DOI: https://doi.org/10.15276/aait.05.2022.17
UDC 519.226

Modeling and forecasting of nonlinear nonstationary processes based on the Bayesian structural time series

Irina A. Kalinina[1)]
ORCID: https://orcid.org/0000-0001-8359-2045; irina.kalinina1612@gmail.com. Scopus Author ID: 57198354193
Aleksandr P. Gozhyj[1)]
ORCID: https://orcid.org/ 0000-0002-3517-580X; alex.gozhyj@gmail.com. Scopus Author ID: 57198358626
[1)] Petro Mohyla Black Sea National University, 10, 68 Desantnykiv Str. Mykolaiv, 54000, Ukraine

ABSTRACT

The article describes an approach to modelling and forecasting non-linear non-stationary time series for various purposes using Bayesian structural time series. The concepts of non-linearity and non-stationarity, as well as methods for processing non-linearity's and non-stationarity in the construction of forecasting models are considered. The features of the Bayesian approach in the processing of nonlinearities and nonstationary are presented. An approach to the construction of probabilistic-statistical models based on Bayesian structural models of time series has been studied. Parametric and non-parametric methods for forecasting non-linear and non-stationary time series are considered. Parametric methods include methods: classical autoregressive models, neural networks, models of support vector machines, hidden Markov models. Non-parametric methods include methods: state-space models, functional decomposition models, Bayesian non-parametric models. One of the types of non-parametric models is Bayesian structural time series. The main features of constructing structural time series are considered. Models of structural time series are presented. The process of learning the Bayesian structural model of time series is described. Training is performed in four stages: setting the structure of the model and a priori probabilities; applying a Kalman filter to update state estimates based on observed data; application of the "spike-and-slab" method to select variables in a structural model; Bayesian averaging to combine the results to make a prediction. An algorithm for constructing a Bayesian structural time series model is presented. Various components of the BSTS model are considered and analysed, with the help of which the structures of alternative predictive models are formed. As an example of the application of Bayesian structural time series, the problem of predicting Amazon stock prices is considered. The base dataset is *amzn_share*. After loading, the structure and data types were analysed, and missing values were processed. The data are characterized by irregular registration of observations, which leads to a large number of missing values and "masking" possible seasonal fluctuations. This makes the task of forecasting rather difficult. To restore gaps in the *amzn_share* time series, the linear interpolation method was used. Using a set of statistical tests (ADF, KPSS, PP), the series was tested for stationarity. The data set is divided into two parts: training and testing. The fitting of structural models of time series was performed using the Kalman filter and the Monte Carlo method according to the Markov chain scheme. To estimate and simultaneously regularize the regression coefficients, the spike-and-slab method was applied. The quality of predictive models was assessed.

Keywords: Bayesian structural time series; forecasting, non-linearities; non-stationarity; forecast estimation

For citation: **Kalinina I. A., Gozhyj A. P. Modeling and forecasting of nonlinear nonstationary processes based on the Bayesian structural time series. *Applied Aspects of Information Technology.* 2022; Vol. 5 No. 3: 240–255. DOI: https://doi.org/10.15276/aait.05.2022.17**

INTRODUCTION

Forecasting of complex systems is one of the important areas of modern science of data analysis and processing [1, 2], [3]. One of the most developed and researched areas is forecasting based on time series. Time series data reflect the dynamic behaviour and cause-and-effect relationships of the main processes in a complex system and provide basic material for making forecasts and studying the development of the system. However, the basis for most forecasting methods is models of linear and stationary processes. But the effective prediction of future states of a real complex system based on time series remains a research challenge, mainly due to the non-linear and non-stationary dynamic behaviour of the system and their various variants in the systems under study. The problem of taking into account and processing nonlinearities and nonstationary in time series forecasting is one of the main tasks of constructing adequate predictive models of the process under study. Models of this type should include a complete representation of the dynamics of nonlinear and non-stationary systems based on observed real data.

One of the methods for processing nonlinearities and nonstationary is the Bayesian approach.

This approach means that a variety of methods are used to solve the following problems:

– construction of probabilistic-statistical models of various types (assessment of the structure and parameters) using statistical data and expert assessments;

– calculation of the final results based on the created model according to the problem statement: estimates of forecasts, control actions, estimates of variables and parameters at the output of filters, pattern recognition, finding solutions for managing the processes and objects under study, etc.;

– analysis of the correctness of the obtained results according to the corresponding sets of statistical quality criteria.

The methodology of the Bayesian approach includes the following methods: recursive Bayesian estimation: filtering, forecasting, smoothing variables; hidden Markov models; optimal recursive Kalman filters (KF); granular (particle) filters (GF); static Bayesian networks (BN); dynamic Bayesian networks (DBN); Markov localization (ML) models; Bayesian maps; Bayesian method of data processing and decision making based on hierarchical models; Bayesian regression, generalized linear models; Bayesian structural time series.

The model *Bayesian structural of time series* (BSTS) was proposed by E. Harvey [4] as a theory of structural models of time series. Unlike traditional statistical ARIMA models, structural time series models consist of unobservable components such as trends and various seasonality components. In addition, models can naturally be extended to include explanatory variables and work with multivariate time series. In the analysis of time series in the case of missing observations, state-space models and methods and recursive equations using the Kalman filter are used [5]. State-space models are based on Markov processes, since each state depends on the previous state. Accordingly, the future state is calculated based on the present. The model parameters are calculated iteratively, and this allows the development of high-dimensional models. From a technical point of view, state-space models and the Kalman filter play a key role in the statistical processing of structural time series models. The structural time series model uses the Markov Chain Monte Carlo (MCMC) sampling algorithm for posterior distribution modelling, which smoothest the predictions obtained using a large number of potential underlying models [6]. The MCMC approach using the Gibbs algorithm limits the preselection, it needs to be paired with probability or the Metropolis-Hastings algorithm to speed up convergence in multivariate models.

The article considers an approach to modelling and predicting non-linear non-stationary processes based on Bayesian structural time series.

ANALYSIS OF LITERARY DATA

Most of the prediction methods described in the literature assume linearity and/or stationarity of the underlying dynamic behavior of the system [7], or consider simple forms of non-stationarity such as well-defined trends and variations. However, real systems exhibit mostly non-linear and non-stationary behavior, which greatly complicates obtaining accurate predictions.

The main concept that characterizes real data is the concept of non-linearity. In this case, the non-linear time series *y(t)* is a signal coming from a non-linear dynamic process. In other words, this is a partial solution of a nonlinear stochastic differential (or difference) equation of the following form

$$dx/dt = F(x,\theta,\varepsilon). \quad (1)$$

This equation controls the development of the process states $x(t)$ from the initial state $x(0)$, where θ - is the process parameter vector and ε - is the system noise. The solution of equation (1) is represented as $x(t) = \varphi(x(0), t)$. Here, $\varphi\ (x(0), t)$ is called a flow or *state transition function*. Many real systems demonstrate such non-linear stochastic dynamics, and the solutions of such systems, called non-linear time series, demonstrate non-Gaussian, multimodality, time irreversibility, and other properties [8].

Most real non-linear dynamic systems operate in transient (non-stationary) conditions. From a statistical point of view, stationarity of time series *y(t)* requires that the joint distribution of each dataset $[y(t + \tau_1), y(t + \tau_2), \ldots, y(t + \tau_k)]$ be invariant τ_i $(i = 1,2, \ldots,k)$ for any k. Even under non-stationary conditions, the dynamics of a complex system can be viewed as a union of much simpler piecewise transitional or near-stationary behaviors. Most often, non-stationarity is explained by certain deterministic and stochastic tendencies in the moments.

Various non-linear and non-stationary time series forecasting methods presented in the literature are considered and classified based on how they are applied to predict time series data for real-world problems [1, 2], [3, 8], [9, 10], [11, 12]. Prediction methods can be classified based on prerequisites or approaches to overcome non-stationarity and non-linearity, as they assume the following features: a known trend shape, piecewise stationarity of signals, progressively varying parameters, or decomposability of a signal into stationary segments in the transformed domain, and they are either parametric or non-parametric, depending on whether the predictor takes a certain form or is built solely in accordance with the data (for example, the number of latent variables may vary).

