https://gospodarkainnowacje.pl

GOSPODARKA I INNOWACJE

LABORATORIUM WIEDZY Artur Borcuch

Volume: 60 | 2025

Economy and Innovation
ISSN: 2545-0573

For more information contact: editor@gospodarkainnowacje.pl

ECONOMETRIC FORECASTING OF LIGHT CAR IMPORTS TO UZBEKISTAN

Kaljanov Bakhadir Dauletbayevich
Tashkent City Department of the Customs Committee senior inspector, independent investigator, tolqindayi2@gmail.com

ARTICLEINFO.

Keywords: Light vehicle imports, ARMA (AutoRegressive Moving Average) model, reliable forecasts.

Abstract

This article studies the issue of forecasting the volume of passenger car imports to the Republic of Uzbekistan based on time series. Developing reliable forecasts for car imports is of urgent importance for the country's economic stability, maintaining market balance and effective management of import policy. The study used statistical data collected on an annual basis for the period 2010–2023, and the ARMA (AutoRegressive Moving Average) model was used to analyze them. The stationarity of the time series was checked using the Dickey-Fuller and extended ADF tests. The results of the study made it possible to identify the seasonal and cyclical characteristics of changes in car imports, as well as to formulate statistically sound forecasts. Based on the choice of the ARMA model, the accuracy of the parameters and the forecasting capabilities, recommendations were developed for optimizing import policy. This approach serves as an important source of information for economic policymakers, investors and business entities in making strategic decisions.

http://www.gospodarkainnowacje.pl/ © 2025 LWAB.

1. INTRODUCTION

The forecast of automobile imports is important for the effective management of the country's economy, meeting market demand and strategic planning. Especially in the context of the development of global economic relations and the growth of demand for automobiles in the domestic market, the need to correctly assess the volume of imports and determine future needs is increasing. The forecast of automobile imports based on time series facilitates the decision-making process for economic policymakers, business entities and investors. This forecasting allows you to foresee possible deficits or excess supply in the market and helps to optimize import policy accordingly.

The ARMA (AutoRegressive Moving Average) model, one of the main methods of forecasting through time series, provides an opportunity to effectively analyze complex and multifactorial processes such as car imports. Using the ARMA model, it is possible to identify future trends based on historical data, assess existing seasonal or cyclical characteristics, and form reliable estimates of import volumes. As a result, this approach serves to make more informed and qualitative decisions by identifying import dynamics and correctly assessing the factors affecting it. In this study, we investigated how forecasting can be performed using [R] software (Figure 1).

Copyright © 2025 All rights reserved International Journal for Gospodarka i Innowacje This work licensed under a Creative Commons Attribution 4.0

