GUS

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

Przegląd Statystyczny
Statistical Review

Vol. 68 | No. 1 | 2021

GŁÓWNY URZĄD STATYSTYCZNY
STATISTICS POLAND

INFORMATION FOR AUTHORS

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

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

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

All work should be submitted to the journal through the ICI Publishers Panel (https://editors.publisherspanel.com/pl.ici.ppanel-app-war/ppanel/index).

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

Przegląd Statystyczny
Statistical Review

Vol. 68 No. 1 2021

EDITORIAL OFFICE'S ADDRESS

University of Gdańsk, ul. Armii Krajowej 101, 81-824 Sopot, Poland

Language editing: Scientific Journals Division, Statistics Poland
Technical editing and typesetting: Statistical Publishing Establishment – team supervised by Wojciech Szuchta

Zakład Wydawnictw Statystycznych

Printed and bound: Statistical Publishing Establishment
al. Niepodległości 208, 00-925 Warsaw, Poland, zws.stat.gov.pl

Website: ps.stat.gov.pl

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

Information on the sales of the journal: Statistical Publishing Establishment
phone no. +48 22 608 32 10, +48 22 608 38 10

Order no. 186/2021 – 220 printed copies

LETTER FROM THE EDITOR

Dear Readers,

In 2020, the Editors and Statistics Poland, the Publisher of *Przegląd Statystyczny. Statistical Review* agreed upon publishing articles exclusively in English. Now, I am pleased to present to you the fifth, fully English-language issue of the journal. I would like to thank all the authors for submitting their papers to *Przegląd Statystyczny. Statistical Review*, and the reviewers for their insightful and comprehensive assessments. It is thanks to your joint effort that fifteen research articles, two anniversary articles and one report from a conference have been published in English so far.

The Editorial Team has been continuously undertaking proper steps to increase the recognition of the journal within the scientific community worldwide. In 2020, we expanded the journal's abstracting and indexing by two databases: ERIH Plus and EBSCO Discovery Service.

We would like to invite members of academia and practitioners to submit papers to *Przegląd Statystyczny. Statistical Review* containing original research on theoretical and empirical subjects relating to statistics, econometrics, mathematical economics, operational research, decision sciences, and data analysis. We would also welcome high-quality papers from all fields of economics, finance and management authored by PhD candidates.

The entire editorial process, from the paper's submission to its publishing, is free of charge. The final decision as to whether an article is accepted for publication is issued within three months.

On behalf of the Board of Editors,
Paweł Miłobędzki
Editor-in-Chief

CONTENTS

Józef Pociecha
The paradigm of statistical inference and the paradigm of statistical learning **1**

Adam Korczyński
Predicting in multivariate incomplete time series. Application of the expectation-maximisation algorithm supplemented by the Newton-Raphson method **17**

Nazarii Kukhar
The influence of the demographic structure on the economic growth of Ukraine **47**

Przegląd Statystyczny. Statistical Review, 2021, vol. 68, 1, 1–16 DOI: 10.5604/01.3001.0015.0375
Received: 18.05.2021, revised: 30.06.2021, accepted: 7.07.2021

The paradigm of statistical inference and the paradigm of statistical learning

Józef Pociecha[a]

Abstract. The starting point for the presentation of the similarities and differences between the principles of conducting statistical research according to the rules of both statistical inference and statistical learning is the paradigm theory, formulated by Thomas Kuhn. In the first section of this paper, the essential features of the statistical inference paradigm are characterised, with particular attention devoted to its limitations in contemporary statistical research. Subsequently, the article presents the challenges faced by this research jointly with the expanding opportunities for their effective reduction. The essence of learning from data is discussed and the principles of statistical learning are defined. Moreover, significant features of the statistical learning paradigm are formulated in the context of the differences between the statistical inference paradigm and the statistical learning paradigm. It is emphasised that the statistical learning paradigm, as the more universal one of the two discussed, broadens the possibilities of conducting statistical research, especially in socio-economic sciences.

Keywords: paradigm theory, learning from data, statistical inference, statistical learning
JEL: C000, C180, C83

1. The statistical inference paradigm

Thomas S. Kuhn, an American physicist, historian and philosopher of science, is the founder of the concept of the scientific paradigm. In his basic work on the philosophy of science entitled *The Structure of Scientific Revolutions* (Kuhn, 1962),[1] he introduced into the philosophy of science the idea of a paradigm as a set of concepts and theories that form the basis of a given science. These theories and concepts are not questioned, at least as long as the paradigm is cognitively creative, i.e. it can be used to create specific theories consistent with the experimental or historical data which the science concerns. The most general paradigm is that of the scientific method, which formulates the criteria for recognising an activity as scientific. The paradigm guides the research effort of scientific communities and is the basic criterion for identifying areas of individual sciences. Kuhn's fundamental claim is that a transition from the old to the new paradigm takes place in the process of scientific revolutions. When a paradigm shift occurs, the scientific world changes qualitatively by enriching it with new facts and theories. Thus, according to Kuhn, the development of scientific theories continues.

In science, and especially in social sciences, different paradigms can occur simultaneously, which can even lead to scientific paradigm wars, involving scientists

[a] Cracow University of Economics, Department of Statistics, ul. Rakowicka 27, 31-510 Kraków, Poland, e-mail: pociecha@uek.krakow.pl, ORCID: https://orcid.org/0000-0003-3140-481X.

[1] Available in Polish, Kuhn (2020).

from different camps to contest each other and deny others their scientific character. Examples of different paradigms in economic sciences are those of classical and Keynesian economics. Parallel paradigms exist in statistics, including descriptive statistics, mathematical statistics, Bayesian statistics or the statistical learning paradigm (Pociecha, 2020).

The paradigm of mathematical statistics (statistical inference) is based on the notions of a (general) population, i.e. a statistical population which we do not study, and a sample, i.e. a subset of the population we are investigating. The paradigm assumes that – on the basis of the sample – we can infer about the population, as long as the sample is representative of the population. A necessary condition for the representativeness of a sample is its random selection. Sample representativeness is a gradual concept; the degree of sample representativeness depends on the sampling method and, for a given sampling scheme, the size of the sample. Quite complicated, layered sampling schemes are generally used in statistical survey practice, especially when the population is very numerous and clearly structured. The sampling theory deals with sampling procedures (see e.g. Bracha, 1996; Steczkowski, 1995; Tillé, 2006).

As part of mathematical statistics, two theories constituting its methodological basis were formulated – the statistical estimation theory and the statistical hypotheses testing theory. The theory of estimation lays down the rules for estimating the parameters of the distribution in the population, based on the obtained sample, with a fixed type of distribution. The theory of statistical hypotheses testing formulates rules for verifying the truthfulness of judgements regarding the parameters of the distribution in a certain population by comparing it with a determined type of distribution, or judgements relating to the compliance of the empirical distribution with a selected theoretical distribution. Moreover, the theory involves examining hypotheses regarding the randomness of a sample or other related judgements about one or more populations.

Statistical inference is of a probabilistic nature and uses two concepts in particular: the concept of a confidence level, defined as a confidence coefficient or interval, understood as a probabilistic measure of an estimation error in parameter estimation, and the concept of a significance level, defined as a predetermined probability (risk) of committing a first type error.

The paradigm of mathematical statistics is based on the stochastic nature of statistical regularities resulting from the indeterministic understanding of the connections between the components of the world around us, on the principles of inductive inference in the incomplete version and the frequency definition of probability. The assumptions of the Aristotelian realist philosophy, assuming that the world (reality) exists objectively (outside our mind) and is knowable constitute the fundamental philosophical basis of the paradigm of statistical inference as a tool for studying mass processes. This philosophy shows that learning about the reality that surrounds us, although difficult, is possible and the scientific effort it entails is deliberate. This justifies the possibility and purposefulness of conducting scientific research also with the use of statistical methods. An important premise of the statistical inference paradigm is Karl Popper's critical rationalism, which forms the philosophical basis for testing scientific hypotheses. The immediate philosophical foundation of this paradigm is probabilism, which originated from ancient sceptics and was developed by neo-positivists. Popper adopts a sceptical understanding of the truth, which we can approach only at a distance acceptable to us (with an error that we accept), and which allows making statistical inferences (Pociecha, 2020).

Conducting statistical research in accordance with the presented paradigm of mathematical statistics is, however, subject to certain limitations and its correct performance – in both theoretical and practical terms – faces a number of difficulties. In particular, it is challenging for socio-economic sciences to put into practice the theoretical requirements for sampling (Cassel et al., 1977). This relates to defining the substantive, spatial and temporal scope of the population, determining the sampling frame, sampling scheme, sampling procedure (quota, group, systematic, stratified sampling), the multi-stage sampling scheme or determining the sample size. Each decision in any of the mentioned areas affects the obtained sample representativeness. However, in the practice of statistical analyses, often no attention is devoted as to how the data set, which we consider a random sample, was obtained and no tests verifying the randomness of a sample are conducted. In effect, the correctness of the obtained results of statistical investigations is often questioned, and the validity of using statistical methods in socio-economic research becomes uncertain.

The statistical estimation theory and the theory of parametric statistical hypotheses verification requires assuming a specific analytical form of distribution in the population. The vast majority of estimation procedures and the verification of parametric hypotheses require the assumption of normal distribution in the population. While in physico-biological studies the assumption of the normality of distribution in the population is in most cases satisfied, in socio-economic studies it is usually not. However, procedures aiming to test the normality of a population distribution are rarely used. There are, of course, also procedures assuming a different than normal analytical form of distribution in the population, but statistical procedures based on such distributions have not been developed and are theoretically complex.

It should also be noted that parametric hypothesis estimation and testing procedures are limited to solving problems which can be effectively parameterised, but there are numerous empirical research problems that cannot. The existing non-parametric tests alleviate the problem of analysing non-parameterised issues only to a certain extent.

Another limitation of the mathematical statistics paradigm is the adoption of an axiomatic in theory and frequentative in practice definition of probability. The Bayesian statistic paradigm extends the understanding of probability to an *a priori probability* and a *probability a posteriori,* which broadens the understanding of this key concept to include its subjective aspect and allows for a clear connection between statistical theory and empirical research. The limitation of classical statistical inference is that the testing of statistical hypotheses is based on minimising the risk of committing the first type error, i.e. rejecting the null hypothesis when it is actually true, which occurs in significance tests. In a large number of cases, it is more important to minimise the risk of making a second type error, i.e. accepting the null hypothesis when it is false. These situations arise, for example, in the process of testing the correctness of financial statements when their audit is performed (Hołda & Pociecha, 2009).

The limitations resulting from the failure of empirical data to meet the theoretical assumptions underlying the methods of the estimation and verification of statistical hypotheses are also highlighted by Wiesław Szymczak in his book on the practice of statistical inference (Szymczak, 2018). In his work, the author critically assesses the role of the paradigm in statistics. Summing up, it should be emphasised that the commonly functioning paradigm of statistical inference does not provide a universal basis for empirical statistical research. It is subject to significant limitations and creates a number of difficulties for the correct implementation of empirical research according to this paradigm.

2. Challenges facing modern statistical research

The rapid development of information technology (IT), encompassing more and more efficient computer hardware and involving an increasingly higher quality and reliability computer software, enables processing great amounts of information, offering new analytical possibilities for contemporary statistical research, unlike ever before. Modern computers have an unimaginable computing power. Currently, the most powerful computer in the world, designated as 1/10 1 – Summit – IBM Power System AC922 has 2,801,644 GB of memory and 2,414,592 cores. Its computing power is at the level of 148,600 teraflops per second, and may even exceed the value of 200,795 teraflops (Onet, n.d.). Thus, it can be concluded that the current computational possibilities for statistical analyses face no technical barriers.

The increased ability to collect, process and store data is now leading to the creation of extremely large data sets for which the term Big Data has been adopted.

Big Data is defined as a dense, continuous and unstructured data stream resulting from interpersonal interactions, interactions between devices being part of the infrastructure of the global computer network, and all other instruments through which this data stream is registered and transmitted (Migdał-Najman & Najman, 2017). The most important source of Big Data, however, are the interactions resulting from humans' connections with IT devices, giving people access to numerous services such as transaction systems, online stores, financial services, mobile services, systems monitoring health, emotions, location, and physical activity. Big Data sets are characterised by the amount of data they contain (*volume*), data processing speed (*velocity*) and data diversity (*variety*). The above-mentioned features include the degree of their reliability (*veracity*), their value for the user (*value*) and the possibility of their visualisation (*visualisation*) (Tabakow et al., 2014).

Even if a Big Data set displays the above-mentioned features, it does not necessarily mean that it is directly useful for conducting statistical analyses. Big Data, in addition to providing useful, up-to-date, accessible, comparable, consistent and accurate data, called *clear data*, consists of inaccurate, repeated, incomplete, wrongly named or non-integrated data, referred to as *dirty data*, as well as *dark data*, whose author, place and time of creation, content and connection with other data remains unidentified (Migdał-Najman & Najman, 2017).

Thus, Big Data contains not only the clear data desired by the analyst, but also dirty data, whose removal often involves complicated cleaning procedures (Kim et al., 2003), and dark data, towards which the analyst should make a decision whether to eliminate them from the given data set. It is difficult to clearly state in what proportions the above-mentioned Big Data components occur, but IT specialists and analysts claim that clear data is a substantial minority within Big Data, which is also reflected in the obvious disproportion between the amount of data collected and the amount of relevant data providing valuable information. IT specialists say that dark data can account for up to 90% of the entire volume of Big Data – it is then this percentage of Big Data that is not fit for analytical purposes. Thus, the technical and IT-related capacity for collecting and storing data are much higher than the ability to analyse and draw conclusions from these data; moreover, this disproportion is growing rapidly (Migdał-Najman & Najman, 2017).

The changes in the acquisition, storage, processing and analysis of data presented above pose new challenges for modern statistical research. It is not the issue of limited data availability or restricted computational possibilities that constitute a barrier to the development and application of statistical methods. On the contrary, today's excess of data and the enormous computing power of computers pose a challenge for the rational application of statistical methods in socio-economic analyses. In consequence, the classical paradigm of mathematical statistics is often insufficiently effective for modern statistical research. This made it necessary to search for a new paradigm of empirical research using statistical tools.

3. Learning from data

The purpose of statistical analysis is to extract information from data, while data analysis involves the processing of data in order to obtain useful information and draw conclusions. Sometimes the term 'data analysis' is understood as a field of knowledge covering the issues of acquiring, storing and processing data, building data warehouses, databases and algorithms; the term may also relate to the knowledge of IT tools such as Excel, Python, R, SQL environment, etc. Data analysis understood in this way is described in many works, e.g. in Alexander & Kusleika (2019). It should be emphasised, however, that the scope of data analysis is significantly broader and involves not only obtaining or processing data with the use of appropriate IT tools, but, above all, inferring about the socio-economic reality which these data come from.

In data analysis, the saying 'let the data speak for itself' means learning from data. The idea is to acquire knowledge not only useful for performing current activities, but also to improve future performance. Learning from data is an inductive inference made on the basis of available observations. Learning outcomes are influenced by three main factors: the components of learning from data, the type of feedback on the basis of which the learning process takes place, and the method of presenting the acquired information (Russel & Norvig, 2003).

In the era of the rapid development of information technology, computers are designed to learn from data. In result, machine learning was created, i.e. self-learning systems based on algorithms which automatically improve through experience (Cichosz, 2000). Machine learning should therefore be understood as the ability of computers to automatically learn from data and transfer this knowledge to the recipient.

Machine learning from data can be realised in three forms: supervised learning, unsupervised learning and reinforcement learning. Supervised learning consists in approximating unknown function f by mapping the input data with the output data by providing individual input data (x_i) and knowing the output data (y_i). The essence of supervised learning is to provide the algorithm with a set of input-output pairs, i.e. (x_i, y_i) pairs in order to approximate unknown function f. According to the supervised learning theory, input-output is entered to find function f that maps the values of x to the value of y. The set of all possible functions that can describe this mapping is called hypothetical space H. Next, function h is selected. This function belongs to hypothetical space H, which, in the author's opinion, approximates unknown function f well and provides the possibility of making rational future decisions. Function h is a hypothesis of the actual course of function f (Russel & Norvig, 2003). In statistical literature, input data are called predictors, or classically – independent (explanatory) variables. In the machine learning terminology output data are called response variables, and classically – dependent variables (Hastie et al., 2009).

Unsupervised learning is a type of machine learning which assumes that there is no exact or even approximate output in the training data. So unsupervised learning involves learning patterns on the basis of the given data when only the input data are known. The aim of unsupervised learning is to either identify the interdependencies between features or to discover the internal structure of a data set. Examples of unsupervised learning include cluster analysis and correspondence analysis. Unsupervised learning methods are taxonomic methods used to classify objects in a multidimensional space of features according to the adopted measure of their similarity or distance (Pociecha et al., 1988).

Reinforcement learning does not use input or output data. It consists in observing the environment by the learning system and selecting activities in order to maximise the rewards and avoid the punishments. The learning system learns on its own the best strategy, called politics, to collectively obtain the highest reward (Géron, 2018).

Function f, connecting the input data with the output data in supervised learning, can be deterministic or indeterministic; consequently, learning can also be deterministic or indeterministic. In the study of socio-economic phenomena, usually indeterministic relationships are observed, therefore supervised learning should be understood as indeterministic learning. Unknown function f is approximated by hypothetical function h. In theory, the more complex the function, the better chance of an exact approximation of function f. The indeterministic learning process involves an inevitable compromise between the complexity of hypothetical function h and the degree of dispersion of the input data. The learning problem is feasible if hypothesis space H contains the actual function f. Unfortunately, it is not always possible to assess whether a given learning problem is feasible because the true function is unknown. One way to bypass this barrier is to use the previously gained knowledge to derive hypothesis h from space H, when it is certain that the actual function f is contained in this space (Russel & Norvig, 2003).

If the actual function f is of a stochastic nature, supervised learning is understood as statistical learning. Bearing in mind that in the vast majority of cases function f, which assigns input data to output data, is defined in a stochastic manner, machine learning is in fact almost entirely statistical learning. However, due to the fact that it was introduced into the literature by computer scientists, the term 'statistical learning' has been dominated by the term 'machine learning'. It is only thanks to the fundamental works of Hastie et al. (2009) and James et al. (2013) that statistical learning begins to occupy its rightful place in the world literature.

The formal definition of statistical learning is presented in numerous studies, including that of Vapnik (2000). According to a popular operational definition, statistical learning is a collection of descriptive statistics, mathematical statistics, and non-parametric and non-algorithmisable procedures for modelling and understanding complex data sets. Statistical learning is a new field of knowledge that has been developing since the turn of the 20th and 21st century, being the product of the development of statistics and computer science. It combines the principles of machine learning with statistical methods (James et al., 2013).

4. Principles of statistical learning

In the general approach to statistical learning, we have dependent variable Y, understood as the *response variable* and k explanatory variables (predictors) $X_1, X_2, \dots, X_k$. We assume that there is a relationship between Y and $X = (X_1, X_2, \dots, X_k)$ which we can generally define as

$$Y = f(X) + \xi, \tag{1}$$

where
f – an unknown function associating Y with X;
ξ – a random component.

The essence of statistical learning is to guess function f using function h, which is one of the hypotheses belonging to hypothesis space H, concerning unknown function f (Hastie et al., 2009).

There are two primary reasons for attempting to guess function f. The first one, of a practical nature, is the prediction of Y based on the knowledge of X. The other reason, of a more cognitive nature, is the inference about Y on the basis of X. The prediction task is when a set of X predictors is available, but the corresponding values of response variable Y are unknown. In this case, we predict Y using the following equation:

$$\hat{Y} = \hat{f}(\mathrm{X}), \tag{2}$$

where $\hat{f}$ is one of the functions falling in hypothetical space H. In this approach, $\hat{f}$ is treated as a black box, in the sense that it is not usually a specific analytical form of $\hat{f}$, on condition that it provides accurate possible forecasts of Y.

If we want to use statistical learning methods for the purpose of inferring about the relationship between the dependent variable and the explanatory variables, then we cannot treat $\hat{f}$ as a black box, but we have to take the specific form of function $\widehat{f}$. Subsequently, we attempt to answer the following questions:

- Which pre-adopted explanatory variables actually affect the response variable?
- What is the direction of the relationship between the response variable and individual explanatory variables?
- What is the appropriate analytical form for $\hat{f}$?
- Is the linear form sufficient?

Statistical learning methods are designed to answer these types of questions (James et al., 2013).

The approximation of the actual f function is the key statistical learning problem. Its estimation is based on a data set, called the training data or training set, containing input and output information (x_{ij}, y_i). In other words, we seek such a function $\hat{f}$, for which

$$Y \approx \hat{f}(X) \tag{3}$$

for any pair of observations from set (X, Y). A parametric or non-parametric approach can be applied here.

The parametric method of statistical learning involves specifying the analytical form of function $\hat{f}$. In the simplest and most common case, we assume it as a linear multivariate model. Then, using the data from the training set, the partial regression coefficients of this model are estimated, most often by means of the least squares method. Of course, in the case of parametric statistical learning there are many options for both the selection of the vector of explanatory variables and the analytical form of the regression function.

Non-parametric statistical learning methods do not make explicit assumptions about the analytical form of the functions for f. Instead, they look for a form of function f which fits as closely as possible to the data from the training set. The non-parametric approach can have a great advantage over the parametric approach, because by avoiding the assumption of a specific analytical form of function f, it can fit the empirical data more accurately. The parametric approach involves the risk that the analytical form of function $\hat{f}$ deviates greatly from the actual function f, which links the predictors with the result variable. Nevertheless, the non-parametric approach has the disadvantage that it does not reduce the number of the estimated parameters to only the significant ones, and thus requires a much larger training set (James et al., 2013).

When selecting the best function $\hat{f}$ belonging to hypothetical space H, one should follow a specific quality criterion of fitting this function to the actual function f. The effectiveness of the statistical learning method with the specified $\hat{f}$ is measured by the mean square error of estimation (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{f}(x_i))^2, \quad (4)$$

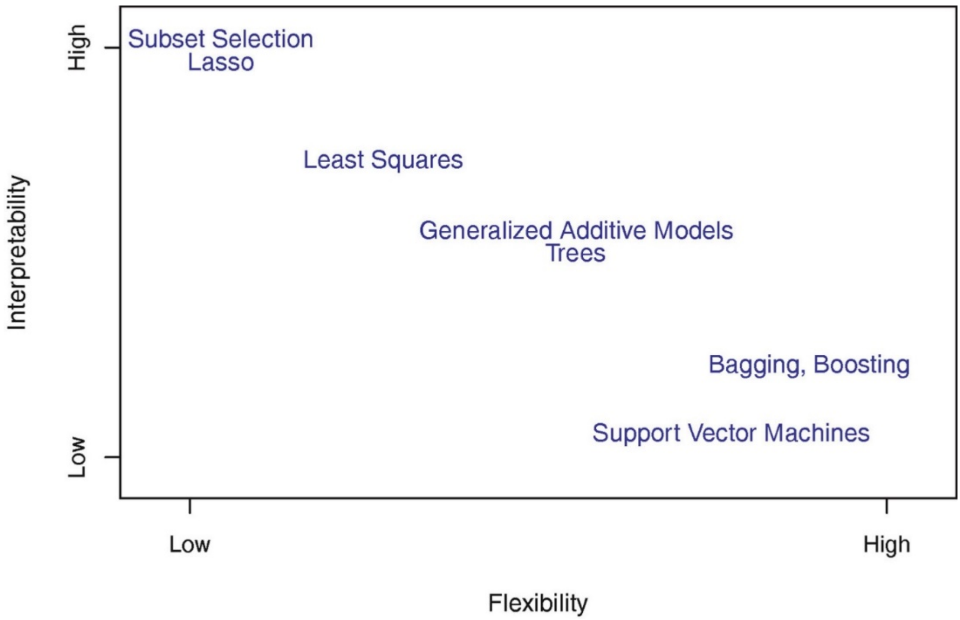
where
$\hat{f}(x_i)$ – the prediction of the actual f for the i-th observation.

