

Predictive autoregressive models using macroeconomic variables: the role of oil prices in the Russian stock market

N. G. Bagautdinova^{a*}, E.I Kadochnikova^b and A.N. Bakirova^c

^aDepartment of Economics Theory and Econometrics, Institute of Management, Economics and Finance, Russian Federation

^bDeputy Director of the Higher School of Information Technologies and Information Systems, Higher School of Information Technology and Information Systems, Kazan Federal University, Russian Federation

^cKazan Federal University, Russian Federation

CHRONICLE

Article history:

Received: February 27, 2021

Received in revised format:

April 29 2021

Accepted: May 21, 2021

Available online:

May 25, 2021

Keywords:

Macroeconomic variables

Autoregressive models

Russian stock market

ABSTRACT

This article evaluates the relationship of macroeconomic variables of the domestic market with the stock index on the Moscow exchange and selects forecast specifications based on an integrated autoregressive model - the moving average. The methods used are included in an integrated autoregressive-moving average model with exogenous variables and seasonal component, Box and Jenkins approach, auto.arima in R function, Hyndman and Athanasopoulos approach, and maximum likelihood method. The results demonstrate that the inclusion of external regressors in the one-dimensional ARIMAX model improves its predictive characteristics. Time series of macro-indicators of the domestic market – the consumer price index, the index of output of goods and services for basic activities are not interrelated with the index of the Moscow exchange, with the exception of the dollar exchange rate. The positive correlation between the Moscow exchange index and macro indicators of the world economy - the S&P stock index, the price of Brent oil, was confirmed. In models with minimal AIC, a rare presence of the MA component was found, which shows that the prevailing dependence of the stock market yield on previous values of the yield (AR component) and thus, better predictability of the yield. It has shown that for stock market forecasting, "manual" selection of the ARIMA model type can give better results (minimum AIC and minimum RMSE) than the built-in auto.arima algorithm in R. It is shown that from a practical point of view, when selecting forecast models, the RMSE criterion is more useful for investors, which measures the standard error of the forecast in points of the stock index. For the scientific novelty, using Russian financial data for the period from March 2000 to March 2018 to measure the connection of macro indicators of domestic and global markets with the Moscow exchange stock index, considering seasonality can be noticed. The comparison of the forecast model's accuracy of the ARIMA type obtained by automatic and "manual" selection by AIC and RMSE is performed in favor of "manual" selection. It could be noted that the main conclusions of the article can be used in scientific and practical activities in the stock markets as a practical significance.

© 2021 by the authors; licensee Growing Science, Canada

1. Introduction

In recent years, the Russian financial market has been characterized by a relatively low capitalization of the stock market. The price/profit ratio of the Russian stock market is four times lower than the US one and two times lower than the Chinese one,

* Corresponding author.

E-mail address: shukribinbakar@gmail.com (N. G. Bagautdinova)

which characterizes the extremely high demand of investors for a premium for Russian risk in the current geopolitical conditions, while at the same time the investment climate and confidence in corporate governance in public joint-stock companies are low. Placement of shares and bonds significantly lags behind Bank lending in terms of the amount of attracted monetary resources. Since the closure of many external sources of Finance and lower commodity prices increase the focus on domestic sources of Finance, the task of stimulating domestic investors and creating favorable conditions for their activities in the financial market comes to the fore (Medvedeva et al., 2016). One of the qualitative directions of its solution can be the formation of methods for analytical forecast estimates of the dynamics of stock indexes. Stock indexes are among the first to react to both positive and negative phenomena occurring in the economy. This highlights the importance of using macroeconomic indicators in their forecast, which is of considerable practical interest.

The dynamics of stock indexes as indicators of the General state of the economy and stock prices have been the subject of numerous discussions in the scientific literature. A significant role in the empirical confirmation of theoretical arguments in favor of the influence of macroeconomic factors on stock indexes was played by the Fama (1981), in which the author argued the following point of view. If the real return on equity is positively related to measures of real activity (capital expenditure, average real rate of return on capital and output) that reflect expected cash flows from investment, then there is a negative relationship between inflation and real activity, which is interpreted in the context of money demand theory and quantitative theory of money. And when these assumptions about the nature of relationships are fulfilled, a negative impact of inflation on real stock prices is expected, called by the author the "proxy effect", which is consistent with the idea of rational expectation, when commodity and securities markets set current prices based on forecasts of the corresponding real variables. This result is quite surprising, since stock market returns should provide a hedge against inflation. Nevertheless, it follows from (Fama, 1981) that stock returns and inflation are endogenous variables that respond to General shocks. The result that was obtained in the work "Stock Returns, Real Activity, Inflation and Money" (1981), were developed by Fama in the work "Stock Returns, Expected Returns and Real Activity" (1990). The author showed that the determinants of stock returns are the expected profit and growth rates of production as an indicator of future cash flows.