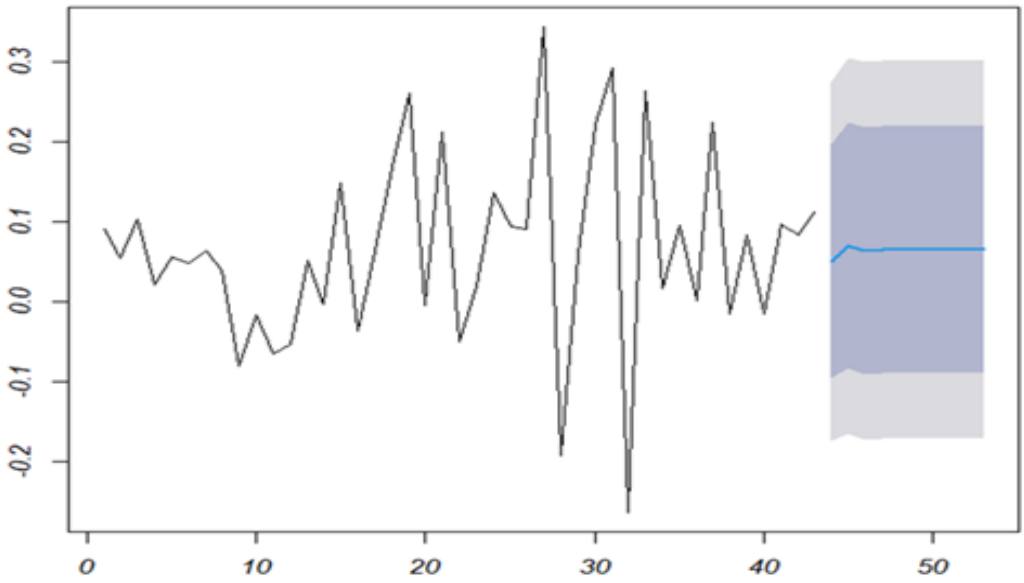

Figure 1 ARIMA from (1,0,0) to zero equal not been average forecasts

2. LITERATURE REVIEW

Cars import the country's transport and logistics infrastructure , population consumption demand satisfy level and macroeconomic to stability noticeable impact shows . Therefore for car import size determination and him/her forecast economic politics important from directions one as International on a scale car import trends analysis in doing timely rows based on statistic approaches wide (Box & Jenkins, 1970; Hamilton, 1994) . Especially in economic variables seasonality and cyclical to the features has was in cases timely row models , including the ARMA (Autoregressive Moving Average) model high accurate forecasts presented (Wei , 2006).

By Box and Jenkins (1970) working ARIMA model developed timely of rows internal structure fundamental tool in assessment as known . This model autocorrelation and random noise components into account to take through economic of processes time according to to pass analysis to do possible will be . Many studies (Stock & Watson, 2002; Enders, 2014) have shown that the ARMA model economic indicators , including imports and export volumes forecast in doing application effective that it is In particular , the ARMA model simple to the structure has to be him/her small time rows also apply for opportunity (Gujarati & Porter, 2009).

Last in years this approaches developed , they software on platforms automated . Including **R** software environment timely rows analysis for comfortable opportunities presented (Cryer & Chan, 2008). Forecast, series , and time series in the R platform packages through stationary verification (ADF tests), model parameters assessment and forecast the results visualization to do opportunity exists . In this respect take **when** viewed empirically **research** statistic from programs use academic of research accuracy and repeatability increases (Hyndman & Athanasopoulos, 2018).

Car import forecast not only economic growth indicators in determining , perhaps infrastructure policy in formation , customs procedures also important in adaptation role The current import policy efficiency assessment , deficit or excess offer risks minimization , as well as foreign investors for predicted numbers based on reliable economic environment create opportunity (OECD , 2011; UNCTAD, 2020).

Also , advanced in research car import variability using log-log transformations (Baltagi , 2005), skewness and kurtosis indicators (Brooks, 2014) in the assessment distribution features deep analysis Through this , extreme in years face giver economic incidents , including the 2016 re- import incident such as situations is determined and their forecast to the model impact is evaluated .

Copyright © 2025 All rights reserved International Journal for Gospodarka i Innowacje This work licensed under a Creative Commons Attribution 4.0

From this outside , temporary of rows stationarity – that is their time according to unchanging to dispersion has to be - forecast of the results accuracy for main factor is . By Dickey and Fuller (1979) offer DF and ADF tests performed this regarding main tests to the point This tests using of the row unity to the root has or absence is determined and accordingly The ARIMA model parameter d is is determined .

This point of view by the way , car import forecast in doing modern statistic approaches , in particular ARMA/ARIMA models application scientific based decisions acceptance in doing important tool This is the case . own in turn state scale economic planning and market demand clear assessment opportunity presented will reach .

3. METHODOLOGY

There are several methods for forecasting time series, the most widely used of which is the ARMA (Autoregressive Moving Average) model. This model combines the AR (Autoregressive) and MA (Moving Average) components, and is widely used due to its simple structure and high flexibility. The ARMA model takes into account the general trends, seasonal variations, and random fluctuations of the time series, thereby increasing the accuracy of the forecast result. The general ARMA model was first presented in the 1951 thesis "Hypothesis Testing in Time Series Analysis" by Peter Whittle, and later became popular in the 1970 book by George Box and Gwilym Jenkins. This model predicts the future based on historical data, assuming other factors are constant. Therefore, historical time series data of a variable are used for forecasting. The longer the series data span, the more accurate the forecast will be. Time series can be formed by year, quarter, or month.

The ARMA model looks like this:

$$Y_t = \alpha + \sum_{i=1}^{p} \gamma Y_{t-1} + \sum_{j=1}^{q} \theta E_{t-j} + E_t$$

In the ARMA model , there are concepts of *p, d, q* . First of all, we need to define *d* . If I(0) is stationary in time, we replace *d with one. p is* partially *autocorrelated q is determined based on* the *autocorrelation* graph. The open source R studio program was used to conduct the research [1].

This analysis process includes general visualization, stationarity checking, determining the optimal order of the ARMA model using various selection criteria, estimating model parameters, performing model diagnostics, and evaluating its forecasting ability. Stationarity checking, determining the appropriate order of the ARMA model using several selection criteria, estimating model parameters, and determining forecasting performance are integral parts of the analysis process.

consists of data from 2010 to 2023 *(annual data) , consisting of a total* **of 14 time series** (Table 1).

