Article

Fuzzy Algorithmic Modeling of Economics and Innovation Process Dynamics Based on Preliminary Component Allocation by SSA Method

Alexey F. Rogachev[1],*, Alexey B. Simonov[2], Natalia V. Ketko[3] and Natalia N. Skiter [4]

[1] Volgograd State Technical University; rafr@mail.ru
[2] Volgograd State Technical University; absimonov@gmail.com
[3] Volgograd State Technical University; gsa-buch@list.ru
[4] Volgograd State Technical University; ckumep@mail.ru

* Correspondence: e-mail@e-mail.com; Tel.: (optional; include country code; if there are multiple corresponding authors, add author initials)

Abstract: In this article, the authors propose an algorithmic approach to building a model of the dynamics of economic and, in particular, innovation processes. The approach under consideration is based on a complex algorithm that includes (1) decomposition of the time series into components using singular spectrum analysis; (2) recognition of the optimal component model based on fuzzy rules, and (3) creation of statistical models of individual components with their combination. It is shown that this approach corresponds to the high uncertainty characteristic of the tasks of the dynamics of innovation processes. The proposed algorithm makes it possible to create effective models that can be used both for analysis and for predicting the future states of the processes under study. The advantage of this algorithm is the possibility to expand the base of rules and components used for modeling. This is an important condition for improving the algorithm and its applicability for solving a wide range of problems.

Keywords: singular spectrum analysis, fuzzy logic, innovative activity, time series, economic development, models.

1. Introduction

An important task arising in many areas of economic science and practice is the analysis and prediction of the dynamics of processes and phenomena. At the same time, there are quite a large number of approaches to solving these problems, in particular, based on the application of statistical methods and methods of data mining. Such methods as trend extraction (for example, based on correlation and regression analysis) and cyclical fluctuations (both based on indices and Fourier series decomposition), as well as the simplest methods based on the use of average values, in combination, can be considered as methods of unsupervised learning. In the same group we can include the majority of methods of data mining, primarily using different types of neural networks, as the simplest direct propagation networks, and usually more effective in the analysis of the dynamics Reccurent Neural Networks; Fuzzy Neural Networks and others. A number of other methods, such as those simulating random walk or using smoothing, moving averages, the more complex ARIMA method, LOESS methods are based on the construction of part of the model totality. Sometimes these parts are highlighted in a preliminary analysis, as in the X-12 method. These approaches make it possible to identify the main regularities in the dynamics of most phenomena and make a qualitative forecast, at least for the short and medium term.

However, in a number of sciences, particularly in economics, when studying the dynamics of some phenomena and processes, modeling becomes more complicated due to their characteristic chaotic behavior, a sharp change of trends, as well as the presence of a large number of different types of fluctuations that may vary over time. The reasons for such behavior may be the presence of a large number of factors affecting the dynamics of the phenomena under study, and a complex, often cyclical nature of the relationship between these factors, as well as between them and the phenomenon under study. In such conditions, known methods do not have the necessary tools to build a predictive model with a high level of reliability of the results. Indeed, methods based on the analysis of the entire training totality (for example, correlation-regression analysis, Fisher's g test, Fuzzy Neural Networks), can be ineffective in some cases due to the fact that the trend and nature of fluctuations changed in the totality, which leads to a low quality of the obtained models. On the other hand, methods based on the analysis of the part of the training aggregate, such as the moving average or X-12, may be ineffective due to the lack of comprehensiveness of the obtained models.

In economics, such phenomena, which are characterized by close to chaotic behavior, often occur in stock exchange analysis, analysis of tourist flows, as well as in many types of long-term analysis, for example, in the long-term analysis of electricity consumption or economic activity of enterprises [1]. The influence of uncertainty is especially great in modeling of innovation activity, what makes modeling in this area especially difficult [2, 3]. Similar phenomena can be observed, for example, in the forecasting of time series (TS) of crop yields, weather phenomena, availability of water resources, and so on [4].

When analyzing such phenomena, it seems effective to combine mathematical methods with methods of expert evaluation of existing trends (fundamental analysis), or to apply methods of intelligent data analysis. However, the results of applying expert methods not only depend significantly on the level of qualification of experts and their availability of the necessary information and tools, but also lead to the risk of systematic errors, for example, in cases where experts are interested in this or that result of predictions. On the other hand, the prerequisites of the obtained results of intelligent analysis methods, such as different types of neural networks, are usually not transparent enough, which can lead to untraceable error in determining the main trends, and also does not exclude system errors. For example, in the case when a bank gives a loan to an agricultural producer on the basis of the bank yield estimate made by artificial intelligence, it, in turn, affects the yield: the lower the bank estimate, the lower the amount of loans issued, which reduces the ability of agricultural producers to use technology to improve yields [4].

Based on the above reasoning, it is obvious the need to improve the algorithms of studying the dynamics of economic phenomena using tools of modern methods of data analysis and statistical models. At the same time, statistical models, supplementing the possibilities of teaching methods with or without a teacher, will provide transparency of conclusions, their verification and validation, facilitate the application of created tools for analysis and forecasting not as "black boxes", but as an understandable in functioning mechanism. The described qualities are especially important in the study of complex systems in which there may be elements of chaotic behavior. Sometimes for such systems, understanding the logic of what is happening is no less important than the accuracy of the resulting forecast. However, such models are built on the basis of certain theoretical hypotheses and do not always allow to react quickly enough to changes in the trend, seasonal and cyclical fluctuations, to reveal hidden regularities in the development of this particular phenomenon. In addition, the effectiveness of such models depends largely on the knowledge of the expert who creates them.

