reporting about achieving the UN sustainable development goals (SDGs) in Montenegro.

The key phase for success of the NSSD 2030 implementation, and therefore for achieving sustainable development of Montenegro, is monitoring and reporting on the results of its implementation. This paper is focused on analysis of readiness of Montenegro to prepare the first monitoring report on the NSSD 2030 implementation scheduled for 2019.

Based on conducted analysis, we can conclude that Montenegro is well advanced in preparation of all preconditions for that process. We based this conclusion on the following achievements analyzed:

- The NSSD 2030 coordination mechanism is already developed, which includes responsible data producers, line ministries, implementing agencies, advisory bodies, working groups and other local stakeholders.
- Twenty-six institutions: data generators are identified and preparatory training for them is scheduled in 2018; 137 SDG indicators will be monitored in Montenegro by 2018 (57% of the total number of indicators); the report process will include collection of 252 national indicators, 20 international indicators produced by selected international organizations and 10 composite indicators.
- Development of a progress report methodology contains four basic elements: (a) preparation of detailed methodologies for reporting on all indicators, (b) producing reporting forms on each individual indicator being monitored ("indicator's passport"), (c) preparation of reporting methodology on the dynamics of the implementation of measures envisaged by the NSOR Action Plan, including mechanism of coordination, and (d) preparation of forms for reporting on individual measures and shares from the NSOR Action Plan. The reporting methodology is in final phase of development and will be ready for the first evaluation process.
- Ministry for sustainable development and tourism has already received the IRIS platform from UNEP and by the early 2018 system will be tested within the Ministry, including

estimation of the level of its interoperability/compatibility with the Government IT system.

- The conducted above research analysis confirmed that there are complementarities between the implementation process of NSSD 2030 and the process of the EU accession; fulfillment of the obligations of accession and later obligations of full membership in the Union, contribute strongly to the achievement of the Montenegro's NSSD 2030. In the final phase of the European integration process of Montenegro it would be very useful to further strengthen recognized inter-linkages between the EU agenda and UN 2030 Agenda (SDGs and negotiation chapters and sub-chapters; closing benchmarks and SDG targets; pre-accession assistance and transposition of SDGs into national development policy context).

In order to be better prepared for development of the first monitoring report for the NSSD 2030, Montenegrin authorities should also organize a series of public awareness sensible campaign on local level about importance of UN SDGs for overall social and economic development of Montenegro.

In addition to all of the above mentioned, the Instrument for pre-accession assistance should be used more for greening of Montenegrin economy and achieving of the SDGs, along with finding additional sources for funding the NSSD 2030 implementation. European investment and structural funds support will follow the pre-accession assistance and could strongly support Montenegro's sustainable economic development projects.

Acknowledgements

The research leading to these results has received funding from the UNDP Montenegro project number 2017/045.

Conflict of interest

The authors declare no conflict of interest.

Author details

Gordana Djurovic*, Milica Muhadinovic, Vasilije Djurovic and Martin M. Bojaj

*Address all correspondence to: gordana@t-com.me

Faculty of Economics, University of Montenegro, Podgorica, Montenegro

References

[1] MONSTAT. Statistical Office of Montenegro [Internet]. 2016. Available from: https://www.monstat.org/eng/index.php [Accessed: December 6, 2017]

[2] Central Bank of Montenegro. Balance of Payment [Internet]. 2017. Available from: http://www.cb-cg.org/eng/index.php?mn1=statistics&mn2=international_economic_relations&mn3=balance_of_payments [Accessed: December 6, 2017]

[3] Government of Montenegro. Montenegro's Development Directions 2018–2021 [Internet]. 2017. Available from: http://www.gov.me/sjednice_vlade_2016/56 [Accessed: December 6, 2017]

[4] Government of Montenegro. National Biodiversity Strategy with the Action Plan for the period 2010–2015 [Internet]. 2010. p. 7. Available from http://www.mrt.gov.me/biblioteka/strategije [Accessed: January 6, 2018]

[5] Constitution of Montenegro. Official Gazette of Montenegro. No. 01/07

[6] Voluntary Montenegro's National Reviews at the UN 2016 High Level Political Forum. New York; 2016. pp. 6-7

[7] UN. Transforming Our World: The UN 2030 Agenda for Sustainable Development. Resolution Adopted by the General Assembly on 25 September 2015, A/RES/70/1 [Internet]. 2015. Available from: http://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_70_1_E.pdf [Accessed: December 6, 2017]

[8] Lucas P, Ludwig K, Kok M, Kruitwagen S. Sustainable Development Goals in the Netherlands: Building Blocks for Environmental Policy for 2030, PBL Netherlands Environmental Assessment Agency, Publication No. 1966. 2016. pp. 8-10. Available from: http://www.pbl.nl/sites/default/files/cms/publicaties/pbl-2016-sustainable-development-in-the-Netherlands_1966.pdf [Accessed: December 15, 2017]

[9] Weitz N, Persson Å, Nilsson M, Tenggre S. Sustainable Development Goals for Sweden: Insights on Setting a National Agenda. Stockholm Environment Institute, Working Paper No. 10. 2015. pp. 1-3. Available from: https://www.sei-international.org/mediamanager/documents/Publications/SEI-WP-2015-10-SDG-Sweden.pdf [Accessed: December 6, 2017]

[10] UN Statistical Commission Agrees on Global Indicators Framework [Internet]. 11 March 2016. Available from: http://www.un.org/sustainabledevelopment/blog/2016/03/un-statistical-commission-endorses-global-indicator-framework/ [Accessed: December 6, 2017]

[11] Adams J, Khan H, Reaside R, White D. Research Methods for Graduate Business and Social Science Students. SAGA Publications; 2007. pp. 161-188

[12] Ministry of Sustainable Development and Tourism of Montenegro, MSDT. National Strategy for Sustainable Development of Montenegro [Internet]. 2016. Available in

Montenegrin from: http://www.mrt.gov.me/odrzivi/165045/Obavjestenje-Nacionalna-strategija-odrzivog-razvoja-do-2030-godine.html [Accessed: November 6, 2017]

[13] UN Statistical Commission. Report of the Inter-Agency and Expert Group on Sustainable Development Goal Indicators. E/CN.3/2017/2 [Internet]. 10 March 2017. Available from: https://unstats.un.org/unsd/statcom/48th-session/documents/2017-2-IAEG-SDGs-E.pdf [Accessed: December 6, 2017]

[14] Djurovic G, Bigovic M, Milovic N. Support for further enlargement of the EU: Statistical analysis of regional differences. Journal of Balkan and Near Eastern Studies. 2016;**19**(3): 243-258. DOI: 10.1080/19448953.2017.1267394

[15] Djurovic G. Building a Sustainable Future for Montenegro Through the EU Accession Process and the UN Sustainable Development Goals. Montenegro: UNDP; 2017. pp. 6-10

[16] Ministry of Sustainable Development and Tourism of Montenegro. The information about initial results in the NSSD 2030 in the period July 2016–December 2017, Document prepared for the National Council for SDCCICZM. 15 December 2017. p. 3

[17] Ministry for Sustainable Development and Tourism. The Rulebook on Internal Organisation and Systematisation of the Ministry for Sustainable Development and Tourism [Internet]. 2017. Available from: http://www.gov.me/sjednice_vlade_2016/56 [Accessed: December 6, 2017]

[18] Official Gazette of Montenegro. Decision on Establishment the National Council for Sustainable Development, Climate Changes and Integrated Coastal Zone Management, Official Gazette of Montenegro, No. 49/13, 39/15 and 66/15. Available from: http://www.sd-network.eu/?k=country%20profiles&s=horizontal%20integration&country=Montenegro [Accessed: November 22, 2017]

[19] Ministry of Sustainable Development and Tourism of Montenegro. National Strategy for Sustainable Development of Montenegro – Summary. 2016. p. 84

[20] Djurovic G. European Union and Montenegro—The Accession Process. EU Info Centre. Podgorica. 2017. pp. 74-78. Available from: http://euic.me/wp-content/uploads/2017/12/EU-i-Crna-Gora-proces-pristupanja-WEB.pdf [Accessed: January 6, 2018]

[21] Djurovic G, Muhadinovic M. European integration priorities from the EC perspective: Towards a new methodology of comparative analysis. In: Yearbook Shipan 2015. The Atlantic Council of Croatia, Centre for International Studies, Zagreb. Vol. XIV. No. 1; 2016. pp. 45-55. Available from: http://www.atlantic-council.com.hr/images/yearbooksipan/Yearbook-2015.pdf [Accessed: January 6, 2018]

[22] Tenth Meeting of the Accession Conference with Montenegro at Ministerial Level Brussels [Internet]. 11 December 2017. Available from: http://www.consilium.europa.eu/en/press/press-releases/2017/12/11/tenth-meeting-of-the-accession-conference-with-montenegro-at-ministerial-level/ [Accessed: December 12, 2017]

[23] Government of Montenegro. The Programme of Accession of Montenegro to the EU [Internet]. 2017. Available from: https://www.eu.me/mn/crna-gora-i-eu/dokumenti/category/218-program-pristupanja-crne-gore-evropskoj-uniji [Accessed: December 22, 2017]

[24] European Commission. Next Steps for a Sustainable European Future—European Action for Sustainability [Internet]. 2016. Available from: http://europa.eu/rapid/press-release_IP-16-3883_en.htm [Accessed: December 12, 2017]

[25] The Council Conclusions. A Sustainable European Future: The EU Response to the 2030 Agenda for Sustainable Development 10370/17 [Internet]. 20 June 2017. Available from: http://www.consilium.europa.eu/en/meetings/gac/2017/06/st10370_en17_pdf/ [Accessed: December 12, 2017]

[26] Djurovic G. Importance of the EU pre-accession assistance in financing of local economic and infrastructure development: Experiences of Montenegro in MFF 2007–2013. In: Proceedings of the International Conference on Local Economic and Infrastructure Development of SEE in the Context of EU. LEIDSEE Academy of Sciences and Arts of Bosnia and Herzegovina. Sarajevo; 20 September 2013. DOI: 10.5644/PI2013-153-21

[27] European Commission. Indicative Strategy Paper for Montenegro 2014–2020. [Internet]. 2014. pp. 2-8. Available from: https://ec.europa.eu/neighbourhood-enlargement/sites/near/files/pdf/key_documents/2014/20140919-csp-montenegro.pdf [Accessed: December 22, 2017]

[28] UN System in Montenegro. Integrated UN Programme for Montenegro 2017–2021 [Internet]. 2014. Available from: http://un.org.me/integrated-un-programme-2017-2021/ [Accessed: December 22, 2017]

[29] European Commission. Reflection paper on the social dimension of Europe. COM 2017(206) 26 April 2017 [Internet]. 2017. Available from: https://ec.europa.eu/commission/sites/beta-political/files/reflection-paper-social-dimension-europe_en.pdf [Accessed: November 29, 2017]

The Application of Discrete Choice Models in Transport

Foued Aloulou

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.74955

Abstract

The discrete choice models are presented as a development and a renovation of the classical theory of choice. They have been based on the premise that the choice of economic agents is most often based on mutually exclusive alternatives or solutions, so that if the individual chooses one, he gives up the choices of the others. In this case, we speak of a discreet choice. Contrary to the microeconomic approach, discrete choice models consider that the environment that shapes the behavior of the choice of an individual is random and specific to each situation. It is influenced by a number of factors in relation to both the socioeconomic characteristics of the individual in question and the attribute being chosen and the circumstances that characterize the environment of choice. This process makes it possible to better disaggregate and personalize the behavior of economic agents and to perceive their preferences according to their motives and characteristics. The objective of this chapter is to highlight the application of these discrete choice models on the transport economy by specifying their contribution to the estimation of the transport demand and the evaluation of the severity of the accidents of the road, after having described the specificities of these models and their main characteristics and methods of application.

Keywords: discrete choice models, random utility, logit model, unordered multinomial logit, binary variables, disaggregated models, microeconometric analysis, probability of road accident, accident gravity, behavior of the modal choice for transport, values of time

1. Introduction

Discrete choice models are presented as a development and a renovation of the classical choice theory. They have overcome the rigidities and inadequacies of consumer behavior study by mentioning the problems of economic agent choices in a random and specific environment for each situation involving the choice between mutually exclusive alternatives.

These situations of choice encountered in reality do not fit with the classical assumptions of consumer theory according to which the goods are perfectly divisible and the problem of choice concerns a continuum of possibilities.

We do not seek to calculate the quantities of the various goods that an individual will need, but to determine the choice between mutually exclusive goods or alternatives from which the individual selects only one that maximizes her utility while taking into account its socioeconomic characteristics conditions and those of the alternative to be chosen.

In addition, the classical microeconomic approach considers that the chosen environment is static, stable, and transparent and that the individuals' decisions are rational and typical, so that the individual choice is deterministic and repetitive.

These hypotheses have limited the field of research in the analysis of demand and individual behavior of consumers. This demand was analyzed using an aggregated approach to macroeconomic variables.

In contrast to this approach, discrete choice models consider that the environment that shapes the individual choice behavior is random and specific to each situation. It is influenced by a number of factors in relation to both the socioeconomic characteristics of the individual in question and the attribute being chosen and the circumstances that characterize the environment of choice.

As a result, the decision-making process of economic agents, which is based on the maximization of an objective function under constraint, is represented by a description of the different characteristics, both the attributes of the alternatives to be considered and the socioeconomic characteristics of consumers as well as the environment of choice.

Each individual has an objective function that seeks to optimize it to achieve the best decision. In a random, uncertain environment, where the choice is not reproducible, this objective function is of random type, and the economic calculation is probabilistic [1].

This process allows to better disaggregate and personalize the economic agent behavior and to perceive their preferences according to their motives and characteristics.

This contribution will enable economists to detect the effect of each element determining the consumer's choice on their consumption behavior, as well as to detail and explain the function of individual and global demand.

These discrete choice models have been very successful thanks to their ability to analyze the random behavior of individuals when making a decision to choose a given solution or to appreciate the valuation of goods or actions.

They have been the subject of several theoretical developments and empirical validations. Their manipulation has become easier thanks to the availability of increasingly disaggregated data and advances in econometric techniques and software.

They were applied for the first time to estimate transport demand. They were subsequently generalized and applied to deal with all the problems of choice concerning mutually exclusive alternatives or also to assess the subjective value of an event.

The transport economy is a privileged domain of application of these probabilistic models. Indeed, the individual who travels every day must choose a particular mode of transport, such departure time or such other, such or such journey, such destination or such other, such or such movement frequency, etc.

The risk analysis of road accidents in terms of frequency and/or severity should then predict the probability that an individual with specific socioeconomic characteristics and driving in a given traffic environment is involved in a road accident and/or that the accident incurred will be of a given severity.

This type of model seeks to study the behavior of transport users regarding their choice of mode of transport or also the risk of transport and to anticipate the modifications brought by changes in the mode characteristics or socioeconomic variables of the decision-maker.

Several families of discrete choice models have been developed and applied (probit, logit, dichotomic logit, multinomial logit, conditional logit, mixed logit, nested logit, etc.), each of which is specified either by the nature of the explanatory variables selected and which characterize the alternatives and/or the individuals or by the statistic distribution law that follows the error terms or its ability to overcome the constraint of independence from irrelevant alternatives (IIA).

The aim of this chapter is to present these discrete choice models while focusing on the unordered multinomial logit model that is most used in empirical studies. This chapter will consist of two parts, the first of which will present the specificities of the multinomial logit model while reviewing its main tools for estimating and testing statistical validation and the interpretation of its coefficients. In the second part, we will try to apply this model to two studies on phenomena related to transport. The first concerns the modal choice of urban transport users for personal travel reasons in the city of Sousse (Tunisia), and the second phenomenon will deal with accidentology by trying to estimate and analyze the severity levels of road accidents in Tunisia.

The general problem of the first application therefore concerns the estimation of the urban passenger transport demand structure for the city of SOUSSE, using discrete choice models.

These models calculate, from a given observations, the probability that an individual selects a particular mode of transportation from a set of possible and mutually exclusive choices. In the second application, we seek to predict the probability of a driver’s exposure to a given gravity accident. This severity may depend on three components: the driver; the vehicle, its condition of use; and the infrastructure.

These various components constitute the traffic system and determine road safety. They interact at a given time and place to explain the occurrence and severity of an accident. Several quantitative and qualitative variables can be identified and measured to describe these components. The purpose of this study is to show how and by how much these explanatory variables affect the severity of a traffic accident. The structure of the estimates is based on disaggregated data collected: on the one hand, the household-displacement survey carried out in 2004 and, on the other hand, the survey sheets proposed by the National Observatory of Circulation (Tunisia). We will then present and interpret the main results of the estimation of these two applications. This interpretation remains a difficult exercise, especially when one addresses an uninitiated public to this type of discrete choice modeling and qualitative econometrics.

2. Presentation of the multinomial logit model

Among the discrete choice models, the multinomial logit model is the most widespread and used in many different fields. This disaggregated model seeks to study the decision of choice or the perception of the value of an event among a set of mutually exclusive alternatives.

Individual choice behavior or the perception of the value of an event is considered as a selection process between several mutually exclusive contingencies that belong to a set of eventualities. The eventuality chosen by an individual will be the one that optimizes its objective function. The decision taken will therefore result from an optimization process reflecting a rational behavior of the individual. As long as the choice of the individual is established in random circumstances that never occur identically, the modeling will be probabilistic. Nobody can correctly predict the choice of the individual, but he can estimate the probability of this choice according to the circumstances of choice and the socioeconomic characteristics of the individual as well as the technique of the alternative to choose.

The multinomial logit model will therefore allow us to estimate the probability that an individual i chooses an alternative j in given circumstances characterizing the environment of choice. This probability can be expressed as a linear (or nonlinear) function of all the variables characterizing this environment of choice (X_k).

Formally, this probability is written according to the following expression:

$$P_{ij} = F_{ij}\left(\sum_{k=1}^{K} \alpha_k X_k\right) \tag{1}$$

P_{ij} is the probability that an individual i establishes the choice j.

The parameters α_k are unknown that we seek to estimate. They, respectively, reflect the weight of each explanatory variable (X_k) in the determination of the probability P_{ij}.

F_{ij} is a distribution function of the explanatory variables and the vector of parameters α_k.

In discrete choice models, the endogenous variable we seek to explain is a qualitative and discrete variable. It illustrates the individual's choice or level of appreciation of the psychological value of a given event. This variable to be explained will take integer values that vary between 1 and J depending on the number of alternatives that make up the entire choice of the individual.

2.1. Specificity of the model

For a more detailed discussion, consider that an individual i of a sample N (such as i = 1 … N) is in front of a set of choices (modes of transport, port of call, types of equipment, place of residence, etc.) or belongs to a given category of population or appreciation of a psychological value of a given phenomenon (risk of accident, time value, etc.) j ($j \in J / j = \{1,2,3\ldots J\}$).

Individual i chooses the alternative j that optimizes (maximizes or minimizes) its objective function (S_i).

The variable to be explained is expressed as follows:

$$\begin{cases} Y_i = 1 \text{ if the individual i chooses option 1, means when } S_{i1} = \underset{j=1\ldots J}{\text{Max}}(S_{ij}) \\ Y_i = 2 \text{ if the individual i chooses option 2 } \left(S_{i2} = \underset{j=2\ldots J}{\text{Max}}(S_{ij})\right) \\ Y_i = J \text{ if the individual i chooses option J } \left(S_{iJ} = \underset{j=1\ldots J}{\text{Max}}(S_{ij})\right) \end{cases} \quad (2)$$

$\mathbf{Y_i}$ designates the choice observed and S_{ij} the level of objective function that the choice of the alternative j gives to the individual i.

The objective function of the individual i is dependent on the socioeconomic characteristics of the individual i (X_{ik}), on the technical ones of the option to be remembered (W_{jh}), and on those of the environment of choice (E_{jm}):

$$S_{ij} = S(X_{ik}, W_{jh}, E_{jm}) \quad (3)$$

It should be emphasized that these variables may be specific to each option *j* and/or to each individual *i*.

A specific variable to the individual is a variable that remains the same regardless of the option chosen by the individual, while a specific variable to the alternative j depends on the specific conditions to the choice.

As long as the objective function is random, we can break it down into two parts: one is determinist ($Vij(X_{ik}, W_{jh}, E_{jm})$) and the other is random (ε_{ij}):

$$S_{ij} = V_{ij}(X_{ik}, W_{jh}, E_{jm}) + \varepsilon_{ij} \quad (4)$$

The deterministic function (Vij) reflects the perception of an average individual of the satisfaction provided by the choice of the alternative j. It can take many forms, but the linear form is the simplest to estimate and interpret:

$$V_{ij} = \alpha_{0j} + \sum_{k=1}^{K} \alpha_{jk} X_{ik} + \sum_{h=1}^{H} \beta_h W_{jh} + \sum_{m=1}^{M} \mu_m E_{jm} \quad (5)$$

The arguments of this deterministic function can be quantitative as well as qualitative variables expressed in the form of a binary variable and/or polytomous.

The weighting coefficients of the explanatory variables α_{jk}, β_h and μ_m reflect the relative importance of each of the explanatory variables relating, respectively, to the socioeconomic characteristics of the individual, the attributes of the alternative, and the environment of choice, in the explanation of the objective function.

However, these coefficients cannot be directly interpreted as the impact of the absolute or relative variation of one of the explanatory variables on the probability of choosing alternative j (or belonging to a population category j). They indicate only the variation direction of this probability but not their amplitude. If they are positive, they positively affect the probability of choice and vice versa. Moreover, the interpretation of these parameters is not identical between the explanatory variable categories [2].

α_{0j} is a constant that can reflect the impact of the other explanatory variables not included in the model for one reason or another and the imbalance observed in the sample between the individuals choices. Probably the individuals who opt for choice 1 will be more numerous than those opting for the second or the jth choice.

The random term of the objective function (ε_{ij}) reflects the not observed behavior of individuals. Thus, two individuals with the same observed characteristics and faced with the same set of choices can make different decisions. It therefore implies the probabilistic nature of discrete choice models. It originates from several sources such as the measurement error on the variables or in the objective function specification, etc. [3].

The specification of the statistic distribution law of this random part makes it possible to define the definitive profile of the choice probability function (P_{ij}). Various specifications of this law were used, but only two were mainly retained: a Weibull distribution (logit model) [4] and a multidimensional normal distribution (probit model) [5].

The individual i will choose the alternative j from a set of alternative J, if and only if $S_{ij} > S_{il}$. The probability of this choice is

$$P_{ij} = \Pr\left(S_{ij} > S_{il}\right) = \Pr\left(V_{ij} + \varepsilon_{ij} > V_{il} + \varepsilon_{il}\right) = \Pr\left(V_{ij} – V_{jl} > \varepsilon_{ij} - \varepsilon_{il}\right) \forall j \neq l \in J. \quad (6)$$

If the error terms are independent and identically distributed according to Weibull[1]'s law, the probability given by the logit model is expressed by the following relation:

$$Pij = \frac{\exp\left(V_{ij}\right)}{\sum_{j=1}^{J} \exp\left(V_{ij}\right)} \quad (7)$$

By respecting the laws of probability such as $0 < P_{ij} < 1$ and $\sum_{j=1}^{J} Pij = 1$, the probability associated with Jth alternative does not need to be specified since it can be calculated from the rest of the calculated probabilities. This excluded alternative of the model will be considered as the reference situation that one seeks to compare it with the observed situation. The coefficients associated with this alternative J will be considered null ($\alpha_{kJ} = \beta_{hJ} = \mu_{mJ} = 0$):

[1]A random variable follows a Weibull or double exponential law or Gumbel distribution, if its cumulative function is written:F $(\varepsilon) = \exp.(-\exp.(-\varepsilon))$

http://dx.doi.org/10.5772/intechopen.74955

$$\mathrm{P_{ij}} = \frac{\exp\left(V_{ij}\right)}{1+\sum_{j=1}^{J-1}\exp\left(V_{ij}\right)} \vee \mathrm{j} = 1\ldots\mathrm{J}-1 \tag{8}$$

$$\mathrm{P_{iJ}} = 1-\sum_{j=1}^{J-1}P_{ij} = \frac{1}{1+\sum_{j=1}^{J-1}\exp\left(V_{ij}\right)} \tag{9}$$

The ratio between Eqs. (8) and (9) gives the following expression: ∨ j = 1…J-1

$$\frac{P_{ij}}{P_{iJ}} = \exp\left(V_{ij}\right) \Rightarrow Log\left(\frac{P_{ij}}{P_{iJ}}\right) = V_{ij} = \alpha_0 + \sum_{k=1}^{K}\alpha_{jk}X_{ik} + \sum_{h=1}^{H}\beta_h W_{jh} + \sum_{m=1}^{M}\mu_m E_{jm} \tag{10}$$

$$\frac{\partial Log\left(\frac{P_{ij}}{P_{iJ}}\right)}{\partial X_{ik}} = \alpha_{jk} \tag{11}$$

2.2. Model interpretation

Unlike linear regression econometric models whose estimated coefficients can be easily interpreted as the elasticity's or the marginal impact of the explanatory variable on the variable to be explained, the interpretation of the coefficients of the logit model is more delicate.