Parametric models. The parametric prediction model defines an explicit functional form with a finite number of parameters θ to describe the relationship between input data consisting of internal and external variables and their autoregressive (retarded) (lag) terms, and output data consisting of future values of the internal variable $y(t + 1)$. Model parameters are estimated from time series implementations. Parametric models include: classical autoregressive models, neural networks, models of support vector machines,

Classical autoregressive models. This group includes models such as autoregressive (AR) or autoregressive moving average (ARMA), which are the most widely studied due to their application in modeling stationary processes. But they are usually unable to accurately predict the development of a non-linear and non-stationary process. Models such as autoregressive integrated moving average (ARIMA) based on the evolution of increment $\Delta y_t = y_{t+1} - y_t$ or $\Delta^2 y_t$, are sometimes used to remove/reduce first order non-stationarity. However, the difference tends to increase high-frequency noise in the time series, and more effort is required to determine the order of the ARMA model. To incorporate non-linearity into the ARMA framework, advanced models such as thresholding AR (TAR) models [9, 10], self-excited AR (SETAR) models [11], and smooth transition AR (STAR) models [12] are used. They were developed for non-linear forecasting. However, these methods are generally limited to non-linear stationary time series forecasting using local linearity assumption implicit with an autoregressive structure.

Neural networks. Neural networks (NN) are used for non-linear time series forecasting in many applications [13, 14], [15, 16], [17]. These models do not require preliminary assumptions about the form of nonlinearity and are universal approximations [16]. Non-linear feed-forward neural network (FNN) models parameterized with the back-propagation algorithm have been used to predict non-linear time series [18]. They are known to outperform traditional statistical methods such as the regression approach and the Box-Jenkins approach in a functional approximation and assume that the dynamics underlying the time series are independent of time. Feedforward neural networks FNNs with recurrent feedback connections have also been considered for time series forecasting [17]. Dynamic recurrent neural network models (RNN) make it possible to predict non-linear time series that occur in various areas [19, 20]. In [18], neural network models (RPNN) with recurrent functions were studied for predicting nonlinear signals.

Models of support vector machines. Support Vector Machine (SVM) model based forecasting methods use a class of generalized regression models such as Support Vector Machine Regression (SVR) and SVM Least Squares (LS-SVM) [22], which are parameterized using convex quadratic programming methods [23]. The SVM displays the input data x_i, which may consist of autoregressive terms of internal and exogenous variables. The scalar product of templates is expressed as a linear combination of the specified kernel functions, on the basis of which SVM are subdivided into linear, Gaussian, polynomial, based on a multilayer perceptron, or radial basis function (RBF) and the corresponding classifiers are built. A linear repressors is then built by structural risk minimization (upper bound on the generalization error), resulting in a better generalization than traditional methods [24]. In [25], the use of SVM for predicting chaotic time series was studied. They showed that SVM have higher prediction accuracy than NN models and use fewer parameters. In [25], predictors based on the least squares (LS) and RBF method were considered and a local SVM (defined in the reconstructed state space) was developed for predicting chaotic time series. Such SVM models can provide higher accuracy for long-term forecasting compared to local polynomial predictors based on LS and RBF.

Hidden Markov Models. Most of the models discussed above involve batch processing, where the model is set up and periodically updated using batches of historical data. However, dimensionality limitation due to excessive computational resources, memory requirements, and large data sizes prevent their applicability to many real-world problems, especially for online process monitoring. A variety of sequential prediction models such as hidden Markov models (HMM) [26] have been explored to overcome this limitation. Some HMM have been used to predict non-linear time series such as extended Kalman filters (EKF) [27] and particle filter (PF) models [28, 29].

Nonparametric models. Parametric models can make accurate predictions when the models are set correctly, but they become suboptimal when the underlying dynamics are unknown or undetectable. In addition, the problem of model displacements persists because the dynamics of most complex real systems are inherently non-linear and non-stationary. Non-parametric models can provide a complete representation of dynamics based on observed data, do not impose any structural assumptions, and simplify modeling efforts. Consequently, the accuracy of modeling and forecasting for non-linear and non-stationary time series is improved. However, compared to parametric models, non-parametric models typically require large datasets from which information about the underlying relationships can be efficiently derived. The most well-known non-parametric models for predicting non-linear and non-stationary time series are state-space neighborhoods, Bayesian non-parametric and functional decomposition models.

State-space based models. State-space-based approaches predict future values by selectively resampling historical observations, with the basic assumption being that future behavior changes smoothly, i.e. observations similar to the target may have similar outcomes. These models are suitable for predicting the dynamics of complex systems due to their simplicity and accuracy [30].

In the nearest neighbor resampling (KNN) approach in [31] with multiple predictor variables, each predictor was assigned an influence weight to identify nearest neighbors. [32] explored a number of approaches based on the KNN method for predicting chaotic time series, for example, the zero-order approximation (single nearest neighbor), nearest neighbors (multiple neighbors), and the model distance weighted approximation (weighted average distance of several neighbors).

For most complex dynamical systems, it is not possible to observe all relevant variables. The state space reconstructed from the time-delay embedding has a strong resemblance to the base state space, as noted in [33] and a new way to predict non-linear time series was proposed [34]. In [24, 35], [36], a local linear model of the reconstructed state space was developed for predicting chaotic time series. The predicted value of the current observation was obtained from the most recent w nesting vectors. Next, k nearest neighbors were determined within the window width w based on the recurrent property of the reconstructed state space.

In [37], a local polynomial regression model was investigated using neighbors and future evolutions in the reconstructed state space. An *ensemble model* was then implemented based on the nearest neighbor model by selecting a set of parameter combinations for the local regression model. This ensemble approach reduced the parameter uncertainties.

Functional decomposition model. Among the non-parametric models of nonlinear and non-stationary forecasting in the literature, attention has recently been paid to functional expansion models. The advantages of this type of model include local characteristic time scales and the use of an adaptive framework that does not require a parametric functional form. These models can be used to capture drifts and non-linear modes of any non-linear and non-stationary processes. Most of the models in this category are mixed or hybrid models that use the decomposition technique. Among the non-parametric decomposition methods, *empirical mode decomposition* (EMD) [38] can decompose non-stationary time series into a finite number of components called *intrinsic mode functions* (IMF), so that the evolution of each IMF can be explored individually using different time scales through classical time. Series prediction methods such as AR or ARMA models [39, 40]. Since EMD allows the original time series to be ideally reconstructed using IMF and to isolate the trend and noise components from the non-stationary process [41], this improves the accuracy of long-term forecasting.

In [42] presented a two-stage EMD model and applied it for long-term forecasting of solvency scores.

Bayesian nonparametric models. Bayesian modeling is the process of incorporating prior information to visualize the subsequent inference, i.e. estimating the conditional distribution $p(\theta|y)$ of the latent model or parameters θ given the observed time series *y(t)* [43]. Unlike other Bayesian methods, Bayesian non-parametric models assume that the hidden structure here grows with the data. In other words, Bayesian non-parametric models look for one model from an infinite set of possibilities (that is, θ can be infinite-dimensional) whose complexity (the number of estimated parameters) adapts according to the data. Among Bayesian nonparametric models, Gaussian process (GP) models have been most widely studied for time series forecasting [44]. The GP model provides not only a point estimate, but also a full forecast distribution. However, models (MP) have two major limitations, namely the computational cost of running the inverse matrix and the assumption of a stationary covariance function. Many attempts to solve these problems have been explored in the literature. Of all the solutions, the non-stationary covariance functions introduced to overcome stationarity assumptions [45, 46] are only suitable for simple non-linear and non-stationary forms such as linear trends and require additional fitting parameters.

One of the types of Bayesian non-parametric models are *Bayesian structural models of time series* (BSTS) [6]. BSTS models have a number of features and allow you to perform simulations taking into account any prior distributions: the BSTS model works with many other options (for example, asymmetric priors); BSTS models can be combined with Bayesian model averaging methods to eliminate the uncertainty associated with model selection; in the BSTS model, you can choose variables yourself. The article [47] describes in detail the Bayesian principles of time series analysis and considers various macroeconomic examples. Paper [48] describes a Bayesian structural hour series approach to predicting the unemployment rate from Google search processing data. A feature of the problem under consideration is that information on unemployment is published periodically with a delay, and search queries are processed continuously. The authors attempted to predict the unemployment rate using a combination of Bayesian structural time series and ensemble methods. The article [49] describes the principles of averaging the Bayesian model, which are used in various methods of data analysis. The article shows that averaging the Bayesian model allows you to get rid of the uncertainty caused by the choice of a non-optimal model. The presented examples make it possible to accurately assess the uncertainty in the forecasts made.