One of the tools which can be used to highlight such changes in trends and identify hidden components is the SSA method [5,6,7,8,9,10]. This method allows us to decompose the original series of dynamics into basic components, which are a combination of a trend, cyclical fluctuations of different frequency, and random deviations. On the basis of

the models built, it is possible to assess both the presence and properties of the main aspects of the dynamics, and the relationship between the phenomena, as well as to build a forecast of the dynamics of the phenomenon.

However, although the SSA method can be used to build interpolation and extrapolation models, its efficiency for solving some economic problems is not high enough. As shown in [5], the SSA method has a lower efficiency than, for example, neural networks when solving problems of electricity consumption forecasting. In this connection, a number of authors suggest combining SSA with other methods (first of all, with methods of intelligent data analysis) in order to improve the quality of extrapolation. The effectiveness of the combination technique is confirmed by the research in [5], where the construction of a hybrid model in which the trend extracted by MSSA (Multi-Channel SSA) is processed by Neural fuzzy network. Similarly, in [8], SSA was used to clean the training data set from noise before training recurrent neural networks (RNN). In some cases, such algorithms are less efficient compared to other hybrid algorithms, for example, when modeling the flow of tourists to Indonesia [9], the combination of SSA and fuzzy time series (FTS) gave worse results than the SARIMA-FTS combination. This can be explained by the fact that in this study SSA was used only to identify the trend and seasonal component, while SARIMA allows also to take into account autoregression. From the point of view of considering the SSA method as a pre-processing tool, of interest is the work [10], in which the Fuzzy C-Means Clustering (FCM) method was used before the SSA method, which allowed to obtain results in the FCM-SSA-ARIMA model that surpass the accuracy of SSA-ARIMA and ARIMA results.

Based on the analysis of the effectiveness of various methods of studying the dynamics of economic processes, the authors in this article explored the possibility of using a combined algorithm for modeling and predicting TS in the economy and, in particular, in innovation, where the influence of uncertainty is particularly high. The proposed algorithm makes it possible to combine a transparent allocation of the main trends on the basis of search methods with an evaluation of the trends selected by expert systems to reduce the influence of human factors, and to build models of the dynamics based on the most significant selected trends (discarding unimportant elements of the dynamics series).

2. Materials and Methods

To solve the problems of forecasting complex economic and innovation processes in conditions of uncertainty, the authors propose to use a combined approach to build a multicomponent model of their dynamics. Similar algorithms are used in radio engineering [11], which makes it possible to increase the efficiency of control and management processes in a competitive environment of functioning of adaptive information transmission systems, as well as to create support for the process of adaptive control of the state of radio devices.

As a similar process control tool for its application in the field of economics, we propose an algorithm, reflected in Figure 1.
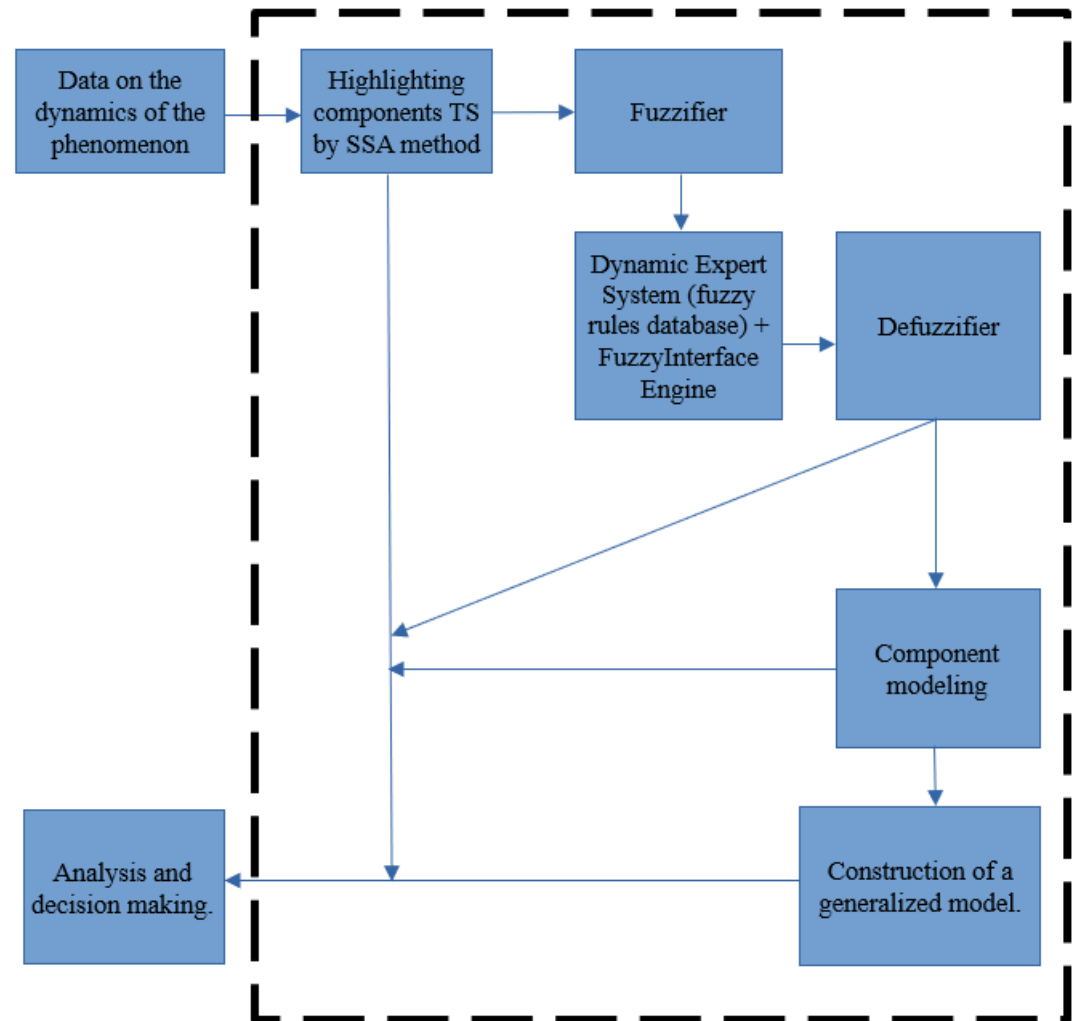
Figure 1. A generalized structural diagram of the dynamics analysis based on fuzzy processing of time series components obtained with SSA.

