COMPUTER SCIENCE

FORECAST OF TOURIST DEMAND IN UKRAINE ON A FAST-FUTURE PROSPECTS

[1]**Viktor Krylov,** *prof.*
[2]***Christina Lipyanina***

[1]*Ukraine, Odesa, Odessa National Polytechnic University*
[2]*Ukraine, Ternopil, Ternopil National Economic University*

DOI: https://doi.org/10.31435/rsglobal_ws/30082018/6047

ARTICLE INFO

Received: 06 July 2018
Accepted: 21 August 2018
Published: 30 August 2018

KEYWORDS

model,
tourist demand,
forecast,
ARMA,
AIC criterion,
BIC criterion,
Leung-Box test.

ABSTRACT

The process of formation of tourist demand was studied and autocorrelation and partial auto-correlation were calculated. Valued behavior of selective ACF and partial PACF, showing the hypothesis about the values of the parameters p and q. Due to the lack of data, several competing ARMA (1.1) and ARMA (2.0) models have been selected. Both models showed a good match with the data, the models are adequate and the errors are random, so the best model is chosen according to the AIC and BIC criterion. The remains of the selected model are checked for the absence of auto-correlation using the Lew Box test. For the selected best model, forecasts were projected for 5 periods ahead. From the forecast of the time series it is clear that the tourist demand in the next 5 years will decline.

Citation: Viktor Krylov, Christina Lipyanina. (2018) Forecast of Tourist Demand in Ukraine on a Fast-Future Prospects. *World Science.* 8(36), Vol.1. doi: 10.31435/rsglobal_ws/30082018/6047

Introduction. The basis of many forecasting methods is the forecasting methodology of Boxing-Jenkins, which does not foresee any special structure in the data of the CR, for which the forecast is made. It uses an iterative approach to determining a valid model among the general class of models. Then the selected model is mapped to historical data to check if it really describes the rows. The model is considered acceptable if the remnants are mostly small, randomly distributed and, in general, do not contain useful information. If the given model is not satisfactory, the process is repeated, but already with the use of a new, improved model. A similar iterative procedure is repeated until a satisfactory model is found. The found model can be used for prediction purposes only from this moment.

Let X_t be given, where t – integer index and X_t – real numbers. Then the ARMA model (p, q) is given as follows:

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)X_t=\left(1+\sum_{i=l}^{q}\theta_i L^i\right)\varepsilon_t$$

where L^i – delay operator, ϕ_i – parameters of the autoregressive part of the model, θ_i – variable mean parameters, ε_t – error value. It is assumed that the errors ε_t are independently equally distributed random variables with a normal distribution with zero mean. To obtain a clearer and more explicit dependency model, the ARIMA model is used. Autoregressive integrated moving average (ARIMA) is a generalization of the autoregressive variable mean model. These models are used when working with numerical rows for a deeper understanding of data or prediction of future points in a row. The model is considered as the ARIMA (p, d, q) procedure, where p, d i q – integral nonnegative numbers that characterize the order of parts of the model (autoregressive, integrated, and alternating mean respectively). ARIMA (p, d, q) is obtained after integrating ARMA (p, q).

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d X_t=\left(1+\sum_{i=l}^{q}\theta_i L^i\right)\varepsilon_t$$

where d – a positive integer that defines the level of differentiation (if $d = 0$, this model is equivalent to the autoregressive variable average). Conversely, by applying the differentiation d times to the ARMA model (p, q), we obtain the ARIMA model (p, d, q), with only autoregressive part to be differentiated. It is important to note that not all combinations of parameters give a "qualitative" model. The choice of ARIMA source model is based on the study of numerical series graphs and the study of autocorrelation coefficients for several time intervals. In particular, the structure of selective autocorrelation coefficients obtained for numeric rows is compared, and the autocorrelation structure associated with a particular ARIMA model is known. The Boxing-Jenkins methodology is based on a set of ARIMA determination, correction, and validation procedures for time series data. The forecast comes directly from the form of a corrected model [6].

Taking into account the above-mentioned problems, the forecasting methodology is used, which aims to:

1) the use of hidden information due to the structuring of numerical series by tensors of pair ranks and the use of their invariants;

2) when structuring the time series, the important information that characterizes this time series should be as low as possible;

3) the constructed model of the structured time series should produce a predicted value with a permissible error.