Various methods of applying Bayesian structural time series make it possible to use them to build models taking into account various prior distributions, reduce uncertainty when choosing a model, and work with different types of distributions [49, 50], [51, 52], [53, 54], [55].

The article will consider the practical aspects of the application of Bayesian structural time series in forecasting financial indicators (stock prices).

PURPOSE AND TASKS OF RESEARCH

The purpose of the article is to study the features of the application of Bayesian structural time series to solve the problems of modeling and predicting nonlinear and non-stationary processes. The objectives of the article are: to determine the approach to building probabilistic-statistical models based on Bayesian structural models of time series, as well as to analyze the various components of the BSTS model, with the help of which the structures of alternative predictive models are formed. Determine the quality of predictive models and make a forecast based on the most effective model. Obtain experimental confirmation of the effectiveness of the proposed approach.

FEATURES OF BUILDING STRUCTURAL MODELS OF TIME SERIES

Structural time series models have three key features for modeling non-linear non-stationary processes:

- The ability to determine the uncertainty in forecasts, in connection with which then to quantify future risks.
- Transparency, to understand the mechanism of the model.
- Ability to include external information for known factors when there is no relationship in existing data.

The structural time series model can be described by a pair of equations [51]. The first, the observation equation, relates the observed data y_t to a vector of latent variables α_t, which is called the "state". The second, the transition equation, describes how the latent state evolves over time:

$$y_t = Z_t^T \alpha_t + \epsilon_t, \epsilon_t \sim N(0, H_t), \quad (2)$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \qquad \eta_t \sim N(0, Q_t). \quad (3)$$

The model matrices Z_t, T_t, and R_t are *structural parameters*. They usually contain a mixture of known values (often 0 and 1) and unknown parameters. The transition matrix T_t is a square matrix, and the matrix R_t can be rectangular if some of the state transitions are deterministic. The presence of R_t in equation (3) makes it possible to work with the full-rank variance matrix Q_t, since any linear dependencies in the state vector can be moved from Q_t to R_t. Often when implemented, H_t is a positive scalar. The residuals ϵ_t and η_t are independent of each other and have a normal distribution with mean 0. It is generally accepted that a model that can be described by equations (2) and (3) is in the form of a state space. A fairly large class of models can be expressed in state-space form, including all varieties of ARIMA and VARMA models.

The main advantages of the state-space time series model are its modularity and flexibility. The independent state components can be combined by combining their Z_t observation vectors and placing other model matrices as elements of a diagonal matrix. When designing a model, this gives you considerable flexibility in choosing the components to model the trend, seasonality, regression effects, and, if necessary, other state components.

Structural time series models are one of the families of state space models. In structural models, the time series is represented as a sum of unobserved components, which can be interpreted as a trend, seasonality, predictor effects, etc.

These components serve as a kind of "building blocks" that can be combined in accordance with the problem being solved and the characteristics of the data.

As an example, we present the *basic structural model* of a time series with predictors as follows [51]:

$$\begin{aligned} y_t &= \mu_t + \tau_t + \beta^T \mathbf{x}_t + \epsilon_t, \\ \mu_t &= \mu_{t-1} + \delta_{t-1} + u_t, \\ \delta_t &= \delta_{t-1} + v_t, \\ \tau_t &= -\sum_{s=1}^{S-1} \tau_{t-s} + w_t, \end{aligned} \qquad (4)$$

where μ_t is the current trend level of the model, and δ_t is the trend growth factor. The seasonal component τ_t can be considered as a set of S dummy variables with dynamic coefficients limited by zero mathematical expectation during a full cycle of S seasons. The independent components of the Gaussian random noise are combined into the vector $\eta_t = (u_t, v_t, w_t)$. The matrix Q_T is diagonal with diagonal entries and $\sigma_u^2, \sigma_v^2, \sigma_w^2$ and H_T is the scalar σ_ϵ^2. The parameters in equation (3) are the variances σ_ϵ^2 , $\sigma_u^2, \sigma_v^2, \sigma_w^2$ and the regression coefficients β. They are subject to evaluation based on the original data. Thus, the presented model contains *trend*, *seasonality* and *regression* components. Vector $\mathbf{x}_t$ is a set of independent factors (predictors).

Bayesian Structural Time Series (BSTS) is related to the linear Gaussian model that is used in Kalman filters. The BSTS model is based on more complex mathematical principles than those used in the linear Gaussian model. The main difference is that Bayesian structural time series allows you to use existing components to build more complex models that reflect known facts or interesting hypotheses about the system. They can be used to design the structure of the model and train on the available data to evaluate the parameters of the model and see how well the model describes and predicts the behaviour of the system.

An approach to modelling and forecasting based on Bayesian structural time series is proposed, consisting of the following stages:

1. *Creation and training of a time series*. The learning process for a High Degree Bayesian Structural Time Series model consists of four steps:

• Setting structures and a priori probabilities.

• Application of the Kalman filter for state estimation.

• Application of the tongue-and-plate sampling method in the design model.

• Bayesian averaging to combine the results to make a prediction.

The flexibility of the BSTS model based on selected modular components is evident in the first two steps. What follows is a learning model for data acquisition using a Bayesian method whose parameters change over time.

2. *Modeling and forecasting*. When solving the problem of predicting a data set, several alternative BSTS models are used, based on the results of data analysis and processing. Each model is completed with components that can reflect the nature of data changes.

The BSTS models are built according to the following algorithm:

• A set of model components.

• If there are no predictors in the model, then the list of prior probabilities corresponds to the prior distribution of the standard distribution of residuals. If the model is with predictors, then the a priori search is carried out using the spike and slab method.

• Setting the number of iterations of the MCMC algorithm and parameters of the random number generator (for reproducibility of calculation results).

• Building BSTS models.

• Assessment of the quality of the model and verification of its adequacy: according to the metric, the speed of fitting models and its components, checking the autocorrelation in the residuals of the models.

All ready-made alternative BSTS models are used for comparison and evaluation in order to select the most appropriate model for the original dataset. The selected models are used for forecasting. The table 1 presents the various components of the BSTS models, using the structure of alternative predictive models [49-52].

FORECASTING USING BAYESIAN STRUCTURAL TIME SERIES

As an example of the application of Bayesian structural time series, the problem of forecasting the prices of Amazon shares is considered. The base dataset is *amzn_share*, which contains the company's stock price values at the close of trading from January 1, 2016 to May 26, 2019.

The data is part of a multivariate time series and was taken from *https://finance.yahoo.com/*.