Data preparation:

To properly assess the performance of our forecasting model, it is necessary to divide the data set into "in-sample" and "out-of-sample" parts. The in-sample part is used to evaluate and calibrate the model, while the "out-of-sample" part forms the basis for assessing the model's forecasting accuracy in the context of unseen data.

In many research papers and scientific articles, it can be observed that **logarithmic data** is used instead of the original data, which helps to break down large decimal numbers into smaller forms.

For example: We can convert the number of car imports in 2022, 3,396,520, into the number log(3,396,520) = 6,531 by taking the logarithm.

[1] https://rstudio.com/products/rstudio/download

Copyright © 2025 All rights reserved International Journal for Gospodarka i Innowacje This work licensed under a Creative Commons Attribution 4.0

of car imports (ie$\Delta log(car)$)

annual growth trend of Car (Number of Car Imports) over time and can be calculated using the log-log method.

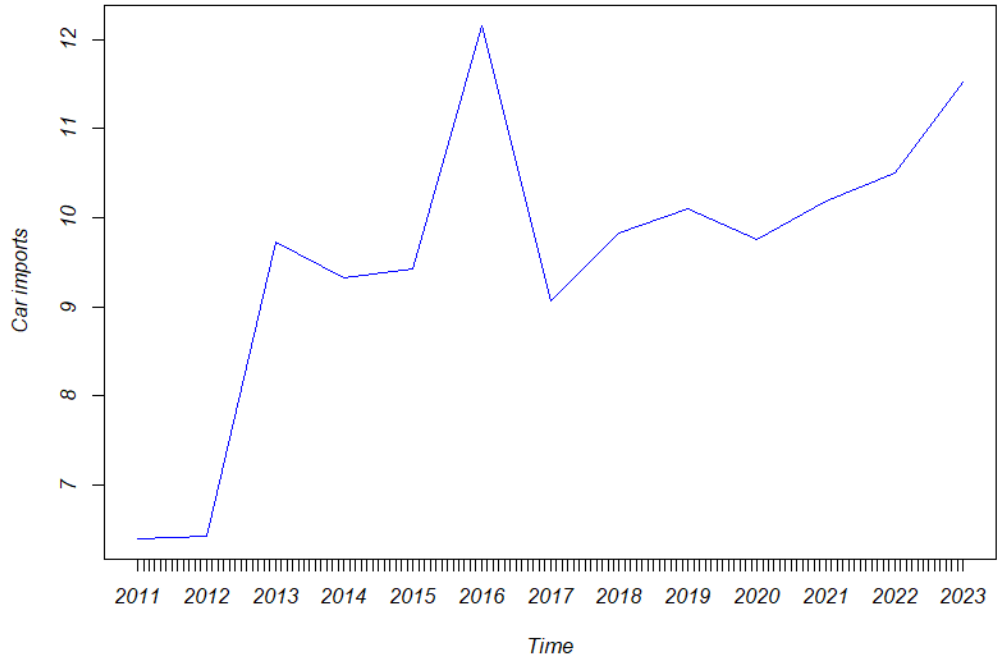

Figure 2. Box plot of growth rate of car imports

A box plot presents the range, median, and quartiles of data in a simple and informative way. The box in the middle of the plot represents the interquartile range (IQR), which is the difference between the first (Q1) and third (Q3) quartiles. Points outside the box are called outliers and are represented by lines extending from the box. The box plot above shows the distribution of the growth rate of the number of passenger cars imported into Uzbekistan. The box to the center relatively symmetrical , but ghost values existence in distribution unusual years that there is is showing . If the right on the side Ghost re-imported in 2016 national cars because of if so , this year extreme high growth pace I think it was to do possible .

Copyright © 2025 All rights reserved International Journal for Gospodarka i Innowacje This work licensed under a Creative Commons Attribution 4.0