To understand the interpretation of these coefficients, we must proceed with a reorganization of the logit model equation (Eq. (8)). It was better to express the probability of each alternative j with respect to the reference situation assumed beforehand (alternative J). For all j = {1, …, J-1}, we must calculate the ratio between the probability of the choice of the alternative j and that of the alternative J (Eq. (10)). When only one explanatory variable varies (we go from X_{k0} to X_{k1}), while keeping the other variables constant, we can measure its effect on the probability ratio between the observed alternative and the reference one:

$$\frac{P_{ij}(X_{k0})}{P_{iJ}(X_{k1})} = \frac{P_{ij}(X_{k0})/_{P_{iJ}(X_{k0})}}{P_{ill}(X_{k1})/_{P_{ij}(X_{k1})}} = \frac{\exp\left(V_{ij}(X_{k0})\right)}{\exp\left(V_{ij}(X_{k1})\right)}$$

$$\mathrm{Log}\left(\frac{P_{ij}(X_{k0})}{P_{ij}(X_{k1})}\right) = \mathrm{V_{ij}}\ (\mathrm{X_{k0}}) - \mathrm{V_{il}}\ (\mathrm{X_{k0}}) = \alpha_{\mathrm{jk}}\ (\mathrm{X_{k0}} - \mathrm{X_{k1}}) \tag{12}$$

α_{jk} measures the effect of changing the X_k variable from variable X_{k0} to X_{k1} on the probability of choosing alternative j rather than the reference alternative J.

When only one explanatory variable varies (e.g., we go from X_{k0} to X_{k1}), while keeping the other variables constant, we can measure its effect on the probability ratio between the observed alternative and the reference one.

In the multinomial logit model, several categories of explanatory variables of both qualitative and quantitative orders can be integrated. The interpretation of continuous variables of a quantitative nature does not pose any problem. The exponential value of the coefficient associated with this variable measures the unit variation impact of this explanatory variable on the probability of choosing the alternative j rather than the reference alternative J.

For qualitative variables, we distinguish between binary ones which will be coded in 0 and 1 and those polytomous which express themselves in several modalities. For example, sex as a variable characterizing the individual can be integrated in the model as a binary variable coded 0 if the individual is male and 1 if he is female. For the professional category variable of the individual, there are more than two functions. In this case, the integrated variable must take an integer from 1 to n according to the number of observed professions. For these explanatory variable categories, a reference situation must always be chosen in order to interpret their estimation coefficients. For binary variables, if the reference situation is the one relating to the code 0 (e.g., male sex), the exponential function of the associated coefficient is interpreted as the effect of the individual passing from the reference situation (0) to that observed (1) on its probability of choosing the alternative j rather than the reference alternative J.

The interpretation of these estimation coefficients becomes more difficult in the presence of a polytomous explanatory variable. These modalities present a collinearity that must be avoided by eliminating a modality and limiting itself to reasoning only according to n-1 modalities which remain. The eliminated modality will be the one of reference.

In this case, the associated coefficient must be interpreted as a function of two references: one relative to the choice J and the other relating to the explanatory variable. For example, if the explanatory variable is the socioprofessional category of the individual and the reference category is the "worker," the exponential function of the coefficient associated with the "manager" variable, for example, indicates the impact of being a "manager" rather than a "worker" on the probability of choosing the alternative j rather than the reference alternative J.

We can also evaluate the impact of the variation of the explanatory variable on the comparative probability of the individual choice by the elasticity. Elasticity is defined as a percentage change in the probability of choosing alternative j rather than alternative J resulting from a 1% change in one of the characteristics of alternative j (W_j) by keeping the other arguments of the probability function constant. The advantage of the interpretation of coefficients in terms of elasticity than unitary variation lies in the fact that the elasticity is calculated independently of the units of measurement of the explanatory variables.

The elasticity calculation constitutes a very indispensable information base for decision-makers to learn about the most influential factors in the individual behavior and determine their optimal action plan in order to achieve their goals.

The elasticity can be calculated with respect to all the arguments of the probability function. We speak of direct elasticity when it is calculated with respect to the arguments relating to the chosen alternative j and of the cross elasticity, when it is calculated with respect to the arguments relative to the other alternatives l # j.

This direct individual elasticity is written:

$$\ell P_j/W_j = \frac{\partial LogP_{ij}(Wjh)}{\partial LogW_{jh}} = \beta_h W_{jh}\left(1 - P_{ij}\right) \tag{13}$$

where W_{jh} is the hth argument of the vector characterizing the alternative j (W_j), β_h being its relative parameter, and P_{ij} is the probability of choice of the eventuality j by the individual i.

The cross elasticity is written:

$$\ell P_j/W_{hl} = \frac{\partial LogP_{ij}}{\partial LogW_{hl}} = -\beta_h W_{hl} P_{ij} \tag{14}$$

2.3. Property of independence from irrelevant alternatives

The logit multinomial model is based on a fundamental assumption but constraining in empirical studies: independence from irrelevant alternatives (IIA).

This hypothesis implies that the choice of the individual will always be the same regardless of the number of alternatives proposed, so that the probability that an individual chooses an alternative j remains constant even if other alternatives are included in the set of considered choices. This assumption imposes the independence between the alternatives, which excludes any possibility of substitution between them. It implies that the ratio of probabilities of choice between two alternatives remains unchanged following the addition or the removal of one or more alternatives from all the choices.

This property (IIA) facilitates estimation and prediction because it implies that the model can be estimated from binomial choice data or by reduced attention to choices in a limited subset of the total set of choice. Therefore, if the assumption (IIA) is verified, the model structure and the estimated parameters for the explanatory variables should remain unchanged when performing the estimate on a small subset of the set of choices.

However, this hypothesis of the logit model has been criticized by several authors, thus limiting its practical relevance. The nested logit model has been developed to overcome this property of IIA. Referring to Eq. (10), we find that the probability ratio between the two alternatives j and J does not depend on the other possible alternatives, hence the property of the independence from irrelevant alternatives (IIA).

3. Case studies

The multinomial logit model has been the subject of several empirical studies on the analysis of various behavioral phenomena of the individual such as the choice of modes of transport [6, 7], the choice of ports of call [8, 9], the choice of the professional function [10], the choice of place of residence [11–13], discrimination in the job market [14], the severity of road accidents [15, 16], the valuation of transport time [17, 18], etc.

In this section, two case studies will be analyzed and interpreted and treated in the case of my research work and supervision on topics particularly related to transport economics. The first study deals with the modal choice problem and the second with road accidentology. These case studies will allow us to better value the practical interest of these models of discrete choice, to account for the diversity of fields of application of these models and to present real results allowing a better understanding of the coefficient interpretation according to the qualitative and quantitative nature of the explanatory variables.

3.1. Modal choice study

In the first case study, we will analyze the transport behavior of transport users by estimating an unordered multimodal logit model on a sample of urban transport users from the city of Sousse (Tunisia). This study will allow us to analyze the transport demand and to identify several information about the direct and indirect elasticities of transport demand in relation to the different attributes of the modes envisaged (transport price, travel time, waiting time, etc.) and to calculate the psychological value of transport time.

The behavior of individual choice in the transport market is considered as a selection process between several modes of transport available (car, bus, metro, two-wheeled vehicle, etc.). The transport user will choose the mode that maximizes its utility.

However, this utility is unobserved. What we actually perceive is the modal choice of the user. In this context the variable to be explained will be the choice established by the transport user and not its utility.

This endogenous variable is thus discrete and qualitative which will take a limited number of integer values, whose each value illustrates a particular choice. This is the foundation of the discrete choice model.

We assume that the choice modal set is composed of three modes such as the private car, bus, and taxi (j = 1, 2, 3).

The variable to be explained is expressed by the following system: ∨ i = 1…n:

$$Y_{ij} = \begin{cases} 1 \text{ if the user i prefers the private car (PC) to other modes} \\ 2 \text{ if the user i prefers the taxi to the other modes} \\ 3 \text{ if the user i prefers the bus to other modes} \end{cases} \quad (15)$$

To avoid collinearity between modal choices, we eliminate the third choice (bus) while considering it as the reference situation. This reference situation will serve us to better interpret our results and evaluate the impact of changing explanatory variables on the probability of choosing the mode j (PC or taxi) rather than the bus mode.

The user i that prefers the private car to the bus mode implies that he gets more satisfaction by using the private car than the bus to get to work. This satisfaction can be systematized by a linear indirect utility function.

$Y_{ij} = 1$ (choice of the PC) if and only if U_i (PC) > U_i (bus) and U_i (PC) > U_i (taxi).

Formally, the indirect utility function U_{ij} depends on a certain number of variables relating to the attributes of the chosen transport mode (W_j) as well as to the user's socioeconomic characteristics (X_i).

Many explanatory variables can be integrated and tested which characterize as well the individual as the attributes of the mode to choose.

For example, four explanatory variables characterizing the transport user such as income, sex, age, and household size and three explanatory variables characterizing the modes such as the kilometric price of the use of each mode, travel time, and access time to each mode. All variables are continuous except the sex will be expressed as a binary variable coded 0 if the user is female and 1 otherwise. The price, travel time, and access time vary for the same individual from one mode of transport to another, while the variables characterizing the user do not vary according to the mode.

With reference to Eq. (10), our model will be expressed by the following relation:

$$\begin{aligned} Log\left(\frac{P_{ij}}{P_{iJ}}\right) &= U_{ij} = \alpha_{0j} + \sum_{k=1}^{4} \alpha_{jk} X_{ik} + \sum_{h=1}^{2} \beta_h W_{jh} \\ &= \alpha_{0j} + \alpha_{1j}\,\mathrm{R_i} + \alpha_{2j} S_i + \alpha_{3j} A_i + \alpha_{4j}\,\mathrm{D_i} + \beta_1 \mathrm{P_j} + \beta_2\,\mathrm{tp_j} + \beta_3\,\mathrm{ta_j} \end{aligned} \tag{16}$$

where R_i, S_i, A_i, D_i, P_j, tp_j, and ta_j are, respectively, income, sex, age, household size i, price, travel time, and access time by the mode j.

α_{jk} and β_h are the coefficients to estimate. The weighting coefficients relating to the socioeconomic characteristics of the users (α_{jk}) are specific to each mode of transport, while those of the attributes (β_h) are constant and do not vary according to the mode or the user.

α_{0j} is a constant that varies from one mode to another.

The estimate of this model requires data by user-displacement couple which are collected through the household-displacement survey database dated 2004 for the city of Sousse (Tunisia). Our sample is made up of 500 households distributed homogeneously over the entire agglomeration.

We are interested to a particular aspect of displacement having a professional motive, on a path home-work that converges to the city center during the morning rush hours by bus, private car, and taxi.

Table 1 presents the results of our estimation. It describes the estimated values of the coefficients associated to the explanatory variables; their standard error (in parenthesis), in a second column; their degree of significance in the third column; and their exponential function in the last column.

All variables are statistically significant for thresholds going from 1–10%; several indicators of quality adjustment of the model were developed to evaluate the predictive ability of the model

Variable	Coefficient	Student's T-test	Exp (coef)
Constant 1	−2.014 (0.31)	−10.66	
Constant 2	−12.82 (0.401)	−15.1	
Income 1	0.218 (0.11)	2.15	1.243
Income 2	0.853 (0.063)	32.14	1.089
Age 1	0.04 (0.013)	14.01	1.04
Age 2	−0.13 (0.0912)	−2.06	0.87
Sex			
Woman	Ref		
Man 1	−0.3341 (0.139)	−3.47	0.715
Man 2	−0.93 (0.173)	−8.7	0.394
Household size 1	0.2943 (0,112)	2.15	1.34
Household size 2	−0.25 (0.125)	−2.25	0.779
Price	−0.0167 (0,078)	−0.67	0.983
Travel time	−0.123 (0,1218)	−2.41	0.884
Access time	−0.38 (0.105)	−4.13	0.68

Standard error in parentheses:
- Number of observations = 500
- Log likelihood = − 116,6517
- Pseudo R^2 = 0.48.

Table 1. Parameter estimates of modal chose model.

(Mc Fadden's pseudo R^2, Estrella indicator, Ben Akiva and Lerman indicator, etc.) [19]. According to the software used (STATA 11), only the Pseudo R^2 and the log likelihood are calculated. Their values show although overall; the explanatory variables selected explain at high degrees the modal choice:

$$Log\left(\frac{P_{i(j=1)}}{P_{i(j=3)}}\right) = -2.014 + 0.218\ \mathrm{R_i} + 0.04\ \mathrm{A_i} - 0.33\ S_i + 0.29\ \mathrm{D_i} - 0.00167\ \mathrm{P} - 0.123\ \mathrm{tp} - 0.38\ \mathrm{ta}$$

The constant parameter illustrates the heterogeneity in the representativeness of the individual choices in our sample. This coefficient is significantly higher for the PC than the taxi, reflecting thus the higher proportion of taxi users compared to those of the PC.

We can interpret the parameter associated with an explanatory variable by fixing the other variables for a given level and varying the said variable. The exponential function of this coefficient indicates the effect of this variation on the probability of choosing the PC mode rather than the bus mode. For example, when a household's income increases by one unit, the probability of choosing the PC mode instead of the bus mode increases by 24.3% (1.243–1).

For the age variable of the users of the PC, odds ratio is 1.04. This implies that a year furthermore increased by 4% the probability of choosing the PC than the bus.

Concerning the coefficient associated with the gender variable, it is interpreted as follows: a man has 28.5% (1–0.715) and 60.6% (1–0.715) of luck less than a woman to choose, respectively, the PC and the taxi rather than the bus, everything else being equal.

The increase of the members of a household of a person increases the probability of choosing PC but brings down the probability of choosing the taxi compared to that of bus. Indeed, one more member in the family increases by 34% the probability of choosing the PC rather than the bus and decreases the probability of choosing the taxi rather than the bus of 22.1%.

In fact, by becoming a householder, we will prefer the car better than the bus thanks to its advantages of availability, flexibility, and accessibility.

For the other explanatory variables characterizing the modes of transport (P, tp, and ta), they negatively affect the probability of choosing both the private car and the taxi to the bus.

The estimated coefficients for these variables are, respectively, −00167, −0.123, and −0.38. This implies that if the cost per kilometer of transport or the travel time or the access time to the mode of transport increases of a unit while keeping all the other variables constant, the probability of choosing the car mode compared to the bus decreases by 1.65% (1-exp (−0.016)), 11.57 and 32%, respectively. The user of the car has a greater sensitivity to the transport time than the cost. This explains well the fundamental reason for the dominance of the car in the modal split, thanks to its quality of service that is better than the bus particularly in terms of access time.

So, the cost and the time of transport play a determining role in the decisions of the modal choice and affect negatively the transport demand as well as the modal sharing between the car and the bus.

The weights of the explanatory variables can be interpreted economically as the marginal utilities of each indirect utility function argument (U_{ij}). They indicate the effect of unitary change of each variable on the utility of the mode (PC).

$$Um_{i1P}(X_k) = \frac{\partial U_{i1}(X_k)}{\partial X_k} = \alpha_{k1}$$

If $X_i = S_i$ is the sex variable, $\alpha_{31} = -0.33$; this implies that the man is less satisfied than the woman by the use of the particular car.

If $X_i = D_i$, $\alpha_{41} = 0.29$; this implies that the more the household is composed of a larger number of individuals, the greater its satisfaction of the use of a private car is important. One more member in the household increases the satisfaction of PC use by 0.29 units.

The weighting coefficients related to the attribute variables of the PC mode are all negative, implying that the increase in both the cost of transport induced by the increase in the fuel price or the cost of acquisition of the PC, as well as the travel time whether it is in traffic or the search for parking caused by congestion, creates a disutility for users of the PC.

We can see that the choice probability of the PC is more sensitive to the search time of parking than the travel time and the costs of displacement. The parking search time provides a triple disutility compared to that caused by the travel time by the user of the PC:

$$\frac{Um_{i1}(ta)}{Um_{i1}(tp)} = \frac{\beta_{31}}{\beta_{21}} = 3.09 = \mathrm{TMS}_{\mathrm{ta/tp}}$$

The ratio of the marginal utilities of the two variables ta and tp measures the marginal rate of substitution of waiting time for travel time. The user agrees to spend 3.09 minutes more on his journey to save an extra minute to search for parking to his PC.

The ratio of marginal utilities of the two variables Tp and P measures the marginal rate of substitution of money for travel time:

$$\frac{Um_{i1}(tp)}{Um_{i1}(P)} = \frac{\partial U_{i1}}{\partial tp} \bigg/ \frac{\partial U_{i1}}{\partial P} = \frac{\beta_{21}}{\beta_{11}} = 73.65$$

The PC user agrees to pay 73.65 currency units to gain a minute in his trips the equivalent of 1.82 USD per hour. The $\mathrm{TMS}_{\mathrm{Tp/P}}$ measures the price of time granted by the user of the PC having given socioeconomic characteristics.

The value of time is defined as the price that the individual is willing to pay to save a unit of time given its motive for displacement and its socioeconomic characteristics.

This value is obtained by comparing the coefficient associated with the time variable and the one associated with the displacement cost variable. It corresponds to the level of disutility associated with the time spent in a given path.

From these results, it is thus possible to detect the most influential determinants on the modal choice of the transport users and consequently determine the function of the transport demand.

3.2. Accidentology study

Discrete choice models were also used to estimate the risk of road accidents. Several authors [15, 16, 20–22] used these models to calculate the probability of occurrence of a road accident and to detect the correlation between driver behavior, the characteristics of the traffic system, and the accident severity. They tried to model the driver's accident risk perception according to a set of factors describing the traffic system. This risk perception expresses a subjective, personal, and psychological assessment of the danger that every motorist seeks to minimize. Usually, the more this risk perception is high, the more lower the accident severity will be. And the more this risk perception is weak, the higher the probability of a serious or fatal accident is high. The risk perception will influence both the occurrence of the accident and the severity of the injuries.

These disaggregated models help to better describe and analyze the risk and severity of an accident by treating each accident separately in Ref. to its circumstances and the driver's individual behavior. The general idea is that the accident severity can be explained according to both the socioeconomic characteristics of the driver who is the victim of a road accident, and of his driving behavior, and the circumstances of the traffic (state of the vehicle, infrastructure, and meteorology).

The objective of this case study is to analyze the severity of road accidents in Tunisia. We seek to estimate a multinomial logit model to predict the probability of a driver's exposure to a given gravity accident. The structure of the estimate is based on disaggregated data collected following the study of survey sheets proposed by the National Observatory of Circulation (Tunisia). Our sample consists of 300 randomly selected traffic accident victims from survey cards dated 2010. In our model, we defined three levels of gravity such as fatal accident, injury accident, and accident-causing material damage.

The endogenous variable is an unordered multinomial variable that will be scored from one to three to indicate the severity level of the observed accident. It will be illustrated by the following system:

$$Y_{ij} = \begin{cases} 1 \text{ if the observed accident is fatal} \\ 2 \text{ if the observed accident only causes injuries} \\ 3 \text{ if the observed accident causes only material damage} \end{cases} \quad (17)$$

The objective function of the driver is his risk perception. Each driver seeks to maximize his risk perception to better estimate the danger of the road and consequently reduce the accident severity.

$Y_{ij} = 1$ if the risk perception is minimal, so that the driver may have a serious accident.

To estimate the probability of exposure of an individual i (such as i = 1, 2, …, 300) to a traffic accident of severity level j (such as j = 1, 2, 3), it is necessary to cross the multinomial variable Y with a number of explanatory variables.

Referring to the accidentology literature, this gravity may depend on three components: the driver, the vehicle and its condition of use, and infrastructure. These various components constitute the road traffic system and determine the road safety. They interact at a given time and place to explain the occurrence and severity of an accident.

Several quantitative and qualitative variables can be identified and measured to describe these components.

We designate by $S_{ij} = S(X_{ik}, V_{ih}, R_{jl}, E_{jm})$ the objective function of the individual i.

It is dependent on both the socioeconomic characteristics of the individual i (X_{ik}) such as sex (X_1), age (X_2), householder (X_3), vigilance level (X_4), and seat belt wearing (X_5); vehicle-operated characteristics (V_{ih}) such as age (V_1), size (V_2), speed (V_3), and airbag equipment (V_4); and those of the borrowed road (R_{jl}) such as road condition (R_1), vision (R_2), lighting (R_3), position of the accident (R_4) and the environment (E_{jm}) such as climate (E_1), time (E_2), and agglomeration (E_3) (**Table 2**).

The obtained results showed that all the variables retained are statistically and theoretically significant and explain at different degrees the severity of an accident.

Explanatory variables	Coefficients	
	Fatal accident	**Injury accident**
Constant	*0.2472(0.006)****	2.5440(0.026)**
Sex (X_1)	0.6233 (0.082)*	0.6549 (0.060)*
Driver's age (X_2)	0.3968 (0.030)**	−0.2925 (0.020)**
Householder (X_3)	−0.1323 (0.744)	0.7078 (0.402)
Vigilance level (X_4)	0.3375 (0.084)*	0.4374 (0.086)*
Seat belt wearing (X_5)	0. 9995 (0.001)***	−0.5509 (0.062)*
Vehicle age (V_1)	*0.6950(0.040)***	−0.4719(0.050)**
Vehicle size (V_2)	0.7906 (0.010)***	−0.5346 (0.006)***
Speed (V_3)	0.1888 (0.049)**	−0.1671 (0.075)*
Airbag equipment (V_4)	0.1022 (0.002)***	−0.6098 (0.020)**
Road condition (R_1)	−0.4400 (0.089)*	0.4722 (0.093)*
Vision (R_2)	0.5127 (0.013)**	−0.5983 (0.004)***
Lighting (R_3)	−0.7897 (0.080)*	0.8874 (0.089)*
Position of the accident (R_4)	−0.1362 (0.060)*	0.2368 (0.068)*
Climate (E_1)	−0.2630 (0.050)**	0.2367 (0.049)**
Time (E_2)	−0.9784 (0.009)***	−0.6947 (0.010)***
Agglomeration (E_3)	−0.7772 (0.012)**	0.1086 (0.015)**

Standard error in parentheses.
***Significant to only one of 1%; **significant to only one of 5%; *significant to only one of 10%.
The incorporeal accident is the reference category:
- Number of observations = 300
- Log likelihood = − 195.969
- Pseudo R^2 = 0.405.

Table 2. Parameter estimates of accidentology study.

Referring to Eq. (10), the estimated value of the weighting coefficient of the sex variable (X_1) corresponds to the ratio of relative probabilities as follows:

$$\log\left(\frac{P(Y=1;X_1=1)/P(Y=3;X_1=1}{P(Y=1;X_1=0)/P(Y=3;X_1=0}\right)=0.62$$
$$\Rightarrow \frac{P(Y=1;X_1=1)/P(Y=3;X_1=1}{P(Y=1;X_1=0)/P(Y=3;X_1=0}=\exp(0.62)=1.86$$

The sign of the coefficient is positive. It implies that the gender variable has a positive effect on the probability of being a victim of a fatal bodily injury rather than an intangible accident. We can interpret this coefficient as follows: a man has a probability of 86% to be the victim of a fatal accident rather than an intangible accident. This probability rate is 92.5% in the case of an injury accident.

For the vigilance variable (X_4), the ratio of relative probabilities is equal to 1.4 (exp (0.33)). This implies that driving without concentration (zero vigilance: alcohol, sleeping while driving, etc.) increases the probability to be the victim of a fatal accident of 40% compared to an intangible accident and increases the probability of being injured in an accident by 55% compared to an intangible accident. Tiredness, driving drowsiness, and alcohol are the factors that increase road insecurity and the probability of having more and more serious accidents. These factors are particularly related to the irresponsible behavior of the driver.