If the estimation error is calculated for the data from the training set, it is a measure of the goodness of fit of function $\hat{f}$ to empirical data y_i. However, we are actually interested in the accuracy of the predictions we obtain when applying a given statistical learning method to a previously unknown set of test data. Whether $y_i \approx \hat{f}(x_i)$ is of no particular interest in this context; what is important is that $\hat{f}(x_0)$ is approximately equal to y_0, where (x_0, y_0) is an observation from the non-processed test set. We select the method which provides the lowest-test MSE as opposed to the lowest-training MSE. If a large number of test observations was available, the average squared prediction error for these test observations (x_0, y_0) could be computed. We select the model for which the average-test MSE is as small as possible. The MSE measure is treated as a generalisation error. In practice, the set of observations is usually divided into two subsets: the training set, on which we train the statistical learning method, and the test set, which is used to verify the effectiveness of the learning method. The proportion of the division of the data set into the training and the test part tends to be problematic, as on the one hand, it is assumed that the training set should be larger than the test set, but on the other hand, the test set cannot be too small. It is recommended that the proportion of the division into the training and test set is 8 to 2 (Géron, 2018). There are numerous studies comparing the prognostic abilities of forecasting models which use different proportions of the division of data sets into the training and testing part (cf. e.g. Pociecha et al., 2014).

Function $\hat{f}$ can fairly flexibly match the actual function f. However, when applying statistical learning procedures, the problem of the overfitting of the $\hat{f}$ function to the data may occur. This is related to the possible overtraining (overfitting) of the data learning model. This is the case when the MSE based on the training set is clearly smaller than the MSE based on the test set. Along with the increase in the flexibility of the statistical learning method, a monotonic decrease in the MSE on the training set is observed. On the other hand, the distribution of the MSE on the test set is U-shaped and with the increasing flexibility of function $\hat{f}$, the MSE first decreases and then increases. These are the basic properties of statistical learning methods, independent of the specific data set and independent of the used learning method (James et al., 2013). This provides an opportunity to determine the optimal degree of flexibility of function $\hat{f}$ relative to the training set. The basic method of determining this optimum is cross-validation (Koronacki & Ćwik, 2005). It consists in separating the training set into mutually complementary subsets, and the models are trained in various combinations of these subsets and evaluated using the remaining, unused subsets; in result, the optimal model is determined.

The search for a compromise between the accuracy of prediction and the interpretation of the statistical learning model is an issue related to the one described above. Statistical learning methods involve functions which fairly flexibly adapt to the data from the training set. Linear regression is an example of an inflexible function, while e.g. spline functions are more flexible. The more restrictive functions, and therefore demonstrating less flexibility, allow for a deeper substantive interpretation of the obtained results. Flexible functions, on the other hand, can lead to such complicated estimates of the shape of the actual function *f* that it is difficult to interpret the relationship between the assumed predictors and the dependent variable. The relationship between the flexibility and interpretability of statistical learning methods is presented by James et al. (2013) as in Figure 1.

Figure 1. The relationship between interpretability and flexibility of statistical learning methods



Source: James et al. (2013, p. 25).

The authors indicate that the relation between the flexibility and interpretability of statistical learning methods is approximately inversely proportional. The Lasso regression is fully interpretable as it allows for the joint selection of explanatory variables and the assessment of their impact on the dependent variable (Kubus, 2014 or Tibshirani, 1996). Generalised additive models (GAM) are interpretable and at the same time more flexible, as they allow non-linear relationships between variables. Fully non-linear models, including bagging or boosting and the support vector method (SVM) are highly flexible, but difficult to interpret in terms of content. To sum up, if the aim of using statistical learning methods is to make the most precise prediction possible, then the most flexible learning method should be selected. If the interpretation of the relationship between the response variable and the explanatory variables is important, then the more classic methods should be applied (James et al., 2013).

Statistical learning methods are used to solve both regression and classification problems. If the result (response) variable is a quantitative (directly measurable) variable, then the explanatory variables' (predictors') influence on it is examined by regression. If the result variable is of a qualitative (directly non-measurable) nature, then the relationship between it and the predictors is examined by means of the classification method (Hastie et al., 2009).

A wide range of statistical learning methods are presented in the literature. There is no one best method in statistics and no method dominates all the others for all possible data sets. For a particular dataset, one method may work best, but another method may prove more efficient in relation to a similar and yet different data set. Therefore, statisticians face an important task of selecting the most effective method which – when applied for a given data set – gives the best results. In conclusion, the choice of the most appropriate statistical learning method is one of the most challenging decisions in statistical research practice (James et al., 2013).

5. Statistical learning paradigm

The previously characterised premises and principles of learning from data allow for the formulation of a statistical learning paradigm. The statistical learning paradigm will be presented against the classical paradigm of statistical inference. The starting point of the mathematical statistics paradigm is the probability theory and its basic concepts, including the random event, the axioms of probability theory, the random variable and its distribution. They are followed by the theorems of the probability of events, Bayes' theorem, the formalisation of particular types of distributions of a random variable and their characteristics in the form of distribution parameters for one- and multi-dimensional variables. The key elements of statistical inference include the concept of the distribution of statistics from a sample, the principles of estimation parameters and the principles of the verification of statistical hypotheses (Kot et al., 2011). The essence of the mathematical statistics paradigm is to start from the theory of probability and statistical inference and to check to what extent the empirical data can fit into the theoretical framework of mathematical statistics.

The statistical learning paradigm involves the opposite – the starting point is the available data set. The theory is based on the 'let the data speak for itself' and 'we learn from the data' concepts, which is consistent with neo-positivist beliefs, according to which all knowledge is based on empirical data, whereas anything that is not based on empirical facts is rejected. Neo-positivists assumed that experience is the source of all knowledge about the real world (Kołakowski, 2004).

The statistical inference paradigm is based on the concept of general population and sample. The condition for the correctness of inference about a population based on a sample is the random selection of the sample. The sampling method focuses strongly on sampling procedures so that the sample is representative of the entire population. The statistical learning paradigm, on the other hand, ignores the notions of population and sample. Instead, it assumes that we have a sufficiently large set of empirical data on the basis of which we can effectively make predictions and infer about the reality which these data come from. In the practice of applying statistical learning procedures, it is often presumed that the training set has the properties of a random sample, but its actual randomness is not verified. Perceiving the training and test set automatically as random samples is in fact an unjustified transfer of the features of the statistical inference paradigm onto the statistical learning paradigm.

In the era of powerful computers, it is possible to collect, process and store large data sets, known as Big Data. However, such sets do not have the characteristics of random samples; they are said to be noisy, i.e. partially random and contain unreliable information which needs to undergo various data cleaning processes. It should be mentioned here that data from a random sample, selected in accordance with the rules of the sampling method, are not subject to 'cleaning' as, by definition, their appearance in the sample is determined by the probability of their occurrence in the population. However, the application of statistical learning procedures should not be limited to Big Data as they are known to be used in training sets with less than one hundred observations (James et al., 2013).

The essence of the statistical learning paradigm is the creation of self-learning systems, i.e. systems which improve automatically through experience. In the case of statistical learning in the supervised version, it involves providing the algorithm with a set of input-output pairs (x_i, y_i) in order to find unknown function f by mapping input data to output data, with the accuracy of the minimised mean square error of the estimate or mean prediction error. The actual function f in the statistical learning paradigm is understood as a black box – it can be a parameterised or non-parameterised function, it can even be a non-algorithmic procedure.

The purpose of statistical learning is to get closer to the real function by estimating it on the training set in which function $\hat{f}$ is taught how to recognise the actual function f as accurately as possible. The statistical learning effect is tested on a test set and its optimisation is performed in the process of cross-validation. The basic difference between the process of statistical estimation and the process of statistical learning is that in the former we estimate the parameters of a pre-determined function, and in the latter the form of this function and its parameters are selected by the learning method.

The statistical learning paradigm includes classical linear regression models, logistic regression, discriminant analysis, polynomial models, splined functions, generalised additive models, kernel classifiers, regression and classification trees, bagging and boosting methods, random forests, neural networks, support vectors machines, the k-means method, and other lesser-known learning procedures. As the list above suggests, the range of statistical learning tools is much wider than that of the classical mathematical statistics tools.

The concept and the theory of probability plays a key role in the paradigm of statistical inference. In the statistical learning paradigm, its role is secondary as there are serious doubts whether the training set could be considered as a random data set. In this sense, the statistical learning paradigm is getting closer to the descriptive statistics paradigm.

The literature also emphasises the difference in terms of the research goals that can be achieved through both paradigms. Statistical research conducted within the mathematical statistics paradigm focuses primarily on explaining the relationships between the studied variables, i.e. on the implementation of analytical goals; thus, the forecasts built on their basis are often imprecise. Empirical research conducted within the statistical learning paradigm involves building on their basis forecasts which would be as accurate as possible; nevertheless, their analytical and interpretative role could be limited.

In conclusion, however, it should be emphasised that the statistical learning paradigm is a more universal research platform, as it has in fact absorbed the statistical inference paradigm at the expense of weakening its original assumptions. The statistical learning paradigm offers a great opportunity to use the computing power of modern computers and large data sets produced by contemporary socio-economic life.

References

Alexander, M., & Kusleika, R. (2019). *Access 2019 PL: Biblia* (R. Meryk & T. Walczak, Trans.). Gliwice: Helion.

Bracha, C. (1996). *Teoretyczne podstawy metody reprezentacyjnej*. Warszawa: Wydawnictwo Naukowe PWN.

Cassel, C. M., Särndal, C. E., & Wretman, J. H. (1977). *Fundations of Inference in Survey Sampling*. New York: John Wiley & Sons.

Cichosz, P. (2000). *Systemy uczące się*. Warszawa: Wydawnictwa Naukowo-Techniczne.

Géron, A. (2018). *Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow* (K. Sawka, Trans.). Gliwice: Helion.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *Elements of Statistical Learning: Data Mining, Inference and Prediction* (2nd edition). New York: Springer. https://doi.org/10.1007/978-0-387-84858-7.

Hołda, A., & Pociecha, J. (2009). *Probabilistyczne metody badania sprawozdań finansowych*. Kraków: Wydawnictwo Uniwersytetu Ekonomicznego.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. New York: Springer. https://doi.org/10.1007/978-1-4614-7138-7.

Kim, W., Choi, B. J., Hong, E.-K., Kim, S. K., & Lee, D. (2003). A Taxonomy of Dirty Data. *Data Mining and Knowledge Discovery*, *7*(1), 81–99. https://doi.org/10.1023/A:1021564703268.

Kołakowski, L. (2004). *Filozofia pozytywistyczna: Od Hume'a do Koła Wiedeńskiego*. Warszawa: Wydawnictwo Naukowe PWN.

Koronacki, J., & Ćwik, J. (2005). *Statystyczne systemy uczące się*. Warszawa: Wydawnictwo Naukowo Techniczne.

Kot, S. M., Jakubowski, J., & Sokołowski, A. (2011). *Statystyka* (2nd edition). Warszawa: Difin.

Kubus, M. (2014). Propozycja modyfikacji metody złagodzonego LASSO. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu / Research Papers of Wrocław University of Economics*, (327), 77–84. https://www.dbc.wroc.pl/dlibra/publication/27745/edition/25111/content?&meta-lang=pl.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kuhn, T. S. (2020). *Struktura rewolucji naukowych*. Warszawa: Wydawnictwo Aletheia.

Migdał-Najman, K., & Najman, K. (2017). Big Data = Clear + Dirty + Dark Data. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu / Research Papers of Wrocław University of Economics*, (469), 131–139. https://doi.org/10.15611/pn.2017.469.13.

Onet. (n.d.). *Komputer Świat*. Retrieved January 21, 2021, from https://www.komputerswiat.pl/onet.

Pociecha, J. (2020). Philosophical foundations of statistical research. *Przegląd Statystyczny. Statistical Review*, *67*(3), 195–211. https://doi.org/10.5604/01.3001.0014.7109.

Pociecha, J., Pawełek, B., Baryła, M., & Augustyn, S. (2014). *Statystyczne metody prognozowania bankructwa w zmieniającej się koniunkturze gospodarczej*. Kraków: Fundacja Uniwersytetu Ekonomicznego w Krakowie. https://im.uek.krakow.pl/wp-content/uploads/2020/03/Statystyczne-metody-prognozowania-bankructwa_online.pdf.

Pociecha, J., Podolec, B., Sokołowski, A., & Zając, K. (1988). *Metody taksonomiczne w badaniach społeczno-ekonomicznych*. Warszawa: Państwowe Wydawnictwo Naukowe.

Russel, S., & Norvig, P. (2003). *Artificial Intelligence: A Modern Approach*. New Jersey: Prentice Hall, Pearson Education, Upper Saddle River.

Steczkowski, J. (1995). *Metoda reprezentacyjna w badaniach zjawisk ekonomiczno-społecznych*. Warszawa–Kraków: Wydawnictwo Naukowe PWN.

Szymczak, W. (2018). *Praktyka wnioskowania statystycznego*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Tabakow, M., Korczak, J., & Franczyk, B. (2014). Big Data – definicje, wyzwania i technologie informatyczne. *Informatyka ekonomiczna / Business Informatics*, (1), 138–153. https://doi.org/10.15611/ie.2014.1.12.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Tillé, Y. (2006). *Sampling Algorithms*. New York: Springer. https://doi.org/10.1007/0-387-34240-0.

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

Przegląd Statystyczny. Statistical Review, 2021, vol. 68, 1, 17–46 DOI: 10.5604/01.3001.0015.0376
Received: 5.03.2021, revised: 4.05.2021, accepted: 17.06.2021

Predicting in multivariate incomplete time series. Application of the expectation-maximisation algorithm supplemented by the Newton-Raphson method

Adam Korczyński[a]

Abstract. Statistical practice requires various imperfections resulting from the nature of data to be addressed. Data containing different types of measurement errors and irregularities, such as missing observations, have to be modelled. The study presented in the paper concerns the application of the expectation-maximisation (EM) algorithm to calculate maximum likelihood estimates, using an autoregressive model as an example. The model allows describing a process observed only through measurements with certain level of precision and through more than one data series. The studied series are affected by a measurement error and interrupted in some time periods, which causes the information for parameters estimation and later for prediction to be less precise. The presented technique aims to compensate for missing data in time series. The missing data appear in the form of breaks in the source of the signal. The adjustment has been performed by the EM algorithm to a hybrid version, supplemented by the Newton-Raphson method. This technique allows the estimation of more complex models. The formulation of the substantive model of an autoregressive process affected by noise is outlined, as well as the adjustment introduced to overcome the issue of missing data. The extended version of the algorithm has been verified using sampled data from a model serving as an example for the examined process. The verification demonstrated that the joint EM and Newton-Raphson algorithms converged with a relatively small number of iterations and resulted in the restoration of the information lost due to missing data, providing more accurate predictions than the original algorithm. The study also features an example of the application of the supplemented algorithm to some empirical data (in the calculation of a forecasted demand for newspapers).

Keywords: missing data, multivariate time series, expectation-maximisation algorithm, Newton-Raphson algorithm
JEL: C13, C19, C61

1. Introduction

Data quality insight is one of the aspects of data science which supports data analysis by providing a framework that allows working with real data. Real data naturally tends to be erroneous, displaying measurement errors, incompleteness of data and other irregularities such as outliers.

The source of a measurement error can be different depending on the application. When analysing environmental and technical problems, it can result from instrument imprecision and differences in the locations of the assessments as shown in a pollution study by Butland et al. (2013). An example of a measurement error in economic analysis is provided by Fukuda (2005) in a study of a business cycle indicator affected by an error resulting from irregular sampling, which arose from

[a] Warsaw School of Economics, Institute of Statistics and Demography, ul. Madalińskiego 6/8, 02-513 Warsaw, Poland, e-mail: akorczy@sgh.waw.pl, ORCID: https://orcid.org/0000-0002-1533-7197.

the need to produce business reports within a short period in order to evaluate the current business conditions. Ghassemi et al. (2015) discuss the issue of noisy data in medical studies where errors result from difficulties related to data collection. In effect, only irregularly sampled heterogeneous clinical records are available.

The problem of missing data has been extensively studied and described in statistical literature. A comprehensive overview of the topics related to the analysis of missing data is provided by Little & Rubin (2002). One of the recent works on missing data and other aspects of data quality and their implications from a survey perspective is by Laaksonen (2018).

The reason behind data incompleteness is twofold. It can be related to data collection issues, such as data entry delays or to the attitude of the respondents who may not be willing to provide certain information. While the former is usually less frequent and driven by completely random events, the latter can result in bias. Depending on the missing data mechanism, the missingness can lead to bias if the probability of observing an outcome would be determined by some subject characteristics or the outcome level itself. Naturally, a reduced amount of information causes the estimation of parameters to be imprecise. Various weighting adjustments and imputation techniques, including multiple imputation, have been implemented to address the above-mentioned problems in different data structures.

One other aspect in data quality assessment is the existence of outliers. Outlying observations can appear in univariate distribution, but when analysed from a multivariate perspective, they are likely to become a more complex problem. The analysis primarily focuses on influential observations, involving such values which affect the estimates or their standard errors. In some cases the outliers are simply erroneous data points and therefore one would expect a correction of such values to be made prior to the analysis. In practice, many erroneous data points can be detected directly by predefined edit checks. If an outlier turns to be a true value or it is not possible to clean the data for operational reasons, e.g. reporting time requirements, robust methods and models are designed to solve such problems. In univariate settings, the natural choice would be to utilise quantiles of the distribution and this concept can be extended to quantile regression models showing robustness against the observations that behave unlike most of the other ones (Koenker, 2005).

The aim of this paper is to assess the extension of the numeric technique for time series analysis outlined in Shumway & Stoffer (1982) in order to address the prediction problem, considering such aspects of data quality as measurement errors and data incompleteness. The model applied for smoothing and forecasting is the Kalman filter (Kalman, 1960), which is estimated through the expectation-maximisation (EM) algorithm. The standard EM algorithm requires closed-form solutions which are not readily available for more complex data structures. This in practice leads to the introduction of constraints imposed upon selected parameters in order to simplify the numeric problem. The extension suggested in this paper involves incorporating the Newton-Raphson method within the regular EM algorithm in order to allow the estimation of parameters which are otherwise set to zero. This solution offers flexibility and reduces constraints when making assumptions for the analysed process, e.g. we can assume that there is a non-zero correlation between the measurement errors in a bivariate time series. The text provides a description of the studied process which uses an autoregressive model with random noise affecting the observed time series data.

2. Model for noisy time series

Kalman (1960) introduced a model for the description of processes with the aim of detecting signals in the presence of random noise. The detection process involves separating the signal from random noises, thereby providing basis for predicting the signals. The model finds its application when the observation is performed independently by two or more individuals or devices, with the goal to measure the same characteristic, although the data from each source are subject to measurement error. An example of such observation is described by Cajner et al. (2019). The purpose of the study was to combine two individual sources of data on the labour market (drawn from separate surveys) in order to reduce a measurement error which was in fact affecting both sources, and thus improve the accuracy of the estimates related to labour market characteristics, including the number of active and paid employees.

In multivariate time series settings, all data sources are utilised to predict the examined signal (e.g. the level of a specific characteristic), but in order to make that prediction, the data must first be cleaned from noise. In the notation, x_t will be the underlying, true level of the analysed parameter over time $t = 1, 2, \dots, n$. Only y_t – the noisy data are observed. A process involving multiple data sources measuring the same characteristic with only limited precision can be expressed as follows (Shumway & Stoffer, 1982, p. 254):

$$\mathbf{y}_t = \mathbf{M}_t \, \mathbf{x}_t + \mathbf{v}_t \quad \text{for} \quad t = 1, 2, \dots, n, \tag{1}$$

where

$\mathbf{y}_t$ – $k \times 1$ is the vector of the observed series at time t;

$\mathbf{M}_t$ – $k \times p$ is the design matrix transforming the unobserved $p \times 1$ vector $\mathbf{x}_t$ into $\mathbf{y}_t$. If applied in incomplete data, $\mathbf{M}_t$ would refer to the missing data indicator as defined in Little & Rubin (2002, Section 1.2), pointing to those elements of observation vector $\mathbf{y}_t$ which are known and which are not;

$\mathbf{v}_t$ – $k \times 1$ is the vector of the process error terms following normal distribution $N(\boldsymbol{\mu}, \mathbf{R})$ with zero-mean vector and $k \times k$ covariance matrix $\mathbf{R}$.

Let us assume that we want to obtain the measurement of $\mathbf{x}_t$, but in fact this variable is measured by two independent sources and we are not able to observe it directly. What we have is 2×1 vectors of measurements from each of the sources over time. Based on (1) the process is then as follows:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} m_{1t} \\ m_{2t} \end{bmatrix} x_t + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} \text{ for } t = 1, 2, ..., n, \tag{2}$$

where

$$\begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} r_1^2 & r_{12} \\ r_{12} & r_2^2 \end{bmatrix}\right). \tag{3}$$

Vector $\mathbf{v}_t$ can be considered as a component aiming to capture the measurement error.

For example, we can assume that there are two sources of data for the same series. In economics it could be demand assessed by two individual sources. In medical statistics the application of the above concept would relate to models for laboratory measurements based on different types of samples (e.g. serum and plasma). Assuming the existence of a correlation between the two specimens (see for example Carey et al., 2016), the two series can be combined in order to describe the changes in the laboratory parameters under study over time.

We further specify the process describing random series $\boldsymbol{x_t}$ as a first-order autoregressive process, thereby introducing a correlation between the adjacent observations, i.e. the outcome at time t would depend on its level at time $t-1$. As the model would simplify the true process, we include the term $\boldsymbol{w_t}$ to reflect deviations of the actual outcome from the modelled series. The model specification is then as follows:

$$\mathbf{x}_t = \mathbf{\Phi}_\mathrm{t}\, \mathbf{x}_{t-1} + \mathbf{w}_t \text{ for } t = 1, 2, ..., n, \tag{4}$$

where

$\mathbf{\Phi}_\mathrm{t}$ – $p \times p$ is the transition matrix expressing the relationship between the adjacent values of the series over time;

$\mathbf{w}_t$ – $p \times 1$ is the vector of model error terms following a normal distribution, $N(\boldsymbol{\mu}, \mathbf{Q})$ is the zero-mean vector and $p \times p$ is covariance matrix $\mathbf{Q}$ for the uncorrelated process. The initial value of process x_0 is assumed to be a normal random vector from $N(\boldsymbol{\mu}_0, \mathbf{\Sigma}_0)$, where $\mathbf{\Sigma}_0$ is a $p \times p$ covariance matrix.

In the simplest case with univariate series, the model would take the following form:

$$x_t = \phi x_{t-1} + w_t \quad \text{for} \quad t = 1, 2, ..., n, \tag{5}$$

where ϕ represents the autocorrelation between the adjacent elements of the series and where the variance of $\mathbf{w}_t$ is univariate and noted by q^2.

Error $\mathbf{w}_t$ captures the uncertainty displayed by the model, which reflects the ability to describe the process creating a series of observations.

In order to obtain the estimates of x_1, x_2 ..., x_n, observed series of vectors $y_1, y_2 \dots, y_n$ are smoothed by using the model specified in (1) and (4). The same framework can be applied to calculate forecast $X_{n+1}, X_{n+2}, \dots, X_{n+l}$, with l representing the number of periods for which the prediction is produced.

The process applicable to bivariate series which begins with an observation, and then proceeds to the smoothing of the series, on to the prediction is depicted in Figure 1.

Figure 1. Smoothing and predicting based on bivariate observed series

t	Y_{1t}	Y_{2t}		X_t
1	y_{1t}	y_{21}		x_1
2	y_{12}	y_{22}		x_2
⋮	⋮	⋮		⋮
n	y_{1n}	y_{2n}	⟶	x_n
				x_{n+1}
				x_{n+2}
				⋮
				x_{n+l}

Source: author's work.