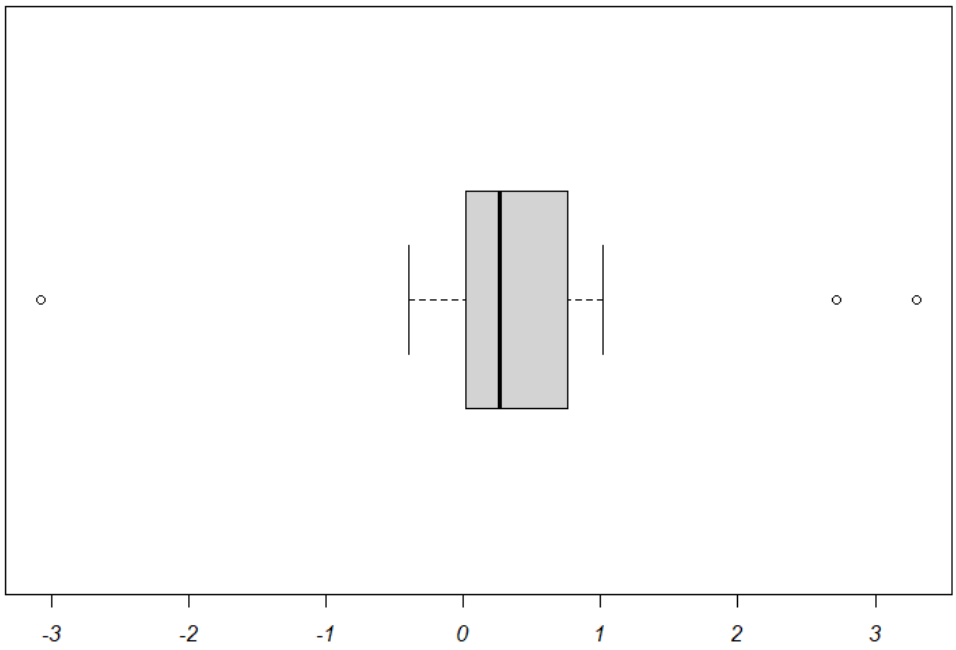

Figure 3. Histogram for growth rate

Histograms are a useful way to visualize the distribution of a data set. This technique helps us see outliers in the data more clearly. By identifying outliers in a histogram, we can investigate their causes and decide whether to remove them from the analysis or handle them differently. However, not all outliers are bad or need to be removed. Some outliers may be valid data points that reflect real events, and removing them may lead to misinterpretation of the results. Therefore, it is important to carefully consider the context and nature of the outliers before making any decisions.

In the graph below, the histogram shows that the data is unevenly distributed — some values are very rare, some are clustered, and some values are centered (around 0) while others are spread out to the edges, which is a sign of a well-fitting ARIMA model. However, some bars are quite high and others are low, indicating some deviation from a normal distribution. This is due to the volume of imports vibrations or uneven distribution that there is to show possible .

Copyright © 2025 All rights reserved International Journal for Gospodarka i Innowacje This work licensed under a Creative Commons Attribution 4.0

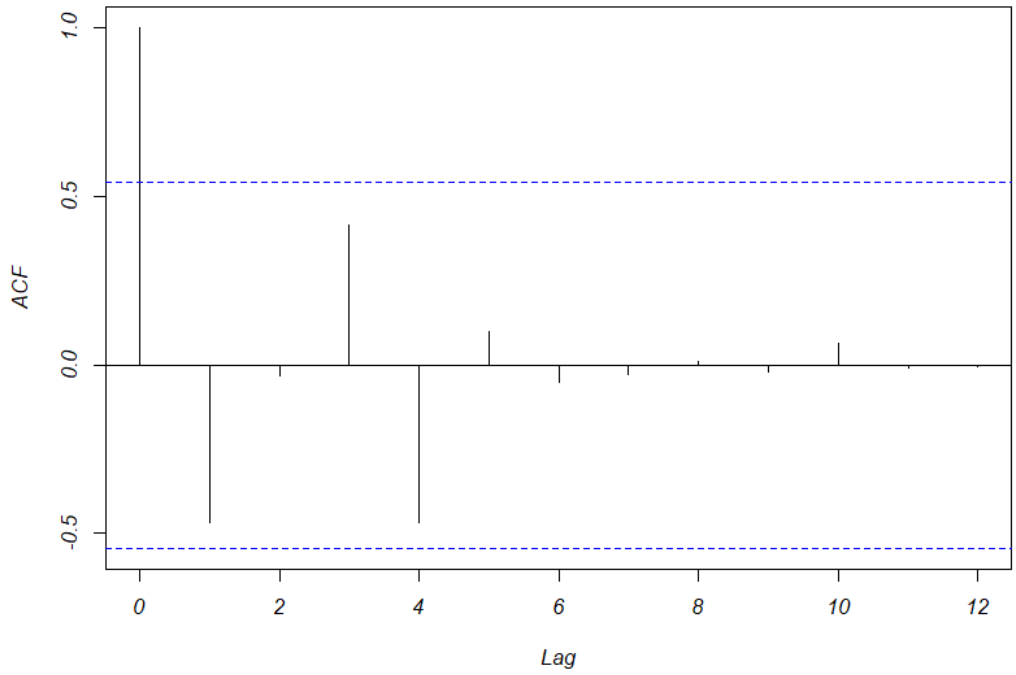

Figure 4. ACF graph of growth rate

This ACF (Autocorrelation Function) plot helps analyze the degree of autocorrelation in the change over time in the number of cars imported into Uzbekistan.