The algorithm consists of the following steps:

1. Selection of time series components using the SSA method. For these calculations we used the Pyts package [12].

2. Creation of fuzzy estimates of the possible character of time series components on the basis of search analysis tools and empirical rules. This estimation takes place in the module "phasifier" and can be performed both in automatic mode and corrected by expert method taking into account the proposed automatic estimations. These calculations were performed using the packages NumPy, ScyPy, and Pandas. The package Matplotlib was used for visualization.

3. The dynamic expert system, based on fuzzy logic, draws conclusions about the optimal classification for each component of the time series. At the same stage, the experts system can draw conclusions from the subject area, concerning, for example, the possible economic content of each component. In addition to the previously mentioned tools, we used FisPro[13] to visualize and verify the calculations.

4. expert system evaluations can be made in the form of fuzzy logic, in which case they must go to the defuzzifier to obtain clear evaluations.

5. Based on clear estimates, the optimal model of each component is selected and constructed, taking into account its type and possible economic content.

6. A generalized model, which can be used for analysis and forecasting of economic and innovative processes, is constructed.

Let us consider the above steps in more detail.

The separation of the dynamics components allows us to consider SSA as a method of exploratory analysis, the application of which is highly effective in cases where a dynamics series may include a large number of different components with insufficiently accurately studied properties. For example, as shown [7], in many time series describing the dynamics of innovation activity in the Russian Federation and Volgograd region for twenty years, we can distinguish a trend, two cyclic components (with the period of about 3 years and about 8 years), as well as various types of autocorrelation and, of course, "white noise". This agrees with the general theoretical ideas about the structure of series of dynamics in the economy, in which the analysis of the dynamics by years traditionally distinguishes short-term cycles of Kitchin (with a period of 2-4 years); medium-term economic cycles of Juglar with a period of 7-11 years and long-term cycles of Kuznets (period 15-25 years) and Kondratiev (period 45-60 years), while long-term cycles in our study [7] were not distinguished because of the limited length of time series.

Before we start to use SSA method, we should set window's length L [14]. Large values of L allow reconstructing long period oscillations, but too large values of L leaves very few observations to make qualitative reconstruction. So L usually takes a value from quarter to half the time series length (N/4 ... N/2) [15]. We can also examine the median value [16] as the optimal L value.

The SSA method consists of two phases, decomposition and reconstruction [17]. Let us consider these stages of the method.

The first stage of decomposition phase is embedding. The original time series F of length N is transformed into a sequence of multidimensional vectors (trajectory matrix X) during this stage. The matrix X has the dimension L × K, where K value calculated as follows:

$$K=N-L+1 \qquad (1)$$

So, K is the number of columns and L is the number of rows of the trajectory matrix X. Each column of this matrix is rolling window of length L, so 1st column has elements of the time series F from the first till L, 2nd column has elements from the second till L+1 and so on.

$$\mathbf{X} = \begin{pmatrix} f_1 & \cdots & f_{N-L+1} \\ \vdots & \ddots & \vdots \\ f_L & \cdots & f_N \end{pmatrix} \qquad (2)$$

X is symmetric matrix with equal elements on all diagonals, so it is a Hankel matrix.

The second stage of the decomposition phase is singular value decomposition (SVD) of trajectory matrix. The result of this stage is the decomposition of the trajectory matrix X into elementary parts, i.e., representation of the matrix X as a sum of elementary matrices. Let λ_1, λ_2, …, λ_L denote the eigenvalues of the matrix $S=X*X^T$ taken in non-decreasing order, and d is number of $\lambda_i > 0$. Let us also denote U_1, U_2, …, U_L the eigenvectors of the matrix S, corresponding to the ordered eigenvalues [18].

Then the singular value decomposition for X is

$$X= \Sigma\ Xi \qquad (3)$$

Xi is the elementary matrix with rank 1. These elementary matrices are calculated by the formula:

$$X_i = \sqrt{\lambda_i} * U_i * V^T, \qquad (4)$$

where elementary vectors Vi can be found after we calculate U and λ_i, using the formula:

$$V_i = X^T * \frac{U_i}{\sqrt{\lambda_i}} \qquad (5)$$

After decomposition we have reconstruction phase. To reconstruct time series, we use diagonal averaging. Each matrix Xi we obtained after decomposition phase is translated into a series of length N. We suppose L * = min (L, K), K * = max (L, K). For each element x of matrix Xi let us also denote $x^*_{ab} = x_{ab}$, if L < K, and $x^*_{ab} = x_{ba}$, if L> K. Diagonal averaging [18] transforms each resulting matrix Xi into a series according to the formula for each element $\widetilde{f_{ik}}$ of vector $\widetilde{F_i}$:

$$\widetilde{f_{ik}} = \begin{cases} \frac{1}{k+1} \sum_{n=1}^{k+1} x^*_{n,k-n+2}, & \forall\, 0 \le k < L^* - 1 \\ \frac{1}{L^*} \sum_{n=1}^{L^*} x^*_{n,k-n+2}, & \forall\, L^* - 1 \le k < K^* \\ \frac{1}{N-k} \sum_{n=k-K^*+2}^{N-K^*+1} x^*_{n,k-n+2}, & \forall\, L^* - 1 \le k < K^* \end{cases} \tag{6}$$

This formula averages the elements along the diagonals. So, applying the diagonal averaging to the resulting matrices Xi, we obtain the series $\widetilde{F_i}$. These series can be considered as additive components of the time series f that was under study. These components may be classified as a trend, cycle or noise component [16], and then we can model it.