Table 1. **Specification of model components**

No.	Component name	Presentation Form	Application features
1	Local level	$y_t = \mu_t + \epsilon_t,$ $\mu_t = \mu_{t-1} + u_t,$ $\epsilon_t \sim N(0, \sigma_\epsilon^2),\ u_t \sim N(0, \sigma_u^2).$	A typical non-stationary process corresponding to the process *"random walk with noise*".
2	Component of the autoregressive process	$y_t = \mu_t + \epsilon_t,$ $\mu_t = \sum_{i=1}^{p} \phi_i\, \mu_{t-i} + u_t,$ $\epsilon_t \sim N(0, \sigma_\epsilon^2),\ u_t \sim N(0, \sigma_u^2).$	Model parameters $\phi_1, \dots, \phi_p$ are subject to estimation. For large values of p, the "*spike-and-slab*" method is used to regularize the parameters $\phi_1, \dots, \phi_p$.
3	Local Linear Trend Component	$y_t = \mu_t + \epsilon_t,$ $\mu_t = \mu_{t-1} + \delta_{t-1} + u_t,$ $\delta_t = \delta_{t-1} + v_t,$ $\epsilon_t \sim N(0, \sigma_\epsilon^2),$ $u_t \sim N(0, \sigma_u^2), v_t \sim N(0, \sigma_v^2)$	The process of "*random walk"* describes both the dynamics of the average level of the time series μ_t,, and the coefficient of its growth δ_t.
4	Robust local linear trend component	$y_t = \mu_t + \epsilon_t,$ $\mu_t = \mu_{t-1} + \delta_{t-1} + u_t,$ $\delta_t = \delta_{t-1} + v_t,$ $\epsilon_t \sim N(0, \sigma_\epsilon^2),$ $u_t \sim N(0, \sigma_u^2), v_t \sim N(0, \sigma_v^2)$	Random fluctuations of this component obey the Student's distribution, and not Gaussian. The component is well suited for *short-term forecasts* based on time series, in which there are sharp jumps in the average level. This will make it possible to obtain more reliable forecasts in the presence of anomalous observations.
5	Component of a semi-local linear trend	$y_t = \mu_t + \epsilon_t,$ $\mu_t = \mu_{t-1} + \delta_{t-1} + u_t,$ $\delta_t = D + \phi \times (\delta_{t-1} - D) + v_t,\ \epsilon_t \sim N(0, \sigma_\epsilon^2),$ $u_t \sim N(0, \sigma_u^2), v_t \sim N(0, \sigma_v^2)$	The growth rate of the average level of the series develops in accordance with AR1. The process described by this component is more stable than the random walk process, which makes the model with this component more suitable for calculating *long-term forecasts*.
6	Seasonal component	$y_t = \gamma_t + \epsilon_t,$ $\gamma_t = -\sum_{s=1}^{S-1} \gamma_{t-s} + w_t,$ $\epsilon_t \sim N(0, \sigma_\epsilon^2),$ $w_t \sim N(0, \sigma_w^2)$	To model processes with clearly defined amplitude and frequency, the seasonal component is used, which is presented as a sum of elementary trigonometric components (*cos* and *sin*) with time-varying coefficients.
7	Component of "holidays" and other important events	$y_t = \beta_{d(t)} + \epsilon_t,$ $\epsilon_t \sim N(0, \sigma_\epsilon^2),$ $\beta_d \sim N(0, \sigma^2)$	The list of events important for the process is formed using auxiliary functions

Source: **compiled by the authors**

After loading, the structure and data types were analyzed, and missing values were processed. The data are characterized by irregular registration of observations, which leads to a large number of missing values and "masking" possible seasonal fluctuations. This makes the task of forecasting rather difficult.

To restore gaps in the *amzn_share* time series, the linear interpolation method was used. The first three missing observations were removed. In Fig. 1, the original row with filled in missing values is visualized.

Fig. 1. **Visualizing Amazon Stock Prices**
Source: **https://finance.yahoo.com/**

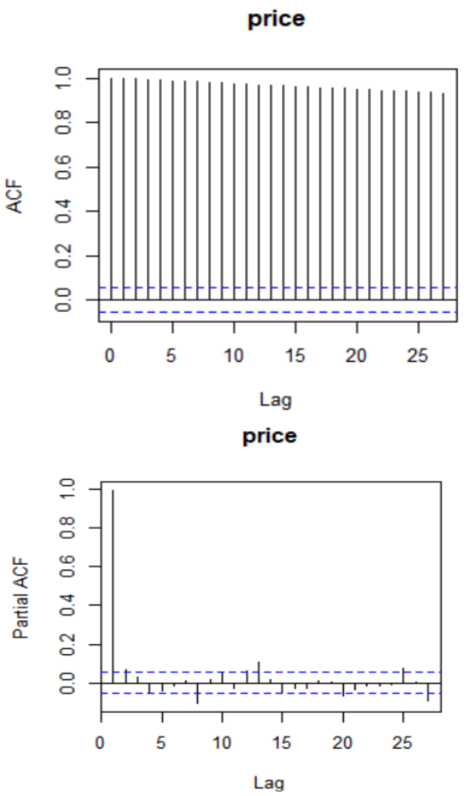

Fig. 2. **ACF and PACF functions for time series**
Source: **compiled by the authors**

Using a set of statistical tests (ADF, KPSS, PP), the original series was tested for stationarity. The result of the check was the conclusion about the non-stationarity of the process, which is reflected by a set of observed values of the time series. The non-stationarity of the process is confirmed by the nature of the values of the sample autocorrelation functions ACF and PACF (Fig. 2).

An important condition for building reliable predictive models based on the method of Bayesian structural time series is the definition and identification of the structure of the time series. The STL method was used to decompose the original series into its constituent components. Fig. 3 shows the results of the decomposition of the original series using the STL method.

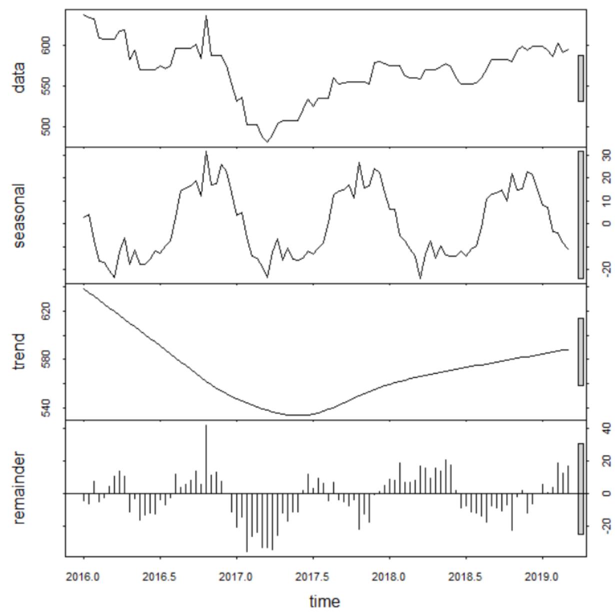

Fig. 3. **STL decomposition of the time series of Amazon stock prices**
Source: **compiled by the authors**

Viewing the data made it possible to determine the principles of modeling. First of all, it is necessary to take into account the dominant role of the trend present in the data (Fig.3), which represents non-linear and non-stationary behavior. There are also patterns that reflect the seasonal behavior of the data (Fig.3) to be reflected in the models. However, their influence is much less.

Before starting the process of building predictive models, the initial data set was divided into two parts: training and testing samples. The test sample consists of 14 observations, which corresponds to a forecast horizon of 14 days for short-term forecasting.

The fitting of structural time series models is performed using the Kalman filter and the Monte Carlo method according to the Markov chain scheme (MCMC). To estimate and simultaneously regularize the regression coefficients, the "spike-and-slab" method is used. This method consists in assigning to each regression coefficient a certain high a priori probability that it is equal to zero ("probability of inclusion" in the model). Using the original data and Bayes' theorem, the inclusion probabilities are updated. Further, during MCMC sampling of the coefficients from the obtained posterior distributions, most of the given values of the coefficients turn out to be equal to zero. Such a regularization mechanism makes it possible to effectively select the most important predictors and simultaneously get rid of multicollinearity, so a large number of predictors can be included in Bayesian structural models without the risk of overfitting.

When forming a structural model based on a preliminary analysis of the time series, various alternative models were considered to select the best one. In Table. 2 shows the studied models.

Table 2. **Description of predictive models**

Model name	Model contents
Model 1	Linear local trend component + annual seasonality component
Model 2	Linear local trend component + autoregressive component
Model 3	Linear local trend component + Weekly seasonality component
Model 4	Linear local trend component + weekly seasonality component + autoregressive component
Model 5	Component of a sustainable local linear trend + autoregressive component
Model 6	Local level component + annual seasonality component
Model 7	Local level component + autoregressive component
Model 8	Local level component + monthly seasonality component

Source: **compiled by the authors**

One of the significant advantages of the Bayesian structural model is the ability to analyze its underlying components. Figure 4 shows the mean values of the MCMC results for the trend and the autoregressive component using the Model 5 model as an example.

The BSTS model makes it possible to test the seasonal components as well. For example, the seasonal component for the days of the week is shown in Fig.5. In weekly seasons, there is a difference in prices by day of the week. However, the distribution of parameters is relatively stable in time for each day of the week.