According to the weighting coefficient of the seat belt-wearing variable (X_5), the nonuse of the seat belt increases the probability of going from an accident with material damage to a fatal accident of 170%. However, the coefficient of this same variable (X_5) relative to the injury accident alternative is negative. This implies that not wearing a seat belt reduces the probability of being an injury accident victim in relation to an intangible accident. Not wearing a seat belt does not prevent injury accident, but it reduces the risk of a fatal accident. Therefore, not wearing a seatbelt is a key factor in the explanation of fatal traffic accidents.

For the age variable (X_2), 1 more year in the driver's age reduces the probability of being a fatal accident victim rather than an intangible accident of 48%. The fatal accident risk decreases with the increase of the driver's age. For the variable speed (V_3), its coefficient is 0.188 in the event of a fatal accident and -0.167 in the event of an injury accident. These coefficients are interpreted as follows: any increase in the circulation speed of 1 Km/h causes an increase in the probability of the fatal accident risk compared to an intangible accident of 20% (exp (0.133) $-$ 1) and a decrease in the injury accident risk of 18% compared to the reference situation. So speed is a risk factor whose excess increases the accident severity.

Concerning the variable airbag equipment (V_4), it positively affects the probability of occurrence of a fatal accident, but negatively that of an injury accident. In other words, a car not equipped with an airbag increases the probability of a fatal accident compared to an injury accident by 10.7%, while it decreases the probability of an injury accident by 45%.

With regard to the infrastructure characteristic vector (road condition, position of the accident, lighting, and vision at the time of the accident), it represents one of the elements that contributes to the explanation of the probability of a fatal accident risk.

All other things being equal, the driver reduces the probability of a fatal accident compared to an intangible accident by 12.7% when it avoids overtaking on a straight-line trajectory, by 35.6% when he takes a good quality road and by 54% when the road is illuminated and the vision is clear.

In terms of environmental factors, we find that the climate (E_1), the time, and the location of the accident negatively affect the probability of occurrence of a fatal accident compared to an intangible accident. In other words, driving in an environment characterized by a normal, sunny day and in an agglomeration zone reduces the probability of a fatal accident in relation to an intangible accident by 30% compared to the driving in the rain, by 60% compared to night driving, and by 54% compared to an out agglomeration driving.

4. Conclusion

Discrete choice models are a valuable tool for analyzing the behavior of individuals when faced with a choice between mutually exclusive alternatives. They are based on the logic of economic rationality which aims at optimizing an objective function while taking into account both the socioeconomic characteristics of individuals and the technical-economic characteristics of the alternative to be chosen, as well as the uncertainty of the environment where the choice reigns.

This objective function is conditional, discrete, and random. It is discrete because the problem of choice is no longer a continuum of possibilities but rather mutually exclusive alternatives, so that if the individual chooses a given alternative, he must renounce others. It is random in that the individual in question does not have perfect knowledge of the value of his objective function dependent on a given choice. This function is not observable. What is known is the choice of the user and not the value of this function. The objective function is conditional because it formalizes the satisfaction of the individual under the condition that he has already chosen the preferred alternative.

The multinomial logit model is the most used in empirical studies. It has the advantage of being able to treat the individual choice between a multitude of options and seeks to estimate the probability of having chosen a given alternative that better meets the requirements of the individual and the specific conditions characterizing the environment of choice.

It predicts the effects of modifying one of the characteristics of the alternative to choose or the individual's socioeconomic variables on the probability of making a relative decision of choice.

It allows better analysis of economic phenomena in relation to human behavior as a decision-making unit such as transport demand, accidentology, and valuation of nonmarket goods (transport time, membership of a given category population, etc.).

The objective of this chapter was to provide the reader with some essential elements for putting this multinomial logit model into practice by presenting in a first part its specificities and the interpretation of its estimated coefficients. In a second part, we tried to apply this model on two cases in relation to transport, one on modal choice and the other on accidentology.

Based on the results of these two applications, several pieces of information can be deduced which may be of great practical interest to individuals and public authorities involved in the transport. They constitute an important information base which guides these economic actors to the best choices of preventive actions and the orientation of the transport policy as well as in the matter of investment, pricing, road safety, etc. They offer us the possibility to calculate a specific time value to each individual according to their socioeconomic characteristics, their modal choice, and the conditions of travel (reason for travel, zone origin destination, time of departure, etc.), to propose the best preventive actions to accidentology, etc.

Author details

Foued Aloulou

Address all correspondence to: alouloufe@yahoo.fr

Higher Institute of Transportation and Logistics, University of Sousse, Tunisia

References

[1] De Palma A, Thisse J-F. Les modèles de choix discrets. Annales d'Economie et de Statistique. 1987;**9**:151-191

[2] Essafi CA. Les modèles logit polytomique non ordonnés: théorie et applications, Institut National de la Statistique et des Etudes Economiques; Série des Documents de Travail de la Direction des Statistiques Démographiques et sociales: Unité Méthodes Statistiques; *Série* des Documents de Travail Méthodologie Statistique; 2004 ;0301. Available from: *http://master.is.free.fr/Ancien_site/Documents/modele_logit.pdf*

[3] Manski CF. The Analysis of Qualitative Choice, PhD Dissertation, MIT; 1973

[4] Mc Fadden D. Conditional Logit Analysis of Qualitative Choice Behaviors. New York: Frantiers in Econometrics; 1974

[5] Hausman J, Wise D. A conditional Probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. Econometrica. 1978;**46**(2): 403-426

[6] McFadden D. The measurement of urban travel demand. Journal of Public Economics. 1974;**3**:303-328

[7] Ben-Akiva M. Structure of passenger travel demand models. Transportation research board record. 1973; 526, Washington DC

[8] Tongzon JL. Port choice and freight forwarders. Transportation Research Part E. 2009; **45**(1):186-195

[9] Malchow M, Kanafani A. Disaggregate analysis of port selection. Transportation Research Part E. 2004;**40**(4):317-338

[10] Boskin MJ. A conditional Logit model of occupational choice. Journal of Political Economy. 1974;**82**(2)

[11] Bhat CR, Guo J. A mixed spatially correlated logit model: Formulation and application to residential choice modeling. Transportation Research Part B. 2004;**38**:47-168

[12] De Palma A, Picard N, Waddell P. Discrete choice models with capacity constraints : An empirical analysis of the housing market of the greater Paris region. Journal of Urban Economics. 2007;**62**:204-230

[13] Gabriel SA, Rosenthal SS. Household location and race: Estimates of a multinomial Logit model. The Review of Economics and Statistics. 1989;**71**(2):240-249

[14] Schmidt P, Strauss RP. The prediction of occupation using multiple Logit models. International Economic Review. 1975;**16**(2)

[15] Aloulou F, sana N. Analyse microéconométrique des accidents routiers en Tunisie. revue économique. 2016;**67**(6):1211-1232

[16] Eluru N, Bhat CR. A joint econometric analysis of Seat Belt use and crash-related injury severity. Accident Analysis and Prevention. 2007;**39**(5):1037-1049

[17] Blayac T, Causse A. Value of travel time; a theoretical legitimization of some nonlinear representative utility in discrete choice models. Transportation Research Part B. 2001;**35**: 391-400

[18] De Lapparent M. Individual demand for travel modes and valuation of time attributes within the regular journey to work framework. L'actualité économique; Revue d'analyse économique. 2003:42

[19] Mc Fadden D, Hausman J. Specification test for the multinomial logit model. Econometrica. 1984;**52**(5):1219-1240

[20] De Lapparent M. Willingness to use safety belt and levels of injury in car accidents. Accident Analysis and Prevention. 2008;**40**(3):1023-1032

[21] Boyer M, Dionne G, et Vanasse C. Infractions au code de sécurité routière, infractions au code criminel et accidents automobiles. Publication CRT, 583, Centre de recherche sur les transports, Université de Montréal. 1988

[22] Abdel Aty MA, et Radwan AE. Modeling traffic accident occurrence and involvement. Accident Analysis and Prevention. 2000;**32**(5):633-642

Section 3

Applications of Statistics on Various Topics

Chapter 7

Application of Principal Component Analysis to Image Compression

Wilmar Hernandez and Alfredo Mendez

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.75007

Abstract

In this chapter, an introduction to the basics of principal component analysis (PCA) is given, aimed at presenting PCA applications to image compression. Here, concepts of linear algebra used in PCA are introduced, and PCA theoretical foundations are explained in connection with those concepts. Next, an image is compressed by using different principal components, and concepts such as image dimension reduction and image reconstruction quality are explained. Also, using the almost periodicity of the first principal component, a quality comparative analysis of a compressed image using two and eight principal components is carried out. Finally, a novel construction of principal components by periodicity of principal components has been included, in order to reduce the computational cost for their calculation, although decreasing the accuracy.

Keywords: principal component analysis, population principal components, sample principal components, image compression, image dimension reduction, image reconstruction quality

1. Introduction

Principal component analysis, also known as the Hotelling transform or Karhunen-Loeve transform, is a statistical technique that was proposed by Karl Pearson (1901) as part of factorial analysis; however, its first theoretical development appeared in 1933 in a paper written by Hotelling [1–8]. The complexity of the calculations involved in this technique delayed its development until the birth of computers, and its effective use started in the second half of the twentieth century. The relatively recent development of methods based on principal components makes them little used by a large number of non-statistician researchers. The purposes of these notes are to disclose the nature of the principal component analysis and show some of its possible applications.

Principal component analysis refers to the explanation of the structure of variances and covariances through a few linear combinations of the original variables, without losing a significant part of the original information. In other words, it is about finding a new set of orthogonal axes in which the variance of the data is maximum. Its objectives are to reduce the dimensionality of the problem and, once the transformation has been carried out, to facilitate its interpretation.

By having p variables collected on the units analyzed, all are required to reproduce the total variability of the system, and sometimes the majority of this variability can be found in a small number, k, of principal components. Its origin lies in the redundancy that there exists many times between different variables, so the redundancy is data, not information. The k principal components can replace the p initial variables, so that the original set of data, consisting of n measures of p variables, is reduced to n measures of k principal components.

The objective pursued by the analysis of principal components is the representation of the numerical measurements of several variables in a space of few dimensions, where our senses can perceive relationships that would otherwise remain hidden in higher dimensions. The abovementioned representation must be such that, when discarding higher dimensions, the loss of information is minimal. A simile could illustrate the idea: imagine a large rectangular plate that is a three-dimensional object, but that for practical purposes, we consider it as a flat two-dimensional object. When carrying out this reduction in dimensionality, a certain amount of information is lost since, for example, opposite points located on the two sides of the rectangular plate will appear confused in a single one. However, the loss of information is largely compensated by the simplification made, since many relationships, such as the neighborhood between points, are more evident when they are drawn on a plane than when done by a three-dimensional figure that must necessarily be drawn in perspective.

The analysis of principal components can reveal relationships between variables that are not evident at the first sight, which facilitates the analysis of the dispersion of observations, highlighting possible groupings and detecting the variables that are responsible for the dispersion.

2. Preliminaries

The study of multivariate methods is greatly facilitated by means of matrix algebra [9–11]. Next, we introduce some basic concepts that are essential for the explanation of statistical techniques, as well as for geometric interpretations. In addition, the relationships that can be expressed in terms of matrices are easily programmable on computers, so we can apply calculation routines to obtain other quantities of interest. It is a basic introduction about concepts and relationships.

2.1. The vector of means and the covariance matrix

Let $\mathbf{X} = \left[X_1 \; \dots \; X_p\right]^t$ be a random column vector of dimension p. Each component, X_i, is a random variable (r.v.) with mean $E[X_i] = \mu_i$ and variance $V[X_i] = E\left[\left(X_i - \mu_i\right)^2\right] = \sigma_{ii}$. Given two r.v., X_i and X_j, we define the covariance between them as $Cov[X_i, X_j] = E[(X_i - \mu_i)(X_j - \mu_j)] = \sigma_{ij}$. The expected values, variances, and covariances can be grouped into vectors and matrices that we will call population mean vector, $\boldsymbol{\mu}$, and population covariance matrix, $\sum$:

$$\boldsymbol{\mu} = E[\mathbf{X}] = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}, \sum = Cov[\mathbf{X}] = E\left[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^t\right] = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{bmatrix} \tag{1}$$

The population correlation matrix is given by $\boldsymbol{\rho} = \left[\rho_{ij}\right]$, where $\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}}$.

In the case of having n values of the r.v.s, we will consider estimators of the previous population quantities, which we will call sample estimators.

Definition 2.1: *Let* $\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{p1} & \cdots & x_{pp} \end{bmatrix}$ *be a simple random sample of a p-dimensional r.v. ordered in the data matrix, with the values of the r.v.s in each column. The p-dimensional sample mean column vector is* $\overline{\mathbf{X}} = [\overline{x}_i]$, *where* $\overline{x}_i = \frac{1}{p}\sum_{m=1}^{p} x_{im}$. *The sample covariance matrix is* $\mathbf{S} = [s_{ij}] = \frac{n}{n-1}\mathbf{S_n} = \frac{n}{n-1}(\mathbf{X}\text{-}\overline{\mathbf{X}})(\mathbf{X}\text{-}\overline{\mathbf{X}})^t$. *The generalized sample variance is the determinant of* $\mathbf{S}$, $|\mathbf{S}|$. *The sample correlation matrix is* $\mathbf{R} = [r_{ij}]$, *where* $r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}}$ *with* $i, j = 1\ldots p$.

Proposition 2.1: *Let* $\mathbf{X}_1, \ldots, \mathbf{X}_p$ *be a simple random sample of a p-dimensional r.v.* $\mathbf{X}$ *with mean vector* $\boldsymbol{\mu}$ *and covariance matrix* $\sum$. *The unbiased estimators of* $\boldsymbol{\mu}$ *and* $\sum$ *are* $\overline{\mathbf{X}}$ *and* $\mathbf{S}$..

2.2. Eigenvalues and eigenvectors

One of the problems that linear algebra deals with is the simplification of matrices through methods that produce diagonal or triangular matrices, which are widely used in the resolution of linear systems of the form $\mathbf{Ax} = \mathbf{b}$.

Definition 2.2: *Let* $\mathbf{A}$ *be a square matrix. If* $\mathbf{v^t A v} \geq 0$ *for any vector* $\mathbf{v}$, $\mathbf{A}$ *is a nonnegative definite matrix. If* $\mathbf{Av} = \lambda\mathbf{v}$, *with* $\mathbf{v} \neq 0$, λ *is an eigenvalue associated with the eigenvector* $\mathbf{v}$.

Proposition 2.2: *Let* $\mathbf{A}$ *be a symmetric p by p matrix with real-valued entries.* $\mathbf{A}$ *has p pairs of eigenvalues and eigenvectors,* $(\lambda_1, \mathbf{e}_1), \ldots, (\lambda_p, \mathbf{e}_p)$, *such that:*

1. *All the eigenvalues are real. Also,*
   a. $\mathbf{A}$ *is positive definite if all the eigenvalues are positive.*
   b. $\mathbf{A}$ *is nonnegative definite if all the eigenvalues are nonnegative.*

2. *The eigenvectors can be chosen with 2-norm equal to 1.*

3. *The eigenvectors are mutually perpendicular.*

4. *The eigenvectors are unique unless two or more eigenvalues are equal.*

5. *The spectral decomposition of* $\mathbf{A}$ *is* $\mathbf{A} = \lambda_1 \mathbf{e}_1 \mathbf{e}_1^t + \cdots + \lambda_p \mathbf{e}_p \mathbf{e}_p^t$.

6. *If* $\mathbf{P} = [\mathbf{e}_1, \ldots, \mathbf{e}_p]$ *is an orthogonal matrix and* Λ *is a diagonal matrix with main diagonal entries* $(\lambda_1, \ldots, \lambda_p)$, *the spectral decomposition of* $\mathbf{A}$ *can be given by* $\mathbf{A} = \mathbf{P}\Lambda\mathbf{P}^t$. *Therefore,* $\mathbf{A}^{-1} = \mathbf{P}\Lambda^{-1}\mathbf{P}^t = \sum_{i=1}^{p} \frac{1}{\lambda_i} \mathbf{e}_i \mathbf{e}_i^t$.

Remark 2.1: *Let* $\mathbf{X}$ *be a matrix with the values of a simple random sample in each column of a p-dimensional r.v., and let* $\mathbf{y}_i^t = (x_{i1}, \ldots, x_{in})$, *with* $i = 1 \ldots p$, *be the ith row of* $\mathbf{X}$. *Let* $\mathbf{1}_n^t = (1, \ldots, 1)$ *be the n by one vector with all its coordinates equal to 1. It can be proven that:*

1. *The projection of the vector* $\mathbf{y}_i^t$ *on the vector* $\mathbf{1}_n$ *is the vector* $\overline{x}_i \mathbf{1}_n$, *whose 2-norm is equal to* $\sqrt{n}|\overline{x}_i|$.

2. *Matrix* $\mathbf{S_n}$ *is obtained from the residuals* $\mathbf{e}_i = \mathbf{y}_i - \overline{x}_i \mathbf{1}_n$, *the squared 2-norm of* $\mathbf{e}_i$ *is equal to* $(n-1)s_{ii}$, *and the scalar product of* $\mathbf{e}_i$ *and* $\mathbf{e}_j$ *is equal to* $(n-1)s_{ij}$.

3. *The sample correlation coefficient* r_{ij} *is the cosine of the angle between* $\mathbf{e}_i$ *and* $\mathbf{e}_j$.

4. *If* $\mathbf{U}$ *is the volume generated by the vectors* $\mathbf{e}_i$, *with* $i = 1 \ldots p$, *then* $|\mathbf{S}| = \frac{\mathbf{U}^2}{(n-1)^p}$. *Therefore, the generalized sample variance is proportional to the square of the volume generated by deviation vectors. The volume will increase if the norm of some* $\mathbf{e}_i$ *is increased.*

2.3. Distances

Many techniques of multivariate statistical analysis are based on the concept of distance. Let $Q = (x_1, x_2)$ be a point in the plane. The Euclidean distance from Q to the origin, O, is $d(Q, O) = \sqrt{x_1^2 + x_2^2}$. If $Q = (x_1, \ldots, x_p)$ and $R = (y_1, \ldots, y_p)$, the Euclidean distance between these two points of $\Re^p$ is $d(Q, R) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_p - y_p)^2}$. All points $(x_1, \ldots, x_p)$ whose square distance to the origin is a fixed quantity, for example, $x_1^2 + \cdots + x_p^2 = c^2$, are the points of the p-dimensional sphere of radius $|c|$.

For many statistical purposes, the Euclidean distance is unsatisfactory, since each coordinate contributes in the same way to the calculation of such a distance. When the coordinates represent measures subject to random changes, it is desirable to assign weights to the coordinates depending on how high or low the variability of the measurements is. This suggests a measure of distance that is different from the Euclidean.

Next, we introduce a statistical distance that will take into account the different variabilities and correlations. Therefore, it will depend on the variances and covariances, and this distance is fundamental in multivariate analysis.

Suppose we have a fixed set of observations in $\mathfrak{R}^p$, and, to illustrate the situation, consider n pairs of measures of two variables, x_1 and x_2. Suppose that the measurements of x_1 vary independently of x_2 and that the variability of the measures of x_1 are much greater than those of x_2. This situation is shown in **Figure 1**, and our first objective is to define a distance from the points to the origin.

In **Figure 1**, we see that the values that have a given deviation from the origin are farther from the origin in the x_1 direction than in the x_2 direction, due to the greater variability inherent in the direction of x_1. Therefore, it seems reasonable to give more weight in the coordinate x_2 than in the x_1. One way to obtain these weights is to standardize the coordinates, that is, $x_1^* = x_1/\sqrt{s_{11}}$ and $x_2^* = x_2/\sqrt{s_{22}}$, where s_{ii} is the sample variance of the variable x_i. Thus, the statistical distance from a point $Q = (x_1, x_2)$ to the origin is $d(Q, O) = \sqrt{\frac{x_1^2}{s_{11}} + \frac{x_2^2}{s_{22}}}$. Therefore, the points that are equidistant from the origin of a constant distance c are on an ellipse centered at the origin, whose major axis coincides with the coordinate that has the greatest variability. In the case that the variability of one variable is analogous to that of the other and that the coordinates are independent, the Euclidean distance is proportional to the statistical distance.

If $Q = (x_1, \dots, x_p)$ and $R = \left(y_1, \dots, y_p\right)$ are two points of $\mathfrak{R}^p$, the statistical distance between them is $d(Q, R) = \sqrt{\frac{(x_1 - y_1)^2}{s_{11}} + \cdots + \frac{(x_p - y_p)^2}{s_{pp}}}$, with s_{ii} being the sample variance of the variable x_i. The statistical distance defined so far does not include most of the important cases where the variables are not independent. **Figure 2** shows a situation where the pairs (x_1, x_2) seem to have an increasing trend, so the sample correlation coefficient will be positive. In **Figure 2**,

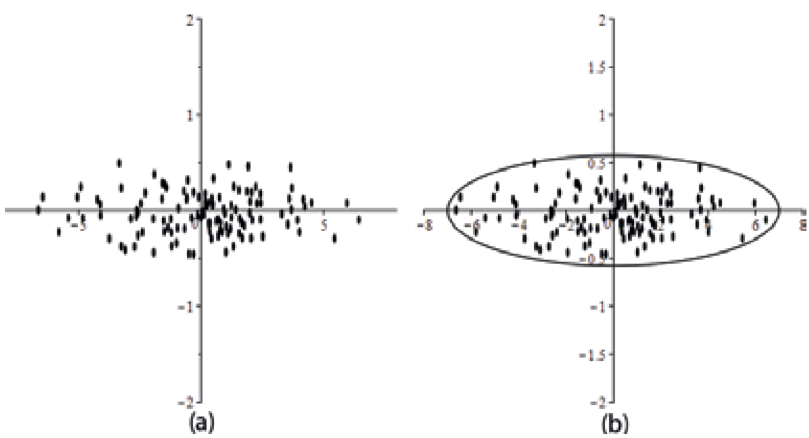

Figure 1. Scatter plot with more variability in x_1 than in x_2. (a)Scatter plot (b) Ellipse of constant distance.

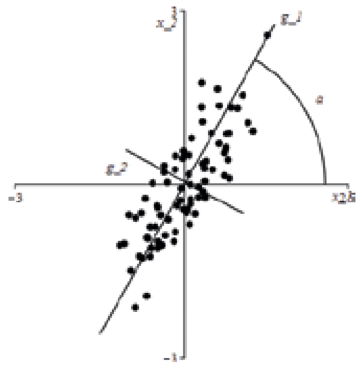

Figure 2. Scatter plot with positive correlation.

we see that if we make a rotation of amplitude α and consider the axes (g_1, g_2) we are in conditions analogous to those of **Figure 1 (a)**. Therefore, the distance from the point $Q = (g_1, g_2)$ to the origin will be $d(Q, O) = \sqrt{\frac{g_1^2}{\tilde{s}_{11}} + \frac{g_2^2}{\tilde{s}_{22}}}$, where $\tilde{s}_{ii}$ is the sample variance of the variable g_i.

The relationships between the original coordinates and the new coordinates can be expressed as

$$\begin{aligned} g_1 &= x_1\cos(\alpha) + x_2\sin(\alpha) \\ g_2 &= -x_1\sin(\alpha) + x_2\cos(\alpha) \end{aligned} \tag{2}$$

and, after some algebraic manipulations, $d(Q, O) = \sqrt{a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2}$, where a_{ij} are values that depend on the angle and the dispersions, and also must meet the condition that the distance between any two points must be positive.

The distance from a point $Q = (x_1, x_2)$ to a fixed point $R = (y_1, y_2)$ in situations where there is a positive correlation is $d(Q, R) = \sqrt{a_{11}(x_1 - y_1)^2 + 2a_{12}(x_1 - y_1)(x_2 - y_2) + a_{22}(x_2 - y_2)^2}$. So, in this case, the coordinates of all points $Q = (x_1, x_2)$ verify the equation $a_{11}(x_1 - y_1)^2 + 2a_{12}(x_1 - y_1)(x_2 - y_2) + a_{22}(x_2 - y_2)^2 = c^2$, which is the equation of an ellipse of center $R = (y_1, y_2)$ and with axes parallel to (g_1, g_2). **Figure 3** shows ellipses with constant statistical distances.