X- axis (Lag) : This is the time indicates the delay (lag) . For example , **Lag = 1** means , previous with period (year , month, quarter) current period between connection indicates .

Y- axis (ACF - Autocorrelation Function) : For each lag autocorrelation value shows .

- **About 1 if** it **is** strong positive dependency .
- **Close to -1 if** it **is** strong negative dependency .
- **Around 0 if** , dependency no
- ✓ **Black lines** : For each lag calculated autocorrelation values indicates .
- ✓ **Blue Significance Bounds** : This **exit borders** statistic in terms of **attention worthy (just random (non - correlation)** that indicates .

General descriptive statistics ((i. e . $\Delta log(Cars)$)

Mean	9.49
Median	9.7 4
Standard deviation	1.50
Skewness	-0.1365
Kurtosis	0.51

The mean (Mean) is the average value of a data set. It is calculated by adding all the values and dividing by the number of observations:

Copyright © 2025 All rights reserved International Journal for Gospodarka i Innowacje This work licensed under a Creative Commons Attribution 4.0

$$\frac{1}{T}\sum_{t=1}^{T} x_t \; t=1,2,3..... T$$

The median is the middle value in a data set when the values are arranged in order of magnitude, and is the 50th quartile of the data set. It is another way to measure central tendency and is especially useful when there are extreme values or outliers in the set. MedianIt is less affected by extreme values than by the mean.

The standard deviation measures how much the data in a distribution deviates from the mean. In other words, it indicates the spread or variability of the data. A standard deviation value of **1.50** indicates that the data is not tightly concentrated around the mean. This means that the data deviates relatively widely from the mean, that is, it has variability. The main reason for this, as we mentioned above, is that the national vehicles exported in 2016 were suddenly re-imported, and the volume of imports this year was artificially high.

In econometrics, α qiyshiqlik(skewness) and **kurtosis**β are two measures of the shape of a probability distribution of a variable. Qiyshiqlikβ measures the degree of symmetry of the distribution, kurtosisand β measures the degree of peaking or flatness of the distribution.

A skewness value of -0.13 means that the distribution is almost symmetric, but slightly skewed to the left.

A kurtosis value of 0.51 indicates that this distribution is platikurtikskewed , meaning that the distribution is flat and has fewer outliers.

Unit Root/Stationarity Tests: DF andADF Tests

Dickey- FullerTechnical description of the (DF) test: The DF ( ) test Dickey-Fullerwas developed by David A. stasionarand WayneA. Fullerin 1979. The main purpose of this test is to determine whether a time series is nonstationary or not. The null hypothesis (H_0) of the test is that the time series is nonstationary (has a unit root), while the alternative hypothesis (H_1) indicates that the time series is stationary Dickey. The DF test is performed by estimating the parameters of the time series avtoregressivmodel and determining the significance of the coefficient indicating the presence of a unit root in the model o.tkaziladi

Advanced Dickey- Fuller(ADF) Test

ADF (Augmented DickeyThe - Fuller) test is an extended form of the DF test and is widely used in economics, finance, and other fields. This test is used to test the unit root hypothesis, meaning that the time series is nonstationary and contains a trend or random walk (random walk) component is considered to be present.

Technical Description of the ADF Test

The validity of the above tests is subject to errors (error term) "white noise" (white noise) is independent of them, i.e., avtokorrelyatsiyait depends on whether they are not among them. However, if the variable (ΔY) of the regression equation is autocorrelated and this is not taken into account, the errors may also be autocorrelated. In such cases, the extended form of the ADF test is used, i.e., the dependent variablep laglarini(delays) avtokorrelyatsiyaniare taken into account by entering.

This approach helps ensure the validity of errors and avtokorrelyatsiyaprevents errors that could occur due to unmeasured variables within the dependent variable.