In spite of its significant advantages, the SSA method requires a fairly good expert analysis of the components of a time series in order to correctly assess their content. This is a rather labor-intensive process due to the large amount of work on expert evaluation in the comparative analysis of the dynamics of a large number of phenomena and processes, presumably related to each other (which is possible and can lead both to the construction of quantitative models and to the identification of fundamental features of the processes under study, as we showed in [7]). Expert assessments are also complicated by possible changes in trends, changes in the nature of fluctuations, as well as the complexity of expert separation based on the visual assessment of short-term cyclical fluctuations and autocorrelation. In addition, as the analysis of the costs of innovation activity in Volgograd region showed, SSA does not always effectively separate the weakly expressed trend and medium-term (long-term) fluctuations, as a result of which one component must be divided into two mathematical models.

To facilitate and partially automate the expert evaluation in our algorithm we propose to use the toolkit of fuzzy sets, which will be implemented in the phasifier and the database of fuzzy rules. This toolkit, based on fuzzy inference rules, will allow you to determine the dynamics model for a particular component (or propose several of the most appropriate models for experts to choose from), as well as to make preliminary conclusions about the possible economic content of specific components on the basis of connected databases and/or knowledge bases.

It should be noted that in [6] a tool for identifying the types of dynamics (trend and cyclic component) on the basis of modeling is proposed for similar problems, however, this approach does not allow expanding the set of rules and including, for example, rules for trend changes or rules for the analysis of cyclic fluctuations, which are not sinusoidal. Another approach, effective from the point of view of extracting the cyclic component, is a combination of SSA(MSSA) and Fisher filter methods[19]. For example, [20] suggests using SSA as a preprocessing method to increase the efficiency of detecting hidden periodicities in the phenomenon under study. On the other hand, in their paper [17] the authors use a logic close to the one proposed by the authors in the algorithm, according to which "Fisher g test <is used> to select the principal components to be aggregated in the reconstruction step". However, even in this approach there are difficulties with the unclear economic content of individual economic cycles reconstructed on the basis of the Fisher's g test, as well as with the impossibility of changing the model used and combining, for example, several approaches to model the cyclicality of different components.

Based on this review, we can conclude that the task of finding a way to extend the toolkit for the recognition of SSA series components is relevant, and the toolkit for phasing and applying fuzzy inference rules, which is easily extensible and customizable, is presumably effective enough to solve this problem.

The fuzzyifier converts the indicators extracted by SSA (in particular, the values of the most significant criteria) into fuzzy linguistic variables belonging to fuzzy sets Aj (individual for each indicator). As a rule, these variables have an affiliation function (AF), for example, of the triangular form

$$\mu_A(s) = \begin{cases} 1 - \frac{|s-c|}{d} & \forall s \in [c-d, c+d] \\ 0 & \forall s \notin [c-d, c+d] \end{cases}, \tag{7}$$

where c is the center point of the triangle;

d is the width of the triangle.

The peculiarity of this transformation is that we can construct fuzzy estimates for any criterion, and, if necessary, make fuzzy estimates based on expert reasoning.

After the transformation, the formation of fuzzy conclusions about the possible type of dynamics of the phenomenon defined by the SSA component under study takes place. For this purpose, additive convolution is used, which is performed on the basis of the rules of the form:

$$\text{if S1 is A11 and S2 is A21, then Z is B,} \qquad (8)$$

where S1 is a fuzzy value for one indicator, S2 for another indicator, and so on;

Z is a fuzzy value for the resulting component type.

For example, "if Spearman criterion value is high and L value obtained by Foster-Stewart method is high, then presence of trend with positive first derivative is very high".

The set of rules contained in the database of fuzzy rules and/or knowledge base, if necessary, can be adjusted, which is an important distinguishing feature of this algorithm, giving it high flexibility.

Based on the obtained fuzzy inferences, the defuzzifier transforms the fuzzy set of obtained linguistic evaluations for the inferences into a ranked list of possible trend types for a given component.

On the basis of the obtained estimations the way of modeling of this component can be automatically chosen. In particular, we used statistical models of trend, cyclical fluctuations and autocorrelation of deviations. The construction of SARIMA models also proved to be effective.

From the authors' point of view, the main advantage of this method is that we can exclude specific values of a random component from the final model, which will reduce the effect of overtraining and allow a better assessment of model error.

It is also necessary to mention the fact that, alternatively, different kinds of neural networks can be used at the modeling stage. In this case, since the data in the components, as a rule, are cleared of random deviations and contain only one trend, the training on such datasets will be highly effective [8].

The peculiarity of the presented algorithm is the possibility to offer economic estimates depending on what components are present in the modeled time series. Thus, the absence of cycles with a period of 7-11 years (Juglar cycles) may indicate difficulties with the effective replacement of production capacity.