This distance can be generalized to $\Re^p$ if $a_{11}, \ldots, a_{pp}, a_{12}, \ldots, a_{p-1,p}$ are values such that the distance from Q to R is given by.

$$d(Q, R) = \sqrt{A + B}, \text{ where } \begin{aligned} A &= a_{11}(x_1 - y_1)^2 + \cdots + a_{pp}\left(x_p - y_p\right)^2 \\ B &= 2a_{12}(x_1 - y_1)(x_2 - y_2) + \cdots + 2a_{p-1,p}\left(x_{p-1} - y_{p-1}\right)\left(x_p - y_p\right) \end{aligned} \tag{3}$$

This distance, therefore, is completely determined by the coefficient a_{ij}, with $i, j \in \{1, \ldots, p\}$, which can be arranged in a matrix given by

http://dx.doi.org/10.5772/intechopen.75007

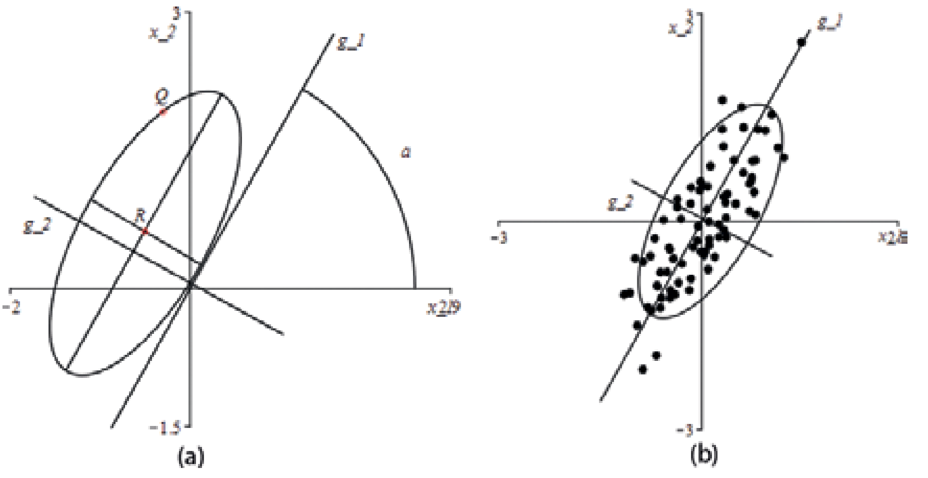

Figure 3. Ellipses of constant statistical distance. (a) Point Q at a constant distance from R. (b) Ellipse $x^2/3 + 4y^2 = 1$ rotated and moved and scatter plot.

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{p1} & \cdots & a_{pp} \end{bmatrix} \tag{4}$$

The elements of Eq. (4) cannot be arbitraries. In order to define a distance over a vector space, Eq. (4) must be a square, symmetric, positive definite matrix. Therefore, the sample covariance matrix of a data matrix, **S**, is a candidate to define a statistical distance.

Figure 4 shows a cloud of points with center of gravity, $(\overline{x}_1, \overline{x}_2)$, at point R. At the first glance, it can be seen that the Euclidean distance from point R to point Q is greater than the Euclidean

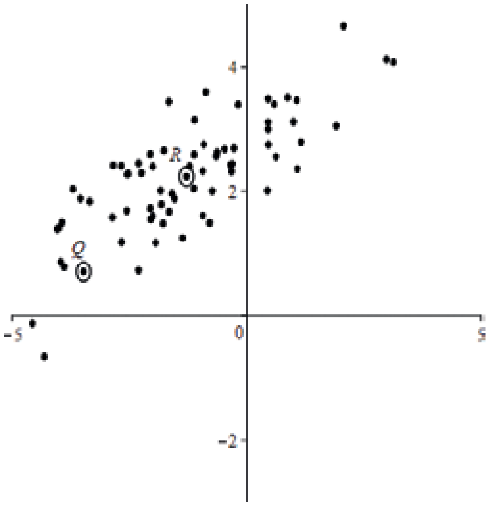

Figure 4. Scatter plot with center of gravity R and a point Q.

distance from point R to the origin; however, Q seems to have more to do with the cloud of points than the origin. If we take into account the variability of the points in the cloud and take the statistical measure, then Q will be closer to R than the origin.

The above given explanation has tried to be an illustration of the need to consider distances other than the Euclidean.

3. Population principal components

Principal components are a particular case of linear combinations of p r.v.s, $X_1, \ldots, X_p$. These linear combinations represent, geometrically, a new coordinate system that is obtained by rotating the original reference system that has $X_1, \ldots, X_p$ as coordinate axes. The new axes represent the directions with maximum variability and provide a simple description of the structure of the covariance.

Principal components depend only on the variance/covariance matrix $\sum$ (or on the correlation matrix $\boldsymbol{\rho}$) of $X_1, \ldots, X_p$, and it is not necessary to assume that the r.v.s follows an approximately normal distribution. In case of having a normal multivariate distribution, we will have interpretations in terms of ellipsoids of constant density, if we consider the distance that defines the $\sum$ matrix, and the inferences can be made from the population components.

Let $\mathbf{X} = \begin{bmatrix} X_1 & \ldots & X_p \end{bmatrix}^t$ be a p-dimensional random vector with covariance matrix $\sum$ and eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. Let us consider the following p linear combinations:

$$\begin{aligned} Y_1 &= l_1^t \mathbf{X} = l_{11} X_1 + \cdots + l_{p1} X_p \\ &\vdots \\ Y_p &= l_p^t \mathbf{X} = l_{1p} X_1 + \cdots + l_{pp} X_p \end{aligned} \tag{5}$$

These new r.v.s verify the following equalities:

$$\begin{aligned} &V[Y_i] = l_i^t \Sigma l_i & & i = 1, \ldots, p \\ &Cov\left[Y_i, Y_j\right] = l_i^t \Sigma l_j & & i, j = 1, \ldots, p \quad i \neq j \end{aligned} \tag{6}$$

Principal components are those linear combinations that, being uncorrelated among them, have the greatest possible variance. Thus, the first principal component is the linear combination with the greatest variance, that is, $V[Y_1] = l_1^t \Sigma l_1$ is maximum. Since if we multiply l_1 by some constant the previous variance grows, we will restrict our attention to vectors of norm one, with which the aforementioned indeterminacy disappears. The second principal component is the linear combination that maximizes the variance and is uncorrelated with the first one, and the norm of the coefficient vector is equal to 1.

Proposition 3.1: *Let $\sum$ be the covariance matrix of the random vector $\mathbf{X} = \begin{bmatrix} X_1 & \ldots & X_p \end{bmatrix}^t$. Let us assume that $\sum$ has p pairs of eigenvalues and eigenvectors, $(\lambda_1, \mathbf{e}_1), \ldots, (\lambda_p, \mathbf{e}_p)$, with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. Then, the ith principal component is given by*

$$Y_i = \mathbf{e}_i^t \mathbf{X} = e_{1i} X_1 + \cdots + e_{pi} X_p \quad i = 1, \dots, p \tag{7}$$

In addition, with this choice it is verified that:

1. $V[Y_i] = \mathbf{e}_i^t \Sigma \mathbf{e}_i = \lambda_i \quad i = 1, \dots, p$.
2. $Cov[Y_i, Y_j] = 0 \quad i, j = 1, \dots, p \quad i \neq j$.
3. *If any of the eigenvalues are equal, the choice of the corresponding eigenvectors as vectors of coefficients is not unique.*
4. $\sigma_{11} + \cdots + \sigma_{pp} = \sum_{i=1}^{p} V[X_i] = \lambda_1 + \cdots + \lambda_p = \sum_{j=1}^{p} V[Y_j]$.

Remark 3.1: *For the demonstration of these results, expressions are used on maximums of quadratic forms between vectors of fixed norm* $\left(\max_{l \neq 0} \frac{l^t \Sigma l}{l^t l} = \lambda_1\right)$. *Also, the Lagrange multipliers method can be used, expressions when the abovementioned maximum is subject to orthogonality conditions and properties on the trace of a matrix (if* $\Sigma = \mathbf{P}\Lambda\mathbf{P}^{\mathrm{t}}$, *then* $tr(\Sigma) = tr(\mathbf{P}\Lambda\mathbf{P}^{\mathrm{t}}) = tr(\Lambda)$*).*

Due to the previous result, principal components are uncorrelated among them, with variances equal to the eigenvalues of $\sum$, *and the proportion of the population variance due to the ith principal component is given by* $\frac{\lambda_i}{\lambda_1 + \cdots + \lambda_p}$.

If a high percentage of the population variance, for example, the 90%, *of a p-dimensional r.v., with large p, can be attributed to, for example, the five first principal components, then we can replace all the r.v.s by those five components without a great loss of information.*

Each component of the coefficient vector $\mathbf{e}_i^t = [e_{1i}, \dots, e_{pi}]$, e_{ki}, also deserves our attention, since it is a measure of the relationship between the r.v.s $X_k$and Y_i.

Proposition 3.2: *If* $Y_1 = \mathbf{e}_1^t \mathbf{X}, \dots, Y_p = \mathbf{e}_p^t \mathbf{X}$ *are the principal components obtained from the covariance matrix* $\sum$, *with pairs of eigenvalues and eigenvectors* $(\lambda_1, \mathbf{e}_1) \quad \dots \quad (\lambda_p, \mathbf{e}_p)$, *then the linear correlation coefficients between the variables* X_k *and the components* Y_i *are given by*

$$\rho_{X_k, Y_i} = \frac{e_{ki}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} \quad i, k = 1, \dots, p \tag{8}$$

Therefore, e_{ki} is proportional to the correlation coefficient between $X_k$and Y_i.

In the particular case that $\mathbf{X}$ has a normal p-dimensional distribution, $\mathrm{N}_p(\mu, \Sigma)$, the density of $\mathbf{X}$ is constant in the ellipsoids with the center at $\boldsymbol{\mu}$ given by $(\mathbf{X} - \boldsymbol{\mu})^t \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu}) = c^2$ that have axes $\pm c\sqrt{\lambda_i}\mathbf{e}_i$ and $i = 1, \dots, p$, where $(\lambda_i, \mathbf{e}_i)$ are the pairs of eigenvalues and eigenvectors of $\sum$.

If the covariance matrix, $\sum$, can be decomposed into $\Sigma = \mathbf{P}\Lambda\mathbf{P}^{\mathrm{t}}$, where $\mathbf{P}$ is orthogonal and Λ diagonal, it can be shown that $\Sigma^{-1} = \mathbf{P}\Lambda^{-1}\mathbf{P}^{\mathrm{t}} = \sum_{i=1}^{p} \frac{1}{\lambda_i} \mathbf{e}_i \mathbf{e}_i^t$. Also, if it can be assumed that $\boldsymbol{\mu} = 0$, to simplify the expressions, then

$$c^2 = \mathbf{x}^t\Sigma^{-1}\mathbf{x} = \frac{1}{\lambda_1}\left(\mathbf{e}_1^t\mathbf{x}\right)^2 + \frac{1}{\lambda_2}\left(\mathbf{e}_2^t\mathbf{x}\right)^2 + \cdots + \frac{1}{\lambda_p}\left(\mathbf{e}_p^t\mathbf{x}\right)^2 \tag{9}$$

If the principal components $y_1 = \mathbf{e}_1^t\mathbf{x}, \ldots, y_p = \mathbf{e}_p^t\mathbf{x}$ are considered, the equation of the constant density ellipsoid is given by

$$c^2 = \frac{1}{\lambda_1}y_1^2 + \frac{1}{\lambda_2}y_2^2 + \cdots + \frac{1}{\lambda_p}y_p^2 \tag{10}$$

Therefore, the axes of the ellipsoid have the directions of the principal components.

Example 3.1: *Let X_1, X_2, X_3 be the three-unidimensional r.v.s and $\mathbf{X} = [X_1, X_2X_3]^t$, with covariance matrix*

$$\Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 8 & -3 \\ 0 & -3 & 2 \end{bmatrix} \tag{11}$$

It can be verified that the pairs of eigenvalues and eigenvectors are $\left(\lambda_1 = 9.243, \mathbf{e}_1^t = [0 \quad 0.924 \quad -0.383]\right)$, $\left(\lambda_2 = 2, \mathbf{e}_2^t = [1 \quad 0 \quad 0]\right)$, and $\left(\lambda_3 = 0.757, \mathbf{e}_3^t = [0 \quad 0.383 \quad 0.924]\right)$. Therefore, the principal components are the following:

$$\begin{aligned} Y_1 &= \mathbf{e}_1^t\mathbf{X} = 0.924X_2 - 0.383X_3 \\ Y_2 &= \mathbf{e}_2^t\mathbf{X} = X_1 \\ Y_3 &= \mathbf{e}_3^t\mathbf{X} = 0.383X_2 + 0.924X_3 \end{aligned} \tag{12}$$

The norm of all the eigenvectors is equal to 1, and, in addition, the variable X_1 is the second principal component, because X_1 is uncorrelated with the other two variables.

The results of **Proposition 3.1** *can be verified for this data, for example, $V[Y_1] = 9.243$ and $Cov[Y_1, Y_2] = 0$. Also, $\sum_{i=1}^{3} V[X_i] = 2 + 8 + 2 = 12 = 9.243 + 2 + 0.757 = \sum_{j=1}^{3} V\left[Y_j\right]$. Thus, the proportion of the total variance explained by the first component is $\lambda_1/12 = 77\%$, and the one explained by the first two is $(\lambda_1 + \lambda_2)/12 = 93.69\%$, so that the components Y_1 and Y_2 can replace the original variables with a small loss of information.*

The correlation coefficients between the principal components and the variables are the following:

$$\begin{array}{lll} \rho_{X_1,Y_1} = 0 & \rho_{X_2,Y_1} = 0.993 & \rho_{X_3,Y_1} = -0.823 \\ \rho_{X_1,Y_2} = 1 & \rho_{X_2,Y_2} = 0 & \rho_{X_3,Y_2} = 0 \\ \rho_{X_1,Y_3} = 0 & \rho_{X_2,Y_3} = 0.118 & \rho_{X_3,Y_3} = 0.568 \end{array} \tag{13}$$

In view of these values, it can be concluded that $X_2$and X_3 individually are practically equally important with respect to the first principal component, although this is not the case with respect to the third

http://dx.doi.org/10.5772/intechopen.75007

component. If, in addition, it is assumed that the distribution of **X** *is normal,* $N_3(\mu, \Sigma)$, *with a null mean vector, ellipsoids of constant density* $\mathbf{x}^t\Sigma^{-1}\mathbf{x} = c^2$ *can be considered. An ellipsoid of constant statistical distance and projections is shown in* ***Figure 5***.

The ellipsoid with $c^2 = 8$ *has been represented in* ***Figure 5 (a)***, *together with its axes and the ellipsoid projections on planes parallel to the coordinate axes. The aforementioned projections are ellipses of red, green, and blue colors that are reproduced in* ***Figure 5 (b)***. *Also, in this figure, the black ellipse obtained by projecting the ellipsoid on the plane determined by the first two main components has been represented. The equation of this ellipse is* $\frac{y_1^2}{a^2} + \frac{y_2^2}{b^2} = 8$, *where* $a = \frac{c}{\sqrt{\eta_1}}$ *and* $b = \frac{c}{\sqrt{\eta_2}}$, *with* η_1 *and* η_2 *being the two smallest eigenvalues of* Σ^{-1}, *and the axes are determined by* Y_1 *and* Y_2. *As can be seen, the diameters of the ellipse determined by the first two components are larger than the others. Therefore, the area enclosed by this ellipse is the largest of all, indicating that it is the one that gathers the greatest variability.*

3.1. Principal components with respect to standardized variables

The principal components of the normalized variables $Z_1 = \frac{X_1-\mu_1}{\sqrt{\sigma_{11}}}, \ldots, Z_p = \frac{X_p-\mu_p}{\sqrt{\sigma_{pp}}}$ can also be considered, which in matrix notation is $\mathbf{Z} = \mathbf{V}(\mathbf{X}\text{-}\boldsymbol{\mu})$, where **V** is the diagonal matrix whose elements are $\frac{1}{\sqrt{\sigma_{11}}}, \ldots, \frac{1}{\sqrt{\sigma_{pp}}}$. It is easily verified that the r.v. **Z** verifies $E[\mathbf{Z}] = 0$ and $Cov[\mathbf{Z}] = \mathbf{V}\Sigma\mathbf{V} = \boldsymbol{\rho}$, where $\boldsymbol{\rho}$ is the correlation matrix of **X**.

Principal components of **Z** are obtained by the eigenvalues and eigenvectors of the correlation matrix, $\boldsymbol{\rho}$, of **X**. Furthermore, with some simplification, the previous results can be applied, since the variance of each Z_i is equal to 1.

Let $W_1, \ldots, W_p$ be the principal components of **Z** and $(v_i, \mathbf{u}_i^t)$, $i = 1, \ldots, p$, the pairs of eigenvalues and eigenvectors of $\boldsymbol{\rho}$, since they do not have to be the same.

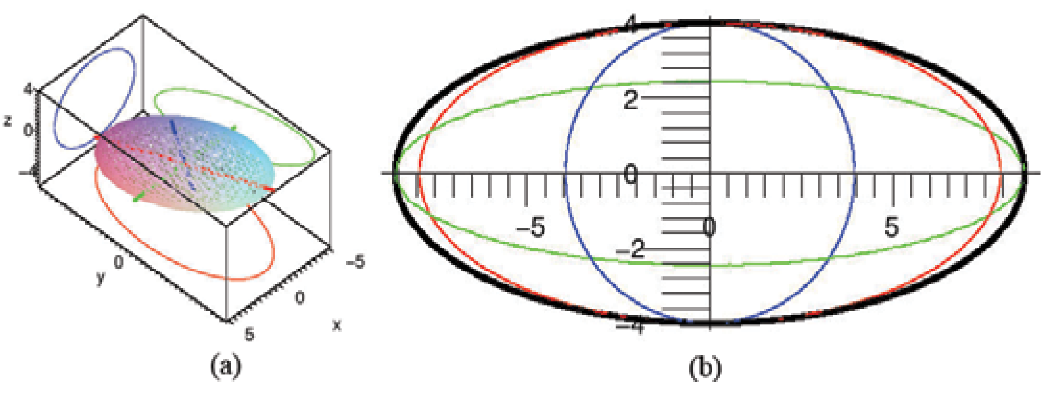

Figure 5. Ellipsoid of constant statistical distance and projections. (a) Ellipsoid of constant density and projections on the coordinate planes. (b) Projections on the coordinate planes and the base plane $\{Y_1, Y_2\}$.

Proposition 3.3: *Let* $\mathbf{Z} = [Z_1, \ldots, Z_p]^t$ *be a random vector with covariance matrix* $\boldsymbol{\rho}$. *Let* $(v_1, \mathbf{u}_1), \ldots, (v_p, \mathbf{u}_p)$ *be the pairs of eigenvalues and eigenvectors of* $\boldsymbol{\rho}$, *with* $v_1 \geq \cdots \geq v_p$. *Then, the ith principal component is given by* $W_i = \mathbf{u}_i^t \mathbf{V}(\mathbf{X}\text{-}\boldsymbol{\mu})$, $i = 1, \ldots, p$. *In addition, with this choice it is verified that:*

1. $V[W_i] = v_i$, $i = 1, \ldots, p$.
2. $Cov[W_i, W_j] = 0$, $i, j = 1, \ldots, p$, $i \neq j$.
3. *If any of the eigenvalues are equal, the choice of the corresponding eigenvectors as vectors of coefficients is not unique.*
4. $\sum_{i=1}^{p} V[W_i] = v_1 + \cdots + v_p = \sum_{j=1}^{p} V[Z_j] = p$.
5. *The linear correlation coefficients between the variables* Z_k *and the principal components* W_i *are* $\rho_{Z_k, W_i} = u_{ki}\sqrt{v_i}$ *and* $i, k = 1, \ldots, p$.

These results are a consequence of those obtained in **Proposition 3.1** and **Proposition 3.2** applied to $\mathbf{Z}$ and $\boldsymbol{\rho}$ instead of $\mathbf{X}$ and $\sum$.

The total population variance of the normalized variables is the sum of the elements of the diagonal of $\boldsymbol{\rho}$, that is, p. Therefore, the proportion of the total variability explained by the ith principal component is $\frac{v_i}{p}$, $i = 1, \ldots, p$.

Example 3.2: *Let* X_1 *and* X_2 *be the two-unidimensional r.v.s and* $\mathbf{X} = [X_1, X_2]^t$ *with the covariance matrix,* $\sum$, *and correlation matrix,* $\boldsymbol{\rho}$, *given by*

$$\Sigma = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$
$$\boldsymbol{\rho} = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix} \tag{14}$$

It can be verified that the pairs of eigenvalues and eigenvectors for S are $(\lambda_1 = 100.04, \mathbf{e}_1^t = [-0.02 \quad -0.999])$ *and* $(\lambda_2 = 0.96, \mathbf{e}_2^t = [-0.999 \quad 0.02])$. *Therefore, the principal components are the following:*

$$Y_1 = \mathbf{e}_1^t \mathbf{X} = -0.02X_1 - 0.999X_2$$
$$Y_2 = \mathbf{e}_2^t \mathbf{X} = -0.999X_1 + 0.02X_2 \tag{15}$$

Furthermore, the eigenvalues and eigenvectors of $\boldsymbol{\rho}$ *are* $(v_1 = 1.2, \mathbf{u}_1^t = [0.707 \quad 0.707])$ *and* $(v_2 = 0.8, \mathbf{u}_2^t = [-0.707 \quad 0.707])$; *hence, the principal components of the normalized variables are the following:*

$$W_1 = \mathbf{u}_1^t \mathbf{Z} = 0.707Z_1 + 0.707Z_2 = 0.707(X_1 - \mu_1) + 0.0707(X_2 - \mu_2)$$
$$W_2 = \mathbf{u}_2^t \mathbf{Z} = -0.707Z_1 + 0.707Z_2 = -0.707(X_1 - \mu_1) + 0.0707(X_2 - \mu_2) \quad (16)$$

Because the variance of X_2 is much greater than that of X_1, the first principal component for Σ is determined by X_2, and the proportion of variability explained by that first component is $\frac{\lambda_1}{\lambda_1+\lambda_2} = 0.99$.

When considering the normalized variables, each variable also contributes to the components determined by $\boldsymbol{\rho}$*, and the dependencies between the normalized variables and their first component are* $\rho_{Z_1,W_1} = u_{11}\sqrt{v_1} = 0.707\sqrt{1.2} = 0.774$ *and* $\rho_{Z_2,W_1} = u_{21}\sqrt{v_1} = -0.707\sqrt{1.2} = -0.774$*. The proportion of the total variability explained by the first component is* $\frac{v_1}{p} = 0.6$.

Therefore, the importance of the first component is strongly affected by normalization. In fact, the weights, in terms of X_i are 0.707 *and* 0.0707 *for* $\boldsymbol{\rho}$*, as opposed to* −0.02 *and* −0.999 *for* Σ.

Remark 3.2: *The above example shows that the principal components deduced from the original variables are, in general, different from those derived from the normalized variables. So, normalization has important consequences.*

When the units in which the different one-dimensional random variables are given are very different and in the case that one of the variances is very dominant compared to the others, the first principal component, with respect to the original variables, will be determined by the variable whose variance is the dominant one. On the other hand, if the variables are normalized, their relationship with the first components will be more balanced.

Principal components can be expressed in particular ways if the covariance matrix, or the correlation matrix, has special structures, such as diagonal ones, or structures of the form $\Sigma = \sigma^2 \mathbf{A}$.

4. Sample principal components

Once we have the theoretical framework, we can now address the problem of summarizing the variation of n measurements made on p variables.

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be a sample of a p-dimensional r.v. $\mathbf{X}$with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. These data have a vector of sample means $\bar{\mathbf{x}}$, covariance matrix $\mathbf{S}$, and correlation matrix $\mathbf{R}$.

This section is aimed at constructing linear uncorrelated combinations of the measured characteristics that contain the greatest amount of variability contained in the sample. These linear combinations are called principal sample components.