Using the conclusion obtained in the article (Fama, 1981), Wasserfallen (1989) used the ARIMA model to separate expected and unexpected components in the observed time series in order to account for unforeseen changes in macro-variables that affect the nominal return on equity: real gross national product, consumption, investment, wages, industrial production, unemployment rate, inflation, and money supply. The author has shown that the impact of macroeconomic news on the stock markets of Great Britain, West Germany and Switzerland for the period 1977-1985 is very small. A similar point of view is presented in the article by Morelli (2002), the author of which showed that the volatility of macroeconomic variables on monthly UK data does not explain the volatility in the stock market.

The globalization of the modern economy creates macro-indicators of the external market that can influence the stock market. In the research of Jones and Kaul (1996), Basher et al. (2012) empirically, it was shown that in developed markets there was a negative relationship between oil prices and the value of shares, while in developing markets there was a positive relationship. In the work of Hayo and Kutan (2002) they investigated the relationship between the American and Russian stock indices.

In order to find the most appropriate model, the paper presents a fairly simple integrated autoregression model - the moving average (ARIMA), its extension – the ARIMAX model with the inclusion of external macroeconomic variables. The models are based on four macroeconomic time series of the Russian domestic market along with the time series of the world market – the S&P stock index, Brent oil prices, for the period from March 2000 to March 2018. According to the prevailing opinion in the econometric literature, RMSE statistics are used to evaluate the predictive qualities of models. ARIMA-type models are classic in obtaining predictive estimates. The ARIMA model for short-term forecasts often shows good results (Bashiri Behmiri et al., 2013; Kadochnikova et al., 2019(a); Kadochnikova et al., 2019(b)) and, thanks to the automatic parameter selection procedure in R, is convenient for forecasting (Hyndman & Khandakar, 2008; Hyndman & Athanasopoulos, 2013).

The main objective of this article is to find the most appropriate monthly levels of the Moscow exchange index based on the comparison of predictive qualities of various autoregression models obtained by automatic and manual selection in the R software environment. The research idea was suggested by Fama (1990), Wasserfalle (1989), Hyndman and Khandakar (2008), Hyndman and Athanasopoulos (2013).

The following results were obtained in the work. For stock market forecasting, the "manual" selection of parameters for ARIMA and ARIMAX models does not exclude better results than the built-in auto.arima algorithm in R. The inclusion of external regressors in the one-dimensional ARIMAX model improves its predictive characteristics. Time series of macro-indicators of the domestic market – the consumer price index, the index of output of goods and services for basic activities are not interrelated with the index of the Moscow Exchange; the exception is the dollar exchange rate, which, although formed on the domestic market, is influenced by a combination of factors of the world market.

The paper includes an introduction, three main sections, and a conclusion. The first section provides an overview of the literature regarding the selection of macroeconomic variables that affect the dynamics of the stock index. In the second section,

the ARIMA and ARIMAX models used are formulated, and the macroeconomic time series used are described. The third section presents the results of evaluating models and compares their predictive qualities. The conclusion contains conclusions and recommendations for further research in the field of analytical econometric instruments of the Russian stock market.

Based on the literature two main research questions were formulated:

1. Is there a connection between the macroeconomic variables of the domestic market and the Moscow exchange index?
2. Can the auto.arima function in R forecast non-stationary time series more accurately than the "manual" parameter selection?

2. Research Methodology

The data sample consists of 6 time series (from March 2000 to March 2018) obtained from the official websites of the Federal state statistics service, AO FINAM Investment holding (Table 1, 2). In order to select the specification and evaluate the predictive properties of models, we form two sets of train set and test set to compare the quality of forecasts. In the first data set, trainset = 39 observations (from January 2015 to March 2018), in the second data set, test set = 15 observations (from January 2017 to March 2018). All time series, except for indicators of the consumer price index, the index of output of goods and services for basic activities are logarithms.

Table 1

Data source

Name of the macroeconomic indicator	Data type	Source	Link
Macroeconomic variables of the domestic market			
Moscow Exchange stock index – IMEX	In points	FINAM	https://www.finam.ru/profile/mirovye-indeksy/micex/export/
consumer price index – P	chain index	FSSS	http://www.gks.ru/free_doc/new_site/prices/potr/tab-potr1.htm
index of output of goods and services by basic types of activity – IQ	base index	FSSS	http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/accounts/#
dollar rate – USD	rub.	FINAM	https://www.finam.ru/profile/kurs-rublya/usd-from-cb/export/
Macroeconomic variables of the world market			
stock index S&P – SP	In points	FINAM	https://www.finam.ru/profile/mirovye-indeksy/sandp-500/export/
Brent oil price – Oil	dollars	FINAM	https://www.finam.ru/profile/tovary/brent/export/