3. Results

The algorithm developed by the authors was initially tested on a generated dataset consisting of a polynomial trend, three harmonics, and a random component with negative autocorrelation with lag 1. The SSA method efficiently extracted all the main components, which allowed us to identify the types of dynamics with a high level of confidence and build a model that described with high accuracy the functions used to generate the dataset.

After the testing, the authors used the developed algorithm to analyze information about GDP and the number of researchers in the Russian Federation. The information for the analysis was taken from the website of the Federal State Statistics Service of the Russian Federation [21].

To build the model of GDP dynamics of the Russian Federation, the data for 1995-2020 were used, the data for 2021 were used as a test data set, on the basis of which the quality of the forecast built by the created model is determined. To eliminate the effect of inflation, the data were adjusted to 2020 prices.
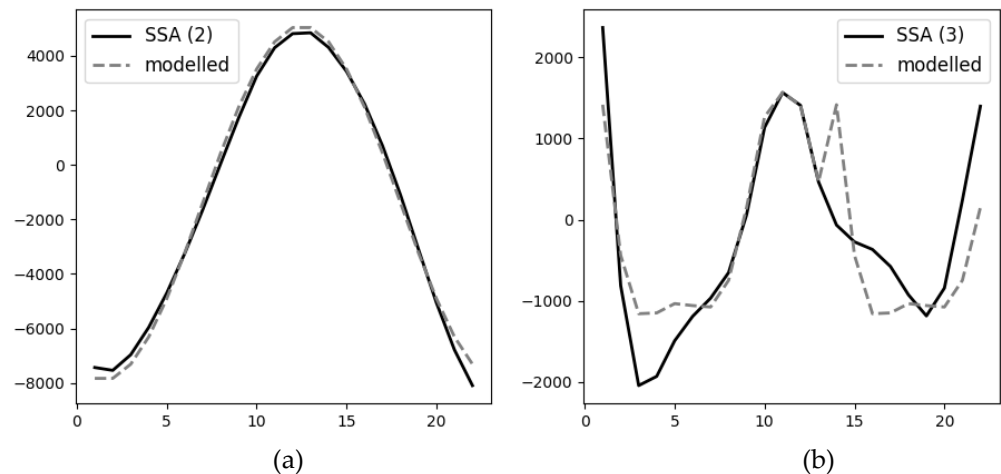
Figure 2. Modeling results for the 2nd (a) and 3rd (b) components extracted by the SSA method.

As a result of the work of the phasifier, it was found that it does not recognize fluctuations whose period is comparable with the length of the studied series of dynamics well enough. In our case, one of the highlighted SSA components was a cycle with a period of about 20 years (the Kuznets cycle), with a total length of the studied period of 26 years, and this cycle was not effectively recognized until we changed the rules to a set of estimates of the presence of the cycle according to data, first and second final difference and peridiogram. After introducing additional rules, the algorithm correctly recognized this long cycle (see Fig. 2a). Then, a cycle model with a period of 22 years based on the sine function was constructed in automatic mode.

Difficulties arose in the study of the third component, which was recognized as random deviations with positive autocorrelation, but was expertly estimated as cyclical fluctuations with a period of about 10 years (Juglar cycles). These cycles have a rather gentle growth and a sharp decline, which complicates their identification, recognition, and modeling. As a result, for effective recognition of such cycles, it was proposed to use in the phasifier an analogue of the Henderson filter to the selected components, which reduced the influence of possibly inaccurately defined values at the beginning and end of the components, but does not significantly affect the overall result of recognition. Then a rule based on detecting the cyclic direction of function change was added. This allowed us to identify this type of cycle and model it by index method (although it was possible to use other methods as well), which gave acceptable accuracy results of component recognition (see Fig. 2b, unexplained variance fraction 48.2%). Since insufficient data do not allow modeling seasonal components by analogy with X-12, an alternative could be the use of neural networks, but the possible quality of training on such a small dataset raises doubts about the reliability of the obtained result. The classical index method used in the study, on the other hand, is fairly easy to automate and reduces the error associated with the decadal cyclic component.
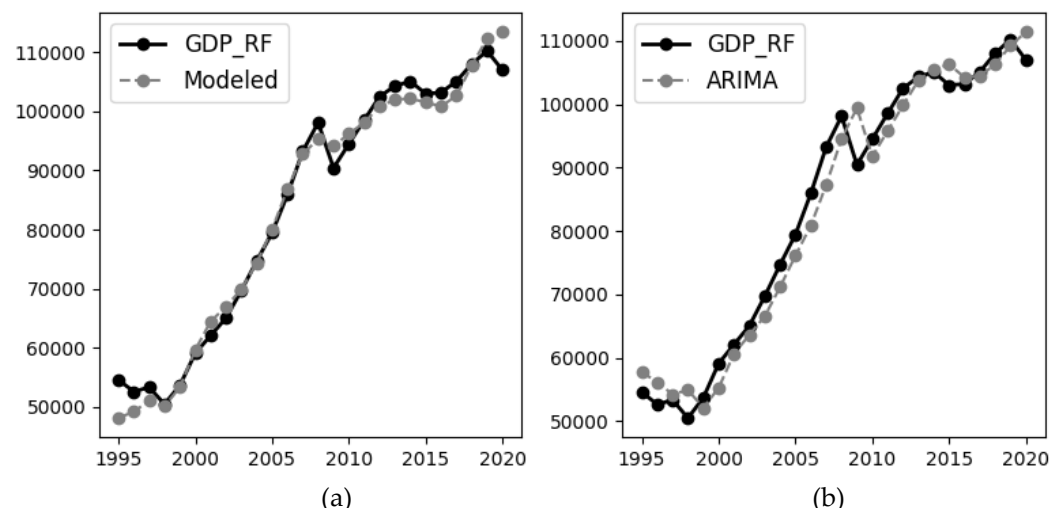
Figure 3. Results of the comparison of the model of GDP dynamics of the Russian Federation (in constant prices), developed with the algorithm (a), and ARIMA (b).