Given n values of any linear combination $l_1^t \mathbf{x}_j = l_{11}x_{1j} + \cdots + l_{p1}x_{pj}$, $j = 1, \ldots, n$, its sample mean is $l_1^t \bar{\mathbf{x}}_j$, and its sample variance is $l_1^t \mathbf{S} l_1$. If we consider two linear combinations, $l_1^t \mathbf{x}_j$ and $l_2^t \mathbf{x}_j$, their sample covariance is $l_1^t \mathbf{S} l_2$.

The first principal component will be the linear combination, $l_1^t \mathbf{x}_j$, which maximizes the sample variance, subject to the condition $l_1^t l_1 = 1$. The second component will be the linear combination, $l_2^t \mathbf{x}_j$, which maximizes the sample variance, subject to the condition that $l_2^t l_2 = 1$ and that the sample covariance of the pairs $\left(l_1^t \mathbf{x}_j, l_2^t \mathbf{x}_j\right)$ is equal to zero. This procedure is continued until the p principal components are completed.

Proposition 4.1: *Let $\mathbf{S} = (s_{ik})$ be the p by p matrix of sample covariances, whose pairs of eigenvalues and eigenvectors are $\left(\hat{\lambda}_1, \hat{\mathbf{e}}_1\right), \dots, \left(\hat{\lambda}_p, \hat{\mathbf{e}}_p\right)$, with $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p \geq 0$. Let $\mathbf{x}$ be an observation of the p-dimensional random variable $\mathbf{X}$, then:*

1. *The ith principal component is given by $\hat{y}_i = \hat{\mathbf{e}}_i^t \mathbf{x} = \hat{e}_{1i} x_1 + \cdots + \hat{e}_{pi} x_p$, $i = 1, \dots, p$.*
2. *The sample variance of $\hat{y}_k$ is $\hat{\lambda}_k$, $k = 1, \dots, p$.*
3. *The sample covariance of $\left(\hat{y}_i, \hat{y}_k\right)$, $i \neq k$, is equal to 0.*
4. *The total sample variance is $\sum_{i=1}^{p} s_{ii} = \hat{\lambda}_1 + \cdots + \hat{\lambda}_p$.*
5. *The sample correlation coefficients between $x_k$and $\hat{y}_i$ are $r_{x_k, \hat{y}_i} = \frac{\hat{e}_{ki}\sqrt{\hat{\lambda}_i}}{\sqrt{s_{kk}}}$, $i, k = 1, \dots, p$.*

In the case that the random variables have a normal distribution, the principal components can be obtained from a maximum likelihood estimation $\hat{\Sigma} = \mathbf{S}_n$, and, in this case, the sampling principal components can be considered as maximum likelihood estimates of the population principal components. Although the eigenvalues of $\mathbf{S}$ and $\hat{\Sigma}$ are different but proportional, with constant proportionality fixed, the proportion of variability they explain is the same. The sample correlation matrix is the same for $\mathbf{S}$ and $\hat{\Sigma}$. We still do not consider the particular case of normal distribution of the variables, so as not to have to include hypotheses that should be verified for the data under study.

Sometimes, the observations $\mathbf{x}$ are centered by subtracting the mean $\overline{\mathbf{x}}$. This operation does not affect the covariance matrix and produces principal components of the form $\hat{y}_i = \hat{\mathbf{e}}_i^t(\mathbf{x} - \overline{\mathbf{x}})$, and in this case $\overline{\hat{y}}_i$ for any component, while the sample variances remain $\hat{\lambda}_1, \dots, \hat{\lambda}_p$.

When trying to interpret the principal components, the correlation coefficients $r_{x_k, \hat{y}_i}$ are more reliable guides than the coefficients $\hat{e}_{ik}$, since they avoid interpretive problems caused by the different scales in which the variables are measured.

4.1. Interpretations of the principal sample components

Principal sample components have several interpretations. If the distribution of $\mathbf{X}$ is close to $N_p(\boldsymbol{\mu}, \Sigma)$, then components $\hat{y}_i = \hat{\mathbf{e}}_i^t(\mathbf{x} - \overline{\mathbf{x}})$ are realizations of the main population components $Y_i = \mathbf{e}_i^t(\mathbf{X} - \boldsymbol{\mu})$, which will have distribution $N_p(\mathbf{0}, \boldsymbol{\Lambda})$, where $\boldsymbol{\Lambda}$ is the diagonal matrix whose elements are the eigenvalues, ordered from major to minor, from the sample covariance matrix. Keeping in mind the hypothesis of normality, contours of constant density, $E_p = \{\mathbf{x} \in \mathfrak{R}^p | (\mathbf{x} - \overline{\mathbf{x}})^t \mathbf{S}^{-1} (\mathbf{x} - \overline{\mathbf{x}}) = c^2\}$, can be estimated and make inferences from them.

Although it is not possible to assume normality in the data, geometrically the data are n points $\mathfrak{R}^p$, and the principal components represent an orthogonal transformation whose coordinate axes are the axes of the ellipsoid E_p and with lengths proportional to $\sqrt{\hat{\lambda}_i}$, with $\hat{\lambda}_i$ being the eigenvalues of $\mathbf{S}$. Since all eigenvectors have been chosen such that their norm is equal to 1, the absolute value of the ith component $[\hat{y}_i = |\hat{\mathbf{e}}_i^t (\mathbf{x} - \overline{\mathbf{x}})|$ is the length of the projection of the vector $(\mathbf{x} - \overline{\mathbf{x}})$ on the vector $\hat{\mathbf{e}}_i$. Therefore, the principal components can be seen as a translation of the origin to the point $\overline{\mathbf{x}}$ and a rotation of the axes until they pass through the directions with greater variability.

When there is a high positive correlation between all the variables and a principal component with all its coordinates of the same sign, this component can be considered as a weighted average of all the variables or the size of the index that forms that component. The components that have coordinates of different signs oppose a subset of variables against another, being a weighted average of two groups of variables.

The interpretation of the results is simplified assuming that the small coefficients are zero and rounding the rest to express the component as sums, differences, or quotients of variables.

The interpretation of the principal components can be facilitated by graphic representations in two dimensions. A usual graph is to represent two components as coordinate axes and project all points on those axes. These representations also help to test hypotheses of normality and to detect anomalous observations. If there is an observation that is atypical in the first variable, we will have that the variability in that first variable will grow and that the covariance with the other variables will decrease, in absolute value. Consequently, the first component will be strongly influenced by the first variable, distorting the analysis.

Sometimes, it is necessary to verify that the first components are approximately normal, although it is not reasonable to expect this result from a linear combination of variables that do not have to be normal.

The last component can help detect suspicious observations. Each observation $\mathbf{x}$ can be expressed as a linear combination of the eigenvectors of $\mathbf{S}$, $\mathbf{x}_j = \hat{y}_{1j}\hat{\mathbf{e}}_1 + \cdots + \hat{y}_{pj}\hat{\mathbf{e}}_p$, with which the difference between the first components $\hat{y}_{1j}\hat{\mathbf{e}}_1 + \cdots + \hat{y}_{qj}\hat{\mathbf{e}}_q$ and the observation $\mathbf{x}_j$ is $\hat{y}_{q-1j}\hat{\mathbf{e}}_{q-1} + \cdots + \hat{y}_{pj}\hat{\mathbf{e}}_p$, which is a vector with square of the norm $\hat{y}_{q-1j}^2 + \cdots + \hat{y}_{pj}^2$, and we will suspect of observations that have a large contribution to the square of the aforementioned norm.

An especially small value of the last eigenvalue of the covariance matrix, or correlation matrix, can indicate a linear dependence between the variables that have not been taken into account. In this case, some variable is redundant and should be removed from the analysis. If we have four variables and the fourth is the sum of the other three, then the last eigenvalue will be close to zero due to rounding errors, in which case we should suspect some dependence. In general, eigenvalues close to zero should not be ignored, and eigenvalues associated with these eigenvalues can indicate linear dependencies in the data and cause deformations in the interpretations, calculations, and consequent analysis.

4.2. Standardized sample principal components

In general, principal components are not invariant against changes of scale in the original variables, as has been mentioned when referring to the normalized population principal components. Normalizing, or standardizing, the variables consists of performing the following transformation $\mathbf{z}_j = \mathbf{D}(\mathbf{x}_j - \overline{\mathbf{x}}) = \left[\frac{x_{1j}-\overline{x}_1}{\sqrt{s_{11}}}, \dots, \frac{x_{pj}-\overline{x}_p}{\sqrt{s_{pp}}}\right]^t$, $j = 1, \dots, p$. If the matrix $\mathbf{Z}$ is the p by n matrix whose columns are $\mathbf{z}_j$, it can be shown that its sample mean vector is the null vector and that its correlation matrix is the sample correlation matrix, $\mathbf{R}$, of the original variables.

Remark 4.1: *Applying that the principal components of the normalized variables are those obtained for the sample observations but substituting the matrix* $\mathbf{S}$ *for* $\mathbf{R}$*, we can establish that if* $\mathbf{z}_1, \dots, \mathbf{z}_n$ *are the normalized observations, with covariance matrix* $\mathbf{R} = (r_{ik})$*, where* r_{ik} *is the sample correlation coefficient between observations* $\mathbf{x}_i$ *and* $\mathbf{x}_k$*, and if the pairs of eigenvalues and eigenvectors of* $\mathbf{R}$ *are* $(\hat{v}_1, \hat{\mathbf{u}}_1), \dots, (\hat{v}_p, \hat{\mathbf{u}}_p)$*, with* $\hat{v}_1 \geq \dots \geq \hat{v}_p \geq 0$*, then*

1. *The ith principal component is given by* $\hat{\omega}_i = \hat{\mathbf{u}}_i^t \mathbf{z} = \hat{u}_{1i} z_1 + \dots + \hat{u}_{pi} z_p$, $i = 1, \dots, p$.
2. *The sample variance of* $\hat{\omega}_k$ *is* $\hat{v}_k$, $k = 1, \dots, p$.
3. *The sample covariance of* $(\hat{\omega}_i, \hat{\omega}_k)$, $i \neq k$, *is equal to 0.*
4. *The total sample variance is* $tr(\mathbf{R}) = p = \hat{v}_1 + \dots + \hat{v}_p$.
5. *The sample correlation coefficients between* z_k*and* $\hat{\omega}_i$ *are* $r_{z_k, \hat{\omega}_i} = \hat{u}_{ki}\sqrt{\hat{v}_i}$, $i, k = 1, \dots, p$.
6. *The proportion of the total sample variance explained by the ith principal component is* $\frac{\hat{v}_i}{p}$.

4.3. Criteria for reducing the dimension

The eigenvalues and eigenvectors of the covariance matrix, or correlation matrix, are the essence of the analysis of principal components, since the eigenvalues indicate the directions of maximum variability and the eigenvectors determine the variances. If a few eigenvalues are much larger than the rest, most of the variance can be explained with less than p variables.

In practice, decisions about the number of components to be considered must be made in terms of the pairs of eigenvalues and eigenvectors of the covariance matrix, or correlation matrix, and different rules have been suggested:

a. When performing the graph $(i, \hat{\lambda}_i)$, it has been empirically verified that with the first values there is a decrease with a linear tendency of quite steep slope and that from a certain eigenvalue this decrease is stabilized. That is, there is a point from which the eigenvalues are very similar. The criterion consists of staying with the components that exclude the small eigenvalues and that are approximately equal.

http://dx.doi.org/10.5772/intechopen.75007

b. Select components until obtaining a proportion of the preset variance (e.g., 80%). This rule should be applied with care, since components that are interesting to reflect certain nuances suitable for the interpretation of the analysis could be excluded.

c. A rule that does not have a great theoretical support, which must be applied carefully so as not to discard any valid component for the analysis, but which has given good empirical results, is to retain those components with variances, $\hat{\lambda}_i$, above a certain threshold. If the work matrix is the correlation matrix, in which case the average value of the eigenvalues is one, the criterion is to keep the components associated with eigenvalues greater than unity and discard the rest.

5. Application to image compression

We are going to illustrate the use of principal components to compress images. To this end, the image of Lena was considered. This photograph has been used by engineers, researchers, and students for experiments related to image processing.

5.1. Black and white photography

The black and white photograph shown in **Figure 6** was considered. First, the image in .jpg format was converted into the numerical matrix **Image** of dimension 512 by 512 (i.e., 2^9x2^9). Second, to obtain the observation vectors, the matrix was divided into blocks of dimension 2^3x2^3, $\mathbf{A}_{ij}$, with which 4096 blocks were obtained, and each of them was a vector of observations.

Figure 6. Black and white photograph of Lena.

$$\mathbf{Image} = \begin{bmatrix} \mathbf{A}_{1,1} & \dots & \mathbf{A}_{1,64} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{64,1} & \dots & \mathbf{A}_{64,64} \end{bmatrix} \tag{17}$$

Third, each matrix $\mathbf{A}_{ij}$ was stored in a vector of dimension 64, $\mathbf{x}$, which contained the elements of the matrix by rows, that is, $\mathbf{x} = [a_{i,1}, \dots, a_{i,8}, a_{i+1,1}, \dots, a_{i+1,8}, \dots, a_{i+8,8}]$. This way, we had the observations $\{\mathbf{x}_k \in \mathfrak{R}^{64} | k = 1, \dots, 4096\}$, which were grouped in the observation matrix $\mathbf{x} = [x_{ij}] \in \mathrm{M}_{4096,64}(\mathfrak{R})$.

Fourth, the average of each column, $\overline{\mathbf{x}} = [\overline{x}_1, \dots, \overline{x}_{64}]$, was calculated obtaining the vector of means, and from each observation x_{ij}, its corresponding mean $\overline{x}_j$ was subtracted. Thus, the matrix of centered observations $\mathbf{U}$ was obtained. The covariance matrix of $\mathbf{x}$ was $\mathbf{S} = \mathbf{U}^t\mathbf{U} \in \mathrm{M}_{64,64}(\mathfrak{R})$.

Fifth, the 64 pairs of eigenvalues and eigenvectors of $\mathbf{S}$, $\left(\hat{\lambda}_i, \hat{\mathbf{e}}_i\right)$, were found, and they were ordered according to the eigenvalues from highest to lowest. The 8 largest eigenvalues are drawn in **Figure 7**. As can be seen, the first eigenvalue is much larger than the rest. Thus, the first principal component completely dominates the total variability.

Sixth, with the theoretical results and the calculations previously made, the 64 principal components $\hat{y}_j = \hat{\mathbf{e}}_j^t \mathbf{x} = \hat{e}_{1,j} x_1 + \cdots + \hat{e}_{64,j} x_p$, $j = 1, \dots, p$, were built. The first principal component was $\hat{y}_1 = -0.1167 x_1 + \cdots - 0.1166 x_{64}$. Therefore, an orthonormal basis of $\mathfrak{R}^{64}$ was built.

Seventh, each vector $\hat{\mathbf{e}}_j = \left[\hat{e}_{1,j}, \dots, \hat{e}_{64,j}\right]^t$ was grouped by rows in a matrix $\mathrm{M}_{8,8}$:

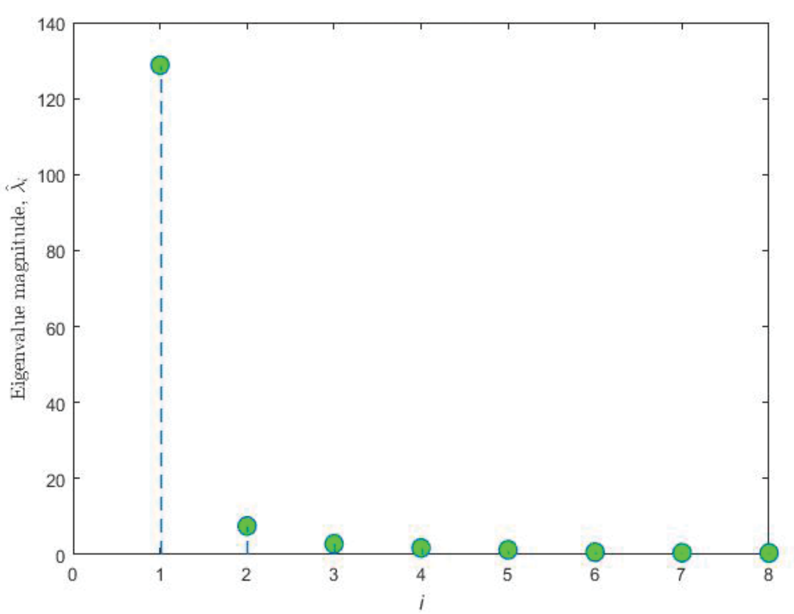

Figure 7. Graph $\left(i, \hat{\lambda}_i\right)$, $i = 1, \dots, 8$, with $\hat{\lambda}_i$ being the eigenvalues ordered from highest to lowest.

$$\hat{\mathbf{E}}_j = \begin{bmatrix} \hat{e}_{1,j} & \cdots & \hat{e}_{8,j} \\ \vdots & \ddots & \vdots \\ \hat{e}_{57,j} & \cdots & \hat{e}_{64,j} \end{bmatrix} \quad (18)$$

Each of the 64 matrices $\hat{\mathbf{E}}_j$ was converted into an image. The images of the first three principal components are shown in **Figure 8**.

At this point, it is important to mention that the data matrix $\mathbf{x}$ has been assumed to be formed by 4096 vectors of $\mathfrak{R}^{64}$ expressed in the canonical base, $\boldsymbol{B}$. Also, the base whose vectors were the eigenvectors of $\mathbf{S}$, $\boldsymbol{B}' = \{\hat{\mathbf{e}}_1, \ldots, \hat{\mathbf{e}}_{64}\}$, was considered. The coordinates with respect to the canonical basis of the vectors of $\boldsymbol{B}'$ were the columns of the matrix $\mathbf{PC} = \left[\hat{\mathbf{e}}_1^t, \ldots, \hat{\mathbf{e}}_{64}^t\right]$. Then, given a vector $\mathbf{v}$ that with respect to the canonical base had coordinates $(x_1, \ldots, x_{64})$ and with respect to the base $\boldsymbol{B}'$ had coordinates $(y_1, \ldots, y_{64})$, the relation between them was $[x_1, \ldots, x_{64}]^t = \mathbf{PC}[y_1, \ldots, y_{64}]^t$. Also, as $\mathbf{PC}$ is an orthogonal matrix, $[y_1, \ldots, y_{64}] = [x_1, \ldots, x_{64}]\mathbf{PC}$. Thus, the coordinates of the 4096 vectors that formed the observations matrix had as coordinates, with respect to the new base, the rows of the matrix of dimension 4096x64 given by $\mathbf{y} = \mathbf{x} \cdot \mathbf{PC}$.

Eight, in order to reduce the dimension, it was taken into consideration that if we keep all the vectors of $\boldsymbol{B}'$, we can perfectly reconstruct our data matrix, because $\mathbf{y} = \mathbf{x} \cdot \mathbf{PC} \Rightarrow \mathbf{x} = \mathbf{y} \cdot \mathbf{PC}^{-1} = \mathbf{y} \cdot \mathbf{PC}^t$. Additionally, for the case under study, to reduce the dimension, if we use the slope change rule, we can consider the first two principal components; five components if we want to explain 97% of the variability, because $\sum_{i=1}^{5} \hat{\lambda}_i / \sum_{j=1}^{64} \hat{\lambda}_j = 97\%$; or eight components if we want to explain 98% of the total variability.

In order to compress the image, the first vectors of the base $\boldsymbol{B}'$ were used. Moreover, supposing that we were left with M, $M < 64$, the matrix $\mathbf{T}_M$ given by Eq. (19) was defined:

$$\mathbf{T}_M = \begin{bmatrix} \mathbf{I}_{M\text{x}M} & \mathbf{0}_{M\text{x}(64-M)} \\ \mathbf{0}_{(64-M)\text{x}M} & \mathbf{0}_{(64-M)\text{x}(64-M)} \end{bmatrix} \quad (19)$$

Therefore, the dimension of $\mathbf{y}_M = \mathbf{y} \cdot \mathbf{T}_M$ was 4096×64.

Ninth, to reconstruct the compressed image, each row of $\mathbf{y}_M$ was regrouped in an 8x8matrix. The ith row of $\mathbf{y}_M$, denoted by $\mathbf{y}_{Mi} = [b_{i,1}, \ldots, b_{i,8}, b_{i,9}, \ldots, b_{i,16}, \ldots, b_{i,64}]$, was transformed into

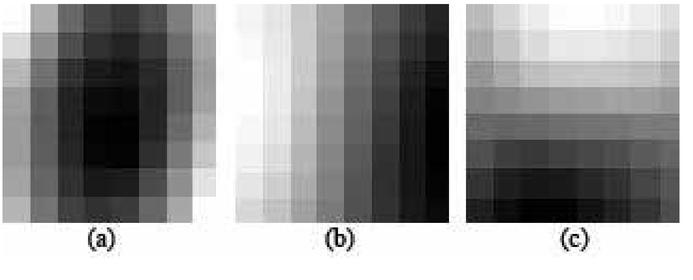

Figure 8. Images of the matrices of the first three principal components. (a) First component. (b) Second component. (c) Third component.

the matrix $\mathbf{B}_i$ given by Eq. (20), and the matrix **Compressed_image** given by Eq. (21) was built:

$$\mathbf{B}_i = \begin{bmatrix} b_{i,1} & \cdots & b_{i,8} \\ b_{i,9} & \cdots & b_{i,16} \\ \vdots & \ddots & \vdots \\ b_{i,57} & \cdots & b_{i,64} \end{bmatrix} \quad i = 1, \ldots, 4096 \tag{20}$$

$$\mathbf{Compressed\_image} = \begin{bmatrix} \mathbf{B}_1 & \cdots & \mathbf{B}_{64} \\ \mathbf{B}_{65} & \cdots & \mathbf{B}_{128} \\ \vdots & \ddots & \vdots \\ \mathbf{B}_{4033} & & \mathbf{B}_{4096} \end{bmatrix} \tag{21}$$

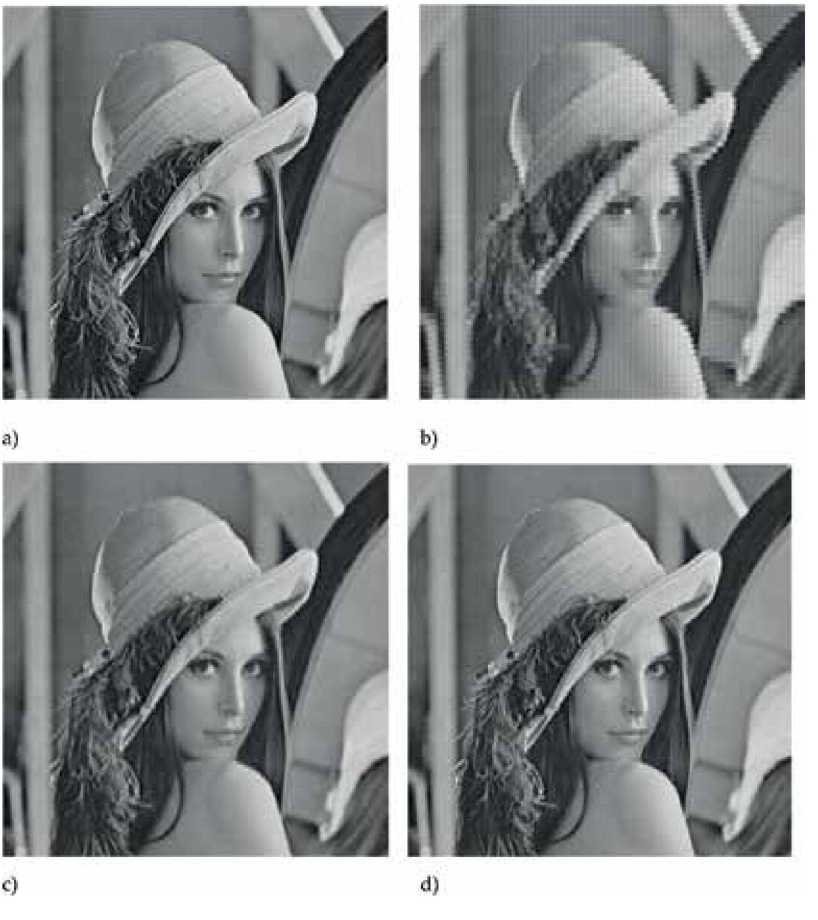

Figure 9. Original and compressed image with two, five, and eight principal components. (a) Original image. (b) Compression with two components. (c) Compression with five components. (d) Compression with eight components.

