Литвин Олег Николаевич – доктор физико-математических наук, профессор, Украинская инженерно-педагогическая академия, г. Харьков; тел.: (057) 771-05-45; e-mail: academ_mail@ukr.net.

Lytvyn Oleg Mykolayovych – Doctor of Physical and Mathematical Sciences, Professor, Ukrainian Engineering and Pedagogical Academy, Kharkov; tel.: (057) 771-05-45; e-mail: academ_mail@ukr.net.

Нечуйвітер Олеся Петрівна – доктор фізико-математичних наук, доцент, Українська інженерно-педагогічна академія, м. Харків; тел.: (057) 771-05-45; e-mail: olesya@email.com.

Нечуйвитер Олеся Петровна – доктор физико-математических наук, доцент, Украинская инженерно-педагогическая академия, г. Харьков; тел.: (057) 771-05-45; e-mail: olesya@email.com.

Nechuiviter Olesia Petrivna – Doctor of Physical and Mathematical Sciences, Assistant Professor, Ukrainian Engineering and Pedagogical Academy, Kharkov; tel.: (057) 771-05-45; e-mail: olesya@email.com.

Кейта Катерина Володимирівна – аспірант, Українська інженерно-педагогічна академія, м. Харків; тел.: (057) 771-05-45; e-mail: chervonakate@mail.ru.

Кейта Катерина Владимировна – аспирант, Украинская инженерно-педагогическая академия, г. Харьков; тел.: (057) 771-05-45; e-mail: chervonakate@mail.ru.

Keita Kateryna Volodymyrivna – PhD student, Ukrainian Engineering and Pedagogical Academy, Kharkov; tel.: (057) 771-05-45; e-mail: chervonakate@mail.ru.

UDC 519.25

T. O. MARYNYCH, L. D. NAZARENKO, N. H. KHOMENKO

COMPARATIVE ANALYSIS OF UNIVARIATE TIME SERIES MODELING AND FORECASTING TECHNIQUES FOR SHORT-TERM UNSTABLE DATA

Проведено емпіричне оцінювання адекватності та прогнозної точності класичних лінійних моделей авторегресії та ковзного середнього, моделей експоненційного згладжування, структурних, нелінійних та непараметричних моделей для одновимірних часових рядів невеликої вибірки з чисельними відхиленнями. Запропоновано метод покращення якості ARMA моделі за рахунок включення фіктивних та пояснювальних змінних, які відтворюють інформацію щодо рідких і аномальних спостережень ряду, та відповідної корекції порядку інтегрування.

Ключові слова: часовий ряд, декомпозиція, прогноз, аномальні відхилення, модель авто регресії та ковзного середнього, експоненційне згладжування.

Проведено эмпирическое оценивание адекватности и прогнозной точности классических линейных моделей авторегрессии и скользящего среднего, моделей экспоненциального сглаживания, структурных, нелинейных и непараметрических моделей для одномерных временных рядов небольшой выборки с многочисленными отклонениями. Предложен метод улучшения качества ARMA модели за счет включения фиктивных и объясняющих переменных, отражающих информацию о редких и аномальных наблюдениях ряда, а также коррекции соответствующего порядка интегрирования модели.

Ключевые слова: временной ряд, декомпозиция, прогноз, аномальные отклонения, модель авторегрессии и скользящего среднего, экспоненциальное сглаживание.

The article summarizes the international experience in univariate time series modeling approaches and methodology. It aims to make empirical assessment of their relevance and forecasting power for short sample volatile data with numerous aberrant observations and structural breaks with the help of the time series R packages. The findings revealed the pitfalls of outliers' neglection including stationarity and model misspecification, biased parameter estimates, deterioration of residuals' properties and prediction accuracy of the models. Empirical research demonstrated the outperformance of the outlier detection methods versus robust approaches that use smaller weights for aberrant observations. We tested a method of improving the forecasting power of the ARMA models by proper identification of hidden patterns and incorporation of additional information about extraordinary events into the model. We also considered frequency domain and nonparametric methods including exponential smoothing, seasonal and trend-cycle decomposition, structural and neural networks models to make comparative forecasting diagnostics. The findings showed slightly worse accuracy of the exponential smoothing and structural state-space models for short prediction horizons and their outperformance for longer forecasting periods. Neural networks showed outstanding in-sample approximation but poor out-of-sample quality. We recommend further studying of the Bayesian regime switching models that have proven to be a comprehensive way to explore hidden patterns in data, as well as dynamic factor multivariate models that can improve explanatory and forecasting power of the time series models in various applications.

Key words: time series, decomposition, forecast, outlier, autoregressive and moving average model (ARMA), exponential smoothing.