Table 2

Descriptive statistics of variables

Variables	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
IMEX	135.7	570.8	1386.9	1179.8	1663.1	2292.2
P	99.46	100.42	100.66	100.83	101.10	103.85
IQ	85.2	101.4	104.4	103.8	107.2	113.5
USD	23.41	28.40	30.31	35.51	33.15	76.33
SP	725.6	1132.0	1320.3	1450.9	1681.5	2816.4
Oil	18.55	37.85	59.99	65.10	94.47	140.43

We forecast the Moscow exchange index based on the ARIMA and ARIMAX models with external regressors of the x matrix – the consumer price index, the index of goods output and services by basic activities, the dollar exchange rate, the price of Brent oil, and the S&P stock index:

$$IMEX_t = c + \sum_{i=1}^p \varphi_i IMEX_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_k b_k X_k + \varepsilon_t \quad (1)$$

The inclusion of external regressors (X_k) can potentially increase the accuracy of the forecast, if there are good estimates of their future values available. ARIMA models provide a convenient and compact description of the process and suggest the selection of an appropriate theoretical design for the actual implementation of the time series of the Moscow exchange stock index. When the R software environment evaluates an ARIMA-type model, it uses maximum likelihood estimation (MLE). In order to choose the most appropriate model for the forecast, we obtained ARMA, ARIMA, ARIMAX, and SERIMAX models based on the initial levels of the IMEX variable, their logarithms, and level modifications with the addition of BoxCox transformations. To construct ARIMA-type models, we applied the approach of Box and Jenkins (1970) and the recommendations of Hyndman and Athanasopoulos (2013) from the following actions:

1. Estimation of the $\varphi_1, \varphi_2 \dots \varphi_p, \theta_1, \theta_2 \dots \theta_q, \Phi_1, \Phi_2 \dots \Phi_P, \Theta_1, \Theta_2 \dots \Theta_Q, b_1 \dots b_k$ coefficients with auto.arima application in R on the original data time series IMEX, the logarithms of the time series IMEX and IMEX time series with the addition of a BoxCox transformation. The auto.arima function in R uses a variation of The Hyndman and Khandakar (2008) algorithm that combines unit root tests, AIC minimization, and MLE to produce a model

Note that models of the ARIMA type assume taking the difference between levels for the transition from a non-stationary series to a stationary one. Therefore, according to (Brooks, 2008) using the logarithms of the IMEX time series, models of this type by taking the difference of logarithms actually get a measure of return on the stock market.

2. Diagnostics of the selected model based on residuals. Building an autocorrelation function (ACF) and a private autocorrelation function (PACF). The Box-Ljung test (Ljung & Box, 1978) was used to check for autocorrelation in the residues. The normality test was performed using the Jarque-Bera test (Jarque & Bera, 1980).
3. Provided that the remainder is not similar to WN, we proceed to manual selection of the model. Identification of ARIMA (p,d,q), ARIMAX (p,d,q) or SARIMA (p,d,q)x(P,D,Q)₁₂, SARIMAX (p,d,q)x(P,D,Q)₁₂ models that combines ACF and PACF for the original IMEX time series, graphical analysis of the time series, ADF tests (Dickey D. A. and Fuller W. A., 1979), KPSS (D. Kwiatkowski et al., 1992) for the logarithm difference IMEX, actually for the stock market yield indicator. Following the recommendation of (Kantorovich G. G., 2002) we do not try to determine exactly p and q, but choose some of their maximum possible values, build all models for $p \leq \max, q \leq \max$, choosing the best model based on the minimum value of the AIC criterion (Hyndman & Athanasopoulos, 2013).
4. Perform diagnostics of selected models for the remainders again: Box-Ljung test, Jarque-Bera test.
5. Return to the forecast of the initial values of the stock index for the test part of the sample and calculation of RMSE, MAE, MAPE errors to assess the predictive properties of the model.

3. Results and Discussion

Fig. 1 shows historical graphs of the actual levels of all variables used in this study.