In the formulation of the model above, there are two random terms capturing the specific parts of the overall variability occurring in the process. Random term $\mathbf{w}_t$ in (4) captures the uncertainty related to the model specification in relation to the actual process, while $\mathbf{v}_t$ in (1) is to express the measurement error or the noise introduced by the measurement technique or displayed by the nature of the observed data.

The model parameters from (1) and (4) are estimated based on an iterative version of the maximum likelihood approach. The usage of the EM algorithm supplemented by Newton-Raphson allows expanding the model complexity, also taking the practical perspective into account, including handling missing data. The following sections describe the application of the estimating algorithm to obtain the maximum likelihood estimates of the model for bivariate incomplete data.

3. Estimation of the model

3.1. The likelihood function and the estimating algorithm

Assuming that the sample is large and the observation units are selected independently, the model specified by (1) and (4) can be estimated using the maximum likelihood method. The method maximises the joint density expressed as a product of individual probabilities given the population distribution (for details on the method see for example Mittelhammer, 2013, Section 8.3; Little & Rubin, 2002, Chapter 6). For the model specified in Section 1 the likelihood is a product of three factors related to the initial condition for the process, the autoregressive part and the part explaining how underlying series x_t $(t = 1, 2, \dots, n)$ are transformed into actual observations (Shumway & Stoffer, 2017, p. 306):

$$f(\mathbf{y}|\boldsymbol{\theta}) = f(\mathbf{x}_0|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \prod_{t=1}^{n} f(\mathbf{x}_t|\mathbf{x}_{t-1}, \boldsymbol{\Phi}, \mathbf{Q}) \prod_{t=1}^{n} f(\mathbf{y}_t|\mathbf{x}_t, \mathbf{R}), \tag{6}$$

where $\boldsymbol{\theta}$ represents the set of parameters describing the population of interest and requiring estimation.

Given joint density (6), the log-likelihood assuming $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, considering $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$, and ignoring the constants is as follows (Shumway & Stoffer, 1982, p. 256):

$$\begin{aligned} \ln L\,(\boldsymbol{\theta}|\mathbf{x}, \mathbf{y}) = & -\frac{1}{2}\ln|\boldsymbol{\Sigma}_0| - \frac{1}{2}(\mathbf{x}_0 - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_0^{-1}(\mathbf{x}_0 - \boldsymbol{\mu}_0) \\ & -\frac{n}{2}\ln|\mathbf{Q}| - \frac{1}{2}\sum_{t=1}^{n}(\mathbf{x}_t - \boldsymbol{\Phi}\mathbf{x}_t)'\mathbf{Q}^{-1}(\mathbf{x}_t - \boldsymbol{\Phi}\mathbf{x}_t) \\ & -\frac{n}{2}\ln|\mathbf{R}| - \frac{1}{2}\sum_{t=1}^{n}(\mathbf{y}_t - \mathbf{M}_t\mathbf{x}_t)'\mathbf{R}^{-1}(\mathbf{y}_t - \mathbf{M}_t\mathbf{x}_t). \end{aligned} \tag{7}$$

The maximum likelihood (ML) estimates of the model given by (1) and (4) are obtained by maximising the log-likelihood function (7) with respect to parameters $\boldsymbol{\mu}_0$, $\boldsymbol{\Sigma}_0$, $\boldsymbol{\Phi}$, $\mathbf{Q}$ and $\mathbf{R}$. In relation to simple problems, the ML estimators can be derived directly from the log-likelihood function. When the level of complexity increases, which may result from the nature of the series (e.g. higher dimensions, more complex model specifications), the direct maximisation is rarely available and in practice referring to optimisation techniques such as Newton-Raphson or EM algorithms becomes inevitable.

Shumway and Stoffer (1982, pp. 256–257) provide formulas for estimating the parameters of (1) and (4) using the EM algorithm. This approach enables addressing two types of data irregularities:

- the expectation step allows the implementation of the smoothing estimator so that the observed data (e.g. two-dimensional series) assessing the same effect but with some level of a measurement error represented by **R** can be transformed into the smoothed series;
- the expectation step can be further extended to incorporate compensation for the missing data, i.e. finding the maximum likelihood estimates through an iterative process replacing the missing values by their expectations drawn from a conditional distribution from a prespecified population.

The algorithms outlined in Shumway & Stoffer (1982) as well as in Little & Rubin (2002, Chapter 8) focus on applications assuming a direct approach to the maximisation step. The suggestion proposed in this paper is to combine the EM algorithm with the Newton-Raphson method, in this case operating as a sub-algorithm. Some examples of such an approach have been described as hybrid maximisation methods in Little & Rubin (2002, pp. 186–188) (see for example Lange, 1995).

The goal of the approach proposed in this paper is to allow more complex model specifications, which is basically equivalent to relaxing some of the constraints used to simplify the computations. In Shumway and Stoffer (1982), it is the assumption on uncorrelated measurement errors in vector $\mathbf{v}_t$ which provides a simplified version of the **R** matrix with a covariance occurring between the measurement errors equal to zero.

The EM algorithm formulation is depicted in Korczyński, 2018, Chapter 4; Little & Rubin, 2002, Chapter 8, and Molenberghs & Kenward, 2007, pp. 93–103. The algorithm consists of an initial step and two main steps: the expectation step, in which the expected value of the log-likelihood is calculated, and the maximisation step, in which updates to the model parameters are found so that they maximise the expected likelihood at the current iteration. In the presented version of the algorithm, we assume that the actual series of $\boldsymbol{y}_t$ $(t = 1, 2, \dots, n)$ is not fully observable, and then data vector $\mathbf{y}$ is split into observed $\boldsymbol{y}^{\text{obs}}$ and missing $\boldsymbol{y}^{\text{mis}}$. The iterative process consists of three steps.

Initial step. Set the parameter vector to certain initial values $\boldsymbol{\theta}^{(0)}$. In a missing data application the complete-case or available-case approach can be used (for details see for example Little & Rubin, 2002, Chapter 8) to obtain the initial estimates.

Expectation step. Calculate the expected value of the log-likelihood function with the current estimates of parameter vector $\boldsymbol{\theta}^{(i)}$:

$$Q\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}^{(i)}\right) = E\left[\ln L\left(\boldsymbol{\theta}|\mathbf{y}\right)|\mathbf{y}^{obs}, \boldsymbol{\theta}^{(i)}\right]. \tag{8}$$

In practice, this step requires finding the expected values of the sufficient statistics and replacing the unknown components by their expected values. The unknown components may be related to the smoothing of the series to identify the underlying unobserved process (Shumway & Stoffer, 1982, Section 2) or to the missing values (Dempster et al., 1977, pp. 3–4).

Maximisation step. Find $\boldsymbol{\theta}^{(i+1)}$ maximising the log-likelihood function considering the current expectations over the log-likelihood function:

$$Q\left(\boldsymbol{\theta}^{(i+1)}\middle|\boldsymbol{\theta}^{(i)}\right) \geq Q\left(\boldsymbol{\theta}^{(i+1)}\middle|\boldsymbol{\theta}^{(i)}\right) \text{ for all } \boldsymbol{\theta}. \tag{9}$$

After having updated $\boldsymbol{\theta}^{(i+1)}$, we need to go back to the first step and the cycle is repeated until convergence is reached, i.e. until the changes observed for the estimates in question are lower than an arbitrarily small value ε.

The EM algorithm, as specified above, ensures convergence to the maximum of the log-likelihood (Molenberghs & Kenward, 2007, p. 95), although the rate of the convergence can be slow.

The suggested extension as outlined in Korczyński (2018, pp. 216–225), involves the replacement of the maximisation step with the Newton-Raphson maximisation of the parameters for which closed-form solutions are not available. In the model outlined by (1) and (4), the Newton-Raphson step would entail estimating the elements of variance-covariance matrix **R,** depicting the behaviour of the measurement error terms. The general notation of the maximisation is as follows:

$$\widehat{\boldsymbol{\theta}}^{(i+1)} = \widehat{\boldsymbol{\theta}}^{(i)} - \left\{E\left[\frac{\partial^2 \ln L\left(\boldsymbol{\theta}^{(i)}|\mathbf{y}\right)}{\partial\boldsymbol{\theta}^{(i)}\partial\boldsymbol{\theta}^{(i)}}\right]\right\}^{-1} E\left[\frac{\partial \ln L\left(\boldsymbol{\theta}^{(i)}|\mathbf{y}\right)}{\partial\boldsymbol{\theta}^{(i)}}\right]. \tag{10}$$

In a regular framework which ignores the noise around the signal and any potentially missing data we would proceed directly to the maximisation of the log-likelihood (7). However, we assume that the true process is only approximated by individual data series and we let some elements of the series be missing, which is unavoidable in statistical practice. An example of this process which will be further discussed in this text is outlined in Figure 2.

Figure 2. Smoothing and predicting based on bivariate observed series with missing data

t	Y_{1t}	Y_{2t}		X_t
1	y_{11}	y_{21}		x_1
2	y_{12}	y_{22}		x_2
3	.	y_{23}		x_3
4	.	y_{24}		x_4
5	y_{15}	y_{25}		x_5
6	y_{16}	.	⟶	x_1
	⋮	⋮		⋮
n	y_{1n}	.		x_n
				x_{n+1}
				x_{n+2}
				⋮
				x_{n+l}

Source: author's work.

Firstly, we need to find the underlying series through the smoothing of the observed data $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ to receive $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$.

Secondly, we need to consider the fact that some parts of the observed series are missing. Both aspects are dealt with by determining the expected value of the log-likelihood function (7).

After calculating the maximum likelihood estimates, we can proceed to the prediction of the next l elements of the series.

In order to simplify the notation, we assume that the actual process is observed through two data sources as described by (2) and (5). In this case the log-likelihood can be written as:

$$\begin{aligned}\ln L\,(\boldsymbol{\theta}|\mathbf{x},\mathbf{y}) = &-\tfrac{1}{2}\ln\sigma_0^2-\tfrac{1}{2\sigma_0^2}(x_0-\mu_0)^2\\ &-\tfrac{n}{2}\ln q^2-\tfrac{1}{2q^2}\textstyle\sum_{t=1}^{n}(x_t-\phi x_{t-1})^2\\ &-\tfrac{n}{2}\ln\begin{vmatrix} r_1^2 & r_{12}\\ r_{12} & r_2^2\end{vmatrix}\\ &-\tfrac{1}{2}\textstyle\sum_{t=1}^{n}\left(\begin{bmatrix}y_{1t}\\ y_{2t}\end{bmatrix}-\begin{bmatrix}m_{1t}\\ m_{2t}\end{bmatrix}x_t\right)'\begin{bmatrix} r_1^2 & r_{12}\\ r_{12} & r_2^2\end{bmatrix}^{-1}\left(\begin{bmatrix}y_{1t}\\ y_{2t}\end{bmatrix}-\begin{bmatrix}m_{1t}\\ m_{2t}\end{bmatrix}x_t\right).\end{aligned} \tag{11}$$

The log-likelihood function (11) can be expressed as:

$$\begin{aligned}\ln L\,(\boldsymbol{\theta}|\mathbf{x},\mathbf{y}) = &-\tfrac{1}{2}\ln\sigma_0^2-\tfrac{1}{2\sigma_0^2}(x_0^2-2\mu_0x_0+\mu_0^2)-\tfrac{n}{2}\ln q^2\\ &-\tfrac{1}{2q^2}\textstyle\sum_{t=1}^{n}(x_t^2-2\phi x_tx_{t-1}+\phi^2x_{t-1}^2)-\tfrac{n}{2}\ln(r_1^2r_2^2-r_{12}^2)\\ &-\tfrac{1}{2(r_1^2r_2^2-r_{12}^2)}\textstyle\sum_{t=1}^{n}[(y_{1t}^2-2m_{1t}y_{1t}x_t+m_{1t}^2x_t^2)r_2^2\\ &-2(y_{1t}y_{2t}-m_{2t}y_{1t}x_t-m_{1t}y_{2t}x_t-m_{1t}m_{2t}x_t^2)r_{12}\\ &+(y_{2t}^2-2m_{2t}y_{2t}x_t+m_{2t}^2x_t^2)r_1^2].\end{aligned} \tag{12}$$

The implementation of the Newton-Raphson step allows assuming the existence of a correlation between the error terms in the y_{1t} and y_{2t} observed series.

In order to apply the EM algorithm, we firstly specify the expectation of loglikelihood (8). This step requires calculating:

- the expectation of the unobserved $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ series . This allows for the smoothing of the observed process to assess the underlying series;
- the expectation of the incomplete elements of the observed $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ series. This part addresses the issue of missing data in the process.

The expected log-likelihood for (12) can be expressed as:

$$
\begin{aligned}
E[\ln L\,(\boldsymbol{\theta}|\mathbf{x},\mathbf{y})] = & -\tfrac{1}{2}\ln\sigma_0^2 - \tfrac{1}{2\sigma_0^2}[E(x_0^2) - 2\mu_0 E(x_0) + \mu_0^2] - \tfrac{n}{2}\ln q^2 \\
& -\tfrac{1}{2q^2}\textstyle\sum_{t=1}^{n}[E(x_t^2) - 2\phi E(x_t x_{t-1}) + \phi^2 E(x_{t-1}^2)] \\
& -\tfrac{n}{2}\ln(r_1^2 r_2^2 - r_{12}^2) - \tfrac{1}{2(r_1^2 r_2^2 - r_{12}^2)}\textstyle\sum_{t=1}^{n}\{[E(y_{1t}^2) \\
& -2m_{1t}E(y_{1t}x_t) + m_{1t}^2 E(x_t^2)]r_2^2 \\
& -2[E(y_{1t}y_{2t}) - m_{2t}E(y_{1t}x_t) - m_{1t}E(y_{2t}x_t) \\
& -m_{1t}m_{2t}E(x_t^2)]r_{12} + [E(y_{2t}^2) \\
& -2m_{2t}E(y_{2t}x_t) + m_{2t}^2 E(x_t^2)]r_1^2\}.
\end{aligned}
\tag{13}
$$

The next step is to find the expectations of $\mathbf{x}_t$ and $\mathbf{y}_t$ in (13). The estimator of the $\mathbf{x}_t$ smoothed series is depicted in Subsection 3.2. The calculation of the expectations of the missing elements of $\mathbf{y}_t$ is presented in Subsection 3.3. Given the expected $\mathbf{x}_t$ and $\mathbf{y}_t$ at a specific iteration of the algorithm, the log-likelihood function is maximised with respect to the $\boldsymbol{\theta}^i$ parameters of interest to update the estimates and obtain $\boldsymbol{\theta}^{i+1}$, which is then utilised to recalculate the expectations of the log-likelihood. The maximisation is outlined in Subsection 3.4. The algorithm runs in cycles, from the expectation step to the maximisation step and over again, until convergence occurs.

3.2. Smoother estimator

For the specified autoregressive model accounting for noise given by (1) and (4), we utilise the Kalman smoother estimator taking relevant steps needed to estimate the studied parameters. The process is outlined in Shumway & Stoffer (1982, pp. 262–263). The expected value of $\mathbf{x}_t$ is noted as:

$$\mathbf{x}_t^n = E[\mathbf{x}_t|\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n], \tag{14}$$

while the error variance-covariance matrix takes the following form:

$$\mathbf{P}_t^n = E[(\mathbf{x}_t - \mathbf{x}_t^n)(\mathbf{x}_t - \mathbf{x}_t^n)'|\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]. \tag{15}$$

As noted in Anderson & Moore (1979, pp. 37–38), error variance-covariance matrix (15) measures how effective the $\hat{\mathbf{x}}_t^n$ estimate is. The trace of $\mathbf{P}_t^n$ from (15) is given by:

$$tr\mathbf{P}_t^n = E\{tr[(\mathbf{x}_t - \mathbf{x}_t^n)(\mathbf{x}_t - \mathbf{x}_t^n)'|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n]\}$$
$$E\{\|\mathbf{x}_t - \mathbf{x}_t^n\|^2|\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}, \tag{16}$$

and it is the conditional error variance for $\hat{\mathbf{x}}_t^n$. The optimisation involves finding $\hat{\mathbf{x}}_t^n$, for which the error variance is minimised.

For $t = 1, 2, \ldots, n$ and with the initial condition for the process of $\mathbf{x}_0^0 = \boldsymbol{\mu}_0$ and $\mathbf{P}_0^0 = \boldsymbol{\Sigma}_0$, the smoothing is performed by means of the following set of recursive equations:

$$\mathbf{x}_t^{t-1} = \boldsymbol{\Phi}\mathbf{x}_{t-1}^{t-1}, \tag{17}$$

$$\mathbf{P}_t^{t-1} = \boldsymbol{\Phi}\mathbf{P}_{t-1}^{t-1}\boldsymbol{\Phi}' + \mathbf{Q}, \tag{18}$$

$$\mathbf{K}_t = \mathbf{P}_t^{t-1}\mathbf{M}_t'(\mathbf{M}_t\mathbf{P}_t^{t-1}\mathbf{M}_t' + \mathbf{R})^{-1}, \tag{19}$$

$$\mathbf{x}_t^t = \mathbf{x}_t^{t-1} + \mathbf{K}_t(\boldsymbol{y}_t - \mathbf{M}_t\mathbf{x}_t^{t-1}), \tag{20}$$

$$\mathbf{P}_t^t = \mathbf{P}_t^{t-1} + \mathbf{K}_t\mathbf{M}_t\mathbf{P}_t^{t-1}. \tag{21}$$

To find estimates $\hat{\mathbf{x}}_t^n$ in (14) and $\widehat{\mathbf{P}}_t^n$ in (15), we carry out backward calculations for $t = n,\ n-1, \ldots, 1$ on:

$$\mathbf{J}_{t-1} = \mathbf{P}_{t-1}^{t-1}\boldsymbol{\Phi}'(\mathbf{P}_t^{t-1})^{-1}, \tag{22}$$

$$\hat{\mathbf{x}}_{t-1}^n = \mathbf{x}_{t-1}^{t-1} + \mathbf{J}_{t-1}(\boldsymbol{x}_t^n - \boldsymbol{\Phi}\boldsymbol{x}_{t-1}^{t-1}), \tag{23}$$

$$\widehat{\mathbf{P}}_{t-1}^n = \mathbf{P}_{t-1}^{t-1} + \mathbf{J}_{t-1}(\mathbf{P}_t^n - \mathbf{P}_t^{t-1})\mathbf{J}_{t-1}'. \tag{24}$$

In expected log-likelihood (13), we need to calculate the expected value of the product of the subsequent elements of unobserved series $E(x_t\, x_{t-1})$. This requires the covariance of the elements, which is calculated for $t = n,\ n-1, \ldots, 2$ using

$$\widehat{\mathbf{P}}_{t-1,t-2}^n = \mathbf{P}_{t-1}^{t-1} + \mathbf{J}_{t-2}' + \mathbf{J}_{t-1}\left(\mathbf{P}_{t,t-1}^n - \boldsymbol{\Phi}\mathbf{P}_{t-1}^{t-1}\right)\mathbf{J}_{t-2}', \tag{25}$$

where

$$\mathbf{P}_{n,n-1}^n = (\mathbf{I} - \mathbf{K}_n\mathbf{M}_n)\boldsymbol{\Phi}\mathbf{P}_{t-1}^{t-1}. \tag{26}$$

The expected values of $\mathbf{x}_t$ are calculated using formulas (17)–(26) recursively. The formulas required to calculate the expected values of the missing elements of $\mathbf{y}_t$ are presented in the next subsection.

3.3. Expectations of incomplete data

3.3.1. General notation for multivariate normal distribution

In order to find the expectations of $\mathbf{y}_t$ for the incomplete series, a reference must be made to the properties of the multivariate normal distribution. The concept is described in Korczyński (2018, pp. 135–137). Under the parametric assumption stating that the process is normal, we can describe it through mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. The process starts from separating the complete and incomplete part of vector $\mathbf{y}_t$ so that the complete part is followed by the missing elements. In effect, the arrangement of elements in the mean vector and variance-covariance matrix takes the form shown in (27), where $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$ represent the known elements for a specific missing data pattern (solid line), while $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}_2$ refer to the missing data elements (dashed line):

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_2 \end{bmatrix}. \tag{27}$$

For example, let us consider the three-element vector $\mathbf{y}_t$ at time $t = 4$ in Figure 3. The observation vector is given by $\mathbf{y}_t = [y_{14} \quad \cdot \quad y_{34}]'$ with the middle element missing. The necessary rearrangement would result in $\mathbf{y}_t^* = [y_{14} \quad y_{34} \quad \cdot]'$, which is equivalent to $\mathbf{y}_t^{(1)} = [y_{14} \quad y_{34}]'$ and $\mathbf{y}_t^{(2)}$ missing,[1] with the respective changes to mean vector and variance-covariance matrix (27): $\boldsymbol{\mu}_1 = [\mu_1 \quad \mu_3]'$, $\boldsymbol{\mu}_2 = \mu_2$, $\boldsymbol{\Sigma}_1 = \begin{bmatrix} \sigma_1^2 & \sigma_{13} \\ \sigma_{13} & \sigma_3^2 \end{bmatrix}'$, $\boldsymbol{\Sigma}_2 = \sigma_2^2$, $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}' = [\sigma_{12} \quad \sigma_{23}]'$. The joint mean vector and variance-covariance matrix for that missing data pattern can be expressed as:

$$\boldsymbol{\mu} = [\, \boldsymbol{\mu}_1 \quad \boldsymbol{\mu}_3 \quad \boldsymbol{\mu}_2 \,]', \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{13} & \sigma_{12} \\ \sigma_{13} & \sigma_3^2 & \sigma_{23} \\ \sigma_{12} & \sigma_{23} & \sigma_2^2 \end{bmatrix}. \tag{28}$$

[1] Notation $\mathbf{y}_t^{(1)}$ is to represent the observed part of vector $\mathbf{y}_t$ at time t. Similarly, $\mathbf{y}_t^{(2)}$ is to represent the missing part. Indices 1 and 2 can refer to various combinations of the three elements of $\mathbf{y}_t$, and therefore it should be noted that y_{1t} (the value of the first element of $\mathbf{y}_t$ at time t) and $\mathbf{y}_t^{(1)}$ represent different notations.

Figure 3. Example of a missing data pattern

Data

t	Y_{1t}	Y_{2t}	Y_{2t}
1	y_{11}	y_{21}	.
2	y_{12}	y_{22}	.
3	y_{13}	.	.
4	y_{14}	.	y_{34}

→

Resulting missing data pattern

s	Y_{1s}	Y_{2s}	Y_{3s}
1	o	o	x
2	o	x	x
3	o	x	o

Source: Korczyński (2018, p. 133).

In the i-th iteration of the EM algorithm the expected value of missing part $\mathbf{y}_t^{(2)}$ is calculated as follows:

$$E\left(\mathbf{y}_t^{(2)}\middle|\mathbf{y}_t^{(1)},\widehat{\boldsymbol{\mu}}^{(i-1)},\widehat{\boldsymbol{\Sigma}}^{(i-1)}\right)=\widehat{\boldsymbol{\alpha}}_{2|1}^{(i)}+\widehat{\boldsymbol{\beta}}_{2|1}^{(i)}\mathbf{y}_t^{(1)}, \tag{29}$$

where:

$$\widehat{\boldsymbol{\alpha}}_{2|1}^{(i)}=\widehat{\boldsymbol{\mu}}_2^{(i-1)}-\widehat{\boldsymbol{\Sigma}}_{21}^{(i-1)}\widehat{\boldsymbol{\Sigma}}_{11}^{-1(i-1)}\widehat{\boldsymbol{\mu}}_2^{(i-1)}, \tag{30}$$

$$\widehat{\boldsymbol{\beta}}_{2|1}^{(i)}=\widehat{\boldsymbol{\Sigma}}_{21}^{(i-1)}\widehat{\boldsymbol{\Sigma}}_{11}^{-1(i-1)}. \tag{31}$$

The variance-covariance matrix of the error term in the model specified by (29) is given by

$$\widehat{\boldsymbol{\Sigma}}_{2|1}^{(i)}=\widehat{\boldsymbol{\Sigma}}_{22}^{(i-1)}-\widehat{\boldsymbol{\Sigma}}_{21}^{(i-1)}\widehat{\boldsymbol{\Sigma}}_{11}^{-1(i-1)}\widehat{\boldsymbol{\Sigma}}_{12}^{(i-1)}. \tag{32}$$

In order to calculate conditional expected values $E\left(y_{jt}^2\middle|w\right)$ and $E\left(y_{jt}y_{kt}\middle|w\right)$ for $j\neq k$, and with w representing the condition, a reference is made to the properties of the variance-covariance matrix (see Härdle & Simar, 2015, pp. 123–125):

$$Cov(\mathbf{x},\mathbf{y})=E(\mathbf{x}\mathbf{y}')-E(\mathbf{x})[E(\mathbf{y})]'. \tag{33}$$

The expectations are conditional on the current parameter estimates and the observed part of vector $\boldsymbol{y}_t$.