The general equation for the ADF test can be written as follows:

Copyright © 2025 All rights reserved International Journal for Gospodarka i Innowacje This work licensed under a Creative Commons Attribution 4.0

$$\Delta x_t = \beta x_{t-1} + \sum_{i=0}^{p} \phi_i \Delta x_{t-i} + \varepsilon_t$$

Here:

$\Delta x_t = x_t - x_{t-1}$ — the first difference of the time series,

β — checks for the existence of a unit root koeffitsient,

ϕ_i — *lags* ,

ε_t — " *white noise* " *errors* .

This model is used nostatsionarlikto test for a time series. If γ\gamma koeffitsientiis not statistically significant, the null hypothesis (nostatsionarlik) is not rejected, meaning that the time series is nonstationary. Conversely, if γ\gamma is significant, the time series statsionaris accepted as nonstationary.

Null and Alternative Hypotheses of ADF Test:

ADF (Augmented The main purpose of the Dickey- Fuller) test is to determine whether a time series is stationary or nonstationary.

Null hypothesis (H_0): The time series has a unit root (p=0) and is nonstationary. This means that the series follows a trend or a random walk (random walk) component.

Alternative hypothesis (H_1*):* The time series is stationary (p<1). This means that the time series has no autocorrelation and is stable around its long-term mean.

The ADF test statistic is calculated based on the regression coefficients and residualthe standard errors of the residuals ( ). These statistics are compared to a specific distribution, which is usually Dickey-Fuller kritikalvalues are taken from the table orMonte Carlocan be calculated using simulation. If the test statistic kritikalis less than the value, we reject the null hypothesis and assume that the time series is stationary. ADF tests are divided into drift**3 types** depending on the presence of a trend in the time series :

1. **No trend or drift (simple ADF model) :**

$$\Delta x_t = \beta x_{t-1} + \sum_{i=1}^{p} y_i \Delta x_{t-i} + \varepsilon_t, \beta = 0 \text{in} H_0 \text{ vs. } \beta < 0 \text{in} H_A$$

In this case, only the lags and the main time series are left. This type of test nostatsionarliknionly checks for variability.

2. Only **drift** (α) is available:

$$\Delta x_t = \alpha + \beta x_{t-1} + \sum_{i=1}^{p} y_i \Delta x_{t-i} + \varepsilon_t, \beta = 0 \text{in}^{H_o} \text{ vs } , \beta < 0 \text{in} H_A$$

In this model, drift (α) is added, which means that the mean of the time series may change, but there is no consistent trend.

3. **a trend and a drift** :

$$\Delta x_t = \alpha + \lambda t + \beta x_{t-1} + \sum_{i=1}^{p} y_i \Delta x_{t-i} + \varepsilon_t , \beta = 0 \text{in } H_0 \text{vs } , \beta < 0 \text{in } H_A$$

This model λthas a trend ( ) and a drift (α), which tests whether the time series has a long-term trend.

Copyright © 2025 All rights reserved International Journal for Gospodarka i Innowacje This work licensed under a Creative Commons Attribution 4.0

	ADF
H_0	x_t~ I (1) ; not stationary
H_1	x_t~ I (0) ; stationary

logConclusion on tests for ( )cars

Test	H_0	t -stat	Critical value	Can he refuse?	Conclusion
ADF	log(cars)has unit root	-4.016	-2. 07	Refused	stationary

*Note: All critical values **(CV)** are at the 5% significance level.*

Thus, the ADF test shows that *the log (cars)* time series is stationary . This series is statistically stationary and ready to be used for ARIMA or other time series models, i.e. we can continue our analysis based on this model.

Model Selection

AIC (AkaikeInformation Criterion) and BIC (Information Criterion) are Bayesstatistics used to compare the goodness of fit lchovlarof time series lumotlariniand lumotlarselect the best-fitting model for a given data oset plami. Both AIC and BIC are based on the probability principle, which states that if two models have the same goodness of fit bo, lsathe better parametrlimodel is kopreferred.proq parametrliagainst the model kidlaydi.

AIC = -2log $\mathcal{L}(\theta)$ *+2k*

BIC = -2log $\mathcal{L}(\theta)$ *+klog(N)*

Here:

log – $\mathcal{L}(\theta)$denotes the maximum objective function value for the model ; log-likelihood*k* – number of

parameters (k= p+ q+ 2);

N – number of samples.