Presentation of the main material. At the stage of identification of the model, it is necessary to perform a time series check for stationary. This is most often used for visual analysis of selective autocorrelation (ACF) and partial auto-correlation (PACF) functions. For stationary time series ACF and PACF quickly fall after several first values. If the graphs slow down, then the time series may turn out to be non-stationary. Non-stationary time series can be transformed into stationary by taking differences. The starting line is replaced by a number of differences. Taking differences can be repeated several times. The number of reps taking the differences needed to obtain the steady-state behavior of the data is indicated by the parameter d. Also, at this stage, statistical tests are used for the presence of a single root (Advanced Dickey-Fuller test [1] - ADF).

After receiving the stationary series, the behavior of the selective ACF and partial PACF is studied and hypotheses about the values of the parameters p and q are put forward. During this, the basic set of ARIMA models is formed. We evaluate the parameters p and q of the ARMA model (p, q), which consists of models AR (p) and MA (q). To do this, it's easiest to use PACF and ACF, respectively. If the selective ACF is quickly cut off and the PACF exponentially moves to zero, the MA (q) should be present in the model. If the selective PACF is quickly cut off and the ACF goes to zero, then the AR (p) should be present in the model. In the event that ACF and PACF are directed to zero, then the modifications of the two types are included. The order of the model AR (p) corresponds to the number of the last non-zero coefficient of PACF, and the model MA (q) is the number of the last non-zero coefficient ACF.

We will show a graph of a series of densely estimated autocorrelation (ACF) and partial auto-correlation (PACF) according to the data of tourist streams of Ukraine in 2000-2017 [1] (Fig. 1).

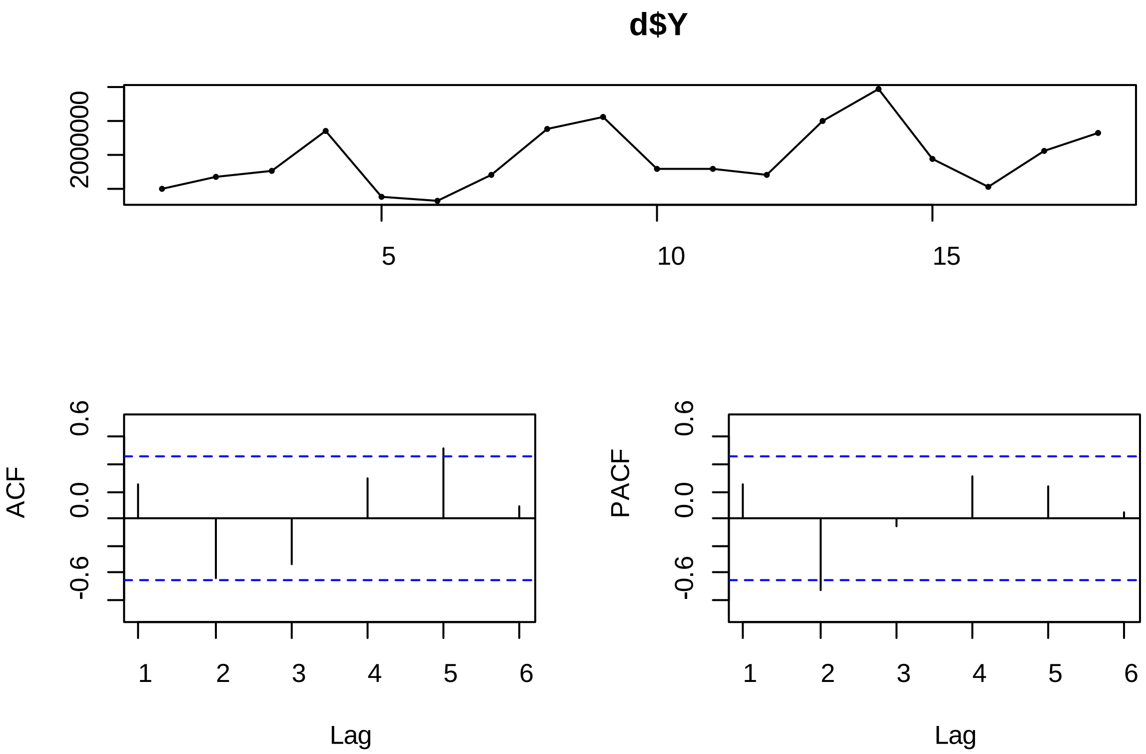

Fig.1 Graph of a series of autocorrelation (ACF) and partial auto-correlation (PACF) according to the data of tourist streams of Ukraine from 2000 to 2017.