We can further adjust (33) to the desired notation using $\mathbf{y}_t^{(1)}$ and $\mathbf{y}_t^{(2)}$:

$$E\left[\left(\mathbf{y}_t^{(2)}|\mathbf{y}_t^{(1)}\right)\left(\mathbf{y}_t^{(2)}|\mathbf{y}_t^{(1)}\right)'\right]=E\left(\mathbf{y}_t^{(2)}|\mathbf{y}_t^{(1)}\right)\left[E\left(\mathbf{y}_t^{(2)}|\mathbf{y}_t^{(1)}\right)\right]'+ Cov\left[\left(\mathbf{y}_t^{(2)}|\mathbf{y}_t^{(1)}\right)\left(\mathbf{y}_t^{(2)}|\mathbf{y}_t^{(1)}\right)\right], \tag{34}$$

where $Cov\left[\left(\mathbf{y}_t^{(2)}|\mathbf{y}_t^{(1)}\right)\left(\mathbf{y}_t^{(2)}|\mathbf{y}_t^{(1)}\right)\right] = \mathbf{\Sigma}_{2|1}$ is the variance-covariance matrix of the error terms in model (29).

With regard to the earlier example presented in Figure 3, we now consider vector $\mathbf{y}_t$ at time t = 3. In this case missing data vector $\mathbf{y}_t^{(2)}$ has two elements representing y_{23} and y_{33}. The expected values of $E(y_{2t}^2|w)$, $E(y_{3t}^2|w)$ and $E(y_{2t}y_{3t}|w)$ can be calculated using formula (34). The respective equation can be expressed as:

$$\begin{bmatrix} E(y_{2t}^2|w) & E(y_{2t}y_{3t}|w) \\ E(y_{2t}y_{3t}|w) & E(y_{3t}^2|w) \end{bmatrix} = \\ = \begin{bmatrix} E(y_{2t}|w)^2 & E(y_{2t}|w)E(y_{3t}|w) \\ E(y_{2t}|w)E(y_{3t}|w) & E(y_{3t}|w)^2 \end{bmatrix} + \mathbf{\Sigma}_{2|1}, \tag{35}$$

where

$$\mathbf{\Sigma}_{2|1} = \begin{bmatrix} \sigma_{y_{2t}|y_{1t}}^2 & \sigma_{y_{2t}y_{3t}|y_{1t}} \\ \sigma_{y_{2t}y_{3t}|y_{1t}} & \sigma_{y_{3t}|y_{1t}}^2 \end{bmatrix}. \tag{36}$$

From the property presented in (34) we can derive well-known equations defining variance and covariance through expected values:

$$D^2X = EX^2 - (EX)^2, \tag{37}$$

and

$$Cov(X,Y) = E(XY) - E(X)E(Y). \tag{38}$$

3.3.2. Adjusting for missing values in an autoregressive model accounting for noise

The calculation of the expected values of $\mathbf{y}_t$ required for the model outlined by (2) and (5) involve the derivation of the mean vector and variance covariance matrix (27) specified for that modelling framework.

For the case of bivariate $\mathbf{y}_t$ and univariate $\mathbf{x}_t$, which is equivalent to the notation given by (2) and (5), the mean vector of $\mathbf{y}_t$ is given by

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} E(y_{1t}) \\ E(y_{2t}) \end{bmatrix} = \begin{bmatrix} E(m_{1t}x_t + v_{1t}) \\ E(m_{2t}x_t + v_{2t}) \end{bmatrix} = \begin{bmatrix} E(m_{1t}x_t) \\ E(m_{2t}x_t) \end{bmatrix} = \begin{bmatrix} m_{1t}\hat{x}_t^n \\ m_{2t}\hat{x}_t^n \end{bmatrix}, \tag{39}$$

for $t = 1, 2, \dots, n$ and based on the fact that the expected value of the error terms in $\mathbf{v}_t$ equals zero (3), and where $\hat{x}_t^n$ (23) is the estimate of x_t^n (14), calculated through the described recursive set of equations. In order to provide the probabilistic part of the distribution specification, i.e. the elements of $\mathbf{\Sigma}$, we should first note that the variance of $\mathbf{y}_t$ is

$$\begin{aligned} \begin{bmatrix} \sigma_1^2 \\ \sigma_2^2 \end{bmatrix} &= \begin{bmatrix} D^2(y_{1t}) \\ D^2(y_{2t}) \end{bmatrix} = \begin{bmatrix} D^2(m_{1t}x_t + v_{1t}) \\ D^2(m_{2t}x_t + v_{2t}) \end{bmatrix} \\ &= \begin{bmatrix} m_{1t}^2 D^2(x_t) + r_1^2 \\ m_{2t}^2 D^2(x_t) + r_2^2 \end{bmatrix} = \begin{bmatrix} m_{1t}^2 \hat{P}_t^n + r_1^2 \\ m_{2t}^2 \hat{P}_t^n + r_2^2 \end{bmatrix}, \end{aligned} \tag{40}$$

for $t = 1, 2, \dots, n$, where r_1^2 and r_2^2 are the variances of error term $\mathbf{v}_t$ in (3). The following equation illustrates the use of the property of variance of the sum of random variables (see for example Jóźwiak & Podgórski, 2006, p. 107):

$$D^2(X + Y) = D^2(X) + D^2(Y) + 2Cov(X, Y), \tag{41}$$

and assuming that two random variables $m_{1t}x_t$ and v_{1t} are independent, which is equivalent to $Cov(m_{1t}x_t, v_{1t}) = 0$. Variance $\hat{P}_t^n$ is calculated using (24).

According to model specification (2), y_{1t} and y_{2t} are strongly linearly correlated. In fact, the two series describe the same process and the differences result from measurement errors. By that, even with a large discrepancy between variances $D^2(y_{1t})$ and $D^2(y_{2t})$, the Pearson correlation coefficient between y_{1t} and y_{2t} is approximately equal to one ($\rho \approx 1$). This leads to the following approximation of the covariance between y_{1t} and y_{2t}:

$$\sigma_{12} = Cov(y_{1t}, y_{2t}) \approx D(y_{1t})D(y_{2t}) = \sigma_1\sigma_2, \tag{42}$$

which comes from a well-known formula for calculating the Pearson correlation coefficient (see for example Jóźwiak & Podgórski, 2006, p. 107):

$$\rho = \frac{Cov(X, Y)}{D(X)D(Y)}. \tag{43}$$

Formulas (40) and (42) provide the following specification of variance-covariance matrix $\mathbf{\Sigma}$ of the joint distribution of $\mathbf{y}_t$ elements in (27):[2]

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}. \tag{44}$$

[2] At this stage $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ are still prior to the rearrangement for the observed and missing parts.

Based on (39) and (44) we can proceed to the description of the steps needed to calculate the expected values of the missing elements of $\mathbf{y}_t$ conditional on the observed data and the current estimates of the mean vector and the variance-covariance matrix at a given iteration of the EM algorithm, denoted as $\boldsymbol{\mu}^{(1i)}$ and $\boldsymbol{\Sigma}^{(i)}$, where i is the iteration number.

We further assume that the observed time series is as shown in Figure 2, i.e. for some parts of the series one of the components can be missing for a number of periods. For the parts of the series where y_{1t} is missing, the mean vector and variance-covariance matrix (27), after undergoing a rearrangement into the observed and incomplete part, take the form of:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_2 \\ \mu_1 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_2^2 & \sigma_{12} \\ \sigma_{12} & \sigma_1^2 \end{bmatrix}. \tag{45}$$

and the expected value of y_{1t} is calculated using (29)

$$E(y_{1t}|w) = E\left(y_{1t}\middle|y_{2t}, \widehat{\boldsymbol{\mu}}^{(i-1)}, \widehat{\boldsymbol{\Sigma}}^{(i-1)}\right) = \widehat{\boldsymbol{\alpha}}_{2|1}^{(i)} + \widehat{\boldsymbol{\beta}}_{2|1}^{(i)}\mathbf{y}_t^{(1)} = \hat{\mu}_1^{(i-1)} + \frac{\sigma_{12}^{(i-1)}}{\sigma_2^{2\,(i-1)}}\left(y_{2t} - \hat{\mu}_2^{(i-1)}\right), \tag{46}$$

where w represents the conditional term in $E\left(y_{1t}\middle|y_{2t}, \widehat{\boldsymbol{\mu}}^{(i-1)}, \widehat{\boldsymbol{\Sigma}}^{(i-1)}\right)$ and is applied here to simplify the further notation.

Expected value $E(y_{1t}^2|w)$ is calculated using (36) and in the considered case the expression is simplified to

$$E(y_{1t}^2|w) = E\left(y_{1t}^2\middle|y_{2t}, \widehat{\boldsymbol{\mu}}^{(i-1)}, \widehat{\boldsymbol{\Sigma}}^{(i-1)}\right) = E(y_{1t}|w)^2. \tag{47}$$

Similarly, for the missing y_{2t} the mean vector and variance-covariance matrix (27) take the following form:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}, \tag{48}$$

and this leads to the expected value of y_{2t}:

$$E(y_{2t}|w) = E\left(y_{2t}\middle|y_{1t}, \widehat{\boldsymbol{\mu}}^{(i-1)}, \widehat{\boldsymbol{\Sigma}}^{(i-1)}\right) = \widehat{\boldsymbol{\alpha}}_{2|1}^{(i)} + \widehat{\boldsymbol{\beta}}_{2|1}^{(i)}\mathbf{y}_t^{(2)}$$
$$= \hat{\mu}_2^{(i-1)} + \frac{\sigma_{12}^{(i-1)}}{\sigma_1^{2\,(i-1)}}\left(y_{1t} - \hat{\mu}_1^{(i-1)}\right). \tag{49}$$

For expected value $E(y_{2t}^2|w)$ we obtain

$$E(y_{2t}^2|w) = E\left(y_{2t}^2\middle|y_{1t}, \widehat{\boldsymbol{\mu}}^{(i-1)}, \widehat{\boldsymbol{\Sigma}}^{(i-1)}\right) = E(y_{2t}|w)^2. \tag{50}$$

The expectations of $\mathbf{x}_t$ determined to smooth the time series, and $\mathbf{y}_t$ to adjust for the missing data described in the preceding subsections are used to calculate the expected log-likelihood function (13), which can be rewritten to

$$E[\ln L\,(\boldsymbol{\theta}|\mathbf{x}, \mathbf{y})] =$$

$$① \quad -\tfrac{1}{2}\ln\sigma_0^2 - \tfrac{1}{2\sigma_0^2}\left[(\hat{x}_0^n)^2 + \hat{P}_0^n - 2\hat{x}_0^n\mu_0 + \mu_0^2\right]$$
$$② -\tfrac{n}{2}\ln q^2 - \tfrac{1}{2q^2}\textstyle\sum_{t=1}^n[(\hat{x}_t^n)^2 + \hat{P}_t^n - 2\phi\hat{x}_t^n\hat{x}_{t-1}^n - 2\phi\hat{P}_{t,t-1}^n + \phi^2(\hat{x}_{t-1}^n)^2 + \phi^2\hat{P}_{t-1}^n] \tag{51}$$
$$③ -\tfrac{n}{2}\ln(r_1^2r_2^2 - r_{12}^2) - \tfrac{1}{2(r_1^2r_2^2 - r_{12}^2)}\, l_1,$$

where l_1 would depend on the missing data pattern. For a complete $\mathbf{y}_t$, we note l_1^{obs}:

$$l_1^{obs} = \textstyle\sum_{t=1}^n[y_{1t}^2 - 2m_{1t}y_{1t}E(x_t) + m_{1t}^2E(x_t^2)]r_2^2$$
$$-2[y_{1t}y_{2t} - m_{2t}y_{1t}E(x_t) - m_{1t}y_{2t}E(x_t) - m_{1t}m_{2t}E(x_t^2)]r_{12} \tag{52}$$
$$+ [y_{2t}^2 - 2m_{2t}y_{2t}E(x_t) + m_{2t}^2E(x_t^2)]r_1^2.$$

For an incomplete $\mathbf{y}_t$, its missing elements in l_1 are replaced by their expected values according to (46)-(47) and (49)-(50).

3.4. Maximisation of the log-likelihood

The EM algorithm involves the maximisation of the log-likelihood function (51) with respect to the parameters of the model given by (2) and (5): μ_0, σ_0^2, ϕ, q^2 and $\mathbf{R}$. The closed-form solutions maximising the log-likelihood can be found for most of the parameters. As we do not assume the error terms in (2) to be uncorrelated, covariance r_{12} leads to the respective log-likelihood being specified as a higher-order polynomial. For the estimation of that parameter, the Newton-Raphson step is proposed as an iterative alternative to direct maximisation.

In order to find both the closed-form and the Newton-Raphson solutions, we first calculate the partial derivatives of the log-likelihood with respect to the model parameters

$$E\left[\frac{\partial \ln L(\boldsymbol{\theta}|\boldsymbol{x},\mathbf{y})}{\partial \theta_j}\right], \tag{53}$$

where $\boldsymbol{\theta} = [\mu_0 \quad \sigma_0^2 \quad \phi \quad q^2 \quad r_1^2 \quad r_2^2 \quad r_{12}]'$. For μ_0, σ_0^2, ϕ, q^2, we can directly maximise the log-likelihood by equating the first derivative of the log-likelihood (53) to zero:

$$E\left[\frac{\partial \ln L(\boldsymbol{\theta}|\boldsymbol{x},\mathbf{y})}{\partial \theta_j}\right] = 0. \tag{54}$$

Therefore, from ① in (51) we follow with the equations below:

$$E\left[\frac{\partial \ln L(\boldsymbol{\theta}|\boldsymbol{x},\mathbf{y})}{\partial \mu_0}\right] = 0, \tag{55}$$

$$E\left[\frac{\partial \ln L(\boldsymbol{\theta}|\boldsymbol{x},\mathbf{y})}{\partial \sigma_0^2}\right] = 0, \tag{56}$$

through which we obtain the following maximum likelihood estimates of μ_0, σ_0^2:

$$\hat{\mu}_0 = \hat{x}_0^n, \tag{57}$$

$$\hat{\sigma}_0^2 = \hat{P}_0^n, \tag{58}$$

where $\hat{x}_0^n$ and $\hat{P}_0^n$ are from (23) and (24).

The two parameters defining the autoregressive process are estimated according to the formula resulting from ② in (51):

$$E\left[\frac{\partial \ln L(\boldsymbol{\theta}|\boldsymbol{x},\mathbf{y})}{\partial \phi}\right] = 0, \tag{59}$$

$$E\left[\frac{\partial \ln L(\boldsymbol{\theta}|\boldsymbol{x},\mathbf{y})}{\partial q^2}\right] = 0. \tag{60}$$

With the use of properties described in (37) and (38), the estimators take the form of:

$$\hat{\phi} = \frac{\sum_{t=1}^{n} \hat{x}_t^n \hat{x}_{t-1}^n + \hat{P}_{t,t-1}^n}{\sum_{t=1}^{n} \left(\hat{x}_{t-1}^n\right)^2 + \hat{P}_{t-1}^n}, \tag{61}$$

$$\hat{q}^2 = \sum_{t=1}^{n} (\hat{x}_t^n - \phi \hat{x}_{t-1}^n)^2 + \hat{P}_t^n - \phi\left(2\hat{P}_{t,t-1}^n + \phi \hat{P}_{t-1}^n\right), \tag{62}$$

where $\hat{x}_t^n$ and $\hat{x}_{t-1}^n$ are from (23), $\hat{P}_t^n$ and $\hat{P}_{t-1}^n$ are given by (24), and $\hat{P}_{t,t-1}^n$ is provided by (25). The variances of error terms r_1^2 and r_2^2 from ③ in (51) can be maximised directly from:

$$E\left[\frac{\partial \ln L(\boldsymbol{\theta}|\boldsymbol{x},\mathbf{y})}{\partial r_1^2}\right] = 0, \quad (63)$$

$$E\left[\frac{\partial \ln L(\boldsymbol{\theta}|\boldsymbol{x},\mathbf{y})}{\partial r_2^2}\right] = 0. \quad (64)$$

First, let us observe that ③ in (51) can be rewritten as:

$$-\frac{n}{2}\ln(r_1^2 r_2^2 - r_{12}^2) - \frac{1}{2(r_1^2 r_2^2 - r_{12}^2)} \sum_{t=1}^{n} a r_2^2 - 2 b r_{12} + c r_1^2, \quad (65)$$

where:
$a = y_{1t}^2 - 2m_{1t}y_{1t}E(x_t) + m_{1t}^2 E(x_t^2),$
$b = y_{1t}y_{2t} - m_{2t}y_{1t}E(x_t) - m_{1t}y_{2t}E(x_t) - m_{1t}m_{2t}E(x_t^2),$
$c = y_{2t}^2 - 2m_{2t}y_{2t}E(x_t) + m_{2t}^2 E(x_t^2).$

After taking the first derivative with respect to the parameter of interest, in this case r_1^2, and multiplying the whole expression by denominator $(r_1^2 r_2^2 - r_{12}^2)^2$, we note that r_1^2 would be cancelled out by the second element of the sum and the resulting maximum likelihood estimator would take the form of

$$\hat{r}_1^2 = -\frac{1}{n r_2^2}\left[n r_{12}^2 + \left(\sum_{t=1}^{n} a r_2^2 - 2 b r_{12} + c \frac{r_{12}^2}{r_2^2}\right)\right]. \quad (66)$$

Similarly, we obtain the following for r_2^2:

$$\hat{r}_2^2 = -\frac{1}{n r_1^2}\left[n r_{12}^2 + \left(\sum_{t=1}^{n} c r_1^2 - 2 b r_{12} + a \frac{r_{12}^2}{r_1^2}\right)\right]. \quad (67)$$

For the estimation of r_{12}, we introduce the Newton-Raphson step, which operates within the EM algorithm according to (10):

$$\hat{\mathrm{r}}_{12}^{i+1} = \hat{\mathrm{r}}_{12}^{i} - \left\{E\left[\frac{\partial^2 \ln L(\boldsymbol{\theta}|\boldsymbol{x},\mathbf{y})}{\partial r_{12}^{(i)} \partial r_{12}^{(i)}}\right]\right\}^{-1} E\left[\frac{\partial \ln L(\boldsymbol{\theta}|\boldsymbol{x},\mathbf{y})}{\partial r_{12}^{(i)}}\right], \quad (68)$$

where

$$E\left[\frac{\partial \ln L(\boldsymbol{\theta}|\boldsymbol{x},\mathbf{y})}{\partial r_{12}^{(i)}}\right] = \frac{n r_{12}}{r_1^2 r_2^2 - r_{12}^2} - \frac{r_{12}}{(r_1^2 r_2^2 - r_{12}^2)^2} \sum_{t=1}^{n} a r_2^2 + c r_1^2 + \frac{1}{r_1^2 r_2^2 - r_{12}^2} \sum_{t=1}^{n} b, \quad (69)$$

and

$$E\left[\frac{\partial^2 \ln L(\boldsymbol{\theta}|\boldsymbol{x},\mathbf{y})}{\partial r_{12}^{(i)}\partial r_{12}^{(i)}}\right] = \frac{n(r_{12}^2+r_1^2 r_2^2)}{(r_1^2 r_2^2-r_{12}^2)^2} - \frac{3r_{12}^2+r_1^2 r_2^2}{(r_1^2 r_2^2-r_{12}^2)^3}\sum_{t=1}^{n} ar_2^2 + cr_1^2 + \frac{2r_{12}(r_{12}^2+3r_1^2 r_2^2)}{(r_1^2 r_2^2-r_{12}^2)^3}\sum_{t=1}^{n} b. \quad (70)$$

The Newton-Raphson step in the EM procedure actually operates as a sub-algorithm, finding the maximum likelihood estimate of r_{12} iteratively, at each run of the EM. It is worth noting that we can apply the use of the Newton-Raphson algorithm to more than one parameter within the EM, which extends the usage of this estimation procedure.

To summarise, the EM algorithm with the Newton-Raphson step operates according to the following scheme (compare with Shumway & Stoffer, 1982, p. 258):

- set the initial values of parameters μ_0, σ_0^2, ϕ, q^2, r_1^2, r_2^2 and r_{12};
- for the current values of the estimates, calculate the expected values of $\mathbf{x}_t$ using $\hat{x}_t^n$ and $\hat{x}_{t-1}^n$from (23), $\hat{P}_t^n$ and $\hat{P}_{t-1}^n$ from (24), and $\hat{P}_{t,t-1}^n$ from (25) and $\mathbf{y}_t$ using (46)–(37) and (49)–(50) for the two missing data patterns, respectively;
- estimate the parameters of the substantive model using formulas $\hat{\mu}_0$, $\hat{\sigma}_0^2$, $\hat{\phi}$, $\hat{q}^2$, $\hat{r}_1^2, \hat{r}_2^2$ and $\hat{r}_{12}$ (57), (58), (61), (62), (66), (67) and (68), noting that (68) for r_{12} is the iterative approach. If the estimates of the parameters are not stabilised, move to step 2, whereas if the convergence has been reached, stop the process. The estimates from the last iteration of the EM algorithm are the maximum likelihood estimates of the model described by (2) and (5).

4. Assessment and application of the EM algorithm with the Newton-Raphson step

4.1. Assessment of the algorithm

The EM algorithm with the Newton-Raphson step (EMNR) as described in the previous section has been applied to the data generated from a distribution imitating the process of autoregression affected by noise, with two observed series measuring the same, unobserved underlying process. The algorithm has been applied using a programme code written in SAS® IML, available upon request.

The sample of n = 200 values was drawn from a population described by two equations – (2) and (5). The parameters of the model from which the sample was drawn are shown in Table 1. The values of the parameters have been selected so as to demonstrate an exemplary autoregressive process with a relatively strong autocorrelation between adjacent elements of time series. The goal was to compare the estimates with the actual values.

Table 1. Parameters of the model used to generate the sample

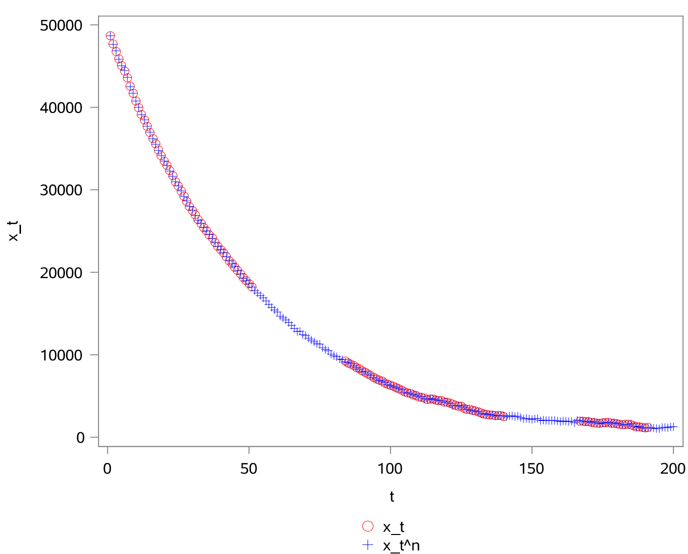
μ_0	σ_0^2	ϕ	q^2	r_1^2	r_2^2	r_{12}
47,000	8,319,328.444	0.98	4,000	4,000	4,000	2,000

Source: author's work.