To evaluate the effectiveness of the algorithm developed in this study, we compared the created model with the ARIMA model. As can be seen from Fig. 3, the model created with the help of the presented algorithm interpolates quite accurately the value of RF GDP in the middle of the series, deviating significantly from the real values in the periods of crises (which is generally characteristic of statistical models). Even in these periods, in general, the model proved to be more efficient than ARIMA. At the same time both models have high enough accuracy, if we evaluate them on the basis of the proportion of the explained variance R^2, it is estimated as 0.973 for ARIMA, and 0.985 for the model created before the change of fuzzy recognition rules. After creating rules that account for the variation in the amplitude of the oscillations and the complex nature of the Juglar cycle, the R^2 estimate rose to 0.992. Moreover, even the model of the first three components provided an R^2 value of 0.976, which shows interpolation better than the ARIMA method. The forecast error for 2021 for the created model was +2.2% compared to -3.4% for ARIMA, but the accuracy of this forecast is not high, also because of the coronary crisis (as mentioned above, crisis phenomena reduce the accuracy of the model).

When analyzing the dynamics of the number of researchers in the Russian Federation, the accuracy of the model was much lower than in the analysis of GDP, which is quite expected - the number of researchers refers to the indicators of innovation activity, which is characterized by high quantitative and qualitative uncertainty.

Obvious fuzzy rules quite effectively distinguished cyclic components, as well as assumed the presence of a quadratic trend. Difficulties appeared in the analysis of the second component, where a sharp change in the trend is modeled (from growth in the number of researchers in 2000-2012. to a sharp decrease in 2012-2020). The proposed rules based on the difference of moving averages with different period showed a possible trend change, but the algorithm highlighted as more preferable the autocorrelation model with lag 1. Theoretically, this could have been changed by adjusting the rules to reduce the preference of the autocorrelation model as a whole, but this would have required a readjustment of the entire algorithm, so a rule tracking a single change in the trend was introduced. This situation confirmed that the question of the optimal approach to trend change tracking using the proposed algorithm remains open.
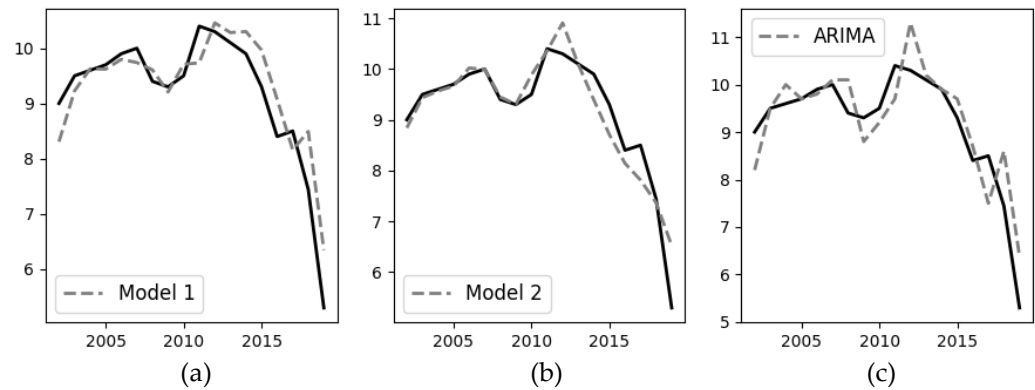
Figure 4. Results of comparison of the model of innovation activity in Russia developed with (a) the algorithm without the rule for trend change (Model1), (b) with trend change (Model2), and (c) ARIMA.

Note that both the autocorrelation model and the trend change model were quite effective in interpolation (see Fig. 4). The model constructed using the original rules, which does not directly incorporate the trend change and indirectly estimates it through autocorrelation, has $R^2=0.810$ fractions of the explained variance. This is higher than a similar estimate of 0.726 for the best ARIMA(0;2;0) model considered. However, the prediction accuracy of both models turned out to be low - they significantly underestimated the share of innovatively active organizations in 2020 (the created model predicted only 2.8% of innovatively active organizations estimated according to the 3rd Oslo Manual methodology, and ARIMA 3.15%), which stabilized and even slightly increased (estimated according to the 3rd Oslo Manual methodology to 6.29% from 5.30% in 2019). This once again proves the low quality of predictions of models of innovation during bifurcations.

At the same time, the model built with the trend change in mind was not only more accurate in interpolation (R2=0.875), but also less pessimistic in estimating the forecast for the next year (at 5.59%). In other words, the trend change model was obviously more effective in making forecasts and from the analytical point of view.

The algorithm created made it possible not only to create a model that interpolates the available data accurately enough and makes it possible to create a relatively accurate forecast; it also highlighted the trend and the three cyclical components. This has obvious analytical value, in particular, making it possible to relate the highlighted cycles to various economic processes and assess the peculiarities of the development of these processes by studying the behavior of these components.