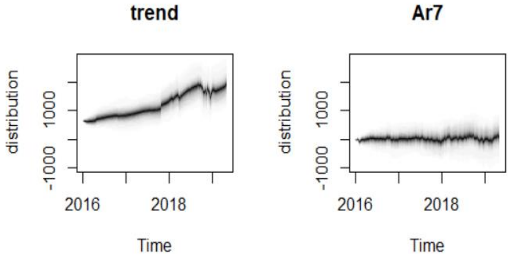

Fig. 4. **Posterior distributions of the components of the Model 5 model.**
Left: local linear trend. Right: autoregressive component
Source: **compiled by the authors**

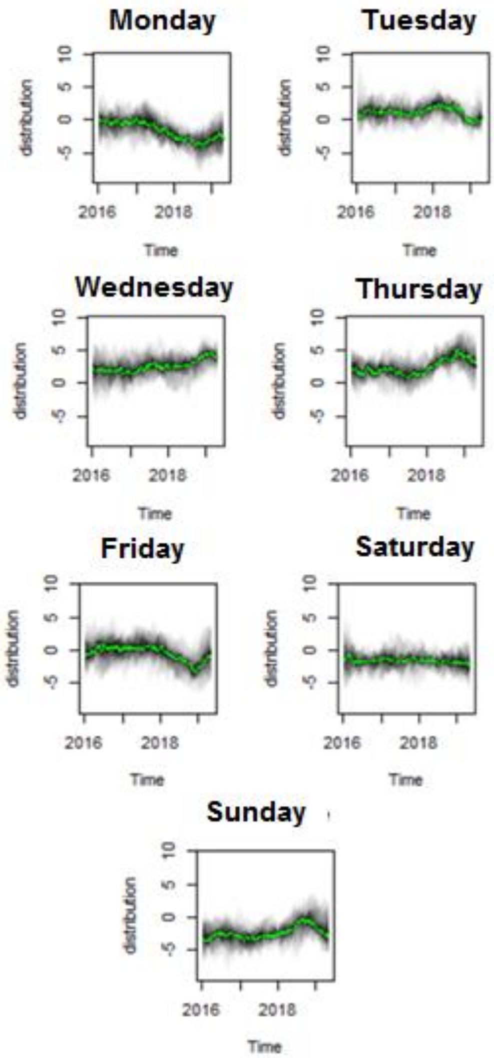

Fig. 5. **Posterior distributions of weekday effects estimated using the Model 4 model.**
The green lines correspond to the medians of these distributions
Source: **compiled by the authors**

On Figure 6 shows that Amazon's stock price was slightly higher on average on Tuesday, Wednesday, and Thursday than on other days, but the effects of each day of the week were not always consistent throughout the historical period. Nevertheless, the contribution of the seasonal component as a whole turned out to be very insignificant compared to the contribution of the trend, which is also consistent with the result of the exploratory analysis of the initial data.

An important property of a good time series model is the absence of autocorrelation in its residuals. For visual verification, a diagram was used, which was built on the basis of a matrix with model residuals and consists of range diagrams for the posterior distributions of the autocorrelation function. Ideally, the centers of these posterior distributions (starting from shift 1 onwards) should be at 0, but in the case of Model 1 this is not the case: the cyclist is clearly visible. A similar property of the original data was discovered during exploratory analysis. The box plots for the Model 7 show a high degree of model fit.

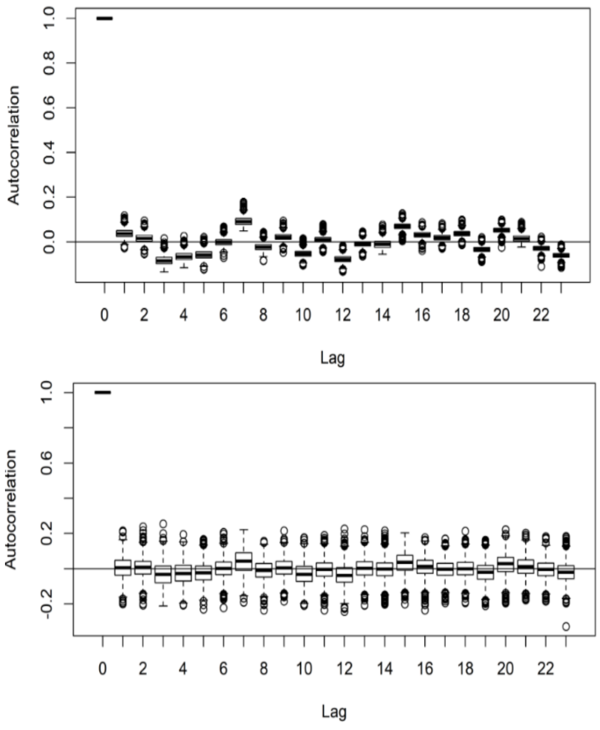

Fig. 6. **Posterior distributions of the autocorrelation function of the residuals of Model 1 and Model 7**
***Source:* compiled by the authors**

The quality of the built BSTS models is simultaneously analyzed using a graph that depicts the accumulated average absolute errors of the next step for each of the compared models (Fig. 7). Below the graph of the accumulated error curves, the original training data is shown, which allows you to better understand where exactly one or another model does not do a good job of describing the data. On Fig.7 curve of accumulated errors Model 7 is below the curve of other models, which further confirms the higher quality Model 7.

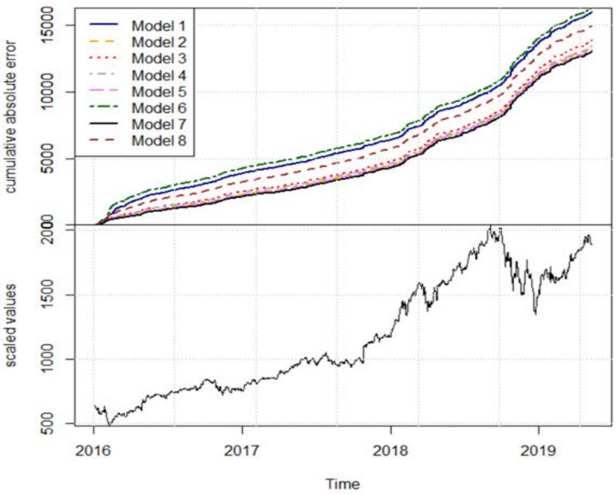

Fig. 7. **An example of comparing the quality of alternative models using the errors of the next step.**
Top: cumulative mean absolute errors of the next step.
Bottom: training dataset
***Source:* compiled by the authors**

Table 3 presents the values of the quality metrics for predictive models. In the table below, the metric *residual.sd* is the mean of the posterior distribution of the standard deviation of the model residuals, and the metric *rsquare* is the usual coefficient of determination (i.e., *the fraction that the variance of the residuals is in the total variance in the data*). The remaining two metrics are calculated using the so-called. "next step errors", which are calculated during model fitting as $y_t - E(y_t|Y_{t-1},\theta)$, where$Y_{t-1} = y_1, y_2, \dots, y_{t-1}$, and θ is a vector with current estimates of model parameters. The *prediction.sd* metric is the standard deviation of the next step errors calculated from the training data, and *relative.gof* is the so-called *Harvey's stats*. The Harvey statistic is similar to the coefficient of determination and is calculated as $R_D^2 = 1 - \sum\frac{v^2}{(n-2)\times\mathrm{var}(\mathrm{diff}(y))}$, where *v* are the errors of the next step, *n* is the number of observations *y* in analyzed time series, and *var* and *diff* are functions for calculating the variance and transition to differences (*differentiation*) of the time series, respectively. An assessment of the quality of predictive models is shown in Fig.8.

Table 3. **Evaluation of the quality of predictive models**

Model name	residual.sd	prediction.sd	rsquare	relative.gof
Model 1	7.753	21.380	0.9997	-0.152
Model 2	6.472	19.952	0.9997	-0.003
Model 3	8.021	20.191	0.9997	-0.027
Model 4	6.697	20.138	0.9997	-0.022
Model 5	5.015	19.970	0.9999	-0.005
Model 6	7.838	22.117	0.9997	-0.233
Model 7	6.483	19.859	0.9998	0.002
Model 8	6.978	20.753	0.9998	-0.087

Source: **compiled by the authors**

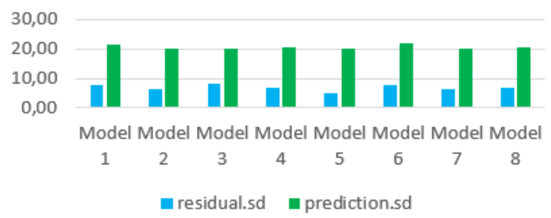

Fig. 8. **Evaluation of the quality of predictive models**
Source: **compiled by the authors**

The adequacy of the models was assessed by how well they described the training data. This approach is accompanied by a high risk of choosing an over trained model as the optimal one. Diagnosing with next-step errors is only part of the insurance against this, and the only objective test of the quality of a model will always be the accuracy of its predictions on an independent data set.