At first, the initial value of process x_0 was drawn from $N(\mu_0, \sigma_0)$. Subsequently, the autoregressive parameter was set equal to ϕ =0.98. Further, an assumption was made that the observed series $\boldsymbol{y}_t$ for $t = 1, 2, \dots, n$ can differ by the error terms, and the transition factor from the unobserved to observed series is one, which is equivalent to $\mathbf{M}_t = [1 \quad 1]'$. Having specified ϕ, the values of x_t were drawn from the autoregressive process with the w_t disturbance term. Following that and having specified $\mathbf{M}_t$, observed series $\mathbf{y}_t$ were drawn from $\mathbf{R} = \begin{bmatrix} 4000 & 2000 \\ 2000 & 4000 \end{bmatrix}$. The draw of the series was followed by the removal of observations from y_{1t} for t between 51 and 84 and t >191, and for y_{2t} for t between 140 and 167. This was to imitate the situation of losing one source of the signal for a specific time period. Overall, this resulted in $r = 133$ complete observations (i.e. 66.5% of data were complete).

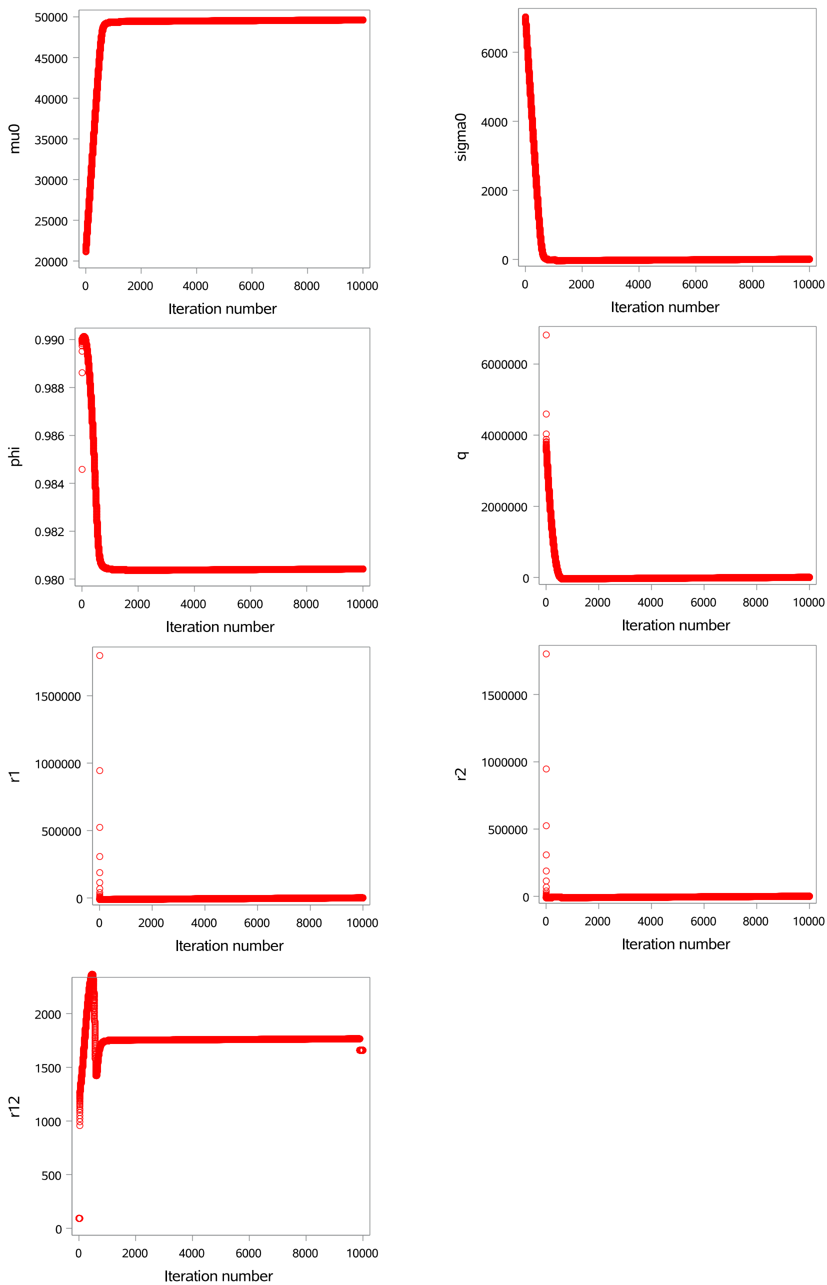
Figure 4 shows the series of the generated x_t values (circles). The breaks in the series represent the time periods with a limited signal, which in this case involved observing only one element – $\mathbf{y}_t$ for some time.

Figure 4. The generated series (circles) and the estimated curve (pluses)



Source: author's work.

Figure 5. Iterations history for the EMNR algorithm



Source: author's work.

The EMNR was applied with the number of iterations equal to 10,000. In order to speed up the processing, parameter r_{12} was set to 0 for the initial 20 iterations, which allowed stabilising at first r_1^2, r_2^2. From the 21st iteration onwards, r_{12} was approximated at each iteration.

The high number of iterations allows observing the behaviour of the algorithm for a longer period. In fact, a far smaller number of iterations was required to reach convergence (see Figure 5). The estimates of x_t for $t = 1, 2, \dots, n$ after smoothing and adjusting for incomplete data are indicated with pluses in Figure 4. By comparing the pluses with the curve showing the actual process, we can see that it was resumed regardless of the missing data which occurred for some periods in both series.

Furthermore, the descriptive statistics for the Absolute Percentage Error (APE) were calculated following an estimation using the EMNR and the standard EM algorithm, in which $r_{12} = 0$. The Mean Absolute Percentage Error (MAPE) was calculated according to the following formula:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{x_t - \hat{x}_t}{x_t}\right|, \tag{71}$$

with the result values multiplied by a 100.

The results are presented in Table 2. We can see that the MAPE expressing the relative deviation from the actual values *ex post* is smaller for the EMNR than for the standard EM algorithm (0.84 vs 0.92) and display a lower standard deviation of APE (1.172 vs 1.283). The prediction error is on average less than 0.9% of the actual values of the underlying process. The forecasts from both algorithms were compared using the Diebold-Mariano (DM) test (1995) with a forecast horizon of three periods, through a one-sided test with the null hypothesis of no difference and an alternative that the EMNR forecast is more accurate than the EM forecast. The DM yielded the test statistic of DM = −3.06 with the p−value = 0.001 and, thus, the null hypothesis was rejected at α = 0.05. The predictions from the EMNR algorithm are significantly more accurate statistically than the ones from the standard EM algorithm. The result is determined by the possibility to relax the assumption for $r_{12} = 0$, which is not required in the EMNR algorithm. The estimates of the standard EM algorithm are presented in the last row of Table 3.

The iteration process of the EMNR is shown in Table 3 and Figure 5, which demonstrate that the algorithm required approximately 1,000 iterations to reach convergence for all the estimated parameters, i.e. μ_0, σ_0^2, ϕ, q^2, r_1^2, r_2^2 and r_{12}. What is more, after reaching convergence the estimates remained stable, which is reflected in the curves representing the estimates at subsequent iterations (Figure 5) and in their comparison with the estimates from the last iterations (Table 3). Autoregressive parameter estimate ϕ is equal to the parameter value. Estimates μ_0, q^2, r_2^2 and r_{12} are close to the true underlying values. A smaller value than expected is obtained for r_1^2. A similar situation is with initial variance σ_0^2, for which it would be the highest deviation from the actual value, however this last parameter is of less importance for the process description. The missing data would be mostly affecting the estimate of the variance of the first error term in (2).

The Newton-Raphson sub-algorithm was set to work in a loop of 50 iterations. The actual convergence was immediate, occurring after only several repetitions of the cycle. An example of sub-iteration for the 10th iteration of the EM algorithm is presented in Figure 6.

Table 2. Descriptive statistics of the APE for the smoothing estimator, and results of the DM test for forecast comparison

Method	n	Mean	Std. Dev.	Minimum	Maximum
EMNR algorithm	200	0.84	1.172	0.0004	6.09
EM algorithm	200	0.92	1.283	0.0012	6.48

Note. A one-sided DM test with an alternative hypothesis stating that EMNR forecasts are more accurate than EM forecasts and a forecast horizon of three periods yielded a test statistic of DM = −3.06 with a p-value = 0.001.

Source: author's calculations. DM test generated using the 'multDM' R package.

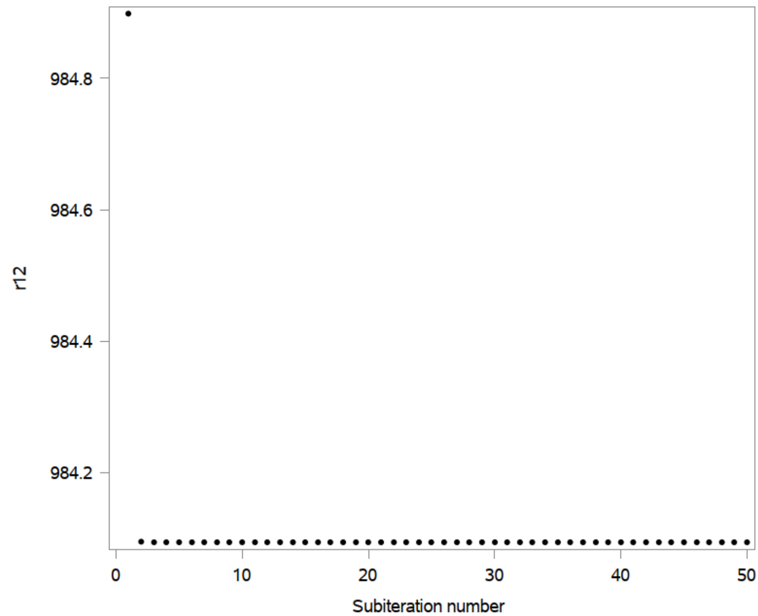
Table 3. Iteration process for the EMNR algorithm

Iteration	μ_0	σ_0^2	ϕ	q^2	r_1^2	r_2^2	r_{12}
1	21,321.20	7,072.78	0.9847	6,863,171.56	1,809,904.50	1,813,848.31	0
2	21,346.27	7,066.46	0.9887	4,633,299.45	958,295.53	961,020.78	0
3	21,384.15	7,056.85	0.9896	4,079,657.86	536,900.08	539,165.70	0
4	21,428.38	7,045.62	0.9898	3,929,662.89	319,062.70	320,625.27	0
5	21,475.19	7,033.73	0.9899	3,871,281.88	197,721.15	198,743.50	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
21	22,271.12	6,831.60	0.9901	3,586,632.42	2,051.60	2,069.82	862.78
22	22,321.24	6,818.88	0.9901	3,572,790.50	2,074.15	2,097.57	902.68
23	22,371.36	6,806.15	0.9901	3,559,551.61	2,106.44	2,133.39	932.70
24	22,421.48	6,793.42	0.9901	3,546,362.07	2,132.28	2,163.73	959.99
25	22,471.60	6,780.69	0.9901	3,533,192.22	2,155.03	2,191.44	984.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
9,991	49,630.05	0.74	0.9804	5,006.55	2,496.28	3,217.49	1,657.64
9,992	49,630.05	0.74	0.9804	5,006.55	2,496.28	3,217.48	1,657.64
9,993	49,630.05	0.74	0.9804	5,006.55	2,496.28	3,217.48	1,657.64
9,994	49,630.05	0.74	0.9804	5,006.55	2,496.28	3,217.48	1,657.64
9,995	49,630.05	0.74	0.9804	5,006.55	2,496.28	3,217.48	1,657.64
9,996	49,630.05	0.74	0.9804	5,006.55	2,496.28	3,217.48	1,657.64
9,997	49,630.06	0.74	0.9804	5,006.55	2,496.28	3,217.48	1,657.64
9,998	49,630.06	0.74	0.9804	5,006.55	2,496.28	3,217.48	1,657.64
9,999	49,630.06	0.74	0.9804	5,006.55	2,496.29	3,217.48	1,657.64
10,000	49,630.06	0.74	0.9804	5,006.55	2,496.29	3,217.48	1,657.64
10,000*	49640.3	0.81	0.9804	6,481.13	1,688.3	2,225.3	0

*The last row includes the estimates from the standard EM algorithm for comparison with the EMNR.

Source: author's calculations.

Figure 6. Sub-iterations history for the Newton-Raphson step within the EM algorithm – 25th iteration of the EMNR algorithm



Source: author's calculations.

4.2. Forecasting demand based on incomplete bivariate time series data

The model given by (2) and (5) has been utilised to analyse the demand for newspapers and to create a forecast for the three months following the last observations to illustrate the application of the EMNR algorithm. The data have been drawn from a website providing the service of monitoring the sales and circulation of the press titles in Poland (Teleskop, n.d.). The analysed dataset consists of two series of the monthly sales and distribution of the printed Polish daily newspaper 'Rzeczpospolita' between January 2016 and May 2018 (Table 4). The sample size equals $n = 29$. Note that for practical applications a larger sample would be expected as the underlying estimation method is the maximum likelihood.

The goal of the example is to demonstrate the application of the EMNR algorithm in the context of a business problem using empirical data. The problem selected for analysis is experienced by newspaper companies and its aim is to predict the correct number of printouts required by particular selling points.

In order to illustrate the functioning of the algorithm, the values in the brackets in Table 4 were considered missing, which, as in sampled data description, is to mimic the break in the reception of one of the signals used to assess the demand in this case.

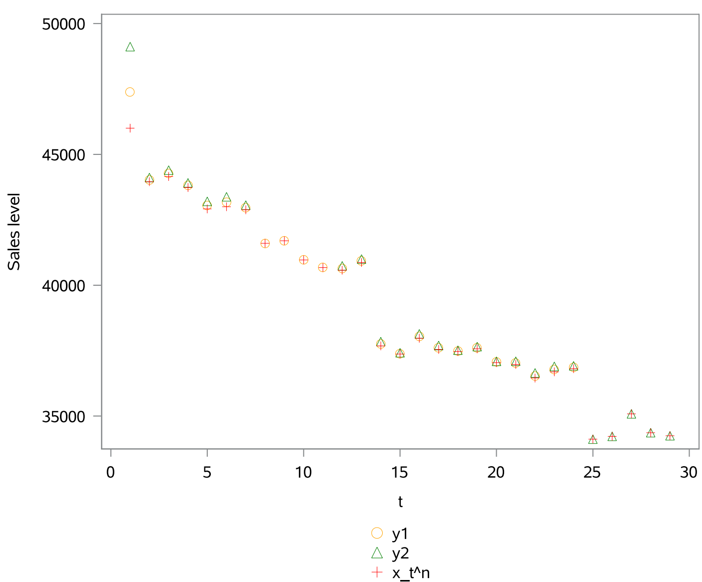
The programme written in SAS® IML and available upon request has been applied to estimate the parameters of the model describing the evolution of the demand for the analysed newspaper. The number of iterations was set to 10,000. Unlike in the case of sampled data, the convergence is slower. The algorithm reached a stable level for the studied parameters close to the end of the predefined iteration number, which might have resulted from the fact that the available sample was very small. The estimates for the early and final iterations of the algorithm are shown in Table 5. The convergence of the Newton-Raphson sub-algorithm was immediate.

Figure 7 presents observed series $\mathbf{y}_t$ along with the estimated smoothed series reflecting underlying process x_t. We can see that the model fit leads to a nearly linear evolution around the shifts captured by the error terms deviations.

Equations (17)–(21) have been used to calculate the forecast for the demand for the 'Rzeczpospolita' newspaper for the three months following the last observation (see Shumway & Stoffer, 1982, p. 262).

Predictions with the standard errors are shown in Table 6. Their values indicate a downward trend observed for the series in the analysis.

Figure 7. Observed series $\mathbf{y}_t$ (circles) and the estimates for $\mathbf{x}_t$ (pluses)



Source: author's calculations.

Table 4. Sales and distribution of the 'Rzeczpospolita' newspaper between January 2016 and May 2018

Time	Sales	Distribution
1	47,389	49,117
2	44,026	44,115
3	44,263	44,398
4	43,812	43,910
5	43,044	43,203
6	43,168	43,379
7	42,961	43,059
8	41,601	(41,694)
9	41,698	(41,813)
10	40,965	(41,071)
11	40,675	(40,780)
12	40,647	40,739
13	40,927	41,000
14	37,753	37,842
15	37,390	37,417
16	38,047	38,139
17	37,612	37,695
18	37,487	37,514
19	37,606	37,652
20	37,064	37,092
21	37,025	37,093
22	36,540	36,648
23	36,782	36,897
24	36,859	36,918
25	(34,080)	34,117
26	(34,191)	34,225
27	(35,058)	35,084
28	(34,297)	34,364
29	(34,156)	34,247

Source: author's calculations based on data from Teleskop (n.d.).

Table 5. Iteration history for the EM algorithm – newspaper dataset

Iteration	μ_0	σ_0^2	ϕ	q^2	r_1^2	r_2^2	r_{12}
1	21,321.2	7,072.78	0.9978	27,289,942.3	2,221,952.9	2,498,100.3	0
2	21,327.8	7,071.02	0.9996	26,186,180.8	1,401,354.4	1,553,840.5	0
3	21,334.8	7,069.17	0.9996	26,164,134.9	904,320.5	1,008,221.3	0
4	21,341.9	7,067.29	0.9995	26,203,873.2	602,394.4	673,174.3	0
5	21,349.1	7,065.41	0.9995	26,224,958.2	413,633.3	463,387.0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
21	21,464.3	7,035.09	0.9995	25,979,283.8	72,433.6	89,516.6	26,224.1
22	21,471.5	7,033.19	0.9996	25,851,652.3	72,421.1	95,489.9	30,575.4
23	21,478.7	7,031.28	0.9996	25,725,415.0	72,710.1	102,398.1	33,939.5
24	21,486.0	7,029.37	0.9996	25,599,466.9	72,555.9	108,888.4	37,129.5
25	21,493.2	7,027.44	0.9997	25,471,929.7	72,184.8	115,224.2	40,116.4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
9,991	46,052.6	111.00	0.9893	767,016.6	61,832.4	311,668.8	138,820.5
9,992	46,052.7	110.98	0.9893	767,014.7	61,832.4	311,668.8	138,820.5
9,993	46,052.7	110.97	0.9893	767,012.9	61,832.4	311,668.8	138,820.5
9,994	46,052.8	110.95	0.9893	767,011.0	61,832.4	311,668.8	138,820.5
9,995	46,052.9	110.93	0.9893	767,009.1	61,832.4	311,668.8	138,820.5
9,996	46,052.9	110.92	0.9893	767,007.2	61,832.4	311,668.8	138,820.5
9,997	46,053.0	110.90	0.9893	767,005.3	61,832.4	311,668.8	138,820.5
9,998	46,053.1	110.89	0.9893	767,003.4	61,832.4	311,668.8	138,820.5
9,999	46,053.1	110.87	0.9893	767,001.5	61,832.4	311,668.8	138,820.5
10,000	46,053.2	110.86	0.9893	766,999.6	61,832.4	311,668.8	138,820.5

Source: author's calculations.

Table 6. Prediction of the demand for the 'Rzeczpospolita' newspaper for June–August 2018

Time	$\hat{x}_{t+1}$	$\sqrt{\hat{P}_{t+1}}$
$t+1$	33,881	875
$t+2$	33,520	1,232
$t+3$	33,162	1,501

Source: author's calculations.

5. Conclusions

Statistical practice is faced with the issue of handling various imperfections resulting from data's nature. Different types of measurement errors need to be modelled and irregularities in the observed data such as missing observations must be taken into account. The Kalman filter is one of the tools that allows modelling noisy time series data. The article focused on exploring the application of the model and the underlying estimation process to situations when empirical data series contain measurement errors and are incomplete. The incompleteness involves situations when one of the sources of the signal is broken for some time, leaving less precise information to estimate the parameters and make predictions. The technique presented in the text was built on the concept described by Shumway & Stoffer (1982), extending the algorithm from the paper to a hybrid version, including the Newton-Raphson sub-algorithm.

The extended version of the algorithm has been verified using sampled data from a model imitating the studied process. The verification showed that the EMNR converged with a relatively small number of iterations and produced stable maximum likelihood estimates. The estimation accounted for incompleteness of the observed data vector, restoring the missing information so that the estimates of the substantive model parameters converged towards the parameters of the data-generating model. The extended EMNR algorithm provided statistically significantly more accurate predictions as compared to the standard EM algorithm. An application to empirical data has also been included jointly with a calculation of the demand for newspapers predicted in subsequent periods closely after the last observation.

The suggested extension of the algorithm could supplement the solution for obtaining maximum likelihood estimates with a better assessment of uncertainty resulting from missing data through multiple imputations within the Bayesian paradigm. The basic scheme for complete data estimation of an autoregressive process can be found in Geweke (2005, Section 7.1). The approach would require a further assessment of the probabilistic characteristics of the estimates, which would lead to a proper assessment of the confidence intervals.

References

Anderson, B. D. O., & Moore, J. B. (1979). *Optimal Filtering*. Hoboken: Prentice-Hall. http://users.cecs.anu.edu.au/~john/papers/BOOK/B02.PDF.

Butland, B. K., Armstrong, B., Atkinson, R. W., Wilkinson, P., Heal, M. R., Doherty, R. M., & Vieno, M. (2013). Measurement error in time-series analysis: a simulation study comparing modelled and monitored data. *BMC Medical Research Methodology*, *13*, 1–12. https://doi.org/10.1186/1471-2288-13-136.

Cajner, T., Crane, L. D., Decker, R. A., Hamins-Puertolas, A., & Kurz, C. (2019). *Improving the accuracy of economic measurement with multiple data sources: The case of payroll employment data* (NBER Working Paper No. 26033). https://doi.org/10.3386/w26033.

Carey, R. N., Jani, C., Johnson, C., Pearce, J., Hui-Ng, P., & Lacson, E. (2016). Chemistry Testing on Plasma Versus Serum Samples in Dialysis Patients: Clinical and Quality Improvement Implications. *Clinical Journal of the American Society of Nephrology*, *11*(9), 1675–1679. https://doi.org/10.2215/CJN.09310915.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x.

Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, *13*(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599.

Fukuda, K. (2005). Forecasting economic time series with measurement error. *Applied Economics Letters*, *12*(15), 923–927. https://doi.org/10.1080/13504850500119161.

Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*. Hoboken: John Wiley & Sons.

Ghassemi, M., Pimentel, M. A. F., Naumann, T., Brennan, T., Clifton, D. A., Szolovits, P., & Feng, M. (2015). *A multivariate timeseries modeling approach to severity of illness assessment and forecasting in ICU with sparse, heterogeneous clinical data*. Proceedings of the AAAI Conference on Artificial Intelligence, Palo Alto. http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9393.

Haïrdle, W. K., & Simar, L. (2015). *Applied Multivariate Statistical Analysis* (4th edition). Berlin, Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-662-45171-7.

Jóźwiak, J., & Podgórski, J. (2009). *Statystyka od podstaw* (6th edition). Warszawa: Polskie Wydawnictwo Ekonomiczne.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, *82*(1), 35–45. https://doi.org/10.1115/1.3662552.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press. https://doi.org/10.1017/CBO9780511754098.

Korczyński, A. (2018). *Screening wariancji jako narzędzie wykrywania zmowy cenowej. Istota i znaczenie imputacji danych*. Warszawa: Oficyna Wydawnicza SGH.

Laaksonen, S. (2018). *Survey Methodology and Missing Data: Tools and Techniques for Practitioners*. Cham: Springer. https://doi.org/10.1007/978-3-319-79011-4.