Tenth and finally, Eq. (21) was converted into a .jpg file. **Figure 9** shows the original image and compressed images with two, five, and eight principal components.

By increasing the number of principal components, the percentage of the variability explained is increased by very small percentages, but, nevertheless, nuances are added to the photo sufficiently remarkable, since they make it sharper, smooth out the contours, and mark the tones more precisely.

5.1.1. Objective measures of the quality of reconstructions

The two methods that we will use are the peak signal-to-noise ratio (*PSNR*) and the entropy of the error image. The PSNR measure evaluates the quality in terms of deviations between the processed and the original image, and the entropy of an image is a measure of the information content contained in that image.

Definition 5.1: *Let N be the number of rows by the number of columns in the image. Let* $\{x_n|n = 1, \ldots, N\}$ *be the set of pixels of the original image. Let* $\{y_n|n = 1, \ldots, N\}$ *be the set of reconstruction pixels. Let* $\{r_n = x_n - y_n|n = 1, \ldots, N\}$ *be the error. The mean square error (MSE) is*

$$MSE = \frac{1}{N}\sum_{n=1}^{N} r_n^2 \tag{22}$$

Definition 5.2: *Let the images under study be the* 8 *bit images. The peak signal-to-noise ratio of the reconstruction is*

$$PSNR = 10\log_{10}\left(\frac{\left(2^8 - 1\right)^2}{MSE}\right) \tag{23}$$

Figure 10 (a) shows *PSNR* of the reconstructions of the image versus the number of principal components used for the reconstruction, together with the regression line that adjusts the said cloud of points. **Figure 10 (b)** shows the values of the PSNR when we use three quarters (black), half (red), quarter (blue), eighth (green), sixteenth (brown), and the thirty-second part (yellow) of the components, which means a corresponding reduction in compression. A behavior close to linearity with a slope of approximately 0.2 can be seen. With the reductions considered, the *PSNR* varies between 27 and 63.

If the entropy is high, the variability of the pixels is very high, and there is little redundancy. Thus, if we exceed a certain threshold in compression, the original image cannot be recovered exactly. If the entropy is small, then the variability will be smaller. Therefore, the information of a pixel with respect to the pixels of its surroundings is high and, therefore, randomness is lost.

Definition 5.3: *Let I be an* 8 *bit image that can take the values* $\{0, \ldots, 255\}$. *Let* p_i *be the frequency with which the value* $i \in \{0, \ldots, 255\}$ *appears. Then, the entropy is*

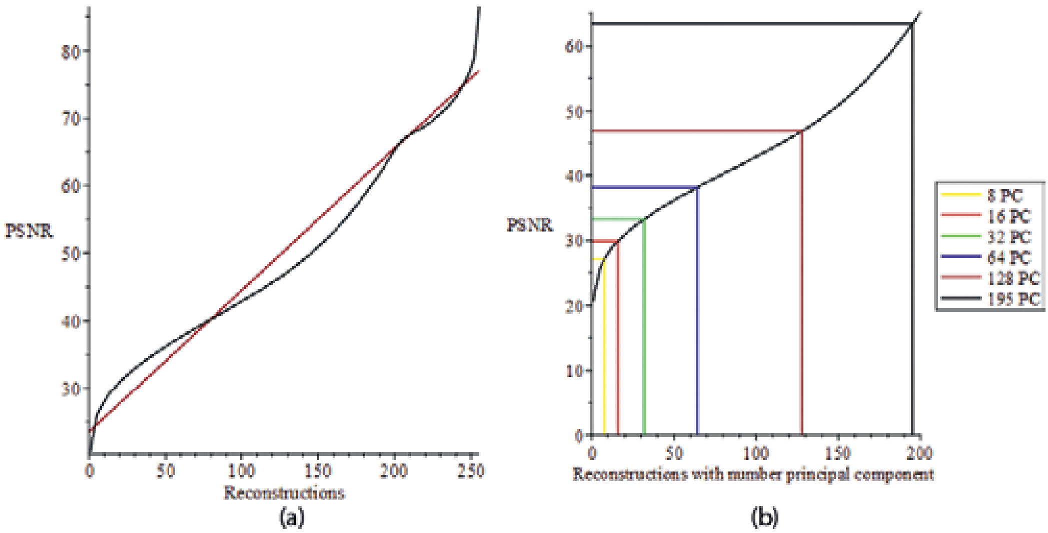

Figure 10. *PSNR* of the reconstructions according to the used principal components. (a) *PSNR*of 256 reconstructions. (b) *PSNR* of some reconstructions.

$$H(I) = -\sum_{i=0}^{255} p_i \log_2(p_i) \tag{24}$$

Figure 11 (a) shows the entropy of the reconstructions from 1 to 256 components. As can be seen, the entropy is increasing until the first 10 components, and then it becomes damped tending asymptotically to the value of the entropy of the image (7.4452). It can be seen that the difference with more than 170 components is insignificant. **Figure 11 (b)** shows the entropy of

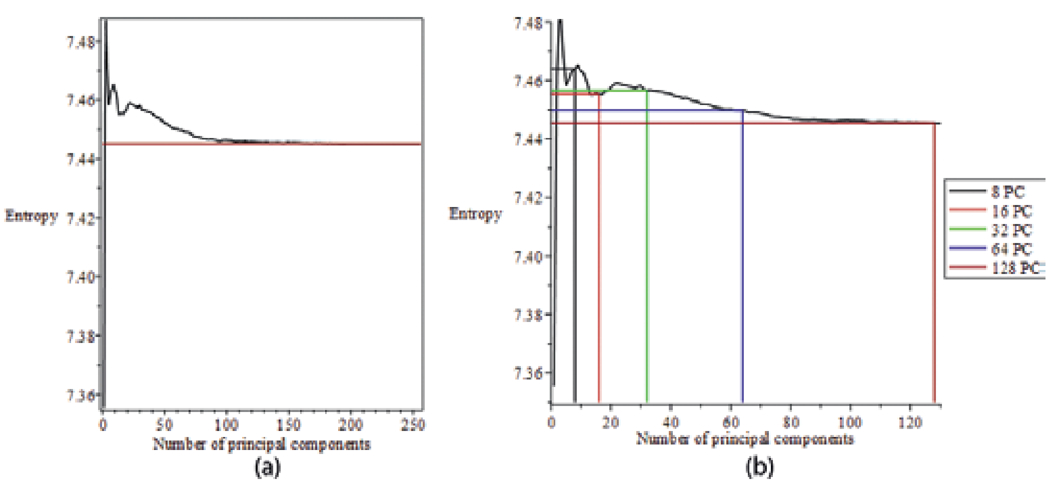

Figure 11. Entropy of reconstructions according to the used principal components. (a) Entropy of reconstructions. (b) Entropy of some reconstructions.

the reconstructions using 8 components (black), 16 components (brown), 32 components (green), 64 components (blue), and 128 components (red), respectively.

Finally, we consider the entropy of the images of the errors. Given an image, I, the value of each of its pixels is an element of the set $\{0, \ldots, 255\}$, and if we have a reconstruction, $\hat{I}$, and consider the error, $E = I - \hat{I}$, then the value of its pixels will be an element of the set $\{-255, \ldots, 255\}$. Therefore, E cannot be considered as an image. Since a pixel of value e_{ij} in E is an error of the same size as $-e_{ij}$, to consider images we denominate image of the error to $\mathrm{Im}(E) = [|e_{ij}|]$, being $E = [e_{ij}]$.

Figure 12 (a) shows the entropy of the error image versus the number of principal components used for the reconstruction, together with an adjusted line of slope $-$ 0.02. **Figure 12(b)** shows the entropy when we use 8 components (black), 16 components (brown), 32 components (green), 64 components (blue), and 128 components (red), respectively. With more than 200 principal components, the entropy of the errors is zero, which means that the errors have very little variability, and with fewer components, the decrease seems linear with slope -0.02.

5.2. Coordinates of the first principal component

In this section, we will consider the coordinates of the first vectors that form the principal components. If we consider that the vectors have been obtained as 2^3x2^3 dimension blocks, vectors will have 64 coordinates. **Figure 13** shows the coordinates of the first six principal components with respect to the canonical base.

As can be seen from **Figure 13**, all coordinates seem to have some component with period 8. This suggests that there may be some relationship with the shape of the blocks chosen and that most vectors are close to being periodic with period 8, because when we consider each of the

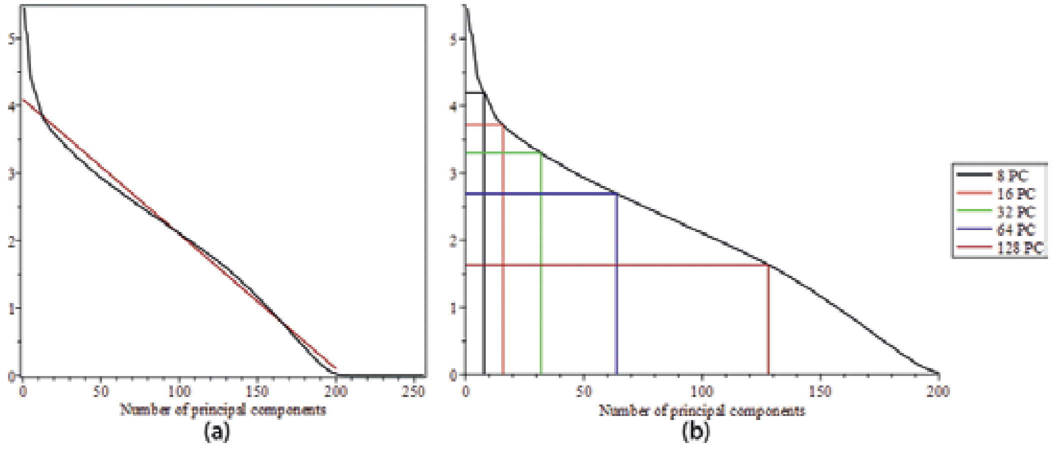

Figure 12. Entropy of the errors of the reconstructions converted into images according to the used principal components. (a) Entropy of differences (b) Entropy of some differences.

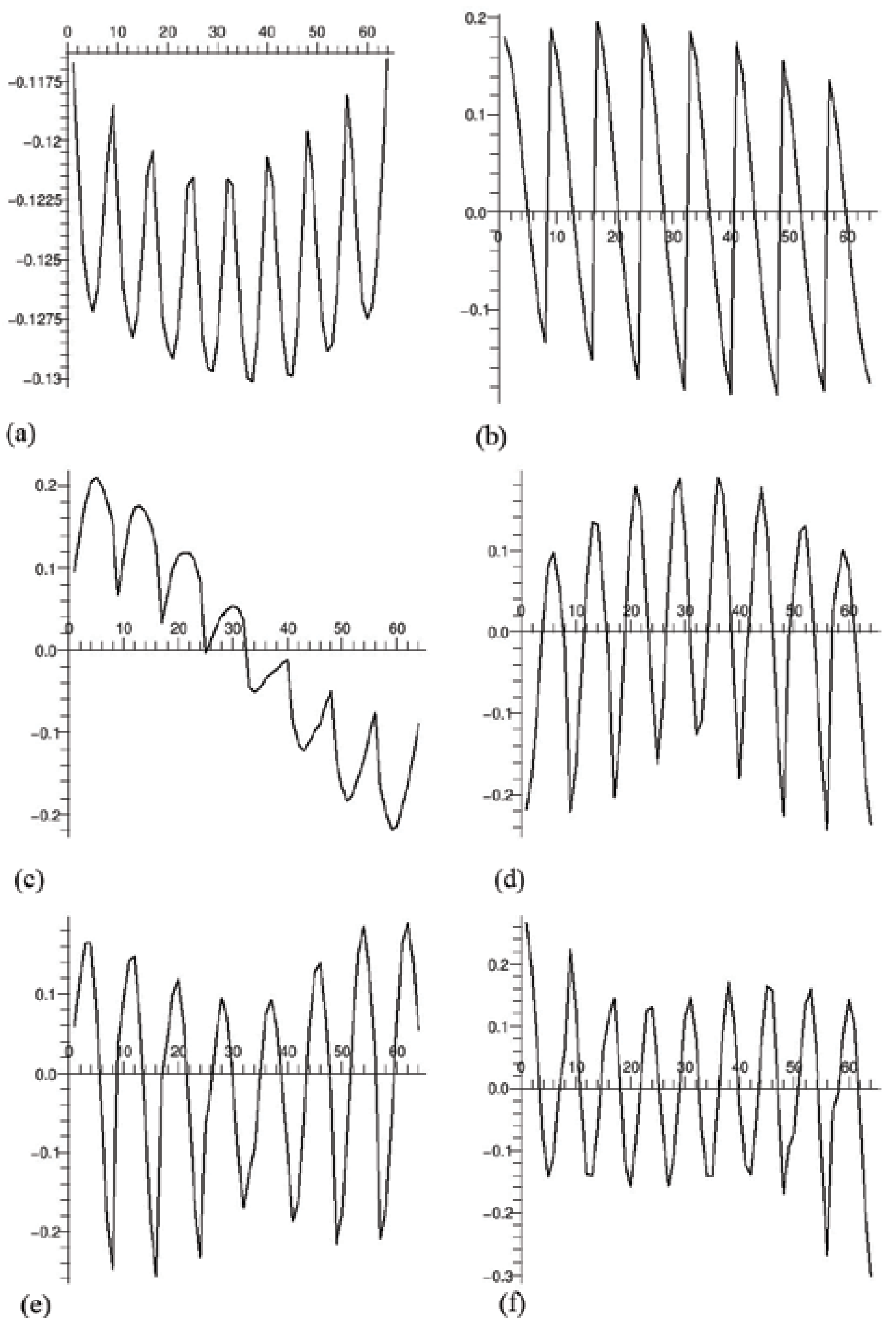

Figure 13. Coordinates of the first six principal components with respect to the canonical base. (a) First component. (b) Second component. (c) Third component. (d) Fourth component. (e) Fifth component. (f) Sixth component.

http://dx.doi.org/10.5772/intechopen.75007

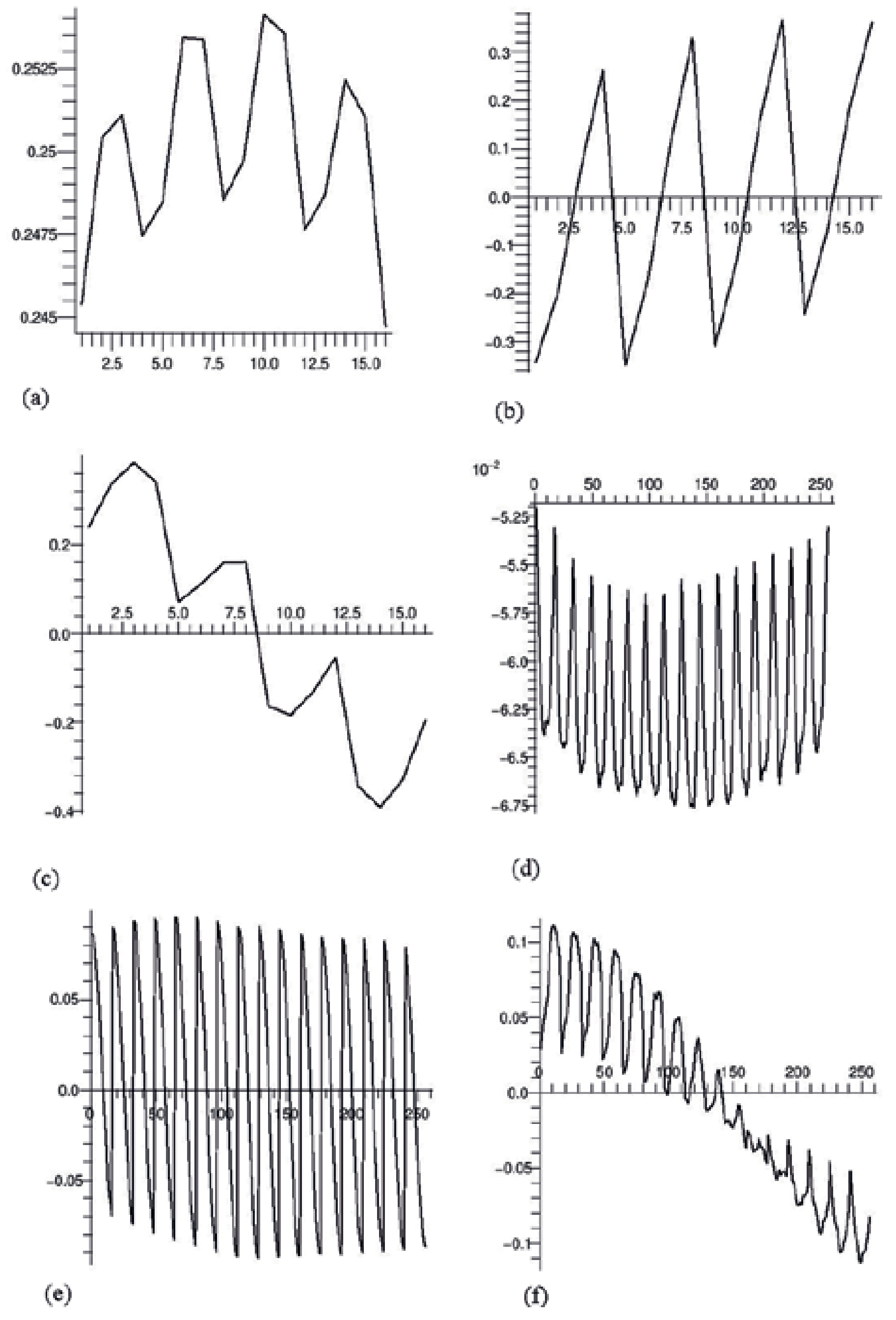

Figure 14. Coordinates of the first three principal components when vectors are constructed from blocks of 2^2x2^2 and 2^4x2^4. (a) First component 2^2x2^2 (b) Second component 2^2x2^2 (c) Third component 2^2x2^2 (d) First component 2^4x2^4 (e) Second component 2^4x2^4 (f) Third component 2^4x2^4.

4096 vectors of 64 components, the first 8 pixels are adjacent to the next 16 pixels, and these are adjacent to the next 8 pixels, and so on, up to 8 times.

Since the first principal components collect a large part of the characteristics of the vectors, it is plausible that they also reflect the periodicity of the vectors. Recall that principal components are linear combinations of vectors and that if all of them had all their periodic coordinates with the same period, then all components would be periodic as well.

In **Figure 14**, the coordinates of the first three principal components are shown when the vectors are constructed from blocks of 2^2x2^2(see **Figure 14 (a-c)**) and from blocks of 2^4x2^4 (see **Figure 14 (d-f)**). As can be seen, the periodicity in the first components is again appreciated.

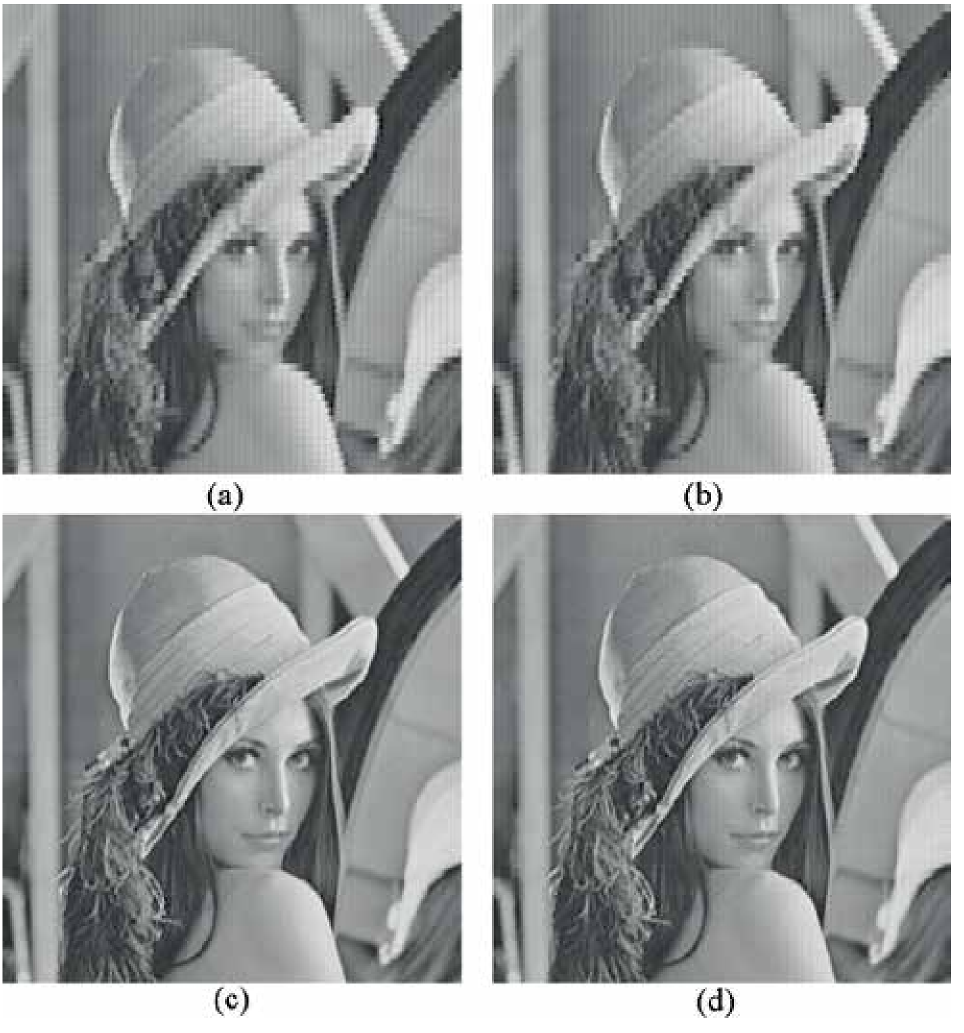

Figure 15. Compression with 2 and 8 original and periodic principal components. (a) Compression with two components (b) Compression with two $components_{per}$. (c) Compression with eight components. (d) Compression with eight $components_{per}$.

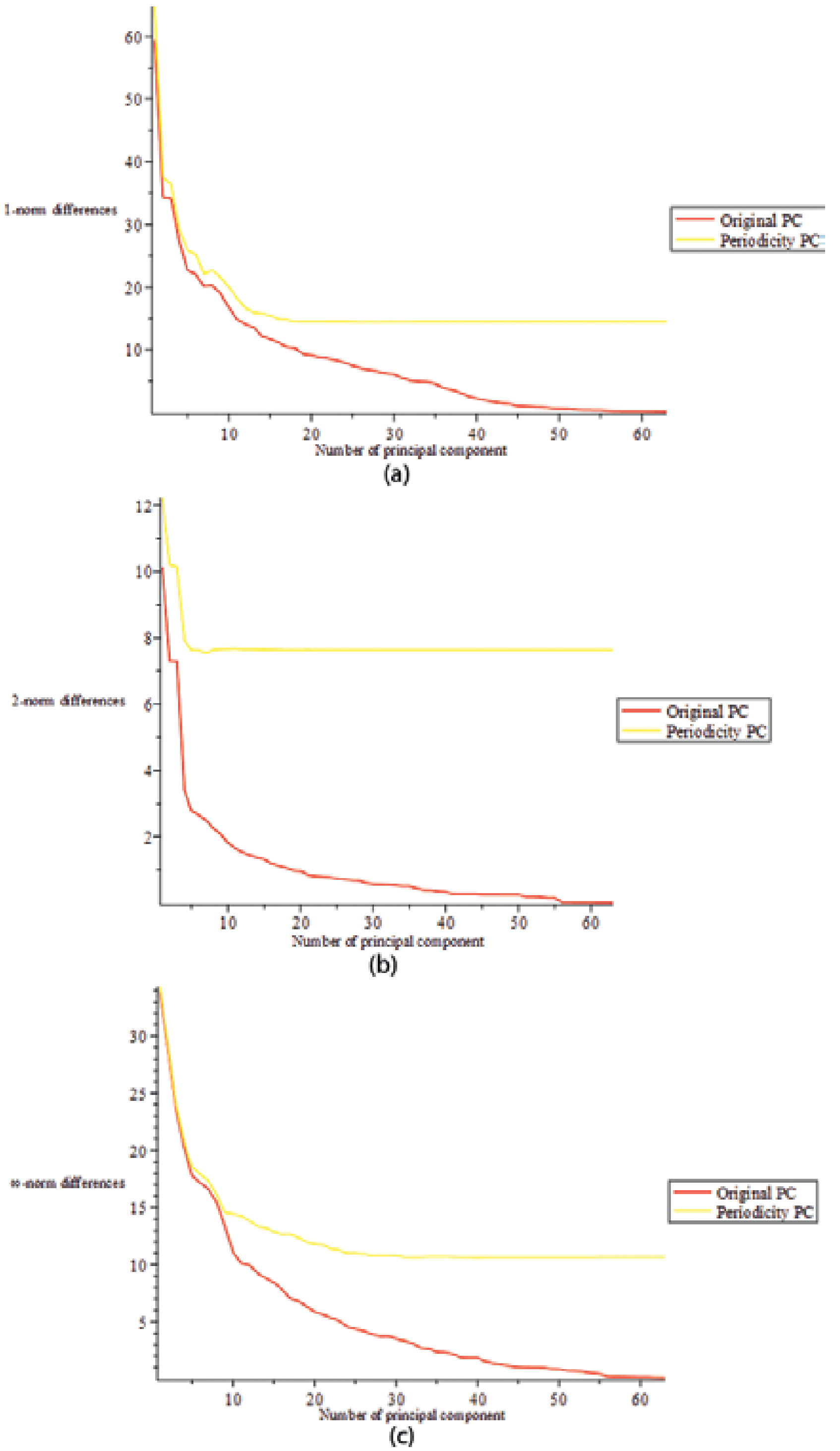

Figure 16. Differences between the image and the reconstruction according to the number of chosen components. (a) 1-norm (b) 2-norm (c) ∞-norm.