The predicted values are calculated for built-in models with the best quality indicators and are presented in Table 4. The predictions are compared with the data from the test sample. Mean absolute specific prediction error (MAPE), mean absolute error (MAE), square root of root mean square error (RMSE), and Theil U-statistics were used as metrics for selecting the optimal model. Since BSTS models predict a large number of possible realizations of future values of the dependent variable, the median values of possible realizations were used to calculate the metrics.

As follows from the above results, the Model 7 should be considered optimal.

Fig.9 shows the visualization of predictive values made on the basis of Model 7. The training data is marked with a black line. The blue line shows the most probable future values of the time series. Around this line, semi-transparent black dots also show other possible implementations of future values. The green dashed lines limit the 95% confidence interval of the predicted values. The initial data were submitted for a time period of 90 days, they were supplemented with forecast values for the next 14 days with quintiles of 5 % and 95 % highlighted. The spread of forecast values increases as the forecast period increases. Thus, the figure depicts only the last 90 observations from the training data. The yellow dots show data from the test sample, which allows you to visually assess the quality of the forecast.

Table 4. **Evaluation of the quality of the forecast**

Model name	MAPE	MAE	RMSE	U-statistics
Model 2	0.3740	686.4784	59.6242	0.0158
Model 3	0.4387	805.9569	68.7171	0.1822
Model 4	0.4164	763.9449	67.0756	0.0178
Model 5	0.3449	634.2707	52.3890	0.0139
Model 7	0.2323	431.2109	36.1843	0.0097

Source: **compiled by the authors**

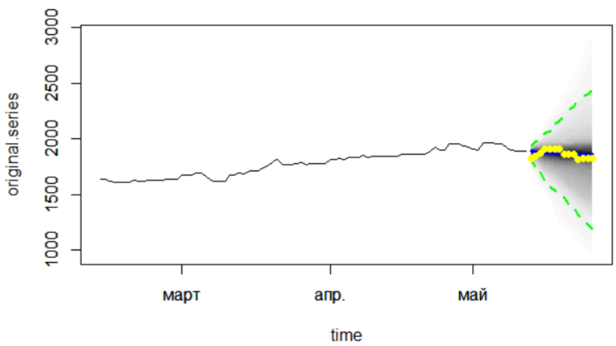

Fig. 9. **Visualization of predictive values based on Model 7**
Source: **compiled by the authors**

FEATURES AND ADVANTAGES OF SOLVING FORECASTING PROBLEMS USING THE BSTS METHOD

In the presented methodology of Bayesian structural time series, there are features that affect the process of building a model:

ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

- Possibility to specify non-standard prior distributions.
- Ability to select repressor using the "spike-and-slab" method.
- Bayesian model averaging.

Along with the presented features, the methodology of Bayesian structural time series has the following advantages:

- When building Bayesian models, a distribution is obtained. Thus, the results are returned (for example, forecasts and components) as matrices or arrays, where the first dimension contains the MCMC iterations
- Models allow modeling with any prior distributions. The default linear Gaussian model is just one variation of the classic prior distribution. Method models work with other variants of distributions (for example, asymmetric priors).
- To build methodology models, it is possible to choose variables on your own.
- Models can be combined with Bayesian model averaging techniques to eliminate the uncertainty associated with model selection.

These advantages are confirmed by the use of six data sets from different application areas for solving forecasting problems.

CONCLUSIONS

The article discusses the features of the Bayesian approach in the processing of nonlinearities and nonstationary in the construction of forecasting models using Bayesian structural time series. Parametric and non-parametric methods for forecasting time series are considered. One of the types of non-parametric models is Bayesian structural time series. An approach to the construction of probabilistic-statistical models based on Bayesian structural models of time series is defined. The main features of constructing structural time series are considered. The process of learning the Bayesian structural model of time series is described. An algorithm for constructing a BSTS model is presented. Various components of the BSTS model are considered and analyzed, with the help of which the structures of alternative predictive models are formed. As an example of the application of Bayesian structural time series, the problem of predicting Amazon stock prices is considered. The data are characterized by irregular registration of observations, which leads to a large number of missing values and “masking” possible seasonal fluctuations. This makes the task of forecasting rather difficult. To restore gaps in the *amzn_share* time series, the linear interpolation method was used. Using a set of statistical tests such as ADF, KPSS, PP, the series was tested for stationarity. The data set was divided into two parts: training and testing samples. The fitting of structural models of time series was performed using the Kalman filter and the Monte Carlo method according to the Markov chain scheme (MCMC). To estimate and simultaneously regularize the regression coefficients, the spike-and-slab method was applied. The quality of predictive models was assessed. Based on the most effective model, a forecast was made for Amazon stock prices. The application of the method of Bayesian structural time series makes it possible to effectively build forecasts taking into account the non-linearity and non-stationarity of the data.

REFERENCES

1. Bidyuk, P., Gozhyj, A., Kalinina, I. & Vysotska, V. “Methods for Forecasting Nonlinear Nonstationary Processes in Machine Learning”. *In: Communications in Computer and Information Science*, 1158. *Springer*. Cham. 2020. p. 470–485. DOI: https://doi.org/10.1007/978-3-030-61656-4_32.

2. Bidyuk, P., Gozhyj, A., Matsuki, Y., Kuznetsova, N. & Kalinina, I. “Modeling and Forecasting Economic and Financial Processes Using Combined Adaptive Models”. *In: Babichev S., Lytvynenko V., Wójcik W., Vyshemyrskaya S. (eds.) Lecture Notes in Computational Intelligence and Decision Making. ISDMCI 2020*. *Advances in Intelligent Systems and Computing*, *Springer*. Cham. 2021; 1246: 395–408. DOI: https://doi.org/10.1007/978-3-030-54215-3_25.

3. Bidyuk, P., Gozhyj, A., Kalinina, I., Vasilev, M. & Malets, R. “Forecasting nonlinear nonstationary processes in machine learning task”. *Proceedings of the 2020 IEEE 3rd International Conference on Data Stream Mining and Processing, DSMP*. 2020. p. 28–32. DOI: https://doi.org/10.1109/dsmp47368.2020.9204077.

4. Harvey, A. C. “Forecasting, Structural Time Series Models and The Kalman Filter”. Cambridge, United Kingdom. *Cambridge University Press*. 1989. DOI: https://doi.org/10.1017/CBO9781107049994.

5. Kalman, R. "A new approach to linear filtering and prediction problems". *Journal of Basic Engineering*. 1960: 82(1): 94–135. DOI: https://doi.org/10.1115/1.3662552.

6. Fragoso, T. M., Bertoli, W. & Louzada, F. "Bayesian model averaging: A systematic review and conceptual classification". *International Statistical Review*. 2017; 86 (1): 1–28. DOI: https://doi.org/10.1111/insr.12243.

7. Hyndman, R. J. & Athanasopoulos, G. "Forecasting: Principles and Practice". 3rd ed. Edition. 2021. 442 p. DOI: https://doi.org/10.4236/ojbm.2016.43042.

8. Tsay, R. S. & Chen, R. "Nonlinear Time Series Analysis". *School of Arts and Sciences, Statistics*. 2018. 512 p. DOI: https://doi.org/10.1002/9781119514312.

9. Nielsen, A. "Practical time series analysis: Prediction with statistics and machine learning". O'Reilly Media, Inc. 2019. 500 p.

10. Chen, H., Chong, T. T. L. & Bai, J. "Theory and applications of TAR model with two threshold variables: The price differences of equivalent assets1". *Econometric Reviews*. 2014; 2 (31): 142–170. – Available from: https://mpra.ub.uni-muenchen.de/54527. – [Accessed: July, 2021].