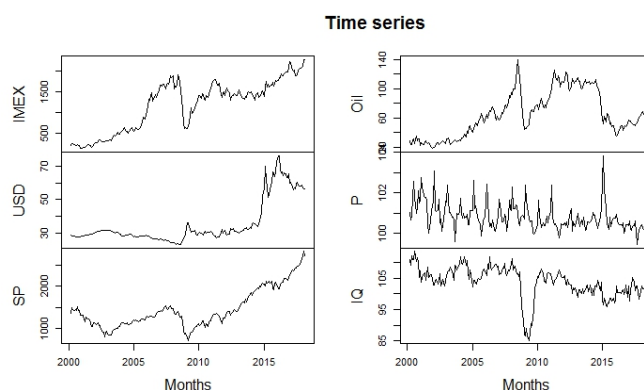


Fig. 1. Time series of key macroeconomic variables

All variables show marked adjustments during the 2008 crisis in response to macroeconomic fluctuations in the global economy. The period of the global financial crisis indicates the presence of structural breaks, there is a sharp decline in the Moscow exchange index, the S&P index, oil prices, the index of output of goods and services for basic activities, a surge in the exchange rate of the dollar, consumer prices. First of all, let's pay attention to the IMEX variable for the time series of the Moscow Exchange stock index. The unsteadiness of the initial levels of the IMEX variable is confirmed by figure 1 and the results of the Dickey-Fuller test and the KPSS test in R: Dickey-Fuller = -2.6077, Lag order = 5, p-value = 0.321; KPSS Unit Root Test - Value of test-statistical is: 1.2548 (critical values 0.463 of 5 pct). From Table 3, you can see that the first logarithm differences for all variables are stationary. Do not forget that the first difference from the logarithms of the Moscow exchange index is the yield of the stock market (Brooks, 2008). As seen in figure 1 and shown in table 3, the $\log(\text{IMEX})$ variable is a Difference Stationary integrated time series, I(1), which allows you to apply ARIMA-type models to predict its dynamics. Monthly average IMEX variables do not have strong differences (Fig.2), which does not show seasonality.

Table 3

Test of macroeconomic variables for stationarity

Augmented Dickey-Fuller test results	t statistics	prob. value
diff(log(IMEX))	-6,6006	0,01
diff(log(SP))	-5,8768	0,01
diff(log(Oil))	-6,2196	0,01
diff(log(USD))	-5,5891	0,01
diff(IQ)	-5,3691	0,01
diff(P)	-8,2614	0,01
Kwiatkowski – Phillips – Schmidt – Shin test results	Value of test-statistic	Critical values (5pct)
diff(log(IMEX))	0,1497	0,463
diff(log(SP))	0,261	0,463
diff(log(Oil))	0,137	0,463
diff(log(USD))	0,161	0,463
diff(IQ)	0,036	0,463
diff(P)	0,036	0,463

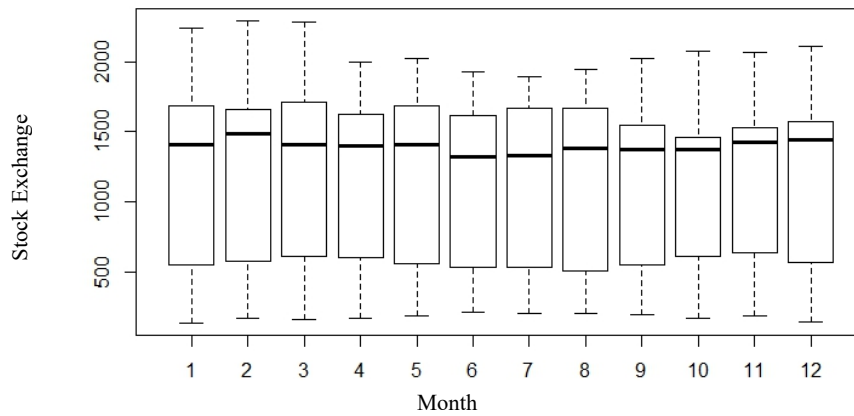


Fig. 2. Box diagram of monthly Moscow stock exchange index levels

Following Hyndman and Athanasopoulos (2013) when selecting the ARIMA specification, based on the results of applying the auto.arima function in R, the minimum value of the AIC criterion belongs to the seasonal model SARIMA(1,1,0)x(2,0,0)₁₂ for the time series of stock market returns (Table 4). The ARIMAX (0,1,0) and SARIMAX(0,0,0)x(2,0,0)₁₂ models have statistically significant coefficients for logarithms of external variables SP, Oil, USD, which confirms the relationship with the dynamics of the Moscow exchange index - the IMEX variable. The relationship of the consumer price index and the index of output of goods and services for basic activities with the index of the Moscow exchange were not found. The dollar exchange rate, despite the fact that it is formed in the domestic financial market, is influenced by a combination of factors of the world market. This is consistent with theoretical expectations. On ACF and PACF, there are no statistically significant values of autocorrelation and partial autocorrelation coefficients for the sarima(1,1,0)x(2,0,0)₁₂ model residues on the early lags (Fig. 3). The Box-Ljung test result does not reject the null hypothesis of no autocorrelation in the residuals of the model, whereas the Jarque-Bera test statistic rejects the null hypothesis of the normal distribution of residues, which does not rely on the accuracy of the constructed confidence interval for the forecast.