Introduction. In the few last decades, a considerable progress was achieved in the time series analysis providing innovative theoretical and methodological tools for time series decomposition, causal inference and prediction. Advances in computer technologies, accumulation of big data sets, and increased availability of the real-time high-frequency data, induced a wide use of numerical and simulation methods, and big interest to nonlinear nonparametric techniques, developing solutions to handle various data, parameter and model uncertainties. Rapid evolution of economic environment and technologies challenged production of the machine learning methods that could be adapted to different applications and objectives. One of the fast-developing fields is financial and macroeconomic time series econometrics which methodology is intensely used in other disciplines. Its biggest issue is a choice of the optimal statistical tools and approaches to model short sample volatile data with numerous outliers and structural breaks. It has been a real challenge for time series analysis in transforming countries where accumulated data sets are small and sometimes misleading.

Publication Review. Recent publications prove increased interest to the mixed modeling techniques, which combine time series domain approach [1, 2], frequency domain approach based on spectral analysis [2, 3] and explanatory approaches [4, p. 351 – 398]. Several econometric methods have been proposed in the field of univariate time series analysis improving detection and assessment of the unobserved components and generating more accurate prediction intervals for observed data. Classical exponential smoothing, structural, and ARIMA techniques have been reconsidered in an innovative State-Space framework of dynamic linear models [4, p. 382]. Introduced first by Kalman (1960) [5] for control of linear systems, state-space representation has been widely used in estimation of the unobserved components or states of time series (level, trend, seasonal, irregular), focusing on the evaluation of uncertainty, shifts, time-varying parameters, non-normal disturbances and nonlinear dynamics of the stochastic systems [2]. To enhance the quality of the forecasts in time series analysis researchers allowed for nonlinearities, outliers' removal and bootstrapping, time-varying parameters, added lagged dependent variables and relevant predictors [3, 6, 7]. Numerous failures of theory-based models with respect to out-of-sample forecasting performance and explanatory ability induced interest to nonlinear nonparametric methods including neural networks and support vector machines, and their hybrids with classical time series techniques [8]. Increased availability of diverse types and sources of data forced innovations in multivariate time series techniques taking advantage of the "data-rich environments". Dynamic factor model was one of such solutions that helped to handle the problem of degrees of freedom summarizing information contained in many economic variables in a small number of factors [9]. Despite a vast variety of sophisticated models and techniques, experimental studies demonstrate that models which rely solely on the past observations of dependent variable, including random walk, cannot be postponed and frequently outperform the multivariate models in predictive accuracy [3, 8].

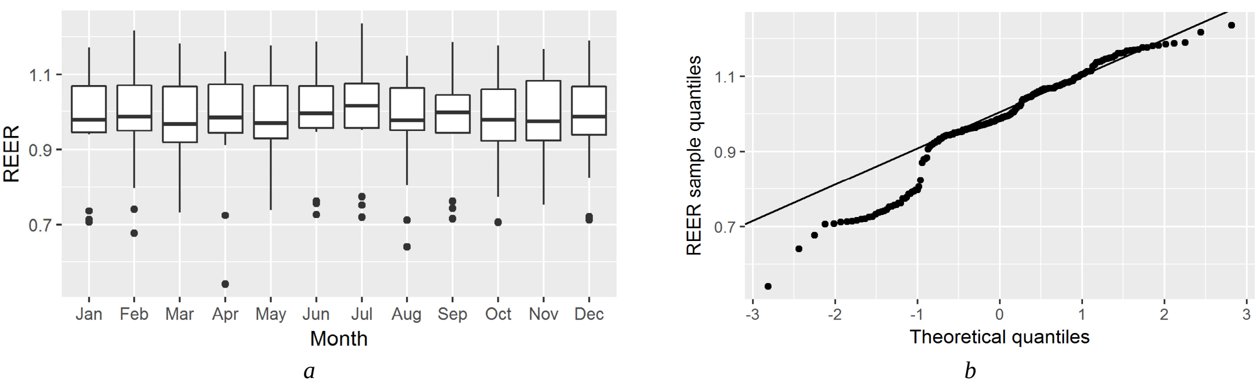
In the last decade researchers produced a set of specialized statistic software and programming languages (EViews, STATA, SAS, SPSS, R, Julia etc.) that provide automated realization of many econometric techniques and approaches. R is the most rapidly developing open source language and environment for statistical computing. This study was conducted with the help of the following R packages for time series analysis:

- "Forecast" package – provides instruments for the Box-Cox data transformation, stationarity analysis, time series decomposition, Fourier analysis, exponential smoothing and bootstrapping, autoregressive integrated moving average (ARIMA) modeling with external regressors, neural networks prediction, cross-validation etc. [10].
- "Stsm" package for structural modeling with Kalman filter – estimates prior values for model parameters, evaluates log-likelihood function by means of the Kalman filter, obtains posterior parameter distribution based on the log-likelihood function maximization, and implements iteration convergence procedure [11];
- "Ttsoutliers" package – helps in detection, estimation and model adjustment for aberrant observations [12].