5.3. Reduction of the first principal component by periodicity

Using the almost periodicity of the first principal component, we can use less information to obtain acceptable reconstructions of the image. If in the first principal component of dimension 64 we repeat the first eight values periodically and use k principal components to reconstruct the image, we go from a reduction of $k/64$ to another of $[(k-1)+8/64]/64$. **Figure 15** shows both the reconstruction of the image with 2 and 8 original principal components and the reconstruction of the image with 2 and 8 principal components, but with the first component replaced by a vector whose coordinates have period 8, we call this $components_{per}$.

The first $components_{per}$ component is not the true one. Therefore, reconstructions from this set cannot be made with total precision. If we use to compare the 1-norm, 2-norm, and ∞-norm of the image and the corresponding reconstruction, with the original principal components and the principal components using their periodicity, we obtain, by varying the number of used principal components, the results shown in **Figure 16**.

With the original principal components (blue), the original image can be completely reconstructed, while if we use only a few components, in this case 10 or less, approximations similar to the original ones are obtained with $components_{per}$ (green).

6. Conclusions

This chapter has been devoted to give a short but comprehensive introduction to the basics of the statistical technique known as principal component analysis, aimed at its application to image compression. The first part of the chapter was focused on preliminaries, mean vector, covariance matrix, eigenvectors, eigenvalues, and distances. That part finished bringing up the problems that the Euclidean distance presents and highlights the importance of using a statistical distance that takes into account the different variabilities and correlations. To that end, a brief introduction was made to a distance that depends on variances and covariances.

Next, in the second part of the chapter, principal components were introduced and connected with the previously explained concepts. Here, principal components were presented as a particular case of linear combinations of random variables, but with the peculiarity that those linear combinations represent a new coordinate system that is obtained by rotating the original reference system, which has the aforementioned random variables as coordinate axes. The new axes represent the directions with maximum variability and provide a simple description of the structure of the covariance.

Then, the third part of the chapter was devoted to show an application of principal component analysis to image compression. An original image was taken and compressed by using different principal components. The importance of carrying out objective measures of quality reconstructions was highlighted. Also, a novel contribution of this chapter was the introduction to the study of the periodicity of the principal components and to the importance of the reduction of the first principal component by periodicity. In short, a novel construction of principal components by periodicity of principal components has been included, in order to reduce the computational cost for their calculation, although decreasing the accuracy. It can be said that using the almost periodicity of the first principal component, less information to obtain acceptable reconstructions of the image can be used.

Finally, we would not like to finish this chapter without saying that few pages cannot gather the wide range of applications that this statistical technique has found in solving real-life problems. There is a countless number of applications of principal component analysis to solve problems that both scientists and engineers have to face in real-life situations. However, in order to be practical, it was decided to choose and develop step by step an application example that could be of interest for a wide range of readers. Accordingly, we thought that such an example could be one related to data compression, because with the advancements of information and communication technologies both scientists and engineers need to either store or transmit more information at lower costs, faster, and at greater distances with higher quality. In this sense, one example is image compression by using statistical techniques, and this is the reason why, in this chapter, it was decided to take advantage of statistical properties of an image to present a practical application of principal component analysis to image compression.

Acknowledgements

This work was supported by the Universidad de Las Americas, Ecuador, and the Universidad Politecnica de Madrid, Spain.

Author details

Wilmar Hernandez[1]* and Alfredo Mendez[2]

*Address all correspondence to: wilmar.hernandez@udla.edu.ec

1 Universidad de Las Americas, Quito, Ecuador

2 Universidad Politecnica de Madrid, Madrid, Spain

References

[1] Jackson JE. A User's Guide to Principal Components. John Wiley & Sons; 1991

[2] Diamantaras KI, Kung SY. Principal Component Neural Networks: Theory and Applications. John Wiley & Sons; 1996

[3] Elsner JB, Tsonis AA. Singular Spectrum Analysis: A New Tool in Time Series Analysis. Plenum Press; 1996

[4] Rencher AC. Multivariate Statistical Inference and Applications. 1st ed. Wiley-Interscience; 1997

[5] Flury B. A First Course in Multivariate Statistics. Springer-Verlag; 1997

[6] Gnanadesikan R. Methods for Statistical Data Analysis of Multivariate Observations. 2nd ed. Wiley-Interscience; 1997

[7] Jolliffe IT. Principal Component Analysis. 2nd ed. Springer; 2002

[8] Wichern DW, Johnson RA. Applied Multivariate Statistical Analysis. 6th ed. Pearson Education Limited; 2014

[9] Strang G. Linear Algebra and its Applications. 4th ed. Thomson; 2006

[10] Larson R, Falvo D. Elementary Linear Algebra. 6th ed. Houghton Mifflin Harcourt Publishing Company; 2009

[11] Lay DC, Lay SR, McDonald JJ. Linear Algebra and its Applications. 5th ed. Pearson; 2015

Chapter 8

Severe Nuclear Accidents and Learning Effects

Thomas Rose and Trevor Sweeting

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.76637

Abstract

Nuclear accidents with core melting as the ones in Fukushima and Chernobyl play an important role in discussing the risks and chances of nuclear energy. They seem to be more frequent than anticipated. So, we analyse the probability of severe nuclear accidents related to power generation. In order to see learning effects of reactor operators, we analyse the number of all known accidents in time. We discuss problems of data acquisition, statistical independence of accidents at the same site and whether the known accidents form a random sample. We analyse core melt accidents with Poisson statistics and derive future accident probabilities. The main part of the chapter is the investigation of the learning effects using generalised linear models with a frequentist and a Bayesian approach and the comparison of the results.

Keywords: nuclear accidents, learning effect, Poisson distribution, generalised linear model, frequentist approach, Bayesian approach

1. Introduction

The Fukushima reactor disaster in 2011 made the question of nuclear safety relevant again. Similar accidents are known to have happened in the Soviet Union in 1986 (Chernobyl) and in the USA in 1979 (Three Mile Island). These core melt accidents are the most severe ones in nuclear reactors. When the rods containing the nuclear fuel and the fission products melt, a huge amount of radioactivity is set free within the reactor and possibly into the atmosphere.

But the rate of such accidents seemed much higher than previously claimed. So, we tried to study the probability of such events empirically by looking at the real events.

This a posteriori approach differs from the a priori approach of Probabilistic Risk Assessment (PRA) which is done during the design phase of a reactor. PRA determines failure probability prior to accidents by analysing possible paths towards a severe accident, rather than using existing data to determine probability empirically.

After an accident very often ‘learning from experience’ is claimed. The luckily low number of severe accidents does not allow for testing this claim. But reactor operators should be interested in reducing all incidents and accidents; so, their frequency should decrease with increasing operating experience. We use the total time reactors are operating, the reactor-years, as a measure of experience, analyse the accidents as a function of this experience with generalised linear models and compare a frequentist and a Bayesian approach.

Accidents can and did happen in several areas of nuclear energy, e.g. military use for weapons or submarine propulsion, medical use or fundamental research. Discussing the risks of nuclear energy involves very different arguments in all these areas. We restricted the study to accidents in nuclear reactors for power generation.

According to our analysis, we have to expect one core melt accident in 3700 reactor-years with a 95% confidence interval of one in 1442 reactor-years and one in 13,548 reactor-years. In a world with 443 reactors, with 95% confidence we have to expect between 0.82 and 7.7 core melt accidents within the next 25 years.

Analysing all known accidents, we can show a learning effect. The probability of an incident or accident per reactor-year decreased from 0.01 in 1963 to 0.004 in 2010. Furthermore, there is an indication of a slightly larger learning effect prior to 1963.

It is well known that the actual number of all incidents and accidents is much higher than the numbers published in scientific journals. Therefore, we studied whether the known incidents and accidents are distributed randomly over the reactors using countries. While the data are random for most of the countries, this is not the case for the USA. From the present data, we cannot decide whether this is due to higher incident rates or to more effective sampling.

After this introduction the second section will explain some basics of the Poisson distribution. In Section 3 we present the data acquisition and its problems. Section 4 contains the discussion of core melt accidents and predictions for future events. The learning effect analysis is presented in Section 5.

While some of the results have been published already elsewhere [1], the underlying statistical work is presented here.

2. Poisson distribution

Rare and random incidents related to a time of reference, an area of reference or similar can be described by the Poisson distribution. Examples are the number of surface defects in body part stamping in the automotive industry or the number of calls in a call centre within a given time.

If the probability of an incident per time is known to be p, then within the time interval T, we expect a total number of $\lambda = pT$ incidents. But the actual number of incidents within T will

http://dx.doi.org/10.5772/intechopen.76637

fluctuate randomly. The Poisson distribution allows us to calculate the probability of a given number x of events within T:

$$\text{Probability of } x \text{ incidents if } \lambda \text{ incidents are expected} = \frac{e^{-\lambda}\lambda^{x}}{x!} \quad (1)$$

If the time of reference, T, is 1 year, then λ is the expected number of incidents within 1 year. If λ is much smaller than one, then it is also the probability of one incident within 1 year and of at least one incident per year. Analysing not only one but many reactors, the expected total number of accidents is simply the sum of the expected number for each single reactor, and, as long as the reactor incidents are independent of each other, the actual number of accidents is Poisson distributed.

In analysing real systems, the number of (statistically fluctuating) incidents x is known, and λ has to be determined. Then, the best estimate for λ is simply this empirical value x. However, this estimate is not necessarily the true value of λ because the incidents occur randomly. Poisson statistics allow us to compute an interval that contains the true value of λ with a confidence level α (typically 90, 95 or 99%), the so-called confidence interval. This is determined by calculating two values, λ_1 and λ_2, for a given number of incidents x. For the 95% confidence interval, we choose $\lambda_1 < x$ such that the probability of observing x or more events is 2.5% and $\lambda_2 > x$ such that the probability of observing x or fewer events is 2.5%. Then, the interval λ_1 to λ_2 is a 95% confidence interval. This means that if we study many cases, then in 95% of these cases, the true value of λ lies within this interval. The more cases we observe the narrower the confidence interval will be and the closer the estimate of λ will be to the true value.

As an example, suppose that the empirical number of events is $x = 4$. Then, the Poisson distribution with a value for λ equal 1.090 gives the probability that the number of events is greater than or equal to 4 to be 2.5%. If λ is 10.242, then the probability that the number of events is less than or equal to 4 is also 2.5%. Thus, for the empirical value of $x = 4$, we say that the true value for λ is between 1.090 and 10.242 with 95% confidence.

A similar measure of the probable distance between the estimated empirical value and the true value is the standard error. In large samples the probability that the distance between the estimated and the true value is less than the standard error is approximately 68%.

3. Data acquisition

3.1. How many reactors?

The International Atomic Energy Agency in Vienna publishes data on all power reactors worldwide [2]. The same and additional information about connection to the grid, shut down, operator, manufacturer and fuel supplier can be found in several Wikipedia entries [3, 4].

It was 1952 when the Soviet Union connected the first nuclear power reactor worldwide to the grid. Two years later the UK followed with Calder Hall. The number of reactors increased steadily until the mid-1980s. After that it grew only from 420 to about 450 in 2011. From this time the number of reactors remained nearly constant.

Table A1 in Appendix shows for all countries worldwide the total amount of nuclear energy produced, of reactor-years and accidents. The total energy in TWh is produced until 31 Dec. 2015. The amount of reactor-years has been calculated from the Wikipedia sources [3, 4] until 31.12.2011 to be comparable with the accident data.

The total operating time of all reactors until the end of 2011 was 14,766 reactor-years.

3.2. How many accidents?

First of all one has to define nuclear incidents or accidents. In 1990, the IAEA introduced the INES scale of incidents or accidents with seven levels [5]. The level 1 event is called an anomaly with, e.g. 'minor problems with safety components…', levels 2–4 are called incidents and levels 5–7 are called accidents. Two of the three destroyed reactors in Fukushima and the accident in Chernobyl were classified as level 7 with 'Major release of radioactive material with widespread health and environmental effects…'. The 1979 Three Mile Island accident in the USA was level 5 with 'Severe damage to the reactor core…' [6].

The USA uses a different scale to classify all, not only nuclear accidents. Major accidents are 'defined as incidents that either resulted in the loss of human life or more than US$50,000 of property damage, the amount the US federal government uses to define major energy accidents that must be reported' [7].

While the reactor data are publicly and easily available, this does not hold for the accident data.

According to the treaty of the International Atomic Energy Agency (IAEA), every member state has to inform the IAEA about events 'at Level 2 or above', but these data are publicly available only for 12 months. So, information about accidents in the past is not easy to get. We found two sources. One set of data has been published by the UK newspaper *The Guardian* [8], and another set published by Benjamin Sovacool in two papers [7, 9] and in his book *Contesting the Future of Nuclear Power* [10]. *The Guardian* list includes INES levels where known. Sovacool lists 'major accidents' according to the USA definition.

The Guardian lists 24 and Sovacool 99 events related to all kinds of nuclear technology. Both lists include the same core melt accidents: Windscale, UK, 1957, in a production plant for military use; Simi Valley, USA, 1959, in a research reactor; Monroe, USA, 1966, in a demonstration breeder reactor; Dumfries, UK, 1967, in a power reactor; Lucens, Switzerland, 1969, in an experimental reactor; Three Mile Island, USA, 1979, in a power reactor; Chernobyl, USSR, 1986, in a power reactor; and Fukushima, Japan, 2011, in three power reactors on the same site. The accidents in the three Fukushima reactors were caused by the same earthquake and the subsequent tsunami so we count them as one. This leaves four core melt accidents in power reactors.

In order to analyse the learning effect, we treated *The Guardian* and Sovacool data separately. From *The Guardian*'s list of 24 incidents, we included only the ones related to power production.

http://dx.doi.org/10.5772/intechopen.76637

This left 16 accidents of INES level 2 and higher. From Sovacool's list, we excluded five accidents not related to power generation.

3.3. Do the accident data represent a random sample?

These lists of publicly known events represent a sample of all incidents and accidents. Only random samples allow to draw conclusions to the underlying population. But are these samples really random? The data had been published by nuclear regulating authorities or collected by scientists, journalists and interested laypeople from a multitude of sources. Depending on the duties of the regulators or the public interest in nuclear energy or the emphasis of the press towards it, events might be detected more often in some countries than in others. So, we compared the number of (known) incidents in each country with its reactor-years.

If the incident probability is the same in all countries and if the probability to detect an accident is also independent of the country, then the number of accidents in a country should be proportional to the number of reactor-years in that country. Plotting the number of accidents versus the reactor-years should result in a straight line. A plot of these data is shown in **Figure 1**. The rightmost point shows the USA data.

So, for all countries except the USA, there seems to be a linear dependence between reactor-years and number of accidents. This is supported by a linear regression for all countries except the USA which gives a slope of 0.0036781 accidents per reactor-year with a standard error of 0.0004785. For each country but the USA, the expected value calculated from the 0.0036781 accidents per reactor-year is within the 95% confidence interval of the empirical accidents. Only for the USA, the empirical accident number of 54 in 3731 reactor-years is far away from the expected number of about 15.2.

While the data for all countries except the USA are compatible with a rate of 3.678 accidents per 1000 reactor-years, the USA data resemble 13.06 accidents per 1000 reactor-years.

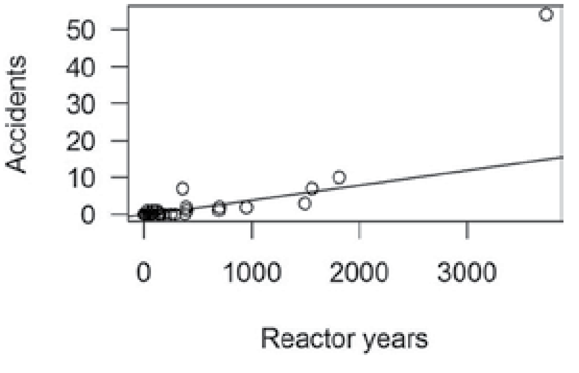

Figure 1. Total number of accidents in several countries versus total number of its reactor-years; the straight line is a linear fit through all data except the rightmost point (data from **Table 1**).

So, with the exception of the USA, there is no indication from the limited available data of non-random sampling or of countries having different overall accident rates. The USA data indicate that here either sampling is not random or the accident rate is higher than in the rest of the world. The present data do not allow us to determine which of these alternatives is the more likely explanation and further studies are needed.

4. Statistics of severe nuclear accidents

4.1. Results of previous PRA calculations

There have been several studies on reactor safety in the past. The first was the reactor safety study or Rasmussen report published in 1975 by the US Nuclear Regulatory Commission as report WASH-1400 or NUREG75/014 [11]. Five years later the German reactors were analysed in the Deutsche Risikostudie Kernkraftwerke [12]. In 1990 *Severe Accident Risks: An Assessment for Five U.S. Nuclear Power Plants* [13] was published. While the first two studies analysed typical reactors in their respective countries, the last one investigated five specified reactors.

What are the results of these studies? WASH-1400 states:

'The Reactor Safety Study carefully examined the various paths leading to core melt. Using methods developed in recent years for estimating the likelihood of such accidents, a probability of occurrence was determined for each core melt accident identified. These probabilities were combined to obtain the total probability of melting the core. The value obtained was about one in 20,000 per reactor per year. With 100 reactors operating, as is anticipated for the U.S. by about 1980, this means that the chance for one such accident is one in 200 per year' [11].

So, the probability for a core melt accident per reactor year is 5×10^{-5}.

The results of NUREG 1150 [14] can be found in Tables 3.2, 4.2, 5.2, 6.2 and 7.2 for the reactors Surry, Peach Bottom, Sequoyah, Grand Gulf and Zion, respectively. The German data are in the Deutsche Risikostudie [12]. The mean values vary between 4×10^{-6} and 3.40×10^{-4} accidents per reactor-year.

4.2. Empirical analysis

Based on the list and information of Sovacool, the following accidents are not included in the present study of severe accidents: Chalk River (1952) showed no core meltdown; Windscale (1957) was a military reactor only used for weapon production; Simi Valley (1959) was an experimental reactor; Monroe (1966) was an experimental reactor; and Lucens (1969) was an experimental reactor and probably showed no core meltdown. In Fukushima three of the six reactors at the site suffered severe destruction with INES ratings of 5–7. This threefold accident is counted as one because all three were triggered by the same cause, the tsunami with subsequent earthquake.

http://dx.doi.org/10.5772/intechopen.76637

There remain four core melt accidents in nuclear reactors for power generation.

Given the number of severe accidents, 4, and the cumulative reactor-years, 14,766, it is straightforward to calculate the probability p of a core melt accident at one reactor in 1 year:

$$p = \frac{4}{14766} = 2.70 \times 10^{-4} = \frac{1}{3700} \tag{2}$$

So, we expect one severe accident in 3700 reactor-years.

This simple calculation contains several uncertainties. Firstly, it is assumed that all reactors at all times have the same failure probability. Secondly, because of the small sample size of four events, it is subject to statistical fluctuations. This can be expressed through the confidence interval. Within a 95% confidence limit, the empirical value of four events leads to a confidence interval of 1.0899 and 10.2416 events in 14,766 reactor-years. Therefore, with a confidence of 95%, the failure rate is between one accident in 1442 and one accident in 13,548 reactor-years. Nevertheless, the most probable value is 1 in 3700 reactor-years.

Based on this value, it is possible to calculate the probability of accidents in the future. In a world with 443 reactors, we should expect 2.99 core melt accidents within the next 25 years with a 95% confidence interval of 0.82 accidents and 7.7 accidents. The USA with 104 reactors have to expect 0.7 core melt accidents within 25 years, with 95% confidence interval between 0.2 and 1.8 accidents.

5. Learning effects

5.1. Introduction

Experience and learning from operating power reactors and from analysing incidents and accidents are important for further reducing accident rates. Increasing operational experience should result in decreasing accident rates. This can be tested empirically by comparing accident rates with the amount of operational experience. In a simple approach, operational experience can be measured by the cumulative number of reactor-years up to a given date.

The small number of core melt accidents makes it difficult to detect any learning effect. Therefore, for this analysis we also included minor accidents and incidents. The two different datasets from *The Guardian* with 35 accidents and from Sovacool with 99 accidents were analysed independently. *The Guardian* data were grouped according to INES levels, and here all incidents of level 2 and above were included. One of the criteria for a level 2 incident is a 'significant contamination within a facility into an area not expected by design'. So, these incidents must be avoided by all means. From Sovacool's data all accidents related to nuclear power generation were included. Some of the basic results given below are summarised in [1], but the analysis here is more detailed.

5.2. Preliminary analysis

In order to analyse the rather low number of accidents, the total number of accidents, which is the cumulative number of accidents that had happened until a given year, was compared to the total number of reactor-years until that year, which is the cumulative reactor-years. Thus, the accident rate is

$$\text{accident rate} = \frac{\text{cumulative number of accidents}}{\text{cumulative reactor years}}. \tag{3}$$

Without any learning effect, the increase in accidents per reactor-year should be the same for every reactor-year; so, this accident rate should remain constant. A learning effect would decrease the accident rate.

We start by investigating *The Guardian* data. As discussed in Section 2, after excluding some accidents from the study, the final number of nuclear power-related incidents or accidents with level 2 and above is 16. The accident rate calculated from these data is plotted against the cumulative reactor-years in **Figure 2**. In order to present the data more clearly, the accident rate is displayed in a logarithmic scale. Every point represents the data of 1 year. The lines are 95% pointwise confidence intervals obtained from Poisson statistics.

A decreasing trend in this plot would indicate the presence of a learning effect. As can be readily seen, the first accident in 1957 resulted in a relatively high accident rate of about 0.05 per reactor-year. The following years saw no (publicly known) accident so the observed rate decreases drastically. Such a decreasing behaviour would be expected if an initial learning effect exists. However, after around 500 reactor-years, the plot appears to stabilise, with the accident rate varying around a constant value of about 1 in 1000 reactor-years. The plot does not indicate a learning effect. We investigate this further using a more detailed statistical analysis in Section 4.2.

Next, the Sovacool data is considered. As discussed in Section 2, after excluding some accidents from the study, the final number of nuclear power-related incidents or accidents with level 2 and above is 99. **Figure 3** is a plot of the log accident rate against cumulative reactor-years for these data, along with 95% pointwise confidence limits.

The slight decreasing trend in the latter portion of the graph along with the confidence limits suggests the possible presence of a small learning effect, with a larger effect apparent in the early years. We investigate this further using a more detailed statistical analysis in Section 5.3.

5.3. Formal statistical analysis

In order to investigate the possibility of a learning effect more formally, we constructed a suitable statistical model. The notation and assumptions below, summarised in the supplementary online material for [1], are common to the analyses of both *The Guardian* and the Sovacool data.