Table 4

The predictive characteristics of the models such as ARIMA, obtained in auto.arima in R

IMEX (train set =39 observations)				
ARIMA (0,1,1) for IMEXtrain : $\theta_1=0,1719$ (s.e.=0,0719), Box-Lj. p-v.: 0,01095, Jarq. Bera p-v.: 5,218E-15				
AIC=2073,39	BIC=2079,74	RMSE=414,569	MAE=368,098	MAPE=18,541
SARIMA(1,1,0)x(2,0,0)₁₂ for log(IMEXtrain): $\phi_1=0,2652, \phi_2=0,0500, \phi_3=-0,1041,$				
(0,0753) (0,0841) (0,0842)				
Box-Lj. p-v.: 0,5608, Jarq. Bera p-v.: 4,661E-06				
AIC=-355,02	BIC=-342,31	RMSE=410,452	MAE=362,332	MAPE=18,226
SARIMA(1,1,0)x(2,0,0) ₁₂ for Box.Cox(IMEXtrain): $\phi_1=0,2653, \phi_2=0,0500, \phi_3=-0,1041,$				
(0,0753) (0,0841) (0,0842)				
Box-Lj. p-v.: 0,5622, Jarq. Bera p-v.: 5,058E-06				
AIC=-351,36	BIC=-338,66	RMSE=410,471	MAE=362,346	MAPE=18,227
ARIMAX (0,1,0) for IMEXtrain : log(USD)=1,7779, log(SP)=0,6424, log(Oil)=5,3918, log(IQ)=0,0345, log(P)=-0,0256				
(4,6586) (0,0991) (0,8918) (0,5674) (1,3427)				
Box-Lj. p-v.: 0,003, Jarq. Bera p-v.: 2,76E-06				
AIC=1989,19	BIC=2001,89	RMSE=357,890	MAE=328,431	MAPE=17,030
SARIMAX(0,0,0)x(2,0,0) ₁₂ for log(IMEXtrain): $\phi_1=1,2037, \phi_2=-0,3755, c=-4,7600, \log(USD)=0,4428,$				
(0,1193) (0,1120) (1,8784) (0,1603)				
log(SP)=1,1039, log(Oil)=0,4532, log(IQ)=0,0345, log(P)=-0,0256				
(0,2017) (0,1141) (0,0688) (1,0785)				
Box-Lj. p-v.: 2,2E-16, Jarq. Bera p-v.: 0,015				
AIC=80,36	BIC=105,81	RMSE=498,773	MAE=461,469	MAPE=24,167

Without sufficient grounds for recognizing "white noise" in the remnants of models obtained in auto.arima in Hyndman and Athanasopoulos (2013), manual selection of the ARIMA type model has been applied.

Let's recall that according to the Dickey-Fuller test and the KPSS test, the variable diff(log(IMEX)) with a first-order difference (d=1) is stationary. On the 12th, 24th and 36th lags, significant coefficients for ACF and PACF are not observed, which does not reveal seasonality (Fig. 4). since auto.arima in R applied seasonal parameters in models, we will not ignore them and use them for seasonal autoregression P=1 and Q=1. Although, it is more likely that manual selection will not be in favor of seasonal parameters.

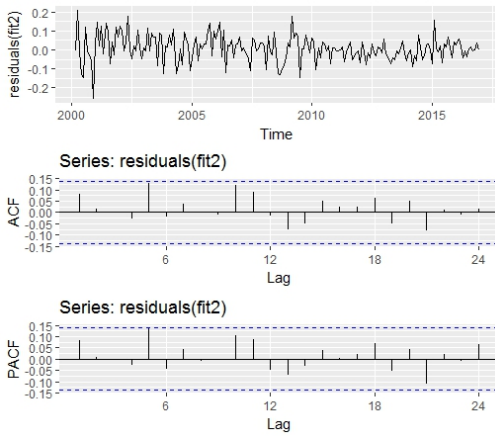


Fig. 3. ACF And PACF charts of the ARMA(1,1,0)×(2,0,0)₁₂ model balances for the stock market yield time series

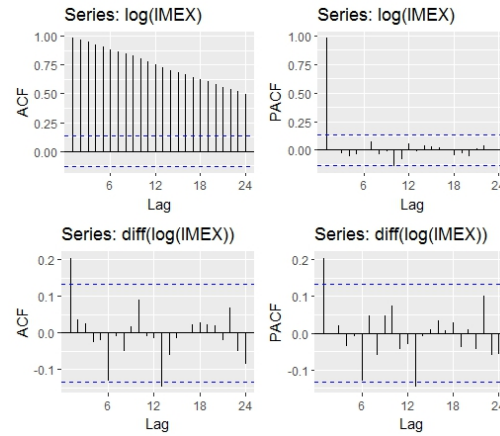


Fig. 4. ACF and PACF Charts for the Moscow exchange index and stock market yield

Evaluating possible sarima and sarima models (for $(p \leq \max, q \leq \max) \times (P \leq \max, Q \leq \max)$, $d=1, D=1$ by the variable $\log(\text{IMEX}_{\text{train}})$) could be considered. 72 models were evaluated for each type; AIC and forecast errors are indicated in Appendix 1. Table 5 shows 10 models of "manual" parameter selection with a minimum AIC.