Problem Statement. This paper is organized as a comparative research of the existing time series forecasting techniques, seeking possible modeling solutions of the high prediction accuracy for unstable short size data with numerous outliers. Empirical research was conducted in R-Studio integrated development environment using 205 monthly data of the Ukrainian real effective exchange rate (*REER*) and other macroeconomic variables (2000:01 – 2017:01) [13].

Modeling results. Exchange rates are explicitly one of the most popular time series used for modeling and forecasting both from the perspective of an economist and a statistician. Any time series analysis requires prior assessment of the series' patterns and distribution. The common features of the exchange rates behavior, revealed by visual inspection of different series, are as follows [7]:

- random walk movement with little or no drift;
- regime switches (from stable to volatile dynamics);
- cyclical evolution that exhibits the periods of sustained growth followed by sudden dramatic falls;
- volatility clustering (periods of tremendous changes are followed by periods of slight changes).



a *b*

Fig. 1 – Analysis of time series distribution: *a* – boxplot; *b* – qqplot.

We consider *REER* to be more representative series than the exchange rate of hryvnia to US dollar for forecasting and modeling in Ukraine. It represents the weighted average of a country's currency relative to a basket of other major currencies, adjusted to the effects of inflation [13] and resolves the problem of the currency regime switches. Analysis of the time series *REER* distribution using box and whiskers plot (*boxplot*) and quantile density function (*qqplot*) [14] (fig. 1) shows existence of numerous outliers and skewness that breaks the normal distribution assumption, assumed by parametric procedures like the $t-$test, $F-$test (ANOVA), and Pearson test. The summary-based distribution, displayed on the boxplot (fig. 1, *a*), considers observations that are 1.5 times interquartile range (*IQR*) below the first quartile to be suspected outliers, and more than 3*IQR* below the first quartile – the true outliers [15]. The skewness of the observed data can be seen in the series' deviation from the normal line on the quantile-quantile plot (fig. 1, *b*).

Decomposing time series into trend-cyclical, seasonal and irregular components gives better understanding of data patterns and can be used to improve the forecast power. Classical additive and multiplicative methods of time series decomposition based on the moving-average procedure are not robust to outliers and changes in seasonal patterns and drop beginning and closing observations of the series [2]. A nonparametric method of "seasonal and trend-cyclical decomposition using locally weighted scatterplot smoothing" (STL), developed by Cleveland et al. [16, 10], gives more efficient estimation of time series' patterns. We used the STL robust decomposition for *REER* to reveal the structure of the series, identify the properties of residuals and aberrant observations (fig. 2). It was found that the functional form of the series' components is additive:

$$Reer_t = S_t + T_t + E_t,$$

where $Reer_t$ – *REER* data at period t; S_t – seasonal component at period t; T_t – trend-cyclical component at period t; E_t – remainder (or irregular or error) component at t. Applying robust decomposition we tried to diminish the influence of outlying observations in 2008, 2009 and 2014 years (fig. 2) to provide a smoothed trend-line. In fig. 2 the trend component exhibits downward movement with cyclical lifts, the remainder component displays clustering volatility and a set of aberrant observations. The relative proportion of the *REER* components is: 7 %, 88 %, 18.5 %.

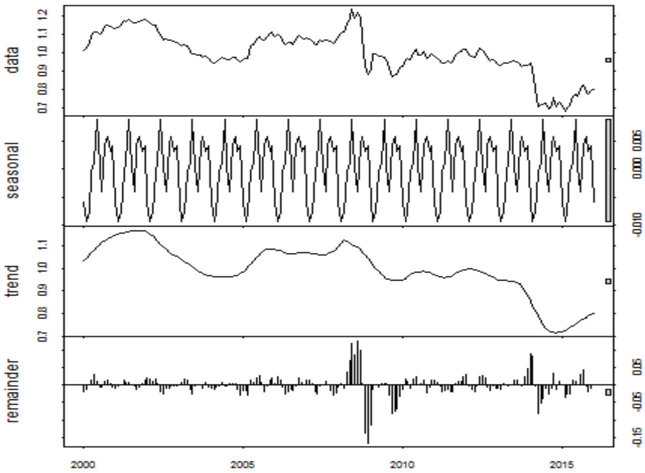


Fig. 2 – *REER* stl decomposition.

Facing the problem of inconstant variances and non-normal error terms we transformed the series using the family of "Box-Cox logarithms and power transformations". The procedure implies estimation of the parameter λ which is the power the series should be raised, to decrease the variance volatility of the residuals [3]:

$$\omega_t = \begin{cases} \log(y_t), & \text{if } \lambda = 0; \\ (y_t^{\lambda} - 1)/\lambda, & \text{otherwise.} \end{cases}$$

The estimated Box-Cox lambda for *REER* is $\lambda = -0.2$. Visual inspection of the transformed series didn't show much improvement in the observed error terms. Though, logarithmic transformation ($\lambda = 0$) did improve the approximation properties of the model.