Lange, K. (1995). A quasi-Newton acceleration of the EM algorithm. *Statistica Sinica*, *5*(1), 1–18. http://www3.stat.sinica.edu.tw/statistica/oldpdf/A5n11.pdf.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd edition). Hoboken: John Wiley & Sons. https://doi.org/10.1002/9781119013563.

Mittelhammer, R. (2013). *Mathematical Statistics for Economics and Business* (2nd edition). New York: Springer. https://doi.org/10.1007/978-1-4614-5022-1.

Molenberghs, G., & Kenward, M. G. (2007). *Missing Data in Clinical Studies*. Chichester: John Wiley & Sons. https://doi.org/10.1002/9780470510445.

Shumway, R. H., & Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, *3*(4), 253–264. https://doi.org/10.1111/j.1467-9892.1982.tb00349.x.

Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples* (4th edition). Cham: Springer. https://doi.org/10.1007/978-3-319-52452-8.

Teleskop. (n.d.). *Polskie badania czytelnictwa*. Retrieved July 14, 2018, from https://www.teleskop.org.pl/zkdp/index.jsp?p=news.

Przegląd Statystyczny. Statistical Review, 2021, vol. 68, 1, 47–74 DOI: 10.5604/01.3001.0015.0377
Received: 6.04.2021, revised: 8.06.2021, accepted: 24.06.2021

The influence of the demographic structure on the economic growth of Ukraine

Nazarii Kukhar[a]

Abstract. The national economy is closely related to the demographic structure of the society. Therefore, in the face of demographic changes, it is necessary to assess the influence of these changes on economic growth. This article presents an estimation of the impact that the future changes in the demographic structure will have on the economic growth of Ukraine, represented by the rate of changes in GDP *per capita*. The decomposition of GDP *per capita* and making the components of this decomposition dependent on the demographic structure allowed an empirical analysis, which used a variety of econometric and statistical techniques and was based on a population forecast prepared by the Ptoukha Institute for Demography and Social Studies of the National Academy of Sciences of Ukraine. As a result, it was determined that the impact of the changes in the demographic structure on Ukraine's long-term economic growth will be highly diverse over the studied period (until 2060). However, considering the entire period of the analysis, the negative effects of the changes in the demographic structure on the economy will be counterbalanced by the positive effects of these changes.

Keywords: GDP *per capita*, demographic determinants of growth, demographic structure, Ukraine

JEL: C53, E17, J11

1. Introduction

The population, its prosperity and social development is the defining feature of a society and constitutes the basis of a state's power. Therefore, the demographic situation has always remained the object of attention of a wide range of stakeholders, including the scientific community. Scientific disciplines are interested in a variety of issues relating to demographics, particularly its impact on different economic and social aspects.

Economists argue that economic conditions affect only the flows of demographic processes (e.g. fertility, life expectancy), which remains insignificant in relation to the demographic structure in the short term. On the other hand, the impact of the demographic conditions on economic processes is immediate and significant (Wasilewska & Pietrych, 2018). There are many channels through which demographic changes affect the economy, but the following are most often distinguished: investments, the labour market and productivity.

In the macroeconomic perspective, demographic issues are studied in the context of economic growth. Demographic conditions have a direct impact on real GDP as they determine the size of the labour force (Jong-Won et al., 2014), which, in turn, is

[a] Jagiellonian University in Kraków, ul. Gołębia 24, 31-007 Kraków, Poland,
e-mail: nazarii.kukhar@doctoral.uj.edu.pl, ORCID: https://orcid.org/0000-0003-0329-9215.

the central factor of production. Thus, the increase in the labour force resulting from the growth of the working age population, directly contributes to the increase in production (Florczak & Przybyliński, 2016). Conversely, a declining population entails a slower output growth, unless it is compensated by an increase in labour productivity. When the productivity level is constant, GDP drops proportionally to the number of workers (Coleman & Rowthorn, 2011).

However, the research results presented in the literature show that the influence of demographic conditions on economic growth is ambiguous. Depending on the economic and social characteristics, and on the development level of individual countries, demographic changes may either accelerate or slow down economic growth, although their impact may as well be neutral (Mączyńska, 2010).

Despite the expected severe economic consequences, demographic change usually remains outside the scope of many macroeconomic policy discussions and debates. For example, most economic growth models assume that population grows at a steady pace, and many business cycle models set the size of the population at a fixed level when analysing aggregate demand (Jong-Won et al., 2014).

Demographic changes are one of the most significant long-term challenges that will have a profound impact on the economy of European countries. The working age population is declining rapidly and this process is expected to continue in the future. The *per capita* income of most Central and Eastern European countries is still below that of Western European countries, putting the former at risk of 'aging before becoming rich' (Batog et al., 2019).

Ukraine is more vulnerable to the negative effects of aging and population decline than many other regions in Europe. A constant downward trend has been observed in this respect ever since Ukraine regained independence. Along with the decline of the whole population of the country, the working age population is also decreasing. Moreover, taking into account the increasing pace of aging, not only an absolute, but also a relative decrease in the number of working population in the overall structure of the Ukrainian population is noted (Kudlak, 2019).

Ukraine is still going through economic, political, and social transformations caused by the collapse of the Soviet Union. In this situation, an aging society poses additional challenges (e.g. maintaining and improving living standards and productivity), which are particularly dangerous, as Ukraine (and other countries in a similar situation) already struggles with underdeveloped institutions, structural weaknesses, a vulnerable economy, and political instability. In addition, compared to higher-income countries, Ukraine has access to fewer financial resources essential to mitigating the negative effects of an aging society (Kupets, 2014). The situation on the Ukrainian labour market is also difficult – the ratio between labour demand and supply fell in all groups of occupations, particularly among those that do not require special training or highly-qualified workers. The most pressing problem is the imbalance between the qualifications and work experience required by employers and the skills and professional experience of those seeking jobs (Stoychik, 2018). Along with this, regional labour market indicators tend to deteriorate. There are regions which are characterised by a critically high number of registered unemployed per one vacancy (Zoidze, 2013). Moreover, almost one-third of the total employed population and more than 25% of those who are actively seeking work are concentrated in five regions (out of 25) (Cymbal & Iarosh, 2020).

The aging of the population in Ukraine is mainly the effect of a decline in fertility and the continuing emigration of the working age population, not due to an enormous increase in life expectancy as observed in developed countries. Considering the number of emigrants from Ukraine, Poland has traditionally taken the leading position in the recent years (Kudlak, 2019). Thus, the results of research on the impact of the changes in demographic conditions on economic processes in Ukraine could also be interesting for Poland.

There are numerous publications in the literature describing the influence of demographic factors on the economic development of various countries. However, it should be noted that in relation to Ukraine, these publications are in most cases limited to the analysis of the general demographic trends and problems, and include only theoretical considerations of the effect demographic conditions have on the economy (Aksonova, 2012; Geyets, 2011; Romanukha, 2016), while only a few studies provide an empirical assessment of this impact.

A study published by the World Bank entitled 'Golden Aging' describes the current demographic situation in Europe and Central Asia (including Ukraine). The expected demographic conditions which are based on a population forecast prepared by the United Nations (UN) are also presented. Considering this forecast, the authors indicate the increasing aging process of the population in the aforementioned regions and, by using the dynamic overlapping generations (OLG) model, determine how this process affects individual countries (Bussolo et al., 2015). The findings are presented below:

- according to the report, due to the aging of the societies of Eastern European countries, a workforce reduction amounting to a total of nearly 30 million people should be expected by 2050;
- the aging of the society will not necessarily reduce the *per capita* income. The authors indicate that even if an increasing dependency ratio is observed, GDP *per capita* may also increase;
- the average age of the workforce in these regions will increase, and yet the average labour productivity will not necessarily decrease;
- as the life expectancy in these regions increases, so will the savings rate.

In a publication entitled 'Demographic Headwinds in Central and Eastern Europe' of the International Monetary Fund (IMF), as well as in a study prepared by the World Bank, the authors look at the demographic pressure in the countries of Central, Eastern and North-Eeastern Europe (including Ukraine) and make an attempt to determine the macroeconomic implications of these demographic trends. According to the authors, demographic changes will trigger the following effects (Batog et al., 2019):

- most countries will experience a significant decline in the workforce by 2050, especially in Ukraine, where this decline will exceed 30% (compared to 2015);
- the aging of the population is likely to entail an increase in expenditure on health care and pensions, in most cases by more than 5% of GDP, and with regard to Ukraine, this increase will reach approximately 10% of GDP (compared to 2015);
- the increase in the average age of the labour force will be related to a decrease in total labour productivity and total factor productivity. In Ukraine the average annual decline in total factor productivity will reach up to 0.6% in the period of 2020–2050;
- the demographic changes will result in an average slowdown of the economic growth by 0.6 percentage points per year in 2020–2050. Ukraine will experience the impact of these changes slightly stronger, as they are likely to cause an annual average decline in GDP *per capita* of about 0.8%.

Kupets (2014) reviews the general demographic trends in Ukraine and analyses the impact of demographic changes on the labour market, particularly on the supply of the labour force and labour productivity. Based on the forecast of the population prepared by the UN, the author argues that, even considering a very optimistic scenario of demographic development, the Ukrainian economy is bound to face a significant reduction in labour force by 2060. However, Kupets also claims that the introduction of substantial improvements in the area of labour and technology productivity is likely to mitigate the negative effects of the demographic changes in the Ukrainian economy. The author argues that if the rate of the increase in labour productivity remains at the level of 0.5% per year, the annual rate of change in GDP *per capita* will be unaffected by the changes in demographic conditions.

Rovný et al. (2021) used Pearson's correlation to examine the strength of relationships between selected economic and demographic indicators. The results of this analysis demonstrate that the *per capita* income is quite strongly correlated with the labour force aged 35–54. On the other hand, a strong positive correlation was found between unemployment and the population aged 0–14 and the elder workforce aged 55–65. Moreover, a very strong negative correlation was also observed between unemployment and the young working population aged 15–34.

In her work, Maksymenko (2009) identifies the mutual channels of influence between fertility and the economy. The author used a vector autoregressive model to determine the short- and long-term responses of fertility to the unemployment rate, the households' level of available financial resources, and economic growth. Maksymenko's research shows that changes in the unemployment rate and the amount of the households' available means explain more than one-third of the variability in fertility.

The current studies, in particular those conducted by the World Bank and the IMF provide a solid analysis of the influence that demographic changes have on economic growth; however, the studies fail to take into account all the main channels of this impact (such as investment or human capital). In result, there is a considerable research gap in terms of the empirical assessment of the impact of demographic conditions on the economy of Ukraine.

Thus, the aim of this work is to present an empirical analysis of the impact of changes in the demographic structure on Ukraine's long-term economic growth in the years 2017–2060 by means of selected econometric techniques (including general equilibrium models, shift-share methods, as well as statistical and econometric methods) suggested by Florczak (2017). The empirical results presented in this article provide information about the scale of the potential long-term changes in the dynamics of GDP *per capita* caused by changes in the demographic structure of the Ukrainian population. The empirical research is based on the medium variant of Ukraine's demographic development forecast until 2060, prepared by the Ptoukha Institute for Demography and Social Studies of the National Academy of Sciences of Ukraine (2014), referred to as IDBSNAN. In the process of constructing the population forecast, such components as fertility, life expectancy and net migration were taken into account.

2. Influence of the demographic structure of Ukraine's society on economic growth

The starting point for the empirical analysis is the decomposition of the influence of the demographic structure on GDP *per capita* (Landmann, 2004; see also Florczak, 2008a):

$$GDPC_t \equiv GDPH_t \cdot AWH_t \cdot RELF_t \cdot RWAP_t, \tag{1}$$

where

$GDPC_t$ = GDP/N – GDP *per capita*, where N is the population;

$GDPH_t$ = GDP / working hours – productivity per hour;

AWH_t = working hours / number of employed persons – number of working hours per one employee;

$RELF_t$ = number of employed persons / number of working age persons – effective employment ratio;
$RWAP_t$ = number of working age persons / total population – population age structure.

By transforming equation (1) from the form reflecting the levels of all components of a given equation to the form showing changes in the growth rate of all components of equation (1), we receive the following formula:

$$\dot{GDPC}_t \equiv \dot{GDPH}_t + \dot{AWH}_t + \dot{RELF}_t + \dot{RWAP}_t, \quad (2)$$

where the dot above the component symbol indicates the growth rate ($\dot{a}_t = \frac{a_t - a_{t-1}}{a_{t-1}}$).

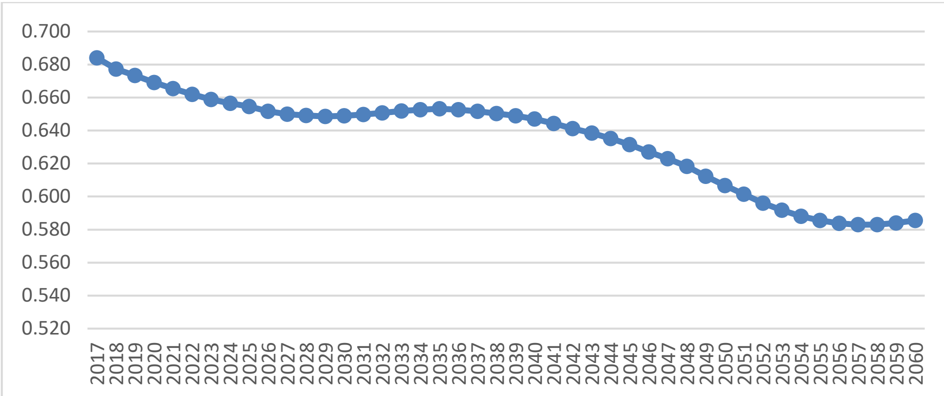
Equation (2) shows that economic growth, approximated by the GDP *per capita* growth rate, is defined by both economic and demographic conditions. Moreover, the strength of the impact of the growth rate changes of all explanatory variables in equation (2) on the dependent variable is the same and amounts to one to one.

Due to the fact that the aim of the paper is to estimate the influence of only the demographic factors on long-term economic growth, and because variable AWH_t is not directly dependent on demographic conditions, the value of AWH_t was fixed at the level noted in 2017 (Florczak et al., 2018). The subsequent parts of the article present the development of all the components of equation (2).

2.1. Development of the age structure and the share of employed persons in the working age population in the years 2017–2060

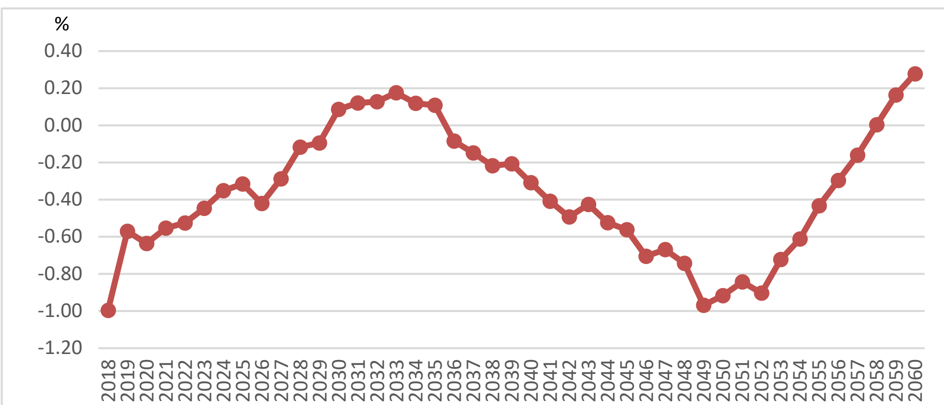
Figure 1 presents the values of the age structure during the analysed period, while Figure 2 demonstrates the annual growth rate for this variable. Due to the fall in the value of the age structure coefficients, a decrease in the level of GDP *per capita* should also be expected in the analysed years. The age structure coefficient will decrease by a seventh, from a level slightly exceeding 0.68 to somewhat below 0.59 within 2017–2060, which corresponds to the average annual decrease in GDP *per capita* in this period by nearly 0.36 percentage points.

Figure 1. Age structure coefficient (RWAP)



Source: author's calculations based on State Statistics Service of Ukraine (SSSU) data (number of working age persons, total population) and IDBSNAN forecasts (number of working age persons, total population).

Figure 2. Annual growth rates of the RWAP

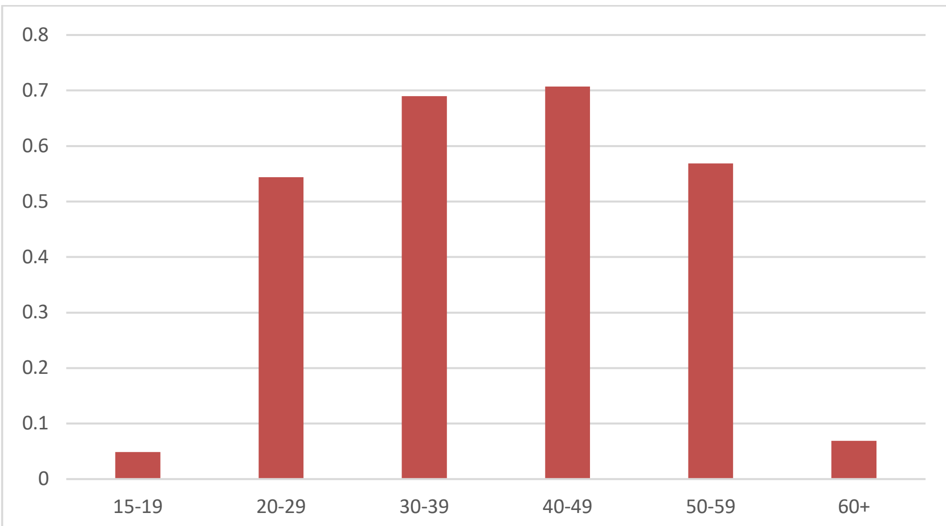


Source: author's calculations based on SSSU data (number of working age persons, total population) and IDBSNAN forecasts (number of working age persons, total population).

As Figures 1 and 2 suggest, the development of RWAP is unevenly distributed over time. Until the beginning of the 2030s, a declining negative impact of this factor is observed, whereas in the years 2030–2035 this influence will be even positive with an average annual increase of nearly 0.12%. However, starting from the mid-2030s until the early 2050s, the increasingly unfavourable impact of this component on economic growth should be taken into account. In the last decade of the analysed period, this negative effect should lessen and changes in the age structure coefficient will contribute to an increase in the level of GDP *per capita* at the end of the studied period.

The next component of equation (2) – the effective employment ratio coefficient – is assumed to be constant for individual age groups: 15–19, 20–29, 30–39, 40–49, 50–59, and 60+, i.e. fixed at the level registered in 2017 (see Figure 3). This division is also related to the explanations of the impact of the age structure of employees on the level of total factor productivity presented in the subsequent parts of this article.

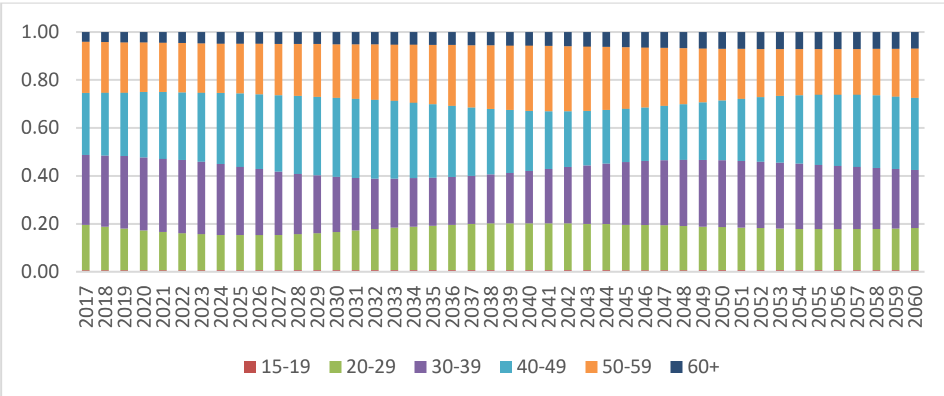
Figure 3. Coefficients of effective employment in Ukraine in 2017 and in the subsequent years of the analysis (by age groups)



Source: author's calculations based on SSSU data (total population, number of working age persons) and the International Labour Organization (ILO) data (number of employed persons) (ILO, n.d.).

Although the effective employment coefficients are constant for individual age groups, it will not be so in the case of the aggregate indicator due to the changes in the demographic structure of the population as seen in Figure 4. Figures 5 and 6 present the development of the effective employment rate in the analysed period.

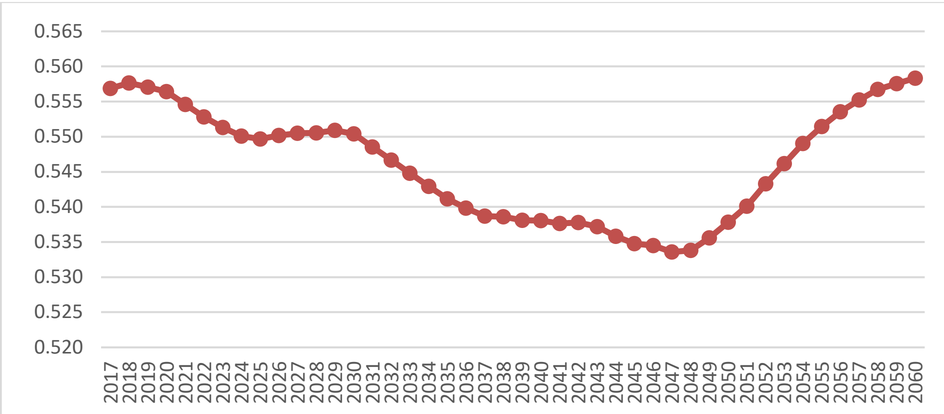
Figure 4. Share of employees by age groups in the total number of employees resulting from the changes in the demographic structure of the population of Ukraine



Source: author's calculations based on SSSU data (number of working age persons, total population) and the IDBSNAN forecast (number of working age persons, total population) and assumptions made regarding the effective employment coefficients in individual age groups.

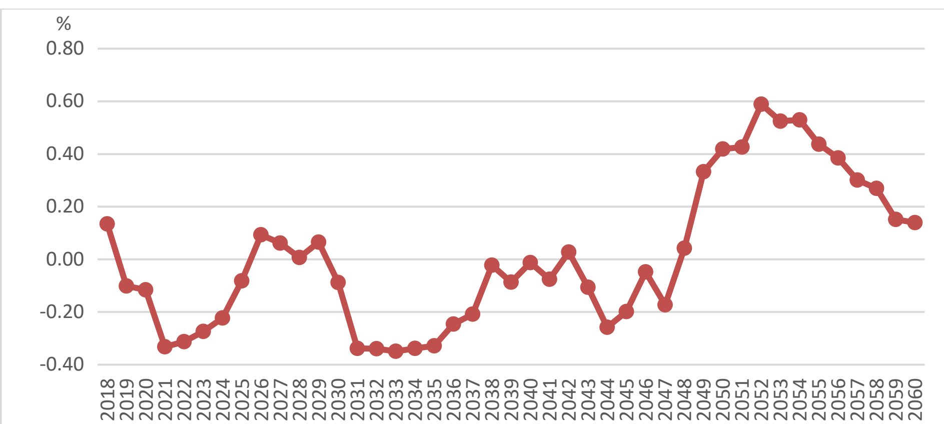
The analysis proves that changes in the effective employment rate caused by changes in the demographic structure of the population will have a diverse impact on the rate of change in GDP *per capita* at different times. Until the first half of the 2020s, this influence will be negative, while in the second half of the 2020s it will be negligible – the average annual rate of change will remain below 0.03%. From the early 2030s until 2047, a negative impact of changes in the effective employment ratio will be observed, whereas starting from 2048, only a positive impact will be recorded, contributing to an increase in GDP *per capita* of more than 0.5% in the years 2052–2053. However, at the end of the analysed period, this positive influence is expected to decrease. When comparing the results of the effective employment ratio and the age structure coefficient, it should be noted that throughout the majority of the analysed period, the direction of the impact of these variables on GDP *per capita* growth remains the same and unfavourable in most cases, whereas in the periods witnessing a different direction of the impact, the strength of the overall effect will also be negative.