The main point is that *the smaller the IC value, the better the model is considered to be* . This is because the minus signs in the expression and IC indicate that the fitness of the model is proportional to its complexity o' lchaydi, which means that lumotlarnioverfitting models (i.e., models with too many koparameters p' bo) lganare penalized and models with few parametrliand good fit are favored. For this analysis, the maximum orders were set to 5, *(p,q ≤ 5)* and a fixed parameter (intercept) was introduced, because o'rtacha Budjetthe lmaganincome ois bozerosishati o' rtachahas the value So' ngra, AIC bo' yichaThe corresponding R code was run to select the best ARMA(p,) model in the sample.q

The smallest AIC value out of all 36 models (6*6) is 41.32 bo' lib, which corresponds to **the ARIMA(1,0,1) model.**

Copyright © 2025 All rights reserved International Journal for Gospodarka i Innowacje This work licensed under a Creative Commons Attribution 4.0

Comparison table

Model	AIC	BIC	sigma² (er. var)	ME	RMSE	MAE	ACF1
ARM(1,1)	**41.32**	43.87	**0.7003**	0.099	0.742	0.447	**-0.013**
ARM(2,1)	43.25	46.45	0.7579	0.109	**0.736**	**0.440**	-0.049
ARM(2,2)	45.03	48.86	0.8329	0.008	0.844	0.528	0.067

Above comparison from the table **ARMY (1.1)** model the most good choice that analysis we do :

1. *Top accuracy :*

- ARMY In the (1,1) model **MAE = 0.447** and **RMSE = 0.742** this is other to models than **less error** with forecast gives .

2. *Good compatibility :*

- In the model **sigma² = 0.700** — this **residue variance** , i.e. the model fits the data **good adapted** indicates .

3. *Short and optimized :*

- Model **plus complicated not** , only one AR (autoregressive) and one MA (moving average) parameter used , but result very good .

1. *Low AIC :*

- **AIC = 41.32** — this other complicated to models than **model selection to the criterion see the most optimal option** is considered .

ARMY (1,1) model his/ her own simplicity , accuracy , low error level and statistic by criteria (AIC, MAE, RMSE, sigma²) advantage because of the most good choice is considered .

Forecasting

ARMY (1,0,0) or AR (1) model for previous periods values based on forecast to do equations as follows expression possible :

$$Y_t = \alpha + \beta Y_{t-1} + \epsilon_t$$

Now using this model future in the eras variables for equations harvest we will do .

$$\hat{Y}_{t+1} = \alpha + \beta Y_t + \epsilon_{t+1}$$

In this, $\hat{Y}_{t+1}$ the forecast for the next period value is , is current in the era Y_t based on the value of . The error term ϵ_{t+1} is unknown. happened because of in the forecast into account not available and his/her expected value accepted as zero will be done .

$$\hat{Y}_{t+2} = \alpha + \beta \hat{Y}_{t+1}$$

$$\hat{Y}_{t+3} = \alpha + \beta \hat{Y}_{t+2}$$

……

Copyright © 2025 All rights reserved International Journal for Gospodarka i Innowacje This work licensed under a Creative Commons Attribution 4.0

$$\hat{Y}_{t+k} = \alpha \sum_{i=0}^{k-1} \beta^{i} + \beta^{k} Y_t$$

Here:

$\sum_{i=0}^{k-1} \beta^{i}$ — permanent terms sum is geometric progression harvest does .

$\beta^{k} Y_t$— and initial $Y_t$of value *k* from the era next the impact shows .

Forecast result :

Through the R program the next 5 years forecast to do done increased , the result following in the table to see possible (Table 2).

Table 2. Through the R program the next 5 years forecast to do done increased , the result

Point_Forecast	Lo.80	Hi.80	Lo.95	Hi.95
131,327.7	50,130,401	156,809.9	137,931.95	229,851.9
201,078.6	80,761,956	232,851.9	165,949.30	287,539.6
259,858.8	88,245.614	311,633.0	171,201. 68	382,601.1
327,123.1	412,237,910	560,479.5	539.26493	624,380,440.8
415,502.7	311,266.821	509,835,052.0	447.47752	723,189,404.3

The following on the graph and To Uzbekistan imported light of cars growth ($\Delta log(cars)$) of forecast reflection (Figure 5).