4. Discussion

The proposed algorithm allows us to build a fairly accurate model of the dynamics of economic phenomena, which can be used both for analytical purposes and to build short-term forecasts. Unlike analogues, the main advantage of this algorithm is the possibility to automate it efficiently, as well as to extend it by adding new rules (in the phasifier) and new models (in the dephasifier). Theoretically, this may allow the user who is not fully familiar with the tools of statistical analysis and data analysis to build highly accurate models. However, as shown in the study, this requires an extended rule base, including tools for analyzing periodic fluctuations that are not sinusoidal. It also makes sense to supplement the base with rules that allow to determine the change in the trend, the appearance and disappearance of periodic fluctuations (for example, in our study the amplitude of short-term fluctuations has been sharply decreasing in the last ten years, which requires an appropriate economic explanation). Perhaps the integration of a knowledge base into the algorithm, which will be able to identify the most typical economic causes of certain types of dynamics components, will make it possible to assess the change in these components.

One of the directions of research that seems promising is the extension of this algorithm to apply it with the MSSA method, which can be effective in solving a number of economic problems. On the one hand, the super-complexity of economic systems leads to the fact that the inclusion of additional factors does not always improve the quality of the economic model, but for a number of problems (where the condition is sufficiently cleared of cross-correlation, in the presence of lag variables, in real-time estimation) the use of MSSA can be quite effective.

Another direction for further improvement of the algorithm is to build special rules to identify and simulate crises. The expected way to realize further improvement of the algorithm may be the diagnosis of significant deviations of real values from those predicted by the model (as has been shown, these deviations tend to grow rapidly during a crisis) and the construction of typical scenarios of crisis development and bifurcations on the basis of historical data. Here, too, the application of the MSSA method to solve the above questions with further automatic recognition of the obtained results seems promising.

Also, an interesting area of research is the application of the considered algorithm for "nowcasting", predictions based on real-time data, similar to [17]. As shown, for example, in works on real-time simulation of the output gap [17;22], the task of real-time simulation is associated with considerable uncertainty, however, its results are highly demanded in the decision-making process.

A special issue is the need to assess the possibility of retraining the model while applying the proposed algorithm. On the one hand, 2 to 4 para-meters are used in modeling each of the components, resulting in a very large total number of parameters. On the other hand, each of the models, in fact, is not a model of the original series, but a model of some of its components. Thus, each model contains random deviations and does not deprive the original data of degrees of freedom, so it makes sense to estimate the degrees of freedom of each component separately, in which case we can argue that model overfitting does not occur. Moreover, even the models of the first three components allow achieving sufficiently high accuracy, and adding additional components to the model makes sense from the point of view of economic analysis rather than ensuring interpolation accuracy.

In general, we can conclude that the algorithm proposed by the authors showed its sufficiently high efficiency in solving a specific economic problem, but for practical applications in various areas of the economy, it requires further development and additional adjustment, on the one hand, in the creation of new rules of recognition of economic trends, often taking into account the specific subject area, and on the other hand, towards studying the best options for preprocessing data and creating effective models (paired with the phasing tools used). The main advantage of this algorithm is that it can be easily modified due to the modular structure and flexibility of fuzzy sets tools used in this algorithm. This is especially important in the current context of striving for sustainable development, including "innovation resilience behavior", which implies ensuring the stability and adaptability of both economic systems and the methodologies used in their management [23]. The created algorithm was also tested for the tasks of modeling the GDP of the Russian Federation and the number of researchers in the Russian Federation, as a result of which qualitative models of the main economic cycles were built and a complex model that allows describing the available data and making a forecast for the next year was created. The accuracy of interpolation and extrapolation was higher than, for example, when using ARIMA method.

5. Conclusions

The study of application of the developed algorithm based on fuzzy logic methods and subsequent statistical modeling for post-processing of the components extracted by SSA showed that the proposed approach is quite effective when working with the generated data. However, when solving the economic task of modeling and forecasting the GDP of the Russian Federation, it was found that more complex rules are required when

evaluating the components extracted by SSA, as well as the use of special preprocessing of these components. The algorithm resulted in an effective descriptive model, which can be used both for analytical purposes and for forecasting purposes. However, to increase the efficiency of this algorithm in solving various economic problems, its further development, tuning and modifications are required.

The models created on the basis of the application of the proposed algorithm confirmed that innovation processes are associated with a higher level of qualitative uncertainty compared to other economic processes. The models built to describe the dynamics of the number of researchers are more complex and require the addition of fuzzy component recognition rules. Moreover, both the constructed models and ARIMA-models (as well as other models of dynamics, which were used for comparison) turned out to be much less accurate than when forecasting the dynamics of GDP of the Russian Federation. All this proves the high degree of complexity of the tasks of the analysis of innovation activity. Nevertheless, the proposed approach of combining SSA and fuzzy estimation has coped quite effectively with this difficult task as well.

Although the SSA method is widely used for solving problems in many fields, its application in economics and, especially, innovativeness, is not wide enough. In our opinion, this is due to considerable complexity of economic processes and regularities, which increases the role of the expert in application of SSA (as well as most other mathematical methods) in economics. However, the use of an expert, in particular for the recognition of the results of the SSA method, raises a number of problems related to its qualification, subjectivity, impartiality, and so on. We hope that the above algorithm, at the expense of the built-in base of fuzzy rules and possible explanation of the economic content of the selected components, which includes some kind of expert system, will help to solve this problem, and allow a wider application of the promising SSA method in further economic research.

6. Patents

Simonov A.B., Rogachev A.F. Certificate of registration of a computer program RU 2021619629. Analysis of dynamic indicators of socio-economic systems by the SSA method. Registered on 15.06.2021. Application No. 2021618587 02.06.2021.