We proceeded our analysis with outliers' detection using "tsoutliers" R package [12]. Its authors distinguish between known effects and outliers interpreting sudden changes in the data dynamics. If the prior behavior of the series is known they suggest intervention variables to be added to the model. Otherwise, the procedure of detecting and correcting the effect of outliers should be incorporated to provide correct model selection and parameter estimation [6, 12]. Depending on the influence on the subsequent series' dynamics, researchers consider five types of outliers: innovational outlier (IO), additive outlier (AO), level shift (LS), temporary change (TC) and seasonal level shift (SLS). An outlier general representation is formulated as:

$$L(B)I(t_j),$$

where $L(B)$ – is a polynomial of lag operators $By_t = y_{t-1}$ and $I(t_j)$ – is an indicator variable that takes the value 1 at time $t = j$ when the outlier jumps and the value 0 otherwise. The polynomial $L(B)$ for each type of outlier is calculate as follows [12]:

$$\text{IO}: L(B) = \frac{\theta(B)}{\alpha(B)\phi(B)};\ \text{LS}: L(B) = \frac{1}{(1-B)};\ \text{AO}: L(B) = 1;\ \text{TC}: L(B) = \frac{1}{(1-\delta B)};\ SLS: L(B) = \frac{1}{(1-B^s)}. \quad (1)$$

Herein $\theta(B)$ represents the moving average polynomial (MA); $\phi(B)$ – the autoregressive polynomial (AR); $\alpha(B)$ – the differencing parameter (d) for data stationarity; $\delta = 0.7$ in [11]; s – the periodicity of the data ($s = 12$ for monthly data).

The ARIMA model for the observed series y_t^* exposed to m outliers with weights ω can be written as [4, 12]:

$$y_t^* = \sum_{j=1}^{m} \omega_j L_j(B) I_t(t_j) + \frac{\theta(B)}{\phi(B)\alpha(B)} a_t, \quad (2)$$

where $a_t \sim \text{NID}(0, \sigma_a^2)$ – is a white-noise process with zero mean and finite variance.

Table 1 displays the results of outliers' detection and location with "tsoutliers" R package [12] by fitting ARIMA model (2) to *REER* training data (2000:01 – 2016:06) (fig. 3). Significance of all types of outliers at different points of time was verified by corresponding t-statistics (table 1).

Table 1 – Results of outliers' detection for *REER*

No	Outlier type	Observation number	Time period	Estimated coefficient	t – statistics
1	LS	64	2005:04	0.06621	4.107
2	TC	102	2008:06	0.06498	4.502
3	TC	104	2008:08	0.07452	4.554
4	AO	105	2008:09	0.04886	4.222
5	LS	107	2008:11	-0.11773	-6.429
6	LS	110	2009:02	0.06000	3.584
7	AO	176	2014:08	-0.04781	-4.754
8	AO	178	2014:10	0.05019	4.999
9	TC	182	2015:02	-0.27213	-19.184
10	LS	194	2016:02	-0.06305	-3.897

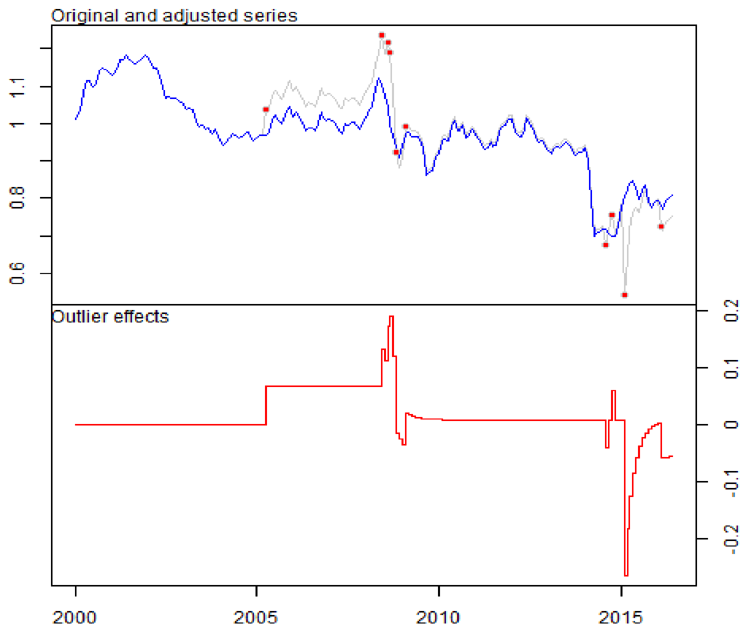


Fig. 3 – *REER* outliers (2000:01 – 2016:06).