Table 5
Predictive characteristics of ARIMA / ARIMAX models with minimal AIC

Models	AIC	RMSE	MAE	MAPE	p-value: Box-Ljung test	p-value: Jarque Bera Test
IMEX _{test} =39 observations						
SARIMA (1,1,0)×(0,0,0) ₁₂	-357,24	405,262	357,8704	18,004762	0,541	2,671E-06
SARIMA (0,1,1)×(0,0,0) ₁₂	-356,56	413,5924	367,0324	18,485539	0,515	1,102E-07
SARIMA (1,1,0)×(0,0,1) ₁₂	-355,56	401,7826	355,2688	17,88169	0,589	2,687E-06
SARIMA (1,1,0)×(1,0,0) ₁₂	-355,5	402,2434	355,599	17,897069	0,582	2,611E-06
SARIMA (1,1,1)×(0,0,0) ₁₂	-355,25	404,2536	356,7594	17,946407	0,544	3,331E-06
SARIMAX (1,1,0)×(0,0,0)₁₂	-427,97	352,3406	317,167	16,5274	0,6554	0,0614
SARIMAX (0,1,1)×(0,0,0) ₁₂	-427,87	354,4579	319,4814	16,64935	0,6586	0,0644
SARIMAX (0,1,0)×(0,0,0) ₁₂	-427,51	356,0189	320,3887	16,71444	0,6122	0,0470
SARIMAX (1,1,0)×(0,0,1) ₁₂	-426,05	351,1251	316,3221	16,49149	0,6577	0,0636
SARIMAX (1,1,0)×(1,0,0) ₁₂	-426,04	351,4884	316,6228	16,50573	0,6566	0,0630

Using AIC, we will choose the model SARIMAX (1,1,0) x (0,0,0)₁₂, which is equivalent to ARIMAX (1,1,0). The Box-Ljung test does not reject the null hypothesis that there is no autocorrelation in the remainder of the model: X-squared = 19.62, p-value = 0.6554. The Jarque-Bera test with a 95% probability without rejecting the null hypothesis about the normal distribution of residuals: X-squared = 8.5265, p-value = 0.0614.

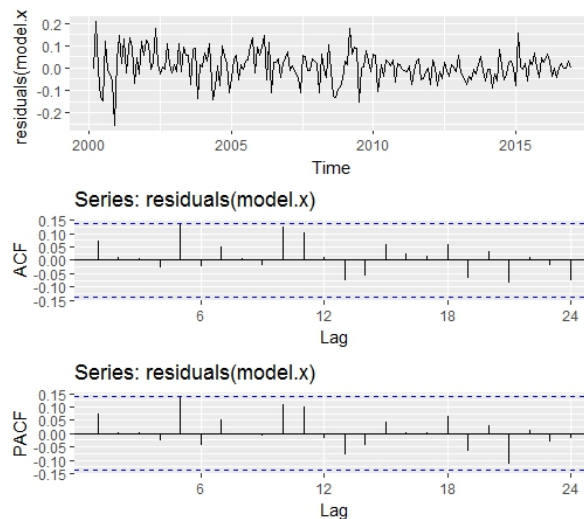


Fig. 6. ACF And PACF graphs of ARIMA (0,1,0) balances for the time series of stock market returns

Among the models selected by the minimum AIC criterion (Table 5) the MA component is rarely found in the dynamics of stock market returns, which indicates a predominant dependence on previous values of returns (AR component) and thus better predictability of returns. Table 6 shows 10 models of "manual" parameter selection with a minimum RMSE error.

Table 6
Predictive characteristics of ARIMA / ARIMAX models with minimum RMSE

Models	RMSE	MAE	MAPE	AIC	p-value: Box-Ljung test	p-value: Jarque Bera test
IMEXtest=39 observations						
SARIMA (0,1,0)×(0,1,2) ₁₂	206,7474	169,0097	8,782359	-286,05	0.0320	7.076E-05
SARIMA (0,1,0)×(1,1,1) ₁₂	207,4164	169,6106	8,814351	-286,04	0.0319	7.64E-05
SARIMA (0,1,0)×(1,1,2) ₁₂	207,6528	170,0274	8,835485	-284,09	0.0321	6.844E-05
SARIMA (0,1,0)×(0,1,1) ₁₂	215,9933	177,4355	9,228461	-287,94	0.0300	0.0001
SARIMA (0,1,1)×(1,1,1) ₁₂	226,4864	184,3966	9,568174	-293,89	0.0319	7.64E-05
SARIMAX(1,1,0)×(1,1,2) ₁₂	248,1999	207,3395	10,8967	-358,13	0,3913	0,1065
SARIMAX (1,1,1)×(0,1,2) ₁₂	252,6572	209,0376	10,98891	-358	0,2923	0,1112
SARIMAX (0,1,1)×(1,1,2) ₁₂	252,9208	210,2522	11,06167	-358,04	0,3934	0,1059
SARIMAX (1,1,0)×(0,1,2) ₁₂	258,8515	213,2178	11,22601	-359,89	0,3080	0,1195
SARIMAX (1,1,0)×(1,1,1) ₁₂	262,0301	215,1333	11,33273	-359,88	0,3000	0,1234