11. Karimuzzaman, S. & Hossain, M. "Forecasting Performance of Nonlinear Time Series Models: An Application to Weather Variable". *Modeling Earth Systems and Environment*. *Statistical Modelling. Springer* 2020; 6 (2): 2451–2463. DOI: https://doi.org/10.1007/s40808-020-00826-6.

12. Terasvirta, T. & Yang, Y. "Specification, estimation and evaluation of vector smooth transition autoregressive models with applications". *School of Economics and Management*. 2014. – Available from: https://repec.econ.au.dk/repec/creates/rp/14/rp14_08.pdf. – [Accessed: July, 2021].

13. Wong, W. K., Xia, M. & Chu, W. C. "Adaptive neural network model for time-series forecasting". *European Journal of Operational Research*. 2010; 207 (2): p. 807–816. DOI: https://doi.org/10.1016/j.ejor.2010.05.022.

14. Veljanovski, G., Atanasovski, M., Kostov, A. & Popovski, P. "Application of neural networks for short term load forecasting in power system of north Macedonia". *IEEE 2020 55th International Scientific Conference on Information, Communication and Energy Systems and Technology (ICEST)*. 2020. p. 99–101. DOI: https://doi.org/10.1109/ICEST49890.2020.9232674.

15. Safi, S. "A comparison of artificial neural network and time series models for forecasting GDP in Palestine". *American Journal of Theoretical and Applied Statistics*. 2016; 5 (2): 56–63. DOI: https://doi.org/10.11648/j.ajtas.20160502.13.

16. Arteaga, C. & Marrero, I. "Universal approximation by radial basis function networks of Delsarte translates". *Neural Networks*. 2013; 46: 299–305. DOI: https://doi.org/10.1016/j.neunet.2013.06.011.

17. Zhang, G. P. "Neural networks for time-series forecasting". *In Handbook of Natural Computing, Springer* Berlin Heidelberg. 2012. p. 461–477. DOI: https://doi.org/10.1007/978-3-540-92910-9_14.

18. Legaard, C. M., et al. "Constructing neural network-based models for simulating dynamical systems". *Machine Learning*. 2021. 1 (1): 1–35. DOI: https://doi.org/10.48550/arXiv.2111.01495.

19. Hrnjica, B. & Bonacci, O. "Lake level prediction using feed forward and recurrent neural networks". *Water Resources Management*: *An International Journal. Published for the European Water Resources Association (EWRA). Springer*. 2019; 33 (7): 2471–2484. DOI: https://doi.org/10.1007/s11269-019-02255-2.

20. Wang, F. "Predicting S&P 500 Market price by deep neural network and Enemble model". *E3S Web of Conferences 214:02040. EBLDM*. 2020; p. 1–4. DOI: https://doi.org/10.1051/e3sconf/202021402040.

21. Rao, K. P., et al. "Real-time identification of incipient surface morphology variations in ultra-precision machining process". *ASME Journal of Manufacturing Science and Engineering*. 2013; 136 (2): 1–11. DOI: https://doi.org/10.1115/1.4026210.

22. Smola, A. J. & Schölkopf, B. "A tutorial on support vector regression". *Statistics and Computing*. 2004; 14: 199–222. DOI: https://doi.org/10.1023/B:STCO.0000035301.49549.88.

23. Balabin, R. M. & Lomakina, E. I. "Support vector machine regression (SVR/LSSVM) – An alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data". *Analyst*. 2011; 136: 1703–1712. DOI: https://doi.org/10.1039/c0an00387e.

24. Ramadevi, B. & Bingi, K. "Chaotic time series forecasting approaches using machine learning techniques: A review". *School of Electrical Engineering, Vellore Institute of Technology, Vellore 632014, India. Special Issue Nonlinear Symmetric Systems and Chaotic Systems in Engineering*. 2022; 14(5): 955. DOI: https://doi.org/10.3390/sym14050955.

25. Sapankevych, N. I., et al. "Time series prediction using support vector machines: A survey". *IEEE Computational Intelligence Magazine*, 2009; 4 (2): 24–38. DOI: https://doi.org/10.1109/MCI.2009.932254.

26. Hanif, M., et al. "Hidden Markov model for time series prediction". *Journal of Asian Scientific Research*. 2017; 7 (5): 1–10. DOI: https://doi.org/10.18488/journal.2.2017.75.196.205.

27. Wang, Z., et al. "An extended Kalman filtering approach to modeling nonlinear dynamic gene regulatory networks via short gene expression time series". *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*. 2009; 6: 410–419. DOI: https://doi.org/10.1109/TCBB.2009.5.

28. Kong, Z., et al. "Process performance prediction for chemical mechanical planarization (CMP) by integration of nonlinear Bayesian analysis and statistical modeling". *IEEE Transactions on Semiconductor Manufacturing*. 2010; 23 (2): 316–327. DOI: https://doi.org/10.1109/TSM.2010.2046110.

29. Hinne, M., Gronau, Q.F. & Wagenmakers, E.-J. "A conceptual introduction to Bayesian model averaging". *Advances in Methods and Practices I,n Psychological Science*. 2020; 3 (2): 200–215. DOI: https://doi.org/10.1177/2515245919898657.

30. Yankov, D., et al. "Ensembles of nearest neighbor forecasts". *In Machine Learning: ECML. Springer Berlin Heidelberg, Lecture Notes in Computer Science*. 2006; 4212: 545–556. DOI: https://doi.org/10.1007/11871842_51.

31. Mehrotra, R. & Sharma, A. "Conditional resampling of hydrologic time series using multiple predictor variables: A K-nearest neighbour approach". *Advances in Water Resources*. 2006; 29 (7): 987–999. DOI: https://doi.org/10.1016/j.advwatres.2005.08.007.

32. Hamid, N. Z. A. &. Noorani, M. S. M. "Modeling of prediction system: An application of nearest neighbor approach to chaotic data". *Applied Mathematics and Computational Intelligence*. 2013; 2 (1): 137–148.

33. Bradley, E. & Kantz, H. "Nonlinear time-series analysis revisited". *ResearchGate*, 2015. p. 1–12. DOI: https://doi.org/10.1063/1.4917289.

34. Flores, J., Cedeño González, J.R., López, R. & Calderon, F. "Evolving nearest neighbor time series forecasters". *Soft Computing. Project: Time series forecasting.* 2019; 23 (8): 1039–1048. DOI: https://doi.org/10.1007/s00500-017-2822-1.

35. Maathuis, H., Boulogne, L. H., Wiering, M. A. & Sterk, A. "Predicting chaotic time series using machine learning techniques". *ResearchGate. Conference: 29th Benelux Conference on Artificial Intelligence (BNAIC'2017). At: Groningen*. 2017. p. 1–16. – Available from: https://www.researchgate.net/publication/320987683. – [Accessed: Jul, 2020].

36. Li, Q. & Lin, R.-C. "A New approach for chaotic time series prediction using recurrent neural network". *Mathematical Problems in Engineering, Hindawi.* 2016; 2016: p. 1–9. DOI: https://doi.org/10.1155/2016/3542898.

37. Regonda, S. et al. "Local polynomial method for ensemble forecast of time series". *Nonlinear Processes in Geophysics*. 2005; 12 (3): 397–406. DOI: https://doi.org/10.5194/npg-12-397-2005.

38. Rilling, G., et al. "On empirical mode decomposition and its algorithms". *In IEEE EURASIP Workshop on Nonlinear Signal and Image Processing NSIP*. 2003. p. 8–11.

39. Wang, X., et al. "Time series forecasting energy-efficient organization of wireless sensor networks". *Sensors*. 2007; 7: 1766–1792. DOI: https://doi.org/10.3390/s7091766.

40. Xu, X., et al. "Forecasting demand of commodities after natural disasters". *Expert Systems with Applications*. 2010; 37 (6): 4313–4317. DOI: https://doi.org/10.1016/j.eswa.2009.11.069.

41. Stroeer, A., et al. "Methods for detection and characterization of signals in noisy data with the Hilbert-Huang transform". *Physical Review. Data Analysis, Statistics and Probability*. 2009; 79(12): 124022. DOI: https://doi.org/10.48550/arXiv.0903.4616.

42. Sangasoongsong, A. & Bukkapatnam, S. T. "Willingness-to-pay prediction based on empirical mode decomposition". *In Industrial Engineering Research Conference*, Reno, NV. 2011.