Figure 5. Effective employment rates in Ukraine



Source: author's calculations based on SSSU data (number of working age persons, total population) and the IDBSNAN forecast (number of working age persons, total population) and assumptions made regarding the effective employment coefficients in individual age groups.

Figure 6. Annual changes in the growth rate of effective employment rates caused by the demographic structure of the population of Ukraine



Source: author's calculations based on SSSU data (number of working age persons, total population) and the IDBSNAN forecast (number of working age persons, total population) and assumptions made regarding the effective employment coefficients in individual age groups.

2.2 Demographic structure and labour productivity in the years 2017–2060

Labour productivity is central to long-term economic growth and largely dependent on the demographic structure. This component is described using the labour productivity model obtained through the transformation of the Cobb-Douglas production function with constant returns to scale (Feyrer, 2007):

$$Y_t = K_t^a \cdot (TFP_t \cdot N_t \cdot h_t)^{1-a}, \tag{3}$$

where
Y_t – GDP in period t;
K_t – fixed assets in period t;
TFP_t – total factor productivity in period t;
N_t – number of employees in period t;
h_t – human capital per employee in period t;
a – productivity of fixed assets (many researchers assume that the value of this component equals 0.3) (Florczak et al., 2018).

Equation (3), after undergoing a two-sided division by the number of employees, takes the following form of the labour productivity function:

$$y_t = k_t^a \cdot (TFP_t \cdot h_t)^{1-a}, \tag{4}$$

where
y_t – labour productivity in period t;
k_t – fixed assets per employee in period t – technical equipment of work.

The technical equipment of work in function (4) can be converted into capital intensity, which will allow the identification of all the channels of the influence of the demographic conditions on labour productivity. Thus, the productivity function will take the following form:

$$y_t = \left(\frac{K}{Y}\right)_t^{\frac{a}{1-a}} \cdot TFP_t \cdot h_t, \tag{5}$$

where
$\frac{K}{Y}$ – capital intensity in period t.

Thus, the labour productivity function encompasses three factors: human capital, total factor productivity, and technical equipment of work or capital intensity. The impact of human capital and total factor productivity on labour productivity, and thus on economic growth, is the same and equals one to one. Therefore, a change of either of these two factors by 1% will also change the level of GDP *per capita* by 1%. On the other hand, a change in the technical equipment of work by one percentage point will contribute to a change in GDP *per capita* by a percentage point. In the case of capital intensity, however, a change in this component by 1% will result in a change in GDP *per capita* by $a/(1-a)$ percentage points.

2.2.1. Human capital and labour productivity

Human capital is understood as all psychophysical features of an individual, such as their abilities and skills, education, knowledge, work experience, health condition, cultural level, etc., which have an impact on the productivity of work and which are inseparably connected with the human being as the holder of these values (Florczak, 2008b). However, when determining the level of human capital, especially in the macro scale, it is not possible to include all the above-mentioned features. For this reason, in most cases empirical research takes into account only a few of them, usually the level of education, work experience, and health conditions (Wößmann, 2003).

There are several methods of quantifying the level of human capital, the most common of which is based on the extended Mincer equation. This particular method quantifies human capital taking into account its main components, i.e. the level of education, work experience, and health conditions. It takes the form of the following equation (Florczak, 2017):

$$HC_t = [(1.433329 \cdot EM_t \cdot LEBM_t + EF_t \cdot LEBF_t) \cdot HCE_t \cdot EE_t]/TE_t, \quad (6)$$

where

HC_t – human capital in period t;
EM_t – number of employed men in period t;
EF_t – number of employed women in period t;
$LEBM_t$ – male life expectancy at birth in period t;
$LEBF_t$ – female life expectancy at birth in period t;
TE_t – total number of employees in period t;
HCE_t – human capital per employee, taking into account education in period t:

$$HCE_t = \frac{1.055274 \cdot EHE_t + ESE_t + EBE_t}{TE_t}, \quad (7)$$

where

EHE_t – number of the employed with higher education in period t;
ESE_t – number of the employed with secondary education in period t;
EBE_t – number of the employed with basic education in period t;
EE_t – work experience sub-index in period t:

$$EE_t = \sum_{i=15}^{69} \left[\frac{N_{it}}{TE_t} \cdot \frac{\exp(0.0228 \cdot E_i - 0.000565 \cdot E_i^2)}{\exp(0.0228 \cdot E_{15} - 0.000565 \cdot E_{15}^2)} \right], \tag{8}$$

where

N_{it} – number of the employed at the age of i and in time period t,

E_i – work experience of the employee in the age of i (calculated as follows: *employee experience = employee age – number of years of education – 5*).

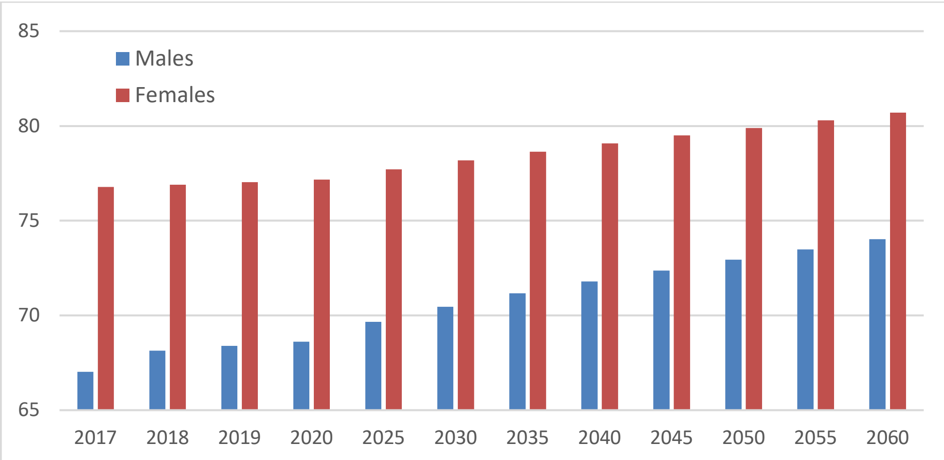
The individual weights used in equations (6), (7) and (8) should be estimated using a representative microeconomic sample based on survey data. Since only a few studies meet this requirement, the relevant estimates were drawn from the results of a study conducted by Vakhitova & Coupe (2013), which is the most recent one in this area.

Thus, to determine the level of human capital in the analysed period, it is necessary to have information on the following: life expectancy at birth by gender (see Figure 7), the employees' education structure (this structure was fixed at the level observed in 2017), the number of years of work experience of a working person (assumed to be fixed at the level registered in 2017 for a particular age of the employees), the share of the employed men and women in the total number of employed persons (fixed at the level observed in 2017), the share of employed persons by individual age groups in the total number of employed persons (for the results of the analysis on the effective employment rate, see Figure 4).[1]

Since the stability of the structure of education was assumed in the analysed period, the HCE sub-index (formula (7)) remains unchanged, and fluctuations of the level of human capital are caused by changes in the health condition of the society (the expected life expectancy of the newly-born) and changes in the sub-index of work experience. The development of the EE sub-index, normalised according to the 2017 level, is presented in Figure 8, while the levels of human capital, also normalised according to the 2017 level are presented in Figure 9. Figures 10 and 11 show the pace of changes of these indicators and Figure 11 presents the impact of the changes in human capital on economic growth.

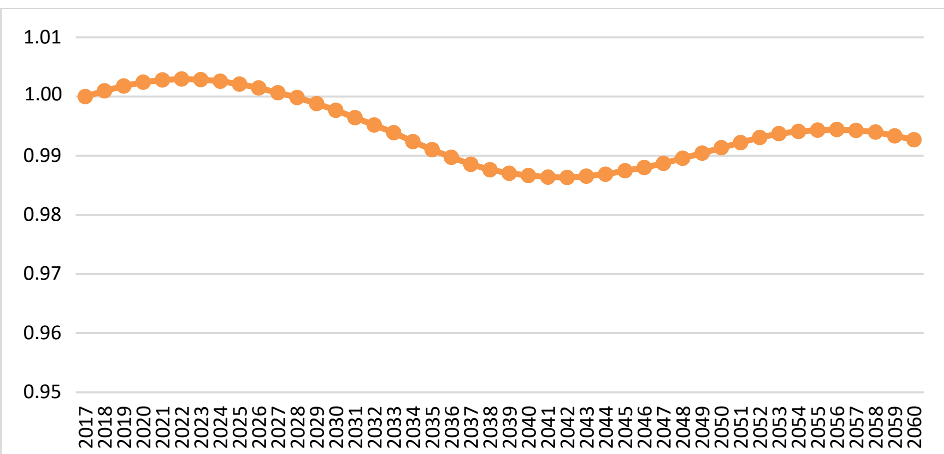
[1] Setting some values at a constant level is necessary to meet the *ceteris paribus* principle with regard to the impact of demographic changes on economic growth.

Figure 7. Life expectancy of men and women in Ukraine



Source: author's work based on IDBSNAN forecasts (life expectancy at birth).

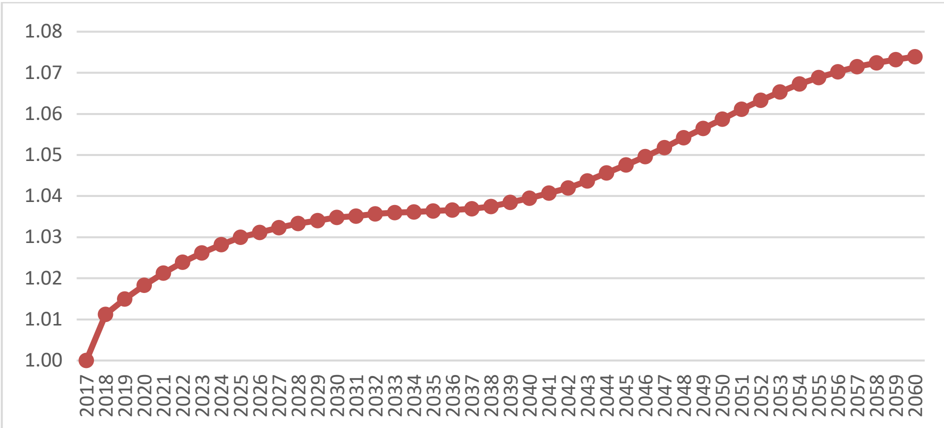
Figure 8. Levels of the work experience sub-index



Source: author's calculations based on the SSSU data (number of the employed persons, education, number of working age persons), ILO data (number of employed persons, education) (ILO, n.d.) and the IDBSNAN forecast (number of working age persons).

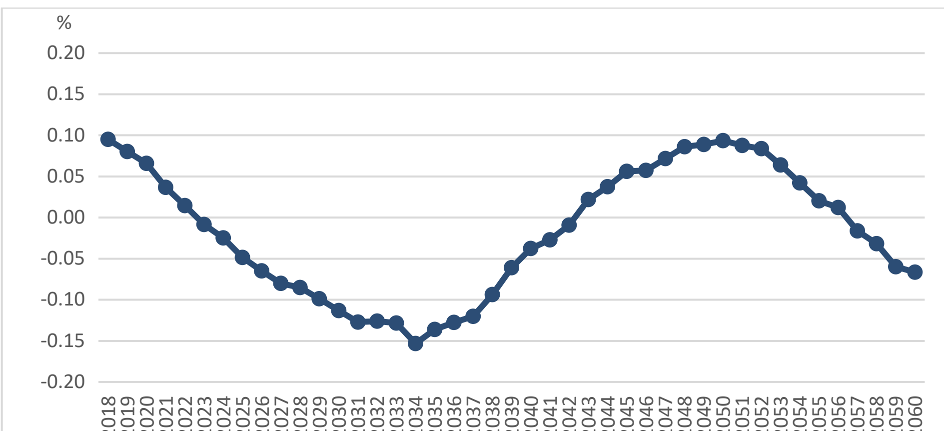
Figures 8 and 9 demonstrate that by 2060 long-term demographic changes in Ukraine will contribute to an increase in the level of human capital by almost one-eighth compared to 2017, which corresponds to its average annual growth of nearly 0.17%. On the other hand, the work experience sub-index will slightly change due to the occurring demographic changes.

Figure 9. Levels of human capital



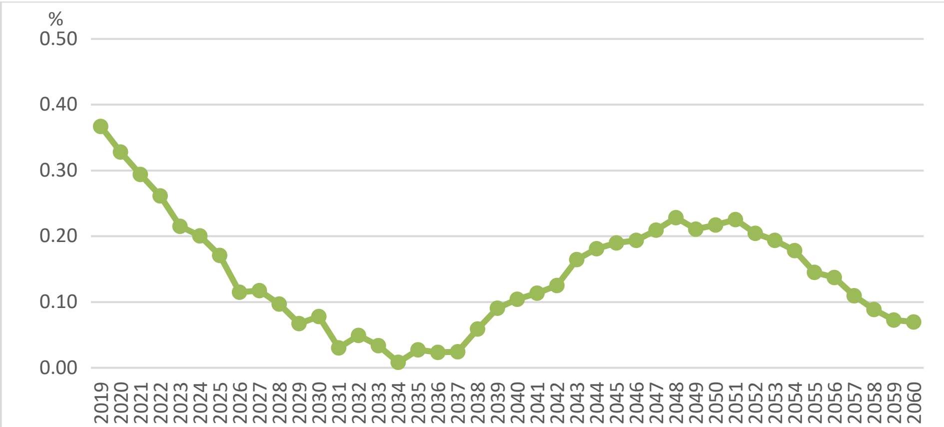
Source: author's calculations based on the SSSU data (number of the employed persons, education, number of working age persons, life expectancy at birth), ILO data (number of employed persons, education) (ILO, n.d.) and the IDBSNAN forecast (number of working age persons, life expectancy at birth).

Figure 10. Changes in the growth rate of the work experience sub-index



Source: author's calculations based on the SSSU data (number of employed persons, education, number of working age persons), ILO data (number of employed persons, education) (ILO, n.d.) and the IDBSNAN forecast (number of working age persons).

Figure 11. Changes in the growth rate of human capital



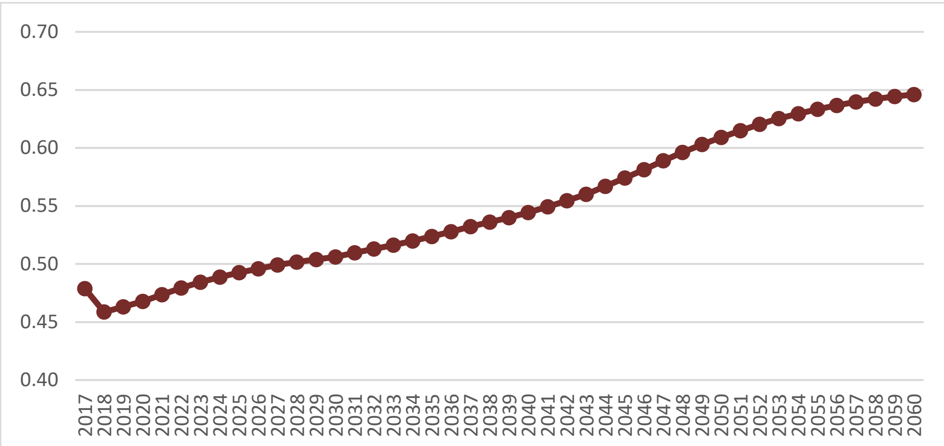
Source: author's calculations based on the SSSU data (number of the employed persons, education, number of working age persons, life expectancy at birth), ILO data (number of employed persons, education) (ILO, n.d.) and the IDBSNAN forecast (number of working age persons, life expectancy at birth).

As Figures 10 and 11 suggest, changes in the values of human capital are to a greater extent caused by changes in the health condition of the Ukrainian society, while the impact of changes in the level of the work experience sub-index is negligible. Moreover, the impact of changes in human capital (resulting from changes in the demographic conditions in Ukraine) on the growth rate of GDP *per capita* in the analysed period is positive, although the impact is not evenly distributed over time. Until the mid-2030s, the growth of human capital will slow down, but from the second half of the 2030s it will gradually accelerate until the end of the 2040s, when another slowdown is expected to occur and remain until the end of the studied period.

2.2.2. Technical equipment of work and labour productivity

The technical equipment of work component also has an impact on work productivity. The decrease in the total number of people employed, which is a consequence of the fall in the number of Ukraine's population in the analysed period (Kupets, 2014), will result in an autonomous increase in the technical equipment of work factor. In order to assess the impact of demographic changes on the amount of capital per working person in the analysed period, the value of fixed assets was assumed to be of the level noted in 2017. Figure 12 presents the results of this experiment.

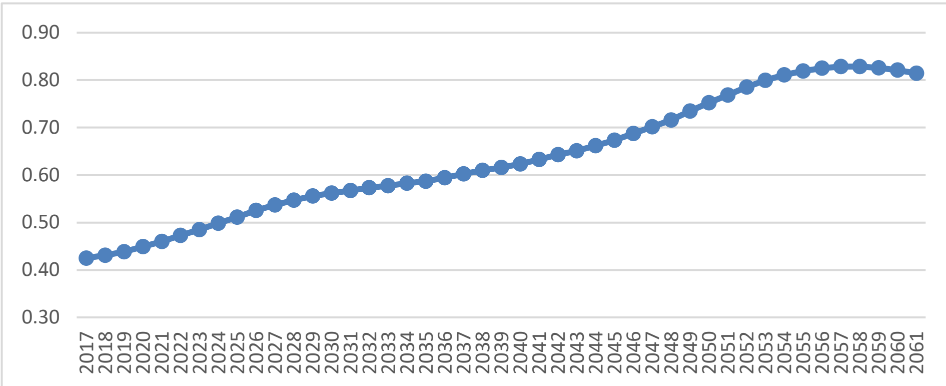
Figure 12. Levels of the technical equipment of work component resulting from changes in the demographic structure of the population of Ukraine (in millions of Ukrainian hryvnias (UAH), at 2017 prices)



Source: author's calculations based on the SSSU data (fixed assets, number of working age persons), ILO data (number of employed persons) (ILO, n.d.) and the IDBSNAN forecast (number of working age persons).

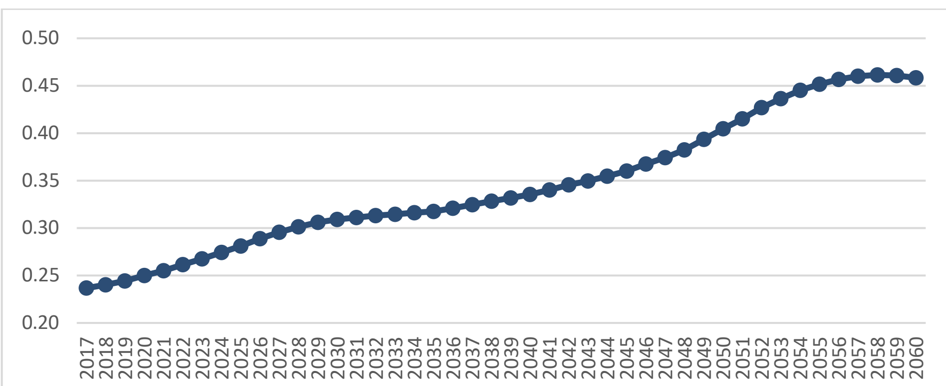
Figure 12 shows that the technical equipment of work will regularly increase (due to the decreasing number of employees and the fixed value of capital). In consequence, the impact of these changes on GDP *per capita* will be positive. However, this effect may be significantly offset by a decrease in the propensity to invest, resulting from e.g. a reduced tendency of the elderly to accumulate savings (Bloom et al., 2010; Liu et al., 2013). Thus, these circumstances should be taken into account when examining the influence that changes in the demographic structure have on the value of tangible assets. For this purpose, the results of the IMF's research on the impact of demographic changes on various macroeconomic categories were used (Jong-Won et al., 2014). The results from the IMF study confirm the statistically significant and negative impact of the old dependency ratio on the investment rate. A change in the old dependency ratio by 1% causes, *ceteris paribus*, a change in the investment rate by 0.332% in the opposite direction. Figures 13 and 14 present the evolution of the demographic dependency ratio and economic dependency ratio of the elderly in Ukraine in 2017–2060.

Figure 13. Values of the economic dependency ratio in Ukraine



Source: author's calculations based on SSSU data (number of working age persons, total population) and the IDBSNAN forecast (number of working age persons, total population) and assumptions made regarding the effective employment coefficients in individual age groups.

Figure 14. Values of the demographic dependency ratio in Ukraine



Source: author's calculations based on SSSU data (number of working age persons, total population) and the IDBSNAN forecast (number of working age persons, total population).

Figures 13 and 14 show that the demographic dependency ratio and the economic dependency ratio for people aged 65+ will almost double in the analysed period. Thus, it will cause unfavourable changes in the propensity to invest and the rate of investment. In order to link the decrease in the investment rate with the amount of tangible assets, the capital accumulation equation was used (Bussolo et al., 2015):

$$FA_t = (1 - \delta) \cdot FA_{t-1} + I_t, \tag{9}$$

where

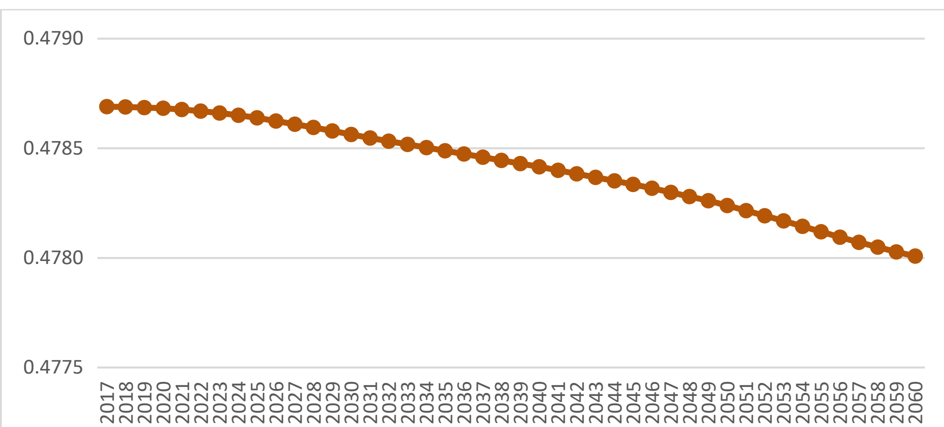
FA_t – amount of tangible property in period t;

δ – tangible assets depreciation rate;

I_t – investment outlays in period t.

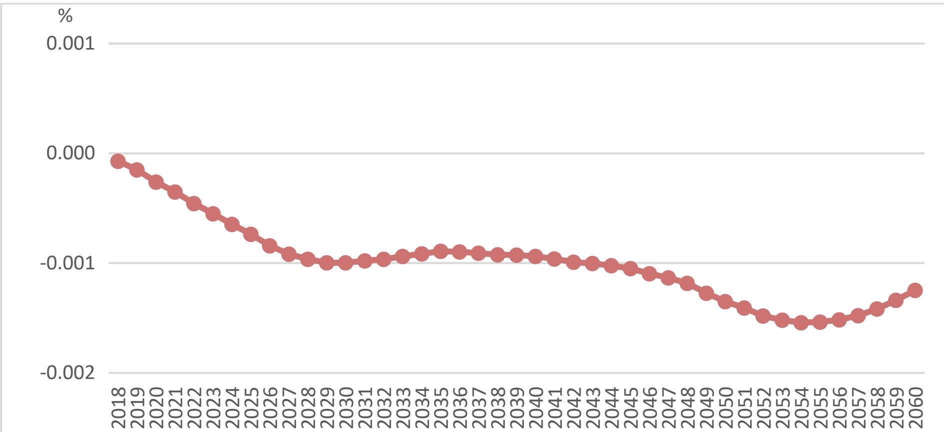
Using the 2017 value of tangible assets as a reference point and adopting the depreciation rate at the level of $\delta = 5\%$ (Filipowicz & Tokarski, 2016), the determination of the value of investments guaranteeing the maintenance of the value of tangible assets at a constant level becomes possible. Thus, if the amount of restitution investments is known, then it is possible to determine their share in the GDP, which will allow linking this category with the results of the research of the IMF. It enables the assessment of the impact of the decline in the propensity to invest, resulting from the changes in the demographic structure, on the level of the tangible capital and, at the same time, on the level of the technical equipment of work. Figure 15 shows the development of the amount of the tangible capital per employee caused by the decline in the propensity to invest, while Figure 16 presents the impact of changes of this category on the growth rate of labour productivity.