Using AIC, we will choose the model SARIMAX (1,1,0)×(0,0,0)₁₂, which is equivalent to ARIMAX (1,1,0). The Box-Ljung test does not reject the null hypothesis that there is no autocorrelation in the remainder of the model: X-squared = 19.62, p-value = 0.6554. The Jarque-Bera test with a 95% probability does not reject the null hypothesis about the normal distribution of residuals: X-squared = 8.5265, p-value = 0.0614. In this study, the AIC criterion is determined from the data set modified in logarithms and serves to select the order p, q, P, Q for the ARIMA model at d=1 for the return on the stock market, while the RMSE is determined from the original test set data and measures the predictive qualities of the ARIMA model directly for the stock index. Obviously, the best model for AIC can be good or bad for RMSE. Before applying RMSE to select a model for a forecast, it is possible to perform cross-validation for a small number of models. Since there are many models, for a quick answer, following Hyndman and Athanasopoulos (2013), one can choose the best model for AIC. Asymptotically, AIC and RMSE with cross validation would probably choose the same model. This question has not yet been fully resolved in the econometric literature. Results comparison of selecting an ARIMA model based on the AIC and RMSE criteria (Table 7) showed that for stock market forecasting, the "manual" selection of an ARIMA model can give better results (minimum AIC and minimum RMSE) than the built-in auto.arima algorithm in R.

Table 7
Comparison of evaluation results of ARIMA-type models for the Moscow exchange stock index

Selection in auto.arima in R			
SARIMA(1,1,0)×(2,0,0) ₁₂ for log(IMEXtrain): $\phi_1=0,2652, \phi_2=0,0500, \phi_3=-0,1041,$ (0,0753) (0,0841) (0,0842)			
Box-Lj. p-v.: 0,5608, Jarq. Bera p-v.: 4,661E-06			
AIC=-355,02	BIC=-342,31	RMSE=410,452	MAE=362,332
MAPE=18, 226			
"Manual" selection in R for log(IMEX)			
SARIMAX (1,1,0)×(0,0,0) ₁₂ : $\phi_1=0,1642$ (s.e.=0,0453)			
log(SP)=1,0364, log(Oil)=0,2428, log(USD)=0,3091, log(IQ)=0,0523, log(P)=-0,0178 (0,1188) (0,0513) (0,1346) (0,0729) (0,2374)			
Box-Lj. p-v.: 0,6554, Jarq. Bera p-v.: 0,0614			
AIC=- 427,97		RMSE= 346,607	MAE=317,167
MAPE=16,5274			
SARIMA (0,1,0)×(0,1,2) ₁₂ : $\theta_1=0,9461, \theta_2=0,7965$ (0,1208) (0,0513)			
Box-Lj. p-v.: 0,0320, Jarq. Bera p-v.: 7.076E-05			
AIC=-286,05		RMSE=206,7474	MAE=169,0097
MAPE=8,7823			

The AIC criterion indicates the optimal order of components in ARIMA models, but this order of components is not necessarily the best forecast for the stock index. From a practical point of view, the RMSE criterion, which measures the root-mean-square error of the forecast in points of the stock index, is more useful for investors. From this position in table 7, the advantage is obtained by "manual" selection in R model SARIMA (0,1,0) ×(0,1,2)₁₂ for log (IMEX).

4. Conclusions and future research

This paper is devoted to forecasting the Moscow exchange stock index taking into account the macro-indicators of the Russian and world economy. This is preceded from empirically proven theoretical arguments in favor of the influence of macroeconomic factors on stock indexes (Fama, 1990, Wasserfallen, 1989). The paper uses monthly data from 2000-2018. For the research, practical recommendations in articles (Hyndman & Khandakar, 2008; Hyndman & Athanasopoulos, 2013) on the methodological approach to constructing ARIMA-type models in R, are applied.