Analysis of *REER* stationarity using the augmented Dicky-Fuller (ADF) (holds strong assumption of the error uncorrelatedness and constant variance), and the Phillipps-Perron (PP) (mitigates the assumption of the normal error distribution) unit root tests [17, p. 344 – 345], and the stationarity test of Kwiatkowski-Phillips-Schmidt-Shin (KPSS) [10] showed that the time series become stationary after the first differencing. Estimation of the autocorrelation (ACF) and partial autocorrelation (PACF) functions and further model identification applying Box and Jenkins approach [1, p. 282 – 319] pointed on the ARIMA (0, 1, 1) selection with AR order $p = 0$, order of differencing $d = 1$, and MA order $q = 1$. Removal of the estimated outliers (table 1) led to the change of ARMA parameters to ($p = 1$, $d = 0$, $q = 1$) (3) implying that *REER* becomes stationary in levels (fig. 3):

$$Reer_t = 95.926 + 96.88 Reer_{t-1} + 49.31 u_{t-1} + 6.62\text{LS}64 + 6.5\text{TC}102 + 7.45\text{TC}104 + 4.89\text{AO}105 - \\ -11.77\text{LS}107 + 6\text{LS}110 - 4.78\text{AO}176 + 5.02\text{AO}178 - 27.21\text{TC}182 - 6.31\text{LS}194 + a_t . \quad (3)$$

The constant in (3) stands for the series' mean value; autoregressive and moving average components account for the autocorrelation properties; outlier dummies comprise the aberrant series' movements. To check the prediction accuracy of the model we split *REER* observations into the training and testing sets of different lengths (7 months and 12 months). Although the ARIMA model (3) was fitted to satisfy strong assumptions of the white noise residuals, its forecasting performance deteriorated with increase of the prediction horizon and corresponding shortening of the training set. Table 2 summarizes modeling results based on both raw data multiplied by 100 and its logarithmic transformation. Model (3), based on the training set 2000:01 – 2016:06, had the lowest maximum likelihood estimator (MLE) of the innovations variance:

$$\sigma^2 = \sum (y - \hat{y})^2 / number\ of\ residuals ,$$

the highest log likelihood value (logarithm of the probability of the observed data from the estimated model [2]) and the smallest Akaike information criterion, adjusted for the sample size (AICc). Training models 1 and 2 in Table 2 showed good approximation features in terms of the mean absolute error (MAE) and the mean absolute percentage error (MAPE) indicators. The out-of-sample median forecast accuracy by MAPE for the raw data model worsened from 1.43 % to 3.19 %, and 10.78 % with increase of the prediction horizon. Logarithm transformation led to the change in the model parameters – ARIMA (2, 0, 0) and detected outliers that improved the forecasting accuracy of the model to 1.86 % MAPE for the 7-month test period, but deteriorated the corresponding results for the 12-month forecast horizon to 12.52 %.

Table 2 – Summary and forecast accuracy of the ARMA model with removed outliers (3)

No	Training model (Testing set)	MLE of the variance σ^2	Log likelihood	Information criteria		Training model accuracy		Forecast accuracy (Testing set)	
				AICc	BIC	MAE	MAPE	MAE	MAPE
1.	2000:01 – 2016:06 (2016:07 – 2017:01)	0.0003532	510.74	-991.18	-947.43	1.3650	1.433	2.3637	3.188
	Log transformed	0.0005057	473.83	-926.48	-894.77	1.0157	1.4087	1.0271	1.8601
2.	2000:01 – 2016:01 (2016:02 – 2017:01)	0.0003552	496.89	-965.75	-925.37	1.3718	1.423	7.9412	10.783
	Log transformed	0.0005073	461.1	-903.22	-874.84	1.0158	1.4089	1.1150	12.5238

We tried to improve the forecast power of the model using intervention variables that account for the movements of REER in different periods of time instead of removing the outliers. The first LS outlier (Table 1) refers to the hryvnia appreciation in April 2005, ruled by the government. The following three outliers (102^{nd}, 104^{th} and 105^{th} observations) exhibit the following short-term strengthening of the national currency in 2008 due to the favorable conjuncture of foreign markets. The 107^{th} (LS), 176^{th} (AO), 182^{nd} (TC) and 194^{th} (LS) outlier observations refer to the hryvnia devaluation in 2008 because of the world financial crisis, in 2014 and following years because of political and macroeconomic instability in Ukraine. We introduced a set of dummy binomial variables to accommodate the above-mentioned outlier effects. They include three structural break variables which incorporate upward level shift starting April 2005, and downward shifts beginning November 2008 and March 2014; dummy variable *dreer* that takes value 1 for additive downward outliers; and seasonal variable *dseas* that accounts for series' increase observed every June. To reduce variance volatility, we removed temporary change outliers No 102 – 105 and 182 (Table 1) using moving average procedure. Identification of the model parameters both by the Box-Jenkins approach [1] and the Hyndman-Khandakar algorithm [2, 10] proved the hypothesis of misspecification of the unit root and stationary tests in case of the outliers' existence, specifying the model ARMA (1, 0, 0) for ln(*REER*) based on additional information on aberrant observations:

$$\ln(Reer_t) = \underset{(0.036)}{4.61} + \underset{(0.019)}{0.96}\ln(Reer_{t-1}) + \underset{(0.018)}{0.06}d2005 - \underset{(0.019)}{0.12}d2008 - \underset{(0.019)}{0.13}d2014 - \underset{(0.005)}{0.05}dreer + \underset{(0.003)}{0.01}dseas + \varepsilon_t . \quad (4)$$

Errors of the estimated coefficients given in brackets in (4) justify significance of all variables by $t-$statistics (estimated coefficient divided by the coefficient error). Diagnostic testing for residual autocorrelation with the Ljung and Box test [4, p. 328 – 329] and the ACF plot showed improved statistics that demonstrated that the residuals were uncorrelated and the model was specified correctly after complete description of outliers with dummy variables (Table 3). Inclusion of additional explanatory variable (*Xreg*) – logarithm of the ratio of the monetary base to the value of international reserves, improved both in-sample and out-of-sample forecast accuracy, and characteristics of the model adequacy (MLE, LL, AIC, BIC, LB). We compared ARMA models of orders (1, 0, 0) and (1, 0, 1) described in (4) and (3) to ARIMA model (0, 1, 0) without any regressors, similar to the random walk model without drift. Our findings revealed better short-term approximation of the ARMA model, and more accurate long-term forecasting performance of the ARIMA model that considered the series as nonstationary in presence of level shifts and additive outliers (Table 4).

Table 3 – Summary and forecast accuracy of the ARMA model with dummy variables (4)

No	Training model (Testing set)	MLE of the variance σ^2	Log likelihood (LL)	Information criteria		Training model accuracy		Forecast accuracy (Testing set)	
				AICc	BIC	LB stat	MAPE	MAE	MAPE
1.	Full sample size 2000:01 – 2017:01	0.0003498	527.06	-1037.38	-1011.53	$Q=17.691$ $p=0.397$	1.3047		
2.	2000:01 – 2016:06 (2016:07 – 2017:01)	0.0003494	509.21	-1001.66	-976.12	$Q=18.061$ $p=0.385$	1.3330	1.2756	1.6769
	with *Xreg*: ln(*mb/rez*)	0.0003298	515.99	-1010.81	-979.1	$Q=15.102$ $p=0.444$	1.3167	1.1927	1.5651
3.	2000:01 – 2016:01 (2016:02 – 2017:01)	0.0003379	500.47	-981.96	-953.58	$Q=18.614$ $p=0.351$	1.3103	4.8390	6.5137
	with *Xreg*: ln(*mb/rez*)	0.0003227	504.69	-990.4	-962.02	$Q=17.388$ $p=0.361$	1.3048	4.6103	6.2101

Empirical estimations prove that to facilitate accurate forecasting with ARMA (4) we need to provide reliable scenarios for explanatory regressors. These predictions should be based on the estimation of the unexpected component of the news announcements rather than on their historic values [7]:

$$n_{k,t} = (A_{k,t} - E_{k,t})/\sigma_k , \quad (5)$$

where $n_{k,t}$ – is a standardized news component for the indicator k up to time t; $A_{k,t}$ – the actual value of the variable in the time period (t; $t+1$); $E_{k,t}$ – the corresponding expected value of the variable; σ_k – variance of the numerator.

Having deduced the ways to facilitate the usage of the time domain ARMA models for unstable short-term data, we tried to compare their performance with other modeling techniques described in publication review. Table 4 summarizes in-sample and out-of-sample accuracy for the following types of models, realized in "forecast" R package [10]: random walk; ARIMA and ARMA with detected outliers (3), (4); seasonal and trend decomposition using Loess (stl); exponential smoothing; exponential smoothing state space model with Box-Cox transformation, ARMA errors, trend and seasonal components (tbats); exponential smoothing with drift (theta); structural state-space MLE models based on a decomposition of the series into a number of components; feed-forward neural networks with a single hidden layer and lagged inputs for forecasting univariate time series (neural). All models were estimated using logarithmically transformed series. Neural network model and ARMA model (4) with dummy variables and explanatory variable *mb/rez* showed the best in-sample accuracy for both training models, ending 2016:06 and 2016:01. The best out-of-sample performance was obtained by the ARMA model (4) and STL model for 7-month horizon period and by the structural and theta models for 12-month horizon period. Although, ARMA model with additional regressors displayed lower bounds of the forecast intervals assuming no correlation and normal distribution of residuals.