Let n_t be the number of reactors that are operational in year *t*, coded as $t = 1, \ldots, T$. For $r = 1, \ldots, n_t$, let Y_{tr} be the number of accidents at reactor *r* in year *t*. It is assumed that accidents at a given reactor in any given year occur independently. Then, accidents at that reactor over a 1-year period will occur according to a (possibly nonhomogeneous) Poisson process, so that Y_{tr} will be distributed as Poisson (λ_{tr}) where λ_{tr} is the expected number of accidents at reactor *r* in year *t* or approximately the probability of at least one accident at the reactor in year *t*. Assuming independence of the Y_{tr} over the reactors' operational time *t*, the total number of accidents $Y_t = \sum_{r=1}^{n_t} Y_{tr}$ in year *t* will be distributed as Poisson (λ_t), where $\lambda_t = \sum_{r=1}^{n_t} \lambda_{tr}$ is the expected total number of accidents in year *t*. If we further assume that the reactors have the same probability of failure in any given year, then $\lambda_{tr} = e_t$, where e_t is the expected number of accidents per reactor in year *t* and $\lambda_t = n_t e_t$. Any variation across reactors will lead to extra-Poisson variation, which can be assessed following model fitting.

Define $N_t = \sum_{u=1}^{t} n_u$ to be the cumulative number of reactor-years at year *t*. We use N_t as a measure of nuclear operational experience in year *t* and postulate that the expected number of accidents per reactor per year, e_t, is some function $e_t = e(N_t)$ of N_t. Without any loss of generality, we may write $e(N) = \alpha \exp\left(-\int_0^N \beta(x)dx\right)$, where $\beta(N)$ is the (instantaneous) rate of learning when the number of reactor-years has reached *N*.

Now, let $X_t = \sum_{u=1}^{t} Y_u$ be the cumulative number of accidents up to time *t*. Assuming independence of the $Y_t's$ over time, X_t will be distributed as Poisson (Λ_t), where $\Lambda_t = \sum_{u=1}^{t} \lambda_u = \sum_{u=1}^{t} n_u e(N_u)$. There is no learning if and only if the function β is identically zero, in which case $e(N_t) = \alpha$, the constant expected number of accidents per reactor per year, and $\Lambda_t = N_t \alpha$. It follows that the expected cumulative accident rate $E(X_t/N_t) = \alpha$. If, however, there is learning, then $\beta > 0$ and $e(N)$ will be a

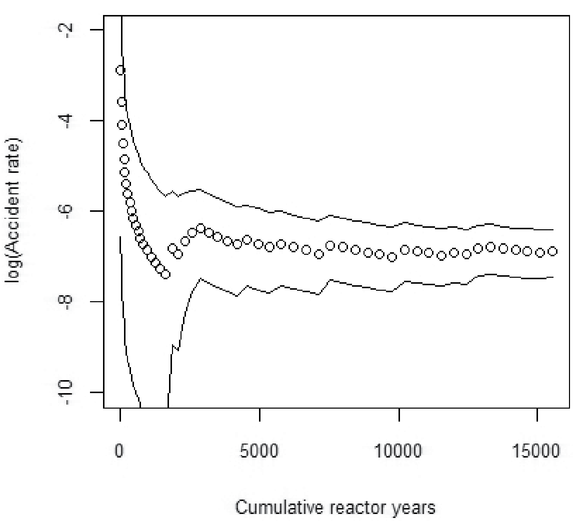

Figure 2. Accident rate = cumulative accidents/cumulative reactor-years on a log scale vs. cumulative reactor years, each data point representing 1 year. The lines are 95% pointwise confidence limits. *Source*: *The Guardian*.

decreasing function of N, so that a plot of X_t/N_t against N_t will exhibit a decreasing trend, as illustrated in **Figures 2** and **3**.

5.3.1. Analysis of The Guardian data

For *The Guardian* data, we took $\beta(N) = \beta$, so that there is either no learning or a constant rate of learning. In this case the expected number of accidents per reactor per year $e_t(N_t) = \alpha \exp(-\beta N_t)$, an exponentially decreasing function of the number of reactor-years. Since $\log\lambda_t = \log n_t + \log \alpha - \beta N_t$, the model is a generalised linear model (GLM) with Poisson family and log link function [15]. The analysis was implemented in the programming language R.

Figure 2 suggests the absence of any learning effect, but to investigate this formally, we set up and tested the null hypothesis $H_0{:}\beta = 0$. Based on the dataset from 1956 to 2011, a likelihood analysis produced a positive estimate of 1.58×10^{-5} for β, but with a standard error of 5.5×10^{-5}, this is far from being statistically significant (with a p-value of 0.78). If all the estimated values were the true values of the parameters, then the probability of a severe accident per reactor-year would reduce from 0.0012 to 0.0009 over the period. If, however, β is taken to be zero, then the estimated probability of a severe accident throughout the period is 0.0010.

Given the erratic behaviour in the early years, with just one accident in 1957 followed by a run of zero accidents over the next 19 years, it is important to investigate the sensitivity of the results to the early data. For the somewhat more informative data discussed in the next section with Sovacool's data, we will proceed more formally by elaborating the model to take into account the possibly different learning behaviour in the early years. In the case of *The*

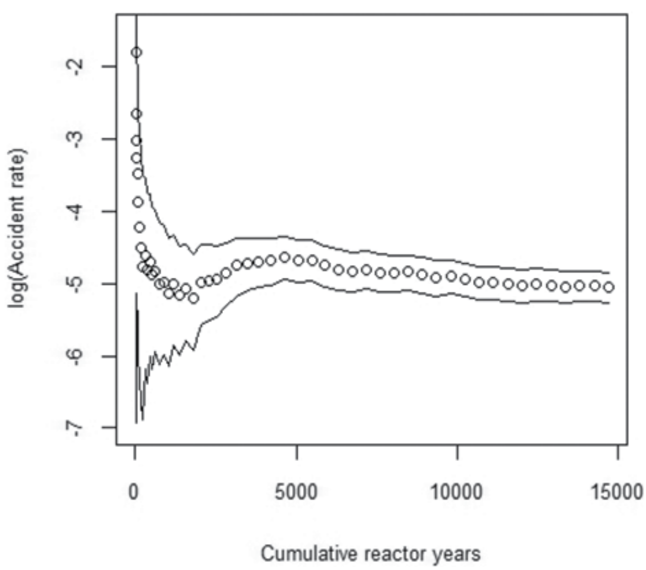

Figure 3. Accident rate = cumulative accidents/cumulative reactor-years on a log scale vs. cumulative reactor-years, each data point representing 1 year. The lines are 95% pointwise confidence limits. *Source*: Sovacool.

http://dx.doi.org/10.5772/intechopen.76637

Guardian data, the GLM results based on the years 1958–2011 produce a negative estimate of -8.61×10^{-5} for β, indicating an increasing accident rate. However, the associated standard error of 5.7×10^{-5} is large, and so again this value of β is far from statistically significant. If β is taken to be zero, then the estimated probability throughout this period is 0.0010, which is the same as the result based on the complete dataset.

Finally, consideration of only the more recent data from 1970 onwards produces a positive estimate of 7.29×10^{-6} for β, which would give rise to a very slight decrease in the accident rate from 0.0011 to 0.0010 over this period. However, again the result is not statistically significant, with a standard error of 6.0×10^{-5}. If β is taken to be zero, then the estimated probability throughout this period is again 0.0010. So, overall, there is no evidence from these data of any learning effect, at least beyond the initial few years of operation.

5.3.2. Analysis of the Sovacool data

The larger size of the Sovacool dataset allows us to elaborate the model to investigate the possibility of a learning effect more formally. To this end we choose a suitable formulation for the function $e(N)$. A change-point model could be used, but we preferred to use a smooth alternative that does no presuppose the existence of a sudden change in the accident rate. A commonly used functional form that models different rates of change at the early and late portions of a series is the biexponential function, given by

$$e(N) = \alpha_0 e^{-\beta_0 N} + \alpha e^{-\beta N}. \tag{4}$$

Here, β is the ultimate rate of learning relevant in the later years. The initial rate of learning β_I, relevant for the early years, can be obtained as a function of all the parameters in the model.

A convenient parameterisation of this function is $e(N) = \alpha e^{-\beta N}\{1 + e^{-\eta(N-\varphi)}\}$, where $\eta = \beta_0 - \beta$ and $\varphi = \{\log(\alpha_0/\alpha)\}/\eta$. With this parameterisation the instantaneous learning rate is

$$\beta(N) = \beta + \frac{\eta}{1 + e^{\eta(N-\varphi)}}. \tag{5}$$

In particular, the initial rate is $\beta_I = \beta + \eta/(1 + e^{-\eta\varphi})$. If the change from the initial to the final rate is quite pronounced, then it can be shown that this model will approximate to a change-point model, with the change-point at $N = \varphi$. We can now set up the likelihood function $L(\theta)$, where $\theta = (\gamma, \beta, \varphi, \eta)$ and $\gamma = \log \alpha$, and carry out a likelihood analysis [16]. Starting values for the computation may be obtained from graphical inspection and/or by fitting a generalised linear model to the data after 1962, using the Poisson family with a log link function.

The main hypothesis of interest is $H_0 : \beta = 0$, which corresponds to no learning in the later years. Another hypothesis of interest is that there is a constant rate of learning throughout the entire period, that is, $H_1 : \beta_I = \beta$. The maximum likelihood estimates and standard errors for various parameters, along with the p-values for the indicated null hypotheses, are exhibited in **Table 1**.

We see that there is some evidence of a learning effect over the latter portion of the data, formally verifying what seems to be indicated in **Figure 3**. Moreover, the rate of learning is fairly constant throughout the period from around 1962 to 2010, as can be seen from **Figures 4** and **5**. In these figures the observed accident rate Y_t/n_t per reactor in year t is plotted against year, in contrast to **Figures 3** and **4** in which the cumulative accident rate is plotted against cumulative reactor-years. The superimposed lines in **Figures 4** and **5** are the estimated theoretical annual accident rates $e(N_t)$ obtained from the biexponential Poisson model. **Figure 5** is the same as in **Figure 4**, except that omitting the data before 1964 allows for a higher resolution of the y axis.

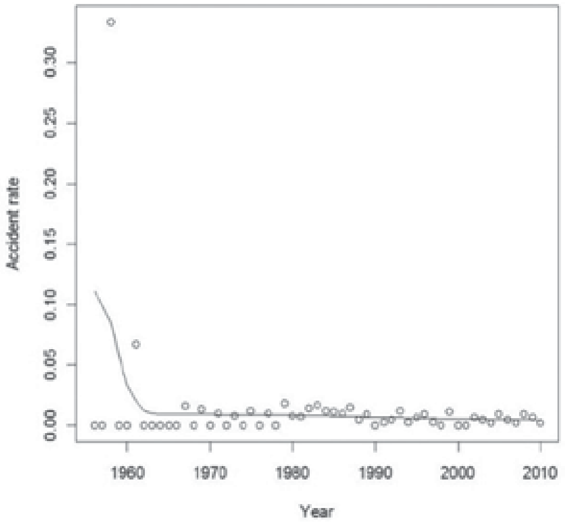

Figure 4. Observed and theoretical annual accident rate per year.

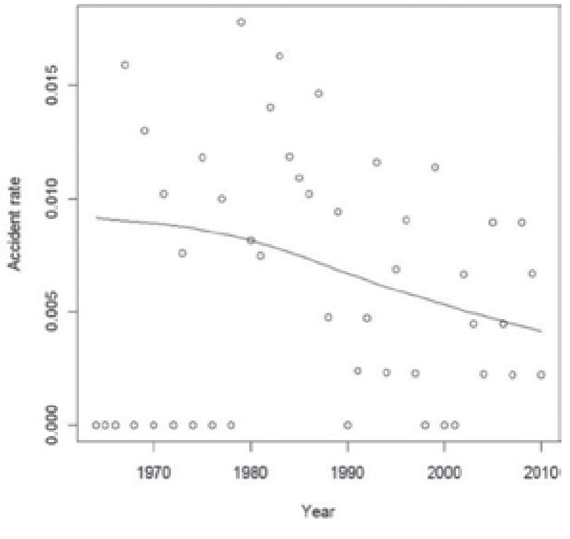

Figure 5. Same as in **Figure 4** for the years 1964–2010.

Although the data indicate a possible nonconstant learning effect over the period, with a larger effect at the beginning of the period up to about 1962, we see from **Table 1** that this is not statistically significant, owing to the highly variable nature of the early data when there were relatively few reactors and only two accidents. If the initial and final rates of learning do differ, then the best estimate of φ, the effective change-point in terms of the number of reactor years, is 43.10, which corresponds to the year 1961. This estimate is highly variable; however, a 90% confidence interval for φ, constructed from the profile likelihood of $\log \varphi$, gave values of φ between 3 and 221, which roughly correspond to the years 1957 and1966, respectively. These change-point results are unreliable, however, and more reliable estimates are obtained later in this section.

The high variability in the change-point contributes to the high degree of error in the estimate of β_I as seen in **Table 1**. However, whether or not there is a change in the rate of learning over the period, the estimated probability of an accident or incident at a reactor in 1 year falls from 0.010 in 1963 to 0.004 in 2010.

As a diagnostic for the model, one may calculate the standardised response residuals $r_t = \left(y_t - \hat{\lambda}_t\right)/\sqrt{\hat{\lambda}_t}$ from the observed values y_t of Y_t and the estimated model values $\hat{\lambda}_t$. When plotted against the year, these showed no unusual pattern. Moreover, the observed standard deviation of these residuals was 0.982, indicating that our initial assumption that λ_{tr} is constant over reactors was a reasonable one. Specifically, if we suppose that there is a positive but constant variation over reactors, so that $\mathrm{var}(\lambda_{tr}) = \sigma^2$, then the theoretical variance of the tth residual at the true parameter values will be $1 + e(N_t)\sigma^2$. Thus, the observed residuals would exhibit extra-Poisson variability, which does not appear to be the case here.

We further carried out a Bayesian analysis of these data. We used a noninformative prior of the form $\pi(\theta) \propto 1/\alpha$. A higher-order asymptotic approximation was computed, using the method in [17]. This was supplemented by the Monte Carlo method described in that paper. The results of the latter analysis, which may be considered to be exact having negligible simulation error, are given in **Table 2**. These are very similar to the asymptotic results.

We see that the Bayesian credible interval for $\beta \times 10^{-5}$ is consistent with the likelihood analysis, providing evidence of a learning effect over the latter portion of the data. The credible interval for $\beta_I - \beta$ provides some evidence of a difference between the initial and final rates of learning, although this difference may be very small. If the initial and final rates of learning do differ, then the Bayes estimate of the change-point φ is 39.37, which corresponds to the year 1961, as in the likelihood analysis. However, the exact Bayesian 90% credible interval

Parameter	Estimate	Standard error	Null hypothesis	p-Value
γ	−4.690	0.194	—	—
$\beta \times 10^{-5}$	5.362	2.476	$\beta = 0$	0.029
$\beta_I - \beta$	0.053	0.056	$\beta_I = \beta$	0.218
φ	43.10	33.42		

Table 1. Likelihood results from the biexponential Poisson model for the Sovacool data.

Parameter	Posterior mean	Posterior credible interval	Posterior probability
γ	−4.67	—	—
$\beta \times 10^{-5}$	5.67	0.990–10.582	95%
$\beta_I - \beta$	0.048	0.016–0.307	90%
φ	39.37	10.18–90.02	90%

Table 2. Bayes results from the biexponential Poisson model for the Sovacool data.

is tighter than the approximate confidence interval produced earlier and corresponds to the years 1959–1963.

Whether or not there is a change in the rate of learning over the period, the estimated probabilities of an accident or incident at a reactor in 1963 and 2010 are identical to those obtained earlier from the likelihood analysis.

6. Summary

Previous Probabilistic Risk Assessments estimated the probability of a core melt accident to be in the range of one in several 10,000 to one in several 100,000 reactor-years. The real core melt accidents in the past happened with a probability of one in 3,700 years. Much more frequent than anticipated before. Thus, a world with 443 reactors has to expect 2.99 core melt accidents within the next 25 years, a country like the USA with 103 reactors 0.7 core melt accidents.

The Guardian data showed that incidents and accidents happen with a probability of approximately $0.001 = 1 \times 10^{-3}$ per reactor-year. The data are consistent with no learning effect on the side of the plant operators. The second investigation based on Sovacool's data shows a decrease of the accident rate from $0.010 = 10 \times 10^{-3}$ per reactor-year in 1963 to $0.004 = 4 \times 10^{-3}$ in 2010. There is also some indication of a stronger learning effect until the beginning of the 1960s, although this is not statistically significant. Between 1963 and 2010, the operating experience increased from 96 to 14,704 reactor-years. So, while operating experience increased by a factor of over 150, the probability of a minor or severe accident at a reactor decreased by merely a factor of 2.5.

It might be interesting to compare the last results with the empirical core melt probability of $1/3700 = 0.27 \times 10^{-3}$. Depending on the dataset, a core melt accident is only 37 times (*The Guardian* data) or 15 times rarer than other accidents or incidents. Regarding the possible outcomes of a core melt accidents, these differences seem to be unexpectedly low and might indicate that the datasets used do not contain all incidents and accidents that happened in the past.

This guess finds support in an article by Phillip A. Greenberg. 'Between 1990 and 1992 the US Nuclear Regulatory Commission received more than 6600 "Licensee Event Reports" because US nuclear plants failed to operate as designed and 107 reports because of significant events (including safety system malfunctions and unplanned and immediate reactor shutdowns)' [18].

Our work shows the possibility of studying learning effects within the nuclear industry. But more detailed results require more analysis and more information from reactor operators and regulators. But this is difficult on an international scale because of the restrictive information policy of the IAEA.

A. Appendix

Country	Total energy produced	Reactor years	Accidents
Argentina	225.63	66.5150685	1
Armenia	90.14	44.1945205	0
Belgium	1422.8	247.421918	0
Brazil	228.57	41.2219178	0
Bulgaria	518.07	151.057534	0
Canada	2907.4	692.589041	1
China	1155.36	125.219178	0
Czechoslovakia	540.62	122.561644	1
East Germany	143.21	79.9534247	1
Finland	719.34	131.328767	0
France	12137.14	1813.29863	10
Germany	4741.16	701.482192	2
Hungary	404.27	106.027397	1
India	459.94	356.736986	7
Iran	12.46	0.3260274	0
Italy	89.78	80.569863	0
Japan	7310.61	1558.76986	7
Lithuania	241.58	43.3972603	0
Mexico	211.94	39.8794521	0
Netherlands	152.9	66.9726027	0
Pakistan	53.56	52.5863014	0
Romania	143.54	19.8849315	0
Russia	4286.29	947.312329	2
Slovakia	433.25	140.19726	0
Slovenia	160.83	30.2657534	1
South Africa	351.43	54.2109589	0

Country	Total energy produced	Reactor years	Accidents
South Korea	3007.33	380.419178	0
Spain	1763.3	284.750685	0
Sweden	2264.97	392.271233	2
Switzerland	870.81	184.484932	0
Taiwan	1178.47	186.80274	0
UK	2698.98	1494.44658	3
Ukraine	2319.26	397.706849	1
USA	24222.89	3731.37808	54

Table A1. Total nuclear energy produced in TWh [3] (until 31 Dec. 2015) and reactor-years [3] (until 31 Dec. 2011) and number of accidents [7] (until 31 Dec. 2011); countries which started to build reactors which were not operating are excluded.

Acknowledgements

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflict of interest

No potential conflict of interest was reported by the authors.

Thanks

Thomas Rose wants to thank Fachhochschule Münster for giving him a sabbatical when this work started and the Department of Science and Technology Studies, University College London for accepting him as an Honorary Senior Research Associate to pursue this work further.

Author details

Thomas Rose[1,2]* and Trevor Sweeting[3]

*Address all correspondence to: thomas.rose@ucl.ac.uk

1 Sensor Technology, Muenster University of Applied Sciences, Steinfurt, Germany

2 Department of Science and Technology Studies, University College London, London, UK

3 Department of Statistical Science, University College London, London, UK

References

[1] Rose T, Sweeting T. How safe is nuclear power? A statistical study suggests less than expected. Bulletin of the Atomic Scientists. 2016;**72**(2):112-115. DOI: 10.1080/00963402.2016.1145910

[2] International Atomic Energy Agency. Nuclear Power Reactors in the World. Vienna: International Atomic Energy Agency; 2016. Reference Data Series 2 [cited 2018 Feb 1]. Available from: http://www-pub.iaea.org/MTCD/Publications/PDF/RDS_2-36_web.pdf

[3] Wikipedia. Liste der Kernkraftwerke [cited 2018 Feb 1]. Available from: https://de.wikipedia.org/wiki/Liste_der_Kernkraftwerke

[4] Wikipedia. Список АЭС мира [cited 2018 Feb 1]. Available from: https://ru.wikipedia.org/wiki/%D0%A1%D0%BF%D0%B8%D1%81%D0%BE%D0%BA_%D0%90%D0%AD%D0%A1_%D0%BC%D0%B8%D1%80%D0%B0

[5] International Atomic Energy Agency. INES: The International Nuclear and Radiological Event Scale. Vienna: International Atomic Energy Agency; 2008. Available from: www.iaea.org/Publications/Factsheets/English/ines.pdf

[6] International Atomic Energy Agency. INES—The International Nuclear and Radiological Event Scale User's Manual: Edition: International Atomic Energy Agency; 2008 [cited 2011 Nov 8]. Available from: http://www-pub.iaea.org/MTCD/publications/PDF/INES-2009_web.pdf

[7] Sovacool BK. A critical evaluation of nuclear power and renewable electricity in Asia. Journal of Contemporary Asia. 2010;**40**(3):369-400

[8] Nuclear accidents; 2011 [cited 2011 Apr 16]. Available from: https://spreadsheets.google.com/ccc?key=0AonYZs4MzlZbdFc0cVRMVDR5c1ZmeC1lR2hac0xjMXc&hl=en#gid=0

[9] Sovacool BK. The costs of failure: A preliminary assessment of major energy accidents, 1907-2007. Energy Policy. 2008;**36**(5):1802-1820

[10] Sovacool BK. Contesting the Future of Nuclear Power: A Critical Global Assessment of Atomic Energy. Singapore u. a.: World Scientific; 2011. Available from: http://www.gbv.de/dms/zbw/656623985.pdf

[11] Reactor Safety Study (WASH 1400): Executive Summary. Oct 1975. Available from: http://www.osti.gov/energycitations/product.biblio.jsp?query_id=6&page=0&osti_id=7134131

[12] Hörtner H. Deutsche Risikostudie Kernkraftwerke Band. 1 Ereignisablaufanalyse. TÜV Rheinland: Köln; 1980

[13] NUREG-1150: Severe Accident Risks: An Assessment for Five U.S. Nuclear Power Plants, Final Summary Report; Vol. 1. Washington; December 1990 [cited 2011 Nov 10]. Available from: http://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/

[14] NUREG-1150: Severe Accident Risks: An Assessment for Five U.S. Nuclear Power Plants, Summary of Plant Results; Vol. 2. Washington; December 1990 [cited 2011 Nov 10]. Available from: URL: http://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/

[15] McCullagh P, Nelder JA. Generalized Linear Models. 2nd ed. London: Chapman and Hall. p. 1999. 532 p

[16] Garthwaite PH, Jolliffe I, Jones B. Statistical Inference. 2nd ed. Oxford: Oxford University Press; 2006. 342 p

[17] Kharroubi SA, Sweeting TJ. Biometrika. 2016;**103**:337-349. DOI: 10.1093/biomet/asw018

[18] Greenberg PA. Safety, accidents, and public acceptance. In: Byrne J, editor. Governing the Atom: The Politics of Risk. New Brunswick, NJ: Transaction Publ; 1996. pp. 127-175

Edited by Türkmen Göksel

Since data grows faster than ever, the role of statistics becomes more and more crucial nowadays, and there is no doubt that statistics will be even more critical in the future.

The application of statistics is extensive, and in our daily lives there is almost no human activity where the use of statistics is not needed. In this limited volume, we try to cover as many as different and multidisciplinary fields in statistics as possible and aim to present recent developments and applications of statistical analysis.

Therefore, this book is organized into three sections: "The Role of Statistics on Quantification," "Applications of Statistics on Economics and Development," and "Applications of Statistics on Various Topics."

Published in London, UK

IntechOpen

ISBN 978-1-83881-354-3