References

1. Rogachev, A.; Darmanyan, A. Computer Modeling and Identification of Seasonal and Cyclical Components of Retrospective Data for Forecasting and Management. XIV International Scientific Conference "INTERAGROMASH 2021". Lecture Notes in Networks and Systems, 2022, vol 246. Springer, Cham. pp 276-283. https://doi.org/10.1007/978-3-030-81619-3_31

2. Pearson, A. Managing innovation: An uncertainty reduction process. In Henrz J., Walker D. [eds.] Managing Innovation, 1991, London: Sage/Oxford University Press, pp 18-27.

3. Freeman, C.; Soete, L. The Economics of Industrial Innovation; 3rd ed; The MIT Press: Cambridge, MA, USA, 1997; 480 p.

4. Rogachev, A. F.; Simonov, A. B. Systematic analysis of retrospective crop yields time series based on their structure identification. IOP Conf. Ser.: Earth Environ. Sci. 1069 012014. 2022, https://doi.org/10.1088/1755-1315/1069/1/012014.

5. Vyalkova, S.A.;Kornykova, O.; Nadtoka, I. Development of mathematical models for the short-term forecasting of daily consumption schedules of active power by Moscow 5th International Scientific and Technical Conference on Mechanical Science and Technology Update, MSTU Omsk, 2021 - https://iopscience.iop.org/article/10.1088/1742-6596/1901/1/012082/pdf

6. Alexandrov, P. I. Development of software for automatic allocation and prediction of additive components of time series in the framework of the Caterpillar -SSA. thesis... Candidate of Physical Mathematics Sciences: 05.13.18., St. Petersburg, Russian Federation, 2006; 152 p

7. Simonov, A.B.; Rogachev, A.F.; Simonova, I.E. Comparative Study of the Innovative Activity Dynamics in the Russian Federation Using the Singular Spectrum Analysis. Communications in Computer and Information Science, 2021, pp. 223–237

8. Xu, Chunhui; Zhang, Anqin; Xu, Chunchen , Chen, Yu. Traffic speed prediction: spatiotemporal convolution network based on long-term, short-term and spatial features. Applied Intelligence 2022 – T. 52, 2, pp 2224 — 2242.

9. Subanara; Sulandari, W. A comparison forecasting methods for trend and seasonal Indonesia tourist arrivals time series. International Conference on Mathematics, Computational Sciences and Statistics, IСоMCoS 2020 29 September 2020, 2020— T. 2329.   URL: https://doi.org/10.1063/5.0042130

10. Nasir, Najah;Samsudin, Ruhaidah;Shabri, Ani Pre-processing Streamflow Data through Singular Spectrum Analysis with Fuzzy C-Means Clustering 2nd Joint Conference on Green Engineering Technology and Applied Computing 2020 – Bangkok,4 February 2020 to 5 February 2020 - T 864 URL: https://iopscience.iop.org/article/10.1088/1757-899X/864/1/012085/pdf

11. Ivashchenko, M. V. ; Nikulin,V. S.; Zlobin,V. I.; Vasiliev, I. V. Modeling of the process of correcting the radio device signals parameters using the principes of the fuzzy sets theory. Proceedings of the Institute of Engineering Physics. 2011. vol. 3(21). pp. 47-51.

12. Faouzi,J. Pyts: a Python package for time series transformation and classification. 2018. URL: htpps://doi.org/10.5281/zenodo.1244152

13. FisPro. Fuzzy Inference System Professional. URL: https://www.fispro.org/en/

14. Hassani, H.; Mahmoudvand, R.; & Zokaei, M. Separability and window length in singular spectrum analysis. Comptes Rendus Mathematique,2011, 349 (17–18), 987–990.

15. Golyandina, N.; Nekrutkin, V.; Zhigljavsky, A.; (2001). Analysis of Time Series Structure: SSA and Related Techniques. Chapman & Hall/CRC. London, 2001; 320 p.

16. Hassani, H.; Mahmoudvand, R.; Zokaei, M.; Ghodsi, M. On the separability between signal and noise in singular spectrum analysis. Fluctuation and Noise Letters 2012 vol. 11. №2, 1250014 11p. URL:https://www.researchgate.net/publication/270725189_On_the_separability_between_signal_and_noise_in_singular_spectrum_analysis

17. de Carvalho,M.; Rua, A. Real-time Nowcasting the US Output Gap: Singular Spectrum Analysis at Work – International Journal of Forecasting, 2017, 33, pp 185-198 URL: https://www.maths.ed.ac.uk/~mdecarv/papers/decarvalho2017.pdf

18. Vokhmyanin,S.V. Testing the algorithm of the method «caterpillar-ssa» for reestablishing of time series (Ispytaniye algoritma metoda "Gusenitsa-SSA" dlya vosstanovleniya vremennogo ryada), Vestnik SibGAU, 2010 vol. 2(28), pp. 59-62. URL: https://www.elibrary.ru/download/elibrary_16347880_55371835.pdf

19. Fisher, R.A. Tests of significance in harmonic analysis. Proceedings of the Royal Society of London, 1929 Ser. A, 125(796), pp. 54-59

20. Yarmohammadi, M. A filter based Fisher g-test approach for periodicity detection in time series analysis. Scientific Research and Essays, 2011, 6 (17), pp. 3717–3723.

21. Federal State Statistics Service. Official Site. - URL: https://rosstat.gov.ru/

22. Edge, R. M., Rudd, J. B. (2012). Real-time properties of the Federal Reserve's output gap. Review of Economics and Statistics, 2016, vol 98(4), pp 785-791

23. Wen-Dong Lv, Dan Tian, Yuan Wei and Rui-Xue X. Innovation Resilience: A New Approach for Managing Uncertainties Concerned with Sustainable Innovation.   Sustainability (Switzerland) 2018, vol. 10, rel. 1011, issue № 3641.