Table 4 – Comparative analysis of the forecasting power for different models

Model	log data 2000:01 – 2016:06				log data 2000:01 – 2016:01			
	Train		test (7 month)		Train		test (12 month)	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
randomwalk	1.0223	1.6402	1.0248	1.7628	1.022	1.628	1.069	4.7266
ARIMA(0,1,0)	1.0219	1.6262	1.0251	1.7719	1.0219	1.6248	1.069	4.7266
ARMA (3)	1.0157	1.4087	1.0271	1.8601	1.0158	1.4089	1.1150	12.5238
ARMA (4)	1.0147	**1.3167**	1.1927	**1.5651**	1.0149	**1.3048**	4.6103	6.2101
Stl	1.0221	1.6338	1.0236	**1.7194**	1.0221	1.6338	1.0744	5.3075
Expsmooth	1.2478	1.6431	1.0249	1.7656	1.0221	1.6306	1.0686	4.6834
Tbats	1.0219	1.6267	1.0247	1.7590	1.0216	1.6132	1.0645	4.2845
Theta	1.023	1.6652	1.0285	1.9096	1.0227	1.6508	1.0606	**3.9287**
Structural	1.6944	11.827	1.0299	1.9717	1.6849	11.678	1.0542	**3.4184**
Neural	1.0046	**1.105**	1.1157	12.655	1.0044	**1.0988**	1.1556	28.742

Prospects for further research. The authors consider promising research methods for time series with aberrant data related to nonlinear and Bayesian regime switching models that have proven to be a comprehensive alternative way to explore hidden patterns in data. For long-term forecasting improvement, the authors recommend further research of nonlinear techniques, including structural and innovative exponential smoothing methods. Another perspective direction concerns multivariate modeling focusing on the ways to describe and predict exogenous explanatory variables and incorporate more significant information without penalty in degrees of freedom. In this respect, we recommend dynamic factor and diffusion index modeling for further research.

Conclusion. A method of improving the forecasting power of the classic time domain autoregressive and moving average model by proper identification and incorporation of explanatory information about aberrant events was implemented in the paper. It was proven that outlier detection methods resulted in narrower forecast intervals and outperformed robust approaches that used low weights for aberrant observations in short-term point predictions. Based on the comparative modeling analysis that revealed improved long-term forecasting characteristics of nonparametric nonlinear approaches, the authors suggested further research perspectives in time series analysis of unstable short sample data.

References

1. *Box G., Jenkins G.* Time Series Analysis: Forecasting and Control. – San Francisco : Holden-Day, 1970. – 575 p.
2. *Hyndman R., Athanasopoulos G.* Forecasting : principle and practice. – Available at : http://otexts.com/fpp/. – Accessed : 25 February 2017.
3. *Bergmeir C., Hyndman R. J., Benitez J. M.* Bagging exponential smoothing methods using STL decomposition and Box-Cox transformation // International journal of forecasting. – 2016. – No. 32. – pp. 303 – 312.
4. *Woodward W. A., Gray H. L., Elliott A. C.* Applied time series analysis. – New York : Taylor & Francis Group, 2012. – 540 p.
5. *Kalman R. E., Bucy R. S.* New results in linear filtering and prediction theory // Journal of Basic Engineering. – 1961. – No. 83. – pp. 95 – 108.
6. *Chen Ch., Liu L.* Joint Estimation of Model Parameters and Outlier Effects in Time Series // Journal of the American Statistical Association. – 1993. – No. 88 (421). – pp. 284 – 297.