43. Gershman, S. J. & Blei, D. M. "A tutorial on Bayesian nonparametric models". *Journal of Mathematical Psychology*. 2012; 56: 1–12. DOI: https://doi.org/10.48550/arXiv.1106.2697.

44. Rasmussen, C. E. "Gaussian processes for machine learning". *Advanced Letures on Machine Learning, Springer* Berlin Heidelberg. 2004. p. 63–71. DOI: https://doi.org/10.7551/mitpress/3206.001.0001.

45. Brahim-Belhouari, S. & Bermak, A. "Gaussian process for nonstationary time series prediction". *Computational Statistics & Data Analysis*. 2004; 47: 705–712. DOI: https://doi.org/10.1016/j.csda.2004.02.006.

46. Plagemann, C., et al. "Nonstationary gaussian process regression using point estimates of local smoothness". *In Machine learning and Knowledge Discovery in Databases*. *Springer*. Berlin: Heidelberg. 2008. p. 204–219. DOI: https://doi.org/10.1007/978-3-540-87481-2_14.

47. Scott, S. L. & Varian, H. R. "Bayesian variable selection for nowcasting economic time series", *Economics of Digitization. NBER Press.* London. 2015. p. 119–136. DOI: https://doi.org/10.7208/chicago/9780226206981.003.0004.

48. Steel, M. "Bayesian time series analysis". *In Macroeconometrics and Time Series Anaysis*. Palgrave Macmillan. 2010. p. 35–45. DOI: https://doi.org/10.1057/9780230280830_4.

49. Scott, S. L. & Varian, H. R. "Predicting the present with Bayesian structural time series". *International Journal of Mathematical Modeling and Numerical Optimisation*. 2014; 5: 4–23. DOI: https://doi.org 10.1504/IJMMNO.2014.059942.

50. Scott, S. L. "Fitting Bayesian structural time series with the BSTS R Package". *The Unofficial Google Data Science Blog*, 2017.

51. Almarashi, M. & Khan, K. "Bayesian structural time series". *Nanoscience and Nanotechnology Letters*. 2020; 12 (1): 54–61. DOI: https://doi.org/10.1166/nnl.2020.3083.

52. Qiu, J., S. Rao Jammalamadaka, S. R. & Ning, N. "Multivariate Bayesian structural time series model". *Journal of Machine Learning Research.* 2018; 19: 1–33. – Available from: http://jmlr.org/papers/v19/18-009.html. – [Accessed: Jul, 2020].

53. Ning, N. & Qiu, J. "The MBSTS package: Multivariate Bayesian structural time series models in R". *Cornell University. Statistics. Methodology*. 2021. DOI: https://doi.org/10.48550/arXiv.2106.14045.

54. Thorakkattle, M. N., Farhin, S. & Ali khan, A. "Forecasting the trends of Covid-19 and causal impact of vaccines using Bayesian structural time series and ARIMA". *Annals of Data Science*. 2022; 9: 1025–1047.

55. Shaffery, P., Yang, R. & Zhang, Y. "Bayesian structural time series for behind-the-meter photovoltaic disaggregation". *2020 IEEE Power & Energy Society Innovative Smart Grid Technology Conference (ISGT).* 2020. DOI: https://doi.org/10.1109/ISGT45199.2020.9087675.

Conflicts of Interest: the authors declare no conflict of interest

Received 16.08.2022
Received after revision 12.10.2021
Accepted 18.10.2021

DOI: https://doi.org/10.15276/aait.05.2022.17
УДК 519.226

Моделювання і прогнозування нелінійних нестаціонарних процесів на основі Байєсівських структурних часових рядів

Калініна Ірина Олександрівна[1)]
ORCID: https://orcid.org/0000-0001-8359-2045; irina.kalinina1612@gmail.com. Scopus Author ID: 57198354193

Гожий Олександр Петрович[1)]
ORCID: https://orcid.org/ 0000-0002-3517-580X; alex.gozhyj@gmail.com. Scopus Author ID: 57198358626
[1)] Чорноморський національний університет імені Петра Могили, вул. 68 Десантників, 10. Миколаїв, 54000, Україна

АНОТАЦІЯ

У статті описано підхід до моделювання та прогнозування нелінійних нестаціонарних часових рядів для різних цілей з використанням байєсівських структурних часових рядів (BSTS). Розглянуто поняття нелінійності та нестаціонарності, а також методи обробки нелінійності та нестаціонарності при побудові моделей прогнозування. Наведено особливості байєсівського підходу в обробці нелінійностей та нестаціонарності. Досліджено підхід до побудови ймовірнісно-статистичних моделей на основі байєсівських структурних моделей часових рядів. Розглянуто параметричні та непараметричні методи прогнозування нелінійних та нестаціонарних часових рядів. До параметричних методів належать методи: класичних авторегресійних моделей, нейронних мереж, моделей опорних векторних машин, прихованих марковських моделей. До непараметричних методів належать методи: моделі простору станів, моделі функціональної декомпозиції, байєсівські непараметричні моделі. Одним із видів непараметричних моделей є байєсівські структурні часові ряди. Розглянуто основні особливості побудови структурних часових рядів. Представлено моделі структурних часових рядів. Описано процес навчання байєсівської структурної моделі часових рядів. Навчання виконується в чотири етапи: завдання структури моделі та апріорних ймовірностей; застосування фільтра Калмана для оновлення оцінок стану на основі спостережених даних; застосування методу "spike-and-slab" для вибору змінних у структурній моделі; Байєсівське усереднення для об'єднання результатів для прогнозування. Наведено алгоритм побудови моделі BSTS. Розглядаються та аналізуються різні компоненти моделі BSTS, за допомогою яких формуються структури альтернативних прогнозних моделей. Як приклад застосування байєсівських структурних часових рядів розглядається задача прогнозування курсів акцій Amazon. Базовим набором даних є *amzn_share*. Після завантаження структура та типи даних були проаналізовані, а відсутні значення оброблені. Для даних характерна нерегулярна реєстрація спостережень, що призводить до великої кількості пропущених значень і «маскування» можливих сезонних коливань. Це ускладнює завдання прогнозування. Для відновлення розривів у часових рядах *amzn_share* використовувався метод лінійної інтерполяції. Використовуючи набір статистичних тестів (ADF, KPSS, PP), ряд перевіряли на стаціонарність. Набір даних розділений на дві частини: навчання та тестування. Підгонку структурних моделей часових рядів проводили за допомогою фільтра Калмана та методу Монте-Карло за схемою ланцюга Маркова (MSMC). Для оцінки та одночасної регулярізації коефіцієнтів регресії застосовано метод "spike-and-slab". Оцінено якість прогностичних моделей.

Ключові слова: Байєсівський структурний часовий ряд (BSTS); прогнозування, нелінійність; нестаціонарність; прогнозна оцінка

ABOUT THE AUTHORS

Irina A. Kalinina - PhD, Associate Professor, Associate Professor Department of Intelligent information systems Petro Mohyla Black Sea National University, 10, 68 Desantnykiv Str. Mykolaiv, 54000, Ukraine
ORCID: https://orcid.org/0000-0001-8359-2045; irina.kalinina1612@gmail.com. Scopus Author ID: 57198354193
Research field: Probabilistic-statistical methods; intelligent web technologies; machine learning; coloured Petri nets

Калініна Ірина Олександрівна - кандидат технічних наук, доцент, доцент кафедри Інтелектуальних інформаційних систем Чорноморський національний університет імені Петра Могили, вул. 68, Десантників, 10. Миколаїв, 54000, Україна

Aleksandr P. Gozhyj - Doctor of Engineering Sciences, Professor, Professor Department of "Intelligent information systems" Petro Mohyla Black Sea National University, 10, 68, Desantnykiv Str.. Mykolaiv, 54000, Ukraine
ORCID: https://orcid.org/ 0000-0002-3517-580X; alex.gozhyj@gmail.com. Scopus Author ID: 57198358626
***Research field*:** Artificial intelligence methods; machine learning; probabilistic-statistical methods; intelligent web technology

Гожий Олександр Петрович - доктор технічних наук, професор, професор кафедри Інтелектуальних інформаційних систем Чорноморський національний університет імені Петра Могили, вул. 68, Десантників, 10. Миколаїв, 54000, Україна