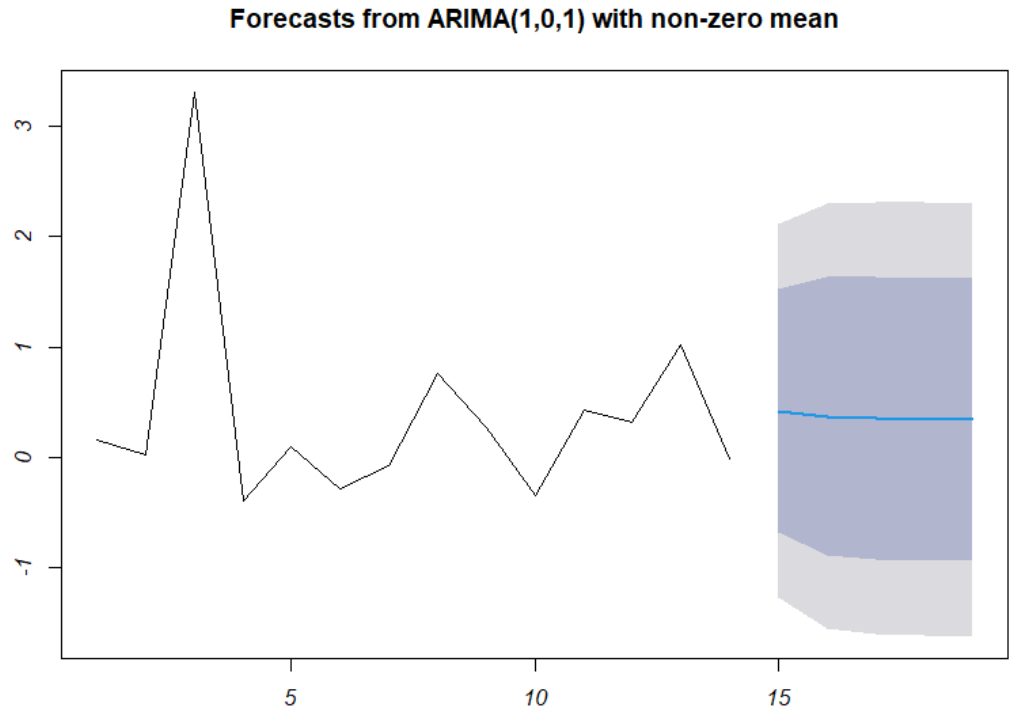

Figure 5. Imported light of cars growth forecast (og(cars))Δl

Copyright © 2025 All rights reserved International Journal for Gospodarka i Innowacje This work licensed under a Creative Commons Attribution 4.0

The blue line in the graph represents **the average forecast value** , and the blue line in the Forecast area represents the average forecast for future periods (starting from period 15). This forecast is relatively stable, meaning that the differentiated logarithmic value *(Δlog(cars) is expected to remain stable at* some level.

Shady regions (Forecast intervals):

Forecast line surrounding dark and hungry shady regions suitable 80% and 95% confidence respectively intervals represents . This The intervals are 80% and 95% confidence intervals of the true values . with this between to fall ko' rsatadi. The widening of intervals over time indicates an increase in uncertainty in forecasting for future periods.

Trust intervals :

80% confidence interval initial forecast during the times approximately -0.5 to 1.5 changes . 95% confidence interval wider approximately -1 to 2.5 was the interval cover This intervals clear The values are given in Table 2 .

Conclusion

This study conducted a time series analysis of automobile imports to Uzbekistan. For the analysis before information logarithmically , by differentiating to inpatient achieved . ADF (Augmented Dickey-Fuller) test through row stationary that confirmed ($p < 0.05$). Of these after various ARMA (AutoRegressive Moving Average) models evaluated and they between **ARM (1,0,1)** model the most was found acceptable because it had lower AIC and BIC values . has forecast mistakes are acceptable at the level It came out .

Using the selected model next 5 periods for car import forecast values Forecast results with together 80% and 95% confidence Also calculate confidence intervals This is the to the forecast was reliability level showing gave .

From the graph apparently as expected , the forecast done values stable and previous periods trend suitable These results are coming . In Uzbekistan car import future status evaluation, economic planning and customs policy in formation important importance has to be possible.

References:

1. Whittle, P., 1951. Hypothesistesting intime series analysis. (No Title).
2. Box, GE, Jenkins, GM, Reinsel, GCand Ljung, GM, 2015.Time series analysis: forecasting and control.John Wiley& Sons.
3. Simonov, J.and Gligorov, Z., 2021.Customs revenues prediction using ensemble methods (statistical modelling vs machine learning).World Customs Journal, 15(2), pp.111-128.
4. Agbenyega, D.A., Andoh, J., Iddi, S.and Asiedu, L., 2022.Modelling Customs RevenueinGhana Using Novel Time Series Methods.Applied Computational Intelligence and Soft Computing, 2022(1), p.2111587.
5. Dr. Mahmoud el Madna.Research Unit"WCO"Forecasting Customs Rvenues&Trade Flow"

Copyright © 2025 All rights reserved International Journal for Gospodarka i Innowacje This work licensed under a Creative Commons Attribution 4.0