7. *Moosa I. A., Bhatti R. H.* The theory and empirics of exchange rates. – Singapore : World Scientific Publishing Co., 2009. – 483 p.
8. *Gupta P., Batra S., Jayadeva*. Sparse Short-Term Time Series Forecasting Models via Minimum Model Complexity. – Neurocomputing, 2017, doi: 10.1016/j.neucom.2017.02.002.
9. *Barhoumi K., Darne O., Ferrara L.* Dynamic factor models: a review of the literature // OECD Journal: Journal of Business Cycle Measurement and Analysis. – 2014. – Vol. 2013/2. – pp. 73 – 107.
10. *Hyndman R., O'Hara-Wild M., Bergmeir C., Razbash S., et. al*. Forecasting Functions for Time Series and Linear Models. – R package "forecast" version 8.0, 2017. – Available at : http://github.com/robjhyndman/forecast. – Accessed : 27 February 2017.
11. *L´opez-de-Lacalle J.* Structural Time Series Models (stsm). – R package version 1.9, 2016. – Available at : https://cran.r-project.org/web/packages/stsm/stsm.pdf. – Aaccessed : 17 February 2017.
12. *L´opez-de-Lacalle J.* Detection of Outliers in Time Series (tsoutliers). – R package version 0.6-5, 2016. – Available at : https://cran.r-project.org/web/packages/tsoutliers/tsoutliers.pdf. – Accessed : 15 February 2017.
13. National Bank of Ukraine (2000 – 2017). Statistical data. – Available at : http://bank.gov.ua/. – Accessed : 15 February 2017.
14. *Wickham H., Chang W.* Elegant Data Visualisations Using the Grammar of Graphics. – R package 'ggplot 2' version 2.2.0, 2017. – Available at : http://docs.ggplot2.org/current/. – Accessed : 13 February 2017.
15. *Chambers J. M., Cleveland W. S., Kleiner B., Tukey P. A.* Graphical Methods for Data Analysis. – New York : Wadsworth & Brooks/Cole, 1983. – 105 p.
16. *Cleveland R. B., Cleveland W. S., McRae J., Terpenning I.* STL: A seasonal-trend decomposition procedure based on loess // Journal of Official Statistics. – 1990. – No. 6 (1). – pp. 3 – 73.
17. *Asteriou D., Hall S. G.* Applied Econometrics (2nd Edition). – New York : Palgrave Macmillan, 2011. – 499 p.

Received 17.03.2017

Бібліографічні описи / Библиографические описания / Bibliographic descriptions

Компаративний аналіз методів моделювання та прогнозування нестабільних часових рядів короткої вибірки / Т. О. Маринич, Л. Д. Назаренко, Н. Г. Хоменко // Вісник НТУ «ХПІ». Серія: Математичне моделювання в техніці та технологіях. – Харків : НТУ «ХПІ», 2017. – № 6 (1228). – С. 63 – 69. Бібліогр.: 17 назв. – ISSN 2222-0631.

Сравнительный анализ методов моделирования и прогнозирования нестабильных временных рядов короткой выборки / Т. А. Маринич, Л. Д. Назаренко, Н. Г. Хоменко // Вісник НТУ «ХПІ». Серія: Математичне моделювання в техніці та технологіях. – Харків : НТУ «ХПІ», 2017. – № 6 (1228). – С. 63 – 69. Бібліогр.: 17 назв. – ISSN 2222-0631.

Comparative analysis of univariate time series modeling and forecasting techniques for short-term unstable data / T. O. Marynych, L. D. Nazarenko, N. H. Khomenko // Bulletin of National Technical University «KhPI» Series: Mathematical modeling in engineering and technologies. – Kharkiv : NTU «KhPI», 2017. – № 6 (1228). – pp. 63 – 69. Bibliog.: 17 titles. – ISSN 2222-0631.

Відомості про авторів / Сведения об авторах / Information about authors

Маринич Тетяна Олександрівна – кандидат економічних наук, старший викладач кафедри прикладної математики та моделювання складних систем, Сумський державний університет, м. Суми; тел.: (066) 66-179-38; e-mail: t.marynych@ssu.edu.ua.

Маринич Татьяна Александровна – кандидат экономических наук, старший преподаватель кафедры прикладной математики и моделирования сложных систем, Сумской государственный университет, г. Сумы; тел.: (066) 66-179-38; e-mail: t.marynych@ssu.edu.ua.

Marynych Tetyana Olexandrivna – Candidate of Economic Sciences (Ph. D.), Sumy State University, Senior Lecturer at the Department of Applied Mathematics and Complex Systems Modeling, Sumy; tel.: (066) 66-179-38; e-mail: t.marynych@ssu.edu.ua.

Назаренко Людмила Дмитрівна – старший викладач кафедри комп'ютерних наук, Сумський державний університет, м. Суми; тел.: (099) 243-00-41; e-mail: nazarenkold@ukr.net.

Назаренко Людмила Дмитриевна – старший преподаватель кафедры компьютерных наук, Сумской государственный университет, г. Сумы; тел.: (099) 243-00-41; e-mail: nazarenkold@ukr.net.

Nazarenko Lyudmyla Dmytrivna – Senior Lecturer at the Department of Computer Science, Sumy State University, Sumy; tel.: (099) 243-00-41; e-mail: nazarenkold@ukr.net.

Хоменко Наталія Григорівна – аспірант, Сумський державний університет, м. Суми; тел.: (066) 950-35-07; e-mail: homenko_nata@ukr.net.

Хоменко Наталия Григорьевна – аспирант, Сумский государственный университет, г. Сумы; тел.: (066) 950-35-07; e-mail: homenko_nata@ukr.net.

Khomenko Nataliya Hryhorivna – graduate student, Sumy State University, Sumy; tel.: (066) 950-35-07; e-mail: homenko_nata@ukr.net.