Figure 15. Amount of technical equipment of work caused by a decline in the propensity to invest, resulting from changes in the demographic structure of the Ukrainian population (in millions of UAH, at 2017 prices)



Source: author's calculations based on the SSSU data (fixed assets, number of working age persons, total population), ILO data (number of employed persons) (ILO, n.d.) and the IDBSNAN forecast (number of working age persons, total population).

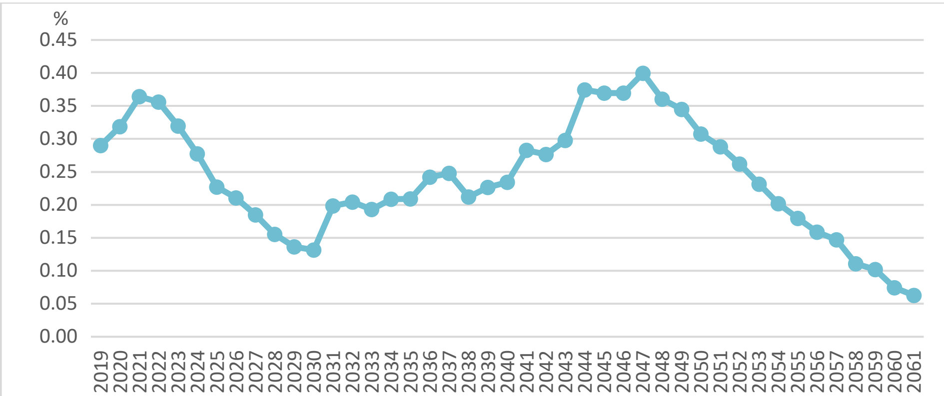
Figure 16. Annual changes in the growth rate of labour productivity resulting from changes in the level of technical equipment of work caused by a decrease in the propensity to invest, which is determined by the demographic structure of the population of Ukraine



Source: author's calculations based on the SSSU data (fixed assets, number of working age persons, total population), ILO data (number of employed persons) (ILO, n.d.) and the IDBSNAN forecast (number of working age persons, total population).

Figures 15 and 16 demonstrate that the decline in the propensity to invest, resulting from the change in the demographic structure of the Ukrainian population will contribute to a decrease in the level of technical equipment of work. However, this will not have any significantly adverse effects on the labour productivity growth (the correction presented in Figure 16 is relatively minor). The total effects of the impact of the changes in tangible capital per worker on labour productivity, and thus on economic growth, are presented in Figure 17.

Figure 17. Annual changes in the growth rate of labour productivity (and GDP *per capita*), resulting from changes in the technical equipment of work, caused by demographic changes in the structure of the population of Ukraine



Source: author's calculations based on the SSSU data (fixed assets, number of working age persons, total population), ILO data (number of employed persons) (ILO, n.d.) and the IDBSNAN forecast (number of working age persons, total population).

As Figure 17 suggests, the negative impact on labour productivity resulting from the decline in the propensity to invest was completely offset by the positive impact from the reduction in the number of employees. However, the total impact of changes in the technical equipment of work is unevenly distributed in the analysed period. Until the beginning of the 2030s, a slowing increase in labour productivity, and thus GDP *per capita*, should be expected, while throughout the 2030s and 2040s an upward trend in the rate of change in labour productivity will occur, with an average annual growth equalling nearly 0.27%. The late 2040s will see another slowdown in the growth, eventually reaching about 0.07% at the end of the analysed period.

2.2.3. Total factor productivity and labour productivity

Total factor productivity is said to be an important factor in long-term economic growth. It is believed that the increase in the total factor productivity contributes to the greatest extent to the improvement of the standard of living of the society. When determining the total factor productivity, it is important to quantify the impact of various conditions, including demographic ones, on the total factor productivity (Florczak, 2011). The starting point is to determine the level of the total factor productivity, which can be done by transforming equation (5) and taking the logs of both sides (Feyrer, 2008):

$$\ln TFP_t = \ln y_t - \frac{a}{1-a} \cdot \ln\left(\frac{K}{Y}\right)_t - \ln h_t. \quad (10)$$

After determining the value of the total factor productivity by using the relationship described in (10), the total factor productivity should be made a function of the age structure of the employees according to the formula below, whose parameter estimates were taken from the work by Feyrer (2008) (for further details, see also Florczak (2017):[2]

$$\begin{aligned} TFP_t = \exp[&(-4.005) \cdot (W1519_t - W1519_{t-5}) + (-2.939) \cdot (W2029_t - W2029_{t-5}) + \\ &+(-2.152) \cdot (W3039_t - W3039_{t-5}) + (-2.038) \cdot (W5059_t - W5059_{t-5}) + \\ &+(-2.044) \cdot (W60UP_t - W60UP_{t-5})] \cdot TFP_{t-5}, \end{aligned} \quad (11)$$

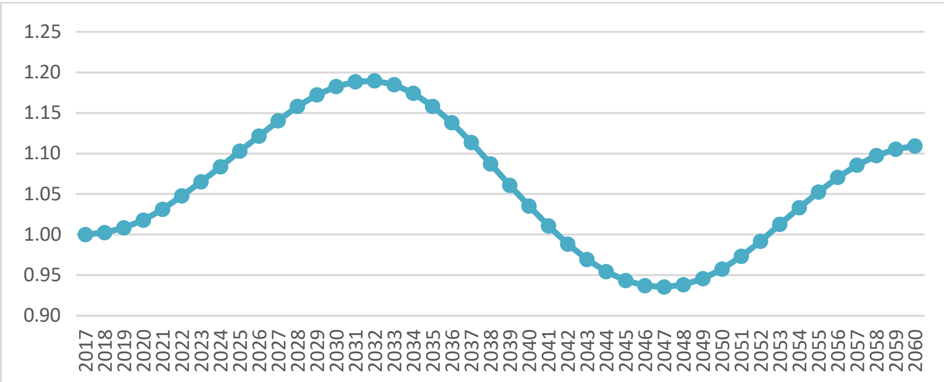
where
*W*1519, *W*2029, *W*3039, *W*5059, *W*60UP – shares of employees in individual age groups in total employment in period t.[3]

Thus, knowing the value of the total factor productivity for 2017 (determined by using formula (10)) and when the development of the share of employees in individual age groups in total employment for 2017–2060 is calculated (see the previous parts of the article), it is possible to estimate the levels of total factor productivity in the analysed period. This will allow the assessment of the impact of changes in the demographic structure on the total factor productivity. Figure 18 shows the levels of total factor productivity normalised with respect to the values relating to 2017, while Figure 19 shows how fast the changes of this component will occur and its impact on the economic growth.

[2] The results regarding the impact of the age on productivity (parameters used in equation (11)) are based on the analysis of selected OECD economies.

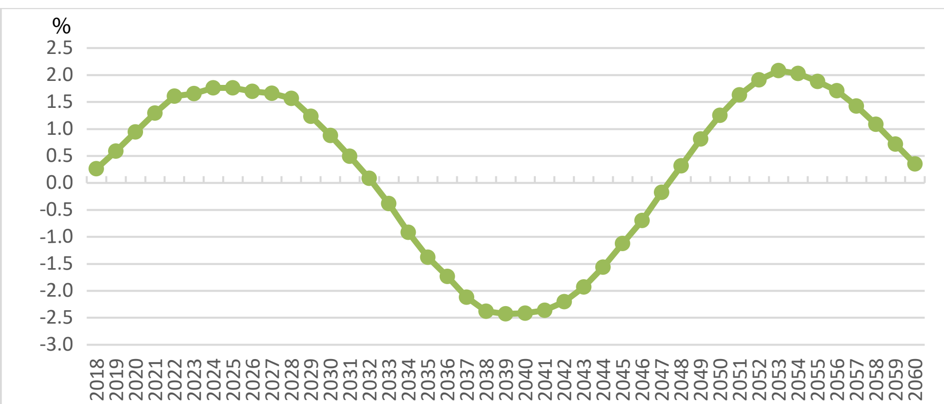
[3] The fifth-order lags in equation (11) result from the fact that in the source study (see Feyrer, 2007), on the basis of which the equation was developed, changes in explanatory variables relate to 5-year periods. Thus, the structural parameters standing next to them measure the strength of the influence over a 5-year period.

Figure 18. Levels of total factor productivity caused by changes in the demographic structure of the Ukrainian population



Source: author's calculations based on the SSSU data (number of the employed persons, number of working age persons), ILO data (number of employed persons) (ILO, n.d.) and the IDBSNAN forecast (number of working age persons).

Figure 19. Annual changes in the growth rate of total factor productivity, caused by changes in the demographic structure of the population of Ukraine



Source: author's calculations based on the SSSU data (number of the employed persons, number of working age persons), ILO data (number of employed persons) (ILO, n.d.) and the IDBSNAN forecast (number of working age persons).

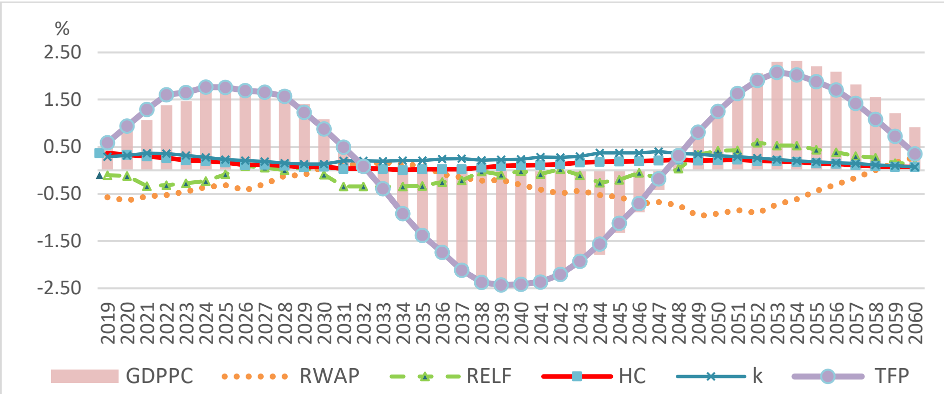
As Figures 18 and 19 show, the impact of the demographic changes on the total factor productivity will not be evenly distributed within the analysed period. Up to the beginning of the 2030s, a favourable, but declining impact of changes in the demographic structure on total factor productivity, and thus on labour productivity, will be noted, whereas the 15 years that follow will be characterised by a significant, negative (average annual decrease by 1.59%) impact of the demo-graphic conditions on economic growth. The mid-2040s will mark the beginning of a positive increase in the pace of changes in total factor productivity, caused by changes in the demographic structure of the society, although the average annual rate (increase by 1.33%) will be slightly lower compared to the decline from the previous period. It should also be noted that the total factor productivity levels, determined by the demographic structure of the society, will mostly be higher in the analysed period than the level registered in 2017, with the exception of the 2040s, when the levels of the total factor productivity will be lower than those of 2017.

3. Cumulative effects and conclusion

Having assessed the impact of the changes in the individual components of relationship (1) on the GDP growth rate *per capita* (where labour productivity has been divided into three components: human capital, technical equipment of work and total factor productivity), their combined effects can be analysed. Figure 20 shows the total effects of the impact of the changes in demographic conditions on Ukraine's economic growth by the year 2060 (being a combination of Figures 2, 6, 11, 17, and 19). On the other hand, the Table presents the average annual rates of change of all formula (1) components for selected periods within the years 2017–2060.

The results show that only human capital exerts a positive influence on the GDP *per capita* growth in the entire analysed period, but this impact, compared to the other factors, is insignificant. Moreover, the technical equipment of work factor, although minor, also has a positive impact almost throughout the entire horizon of the analysis (except for the first year). On the other hand, the age structure coefficient will exert a negative and significant impact on GDP *per capita* growth throughout nearly the entire analysed period (except for the first half of the 2030s). The influence of other factors on GDP *per capita* varies over time. The change in demographic conditions will generally have a negative impact on the effective employment rate in the years 2017–2047 (with the exception of the second half of the 2020s, when this impact will be almost neutral), and starting from the second half of the 2040s a positive impact of demographic changes on this category, thus on the economic growth, should be expected. Total factor productivity has the greatest impact on the GDP *per capita* growth rate in the analysed period. The impact of this factor is so substantial that the GDP *per capita* changes are consistent with this component's shape of changes.

Figure 20. Annual total effects of the impact of changes in the demographic structure of the Ukrainian population on GDP *per capita*



Source: author's calculations based on the SSSU data (fixed assets, number of working age persons, total population, life expectancy at birth, education), ILO (number of the employed persons, education) data (ILO, n.d.) and the IDBSNAN forecast (number of working age persons, total population, life expectancy at birth).

Table. Decomposition of the impact of demographic conditions on Ukraine's GDP *per capita* (average annual rates of changes, in percentages)

Years	Category					
	Total factor productivity (TFP)	Human capital (HC)	Technical equipment of work (k)	Effective employment ratio (RELF)	Population age structure (RWAP)	GDP *per capita* (GDPPC)
2017–2020	0.5963	0.6054	−0.2266	−0.0278	−0.7343	0.2237
2020–2025	1.6159	0.2284	0.3087	−0.2447	−0.4388	1.4698
2025–2030	1.4065	0.0950	0.1637	0.0276	−0.1673	1.5262
2030–2035	−0.4221	0.0297	0.2026	−0.3387	0.1301	−0.3960
2035–2040	−2.2157	0.0603	0.2324	−0.1153	−0.1933	−2.2311
2040–2045	−1.8368	0.1549	0.3202	−0.1223	−0.4827	−1.9656
2045–2050	0.2990	0.2119	0.3564	0.1143	−0.8006	0.1836
2050–2055	1.9051	0.1894	0.2324	0.5013	−0.7028	2.1257
2055–2060	1.0562	0.0956	0.1185	0.2494	−0.0026	1.5185

Source: author's calculations based on the SSSU data (fixed assets, number of working age persons, total population, life expectancy at birth, education) (SSSU, n.d.), ILO data (number of employed persons, education) (ILO, n.d.) and the IDBSNAN forecast (number of working age persons, total population, life expectancy at birth).

The overall impact of the changes in the demographic conditions of the society on the long-term economic growth of Ukraine will be significant, but also significantly diversified over time. Until the beginning of the 2030s, a positive impact of demographic changes on the GDP *per capita* growth will be noted, with an average annual increase of nearly 1.1% and a peak observed in the middle of the second decade of the 21st century. The next 15 years will be characterised by a strong negative impact of the demographic conditions on economic growth, with an average annual decline of almost 1.7% and a local extreme occurring at the turn of the 2030s and 2040s. However, in the remaining years of the analysed period, significant positive effects of the changes in the demographic structure of the society on the economic growth of Ukraine should reoccur, with an average annual increase of almost 1.6% and the greatest intensity observed at the end of the first half of the 2050s. Thus, the negative influence of the demographic changes on the economic growth of Ukraine over the analysed period should be neutralised by the positive impact that these changes are likely to entail.

Comparing these results with the analysis of the impact of demographic changes on Poland's economic growth (Florczak, 2017), numerous similarities can be observed. As in the case of Ukraine, Poland's economic growth is most strongly influenced by the total productivity of the factors of production. As in the case of Ukraine, Poland's economic growth is most strongly influenced by the total factor productivity. There is also a similarity in the pace of changes in the GDP *per capita*, caused by the changes of the demographic structure. In the light of these similarities, it seems that the Polish economy cannot count on a permanent and substantial labour emigration of Ukrainians in the long run, since Ukraine itself is bound to experience similar demographic problems of a nearly equal intensity and almost identical synchronisation as Poland.

References

Aksonova, I. V. (2012). Statistical Analysis of the Influence of Demographic Processes for Economic Development in Ukraine. *Business Inform,* (6), 86–89. http://www.business-inform.net/export_pdf/business-inform-2012-6_0-pages-86_89.pdf.

Batog, C., Crivelli, E., Ilyina, A., Jakab, Z., Lee, J., Musayev, A., Petrova, I., Scott A., Shabunina, A., Tudyka, A., Xu, X. C., & Zhang, R. (2019). *Demographic Headwinds in Central and Eastern Europe* (IMF Departmental Paper No. 19/12). https://www.imf.org/en/Publications/Departmental-Papers-Policy-Papers/Issues/2019/07/11/Demographic-Headwinds-in-Central-and-Eastern-Europe-46992.

Bloom, D. E., Canning, D., & Fink, G. (2010). Implications of population ageing for economic growth. *Oxford Review of Economic Policy*, *26*(4), 583–612. https://doi.org/10.1093/oxrep/grq038.

Bussolo, M., Koettl, J., & Sinnott, E. (2015). *Golden Aging: Prospects for Healthy, Active, and Prosperous Aging in Europe and Central Asia*. Washington: World Bank. https://openknowledge.worldbank.org/handle/10986/22018.

Coleman, D., & Rowthorn, R. (2011). Who's Afraid of Population Decline? A Critical Examination of Its Consequences. *Population and Development Review*, *37*(s1), 217–248. https://doi.org/10.1111/j.1728-4457.2011.00385.x.

Cymbal, O., & Iarosh, O. (2020, May 26). *Ukrainskyi rynok pratsi u dzerkali ofitsiinykh ta neofitsiinykh dzherel*. https://voxukraine.org/ukrayinskij-rinok-pratsi-u-dzerkali-ofitsijnih-ta-neofitsijnih-dzherel/.

Feyrer, J. (2007). Demographics and productivity. *The Review of Economics and Statistics*, *89*(1), 100–109. https://doi.org/10.1162/rest.89.1.100.

Feyrer, J. (2008). Aggregate Evidence on the Link Between Age Structure and Productivity. *Population and Development Review*, *34*, 78–99. http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=259EB4445A187371450AC75976FAD182?doi=10.1.1.303.3713&rep=rep1&type=pdf.

Filipowicz, K., & Tokarski, T. (2016). Zróżnicowanie wydajności pracy w Europie: na podstawie grawitacyjnego modelu wzrostu gospodarczego. In A. Nowosad & R. Wisła (Eds.), *Zróżnicowanie rozwoju współczesnej Europy* (pp. 311–377). Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.

Florczak, W. (2008a). Efektywna podaż pracy a wzrost gospodarczy. *Gospodarka Narodowa*, *228*(11–12), 21–46. https://doi.org/10.33119/GN/101300.

Florczak, W. (2008b). Kapitał ludzki w empirycznych modelach wzrostu. *Ekonomista*, (2), 169–200.

Florczak, W. (2011). Produktywność czynników wzrostu PKB. *Wiadomości Statystyczne*, *56*(2), 8–26. https://stat.gov.pl/czasopisma/wiadomosci-statystyczne/archiwum/.

Florczak, W. (2017). Wpływ starzejącego się społeczeństwa na długookresowy wzrost gospodarczy Polski do roku 2050. *Gospodarka Narodowa*, *291*(5), 73–104. https://doi.org/10.33119/GN/100743.

Florczak, W., & Przybyliński, M. (2016). Zmiany w liczebności i strukturze populacji a rozwój społeczno-ekonomiczny. *Studia Ekonomiczne*, *3*(90), 396–422. https://inepan.pl/wp-content/uploads/2016/08/SE_2016_3_Florczak.pdf.

Florczak, W., Przybyliński, M., Świeczewska, I., Tomaszewicz, Ł., & Trębska, J. (2018). *Wybrane zjawiska i procesy wpływające na rozwój polskiej gospodarki w pierwszej połowie XXI wieku. Projekcje na podstawie systemu modeli makroekonomicznych*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego. https://wydawnictwo.uni.lodz.pl/wp-content/uploads/2019/01/Florczak-i-in._Wybrane-zjawiska-.pdf.

Geyets, V. M. (2011). The consequences of demographic challenges for economic growth in Ukraine. *Demography and social economy*, (1), 3–23. https://doi.org/10.15407/dse2011.01.

International Labour Organization. (n.d.). *ILO Data Explorer* [Data sets]. Retrieved February 3, 2021, from https://www.ilo.org/shinyapps/bulkexplorer9/.

Jong-Won, Y., Jinill, K., & Jungjin, L. (2014). *Impact of Demographic Changes on Inflation and the Macroeconomy* (IMF Working Paper No. 14/210). https://www.imf.org/external/pubs/ft/wp/2014/wp14210.pdf.

Kudlak, V. (2019). Dynamics of dymographic processes in Ukraine and its impact on national economy. *Galician economic journal (Tern.)*, *4*(59), 98–104. https://doi.org/10.33108/galicianvisnyk_tntu2019.04.098.

Kupets, O. (2014). Labor Market Challenges of an Aging and Shrinking Population in Ukraine. *Journal of Comparative Economic Studies*, *9*, 99–134. http://www.ces.kier.kyoto-u.ac.jp/jces/09_jces_2014/08_Kupets.pdf.

Liu, Y., Xiao, H., & Zhu, S. (2013). *Population Aging, Saving Rates and Long-term Economic Growth in China: Based on Dynamic GGE Model.* Annual Conference on Global Economic

Analysis, Shanghai. https://www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=4126.

Maksymenko, S. (2009). Fertility, Money Holdings, and Economic Growth: Evidence from Ukraine. *Comparative Economic Studies*, *51*(1), 75–99. https://doi.org/10.1057/ces.2008.45.

Mączyńska, E. (2010). Przełom cywilizacyjny a wzrost gospodarczy. Niedoceniane aspekty demograficzne. In A. Potrykowska & E. Orzełek (Eds.), *Biuletyn nr 55* (pp. 18–28). Warszawa: Rządowa Rada Ludnościowa. https://stat.gov.pl/cps/rde/xbcr/bip/BIP_biuletyn55.pdf.

Ptoukha Institute for Demography and Social Studies of the National Academy of Sciences of Ukraine. (2014). *National Population Projections: Revision*. Retrieved February 3, 2021, from https://idss.org.ua/forecasts/nation_pop_proj_en.

Romanukha, O. M. (2016). Social-economic consequences of Ukrainian population ageing. *Scientific Bulletin of KSU. Series Economic Sciences, 17*(3), 91–94. http://www.ej.kherson.ua/journal/economic_17/3/25.pdf.

Rovný, P., Moroz, S., Palkovič, J., & Horská, E. (2021). Impact of Demographic Structure on Economic Development of Ukrainian Coastal Regions. *Sustainability*, *13*(4), 1–19. https://doi.org/10.3390/su13041798.

Stoychik, T. (2018). Forming the modern labour market: analysis of request and proposal of the working force. *Professional Pedagogics*, (17), 101–105. https://doi.org/10.32835/2223-5752.2018.17.101-105.

Vakhitova, G., & Coupe, T. (2013). *The Relations between Education and Migration in Ukraine*. Budapest: International Labour Organization. http://legacy.iza.org/en/papers/7990_19032014.pdf.

Wasilewska, E., & Pietrych, Ł. (2018). Starzenie się społeczeństwa a wzrost gospodarczy w krajach Unii Europejskiej. *Zeszyty Naukowe Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie. Problemy Rolnictwa Światowego*, *18*(4), 481–492. http://sj.wne.sggw.pl/pdf/PRS_2018_T18(33)_n4.pdf.

Wößmann, L. (2003). Specifying human capital. *Journal of Economic Surveys*, *17*(3), 239–270. https://doi.org/10.1111/1467-6419.00195.

Zoidze, D. R. (2013). Modern labour market: dismissal of employees as a factor of structural discrepancy between supply and demand. *Problems of Economy,* (2), 59–65. https://www.problecon.com/export_pdf/problems-of-economy-2013-2_0-pages-59_65.pdf.