The advantage of applying the ARIMA model to the time series of the Moscow exchange index modified in logarithms is the ability to model both the monthly dynamics of returns on the stock market (based on the first logarithm differences) and the stock index (when switching to the original data). An analysis of the predictive characteristics of various autoregression models has shown that the simplest ARIMA models are the best, in particular with AR(1), which shows a predominant dependence on previous returns on the stock market and thus a better predictability of monthly levels. Calculations based on data for the study period showed that manual selection of ARIMA-type models does not exclude obtaining better predictive characteristics than auto. arima in R. The ARIMAX model confirmed the expected result for the Russian economy on the statistical significance of parameters under external regressors – the S&P index, the price of Brent oil, the dollar exchange rate. The expected assumption was confirmed that the macroeconomic variables of the domestic market – the index of output of goods and services by basic activities, the consumer price index, with the exception of the dollar exchange rate, do not have a statistically significant relationship with the index of the Moscow exchange. To continue this research in the direction of finding the best predictive model, cross-validation can be proposed before applying RMSE to select the model. Also, taking into account the criticism of RMSE (Armstrong, J. S., Collopy, F., 1992), we can try to apply the MASE criterion (Hyndman R. J., Koehler A. B., 2006) for model selection and compare the results with the AIC selection. In the direction of the analytical econometric tools development of the Russian stock market, it is possible to supplement ARIMAX models with macroeconomic indicators of the money market: money supply, MIACR rate.

Acknowledgements

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

References

- Armstrong, J. S., Collopy, F. (1992). *Error Measures for Generalizing About Forecasting Methods: Empirical Comparisons*. Retrieved from http://repository.upenn.edu/marketing_papers/69
- Basher, S. A., Haug, A. A., Sadorsky, P. (2012). Oil Prices, Exchange Rates and Emerging Stock Markets. *Energy Economics*, 34(1), 227–240.
- Bashiri Behmiri, N., & Pires Manso, J. R. (2013). Crude oil price forecasting techniques: a comprehensive review of literature. Available at SSRN 2275428.
- Box, G., & Jenkins, G. (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- Brooks, C. (2008). *Introductory Econometrics for Finance*. Cambridge University Press, 674 p.
- Dickey, D.A., & Fuller, W.A. (1979). Distribution of the Estimators for Autoregressive Time-Series with a Unit Root. *Journal of the American Statistical Association*, 74, 427–431.
- Fama, E.F. (1981, Sep.). Stock Returns, Real Activity, Inflation and Money. *The American Economic Review*, 71(4), 545–565.
- Fama, E.F. (1990, Sep.) Stock Returns, Expected Returns and Real Activity. *The journal of finance*, XLV(4),1089-1108.
- Hayo, B., & Kutan, A. M. (2002). The Impact of News, Oil Prices and International Spillovers on Russian Financial Markets. *Center for European Integration Studies Working Paper*, 20.
- Hyndman, R.J. (2014, 23-25 September). *Forecasting: Principles & Practice*. Leader: University of Western Australia, 138 p.
- Hyndman, R.J., & Athanasopoulos, G. (2016, 1 March) *Forecasting: principles and practice*. OTexts: Melbourne, Australia. <http://otexts.org/fpp/>
- Hyndman, R.J., & Khandakar, Y. (2008, July) Automatic Time Series Forecasting: the forecast Package for R. *Journal of Statistical Software*, 27(3),1-23.
- Hyndman, R.J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.
- Jarque, C. M., & Bera, A. K. (1980). Efficient test for normality, homoscedasticity and serial independence of residuals. *Economic Letters*, 6(3), 255-259.
- Jones, C. M., & Kaul, G. (1996). Oil and the Stock Markets. *The Journal of Finance*, 51(2), 463–491.
- Kadochnikova, E., Erina, T., & Zainulli, S. (2019). The use of arima model in forecasting accounts payable. *SCOPUS09758364-2019-10-2-SID85074629367*.
- Kadochnikova, E., Erina, T.V., & Margushova, A. (2019). Forecasting the enterprise tax base through regression of one-dimensional time series. *International Journal on Emerging Technologies*, 10(2), 232-236.
- Kantorovich, G.G. (2002). The time series analysis. *HSE Economic Journal*, 2,251-273.
- Kantorovich, G.G. (2002). The time series analysis. *HSE Economic Journal*, 1, 85-116.
- Kwiatkowski, D., P. C. B., Phillips, P., Schmidt, & Y. Shin (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. *Journal of Econometrics*, 54, 159-178.
- Ljung, G. M., & Box, G. E. P. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika*, 65, 297–303.
- Medvedeva, O.E., Ajupov, A.A., & Bagautdinova, N.G. (2016). Speculation strategies for the Russian stock market in the conditions of crisis. *International Business Management*, 10(23), 5513-5516.
- Morelli, D. (2002). The relationship between conditional stock market volatility and conditional macroeconomic volatility: empirical evidence based on UK data. *International Review of Financial Analysis*, 11, 101–110.



© 2021 by the authors; licensee Growing Science, Canada. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).