It is difficult to accurately estimate ACF and PACF from the graph, because there is a lack of data; therefore, we estimate several competing models:

– Model-1: Evaluate ARMA (1,1) or ARIMA (1,0,1) (Fig. 2)

– Model-2: Evaluate ARMA (2.0) or ARIMA (2,0,0) (Fig. 3)

```
Series: y
ARIMA(1,0,1) with non-zero mean

Coefficients:
          ar1     ma1       mean
      -0.3624  0.9999  2442002.7
s.e.   0.2566  0.3334   126052.8

sigma^2 estimated as 1.624e+11:  log likelihood=-257.34
AIC=522.68   AICc=525.76   BIC=526.24

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
Training set 11680.17 367897.4 304134.7 -1.770237 12.34609 0.6979396 -0.05879461
```

Fig. 2. Evaluation of model-1 (ARMA (1,1))

Thus, the estimated model-1 ARMA (1,1) (Fig. 2) equation has the form:

$$\begin{cases} z_t = y_t - 2442002.7 \\ z_t = -0.36 \cdot z_{t-1} + \varepsilon_t + 0.99\varepsilon_{t-1} \end{cases}$$

```
Series: y
ARIMA(2,0,0) with non-zero mean

Coefficients:
         ar1      ar2        mean
      0.3895  -0.5264  2459307.76
s.e.  0.2012   0.1890    75894.87

sigma^2 estimated as 1.503e+11:  log likelihood=-255.88
AIC=519.76   AICc=522.84   BIC=523.32

Training set error measures:
                    ME     RMSE    MAE       MPE     MAPE      MASE        ACF1
Training set -9200.686 353881.4 297658 -2.456658 12.41248 0.6830767 -0.01878385
```

Fig. 3. Estimation of model-2 (ARMA (2,0))

Evaluation of model-2 ARMA (0.1) equation has the form (Fig. 3):

$$\begin{cases} z_t = y_t - 2459307.8 \\ z_t = 0.39 \cdot z_{t-1} + \varepsilon_t - 0.53\varepsilon_{t-1} \end{cases}$$

Both models showed a good match with the data. The models are adequate and the errors are random. The residual quadratic errors are almost the same. For such cases, several approaches to choosing a model were developed that take into account both the quality of fitting the model and the number of its parameters. Information criterion Akaike (Akaike), or AIC, allows you to choose the best model from the group of models of applicants.

Let's select the best model by penalty criterion AIC:

$$AIC = -2 \cdot lnL + 2 \cdot k$$

Where lnL - logarithm of the likelihood function, and k - number of parameters of the model.

The more parameters, *k*, the more complex the model, the higher the AIC. The lower the likelihood function, *L*, that is, the lower the probability of the data obtained in this model, the higher the AIC.

In accordance with the Beesovsky information criterion developed by Schwarz, we will calculate the BIC criterion:

$$BIC = -2 \cdot lnL + lnn \cdot k$$

Table 1. Model penalty estimates

Criterion / Model	AIC	BIC
Model 1	522.6831	526.2446
Model 2	519.7599	523.3214

By AIC and BIC criteria (see table.1) the best was the model 1 - ARMA (1,1).

Check the remnants of the selected model for the lack of auto-correlation using the Lew Box test:

$$H0: \rho 1 = \rho 2 = \cdots = \rho k$$

H_a: At least one of the correlations isn't equal to zero

$$\text{LB=n(n+1)} \sum_{k=1}^{h} \frac{\hat{e}_k^2}{n-k}$$

If to use the statistics of LB to the output series, then with the correct H_0 statistics has x^2 distribution with h degrees of freedom. If the remainder of the ARMA (p, q) models, then the number of degrees of freedom falls to $h - (p + q)$.

We evaluated the AR (2) model, so the degrees of freedom fall on $p + q = 2$ (Fig.4).

```
        Box-Ljung test

data:  resid_mod_1
X-squared = 18.007, df = 8, p-value = 0.02117
```

Fig. 4 Evaluation of the model by the Lew Box test

We do not discard H_0, so we can assume that the model correctly describes the structure of the correlation.

For off-season rows, Rob Hyndman recommends taking $lag, h = 10$, for seasonal $h = 2m$, where m - periodicity of seasonality, i.e. $h = 24$ for lunar data.

For ARMA (p, q) you can see visually, where the roots of the AR and MA are (Fig. 5)

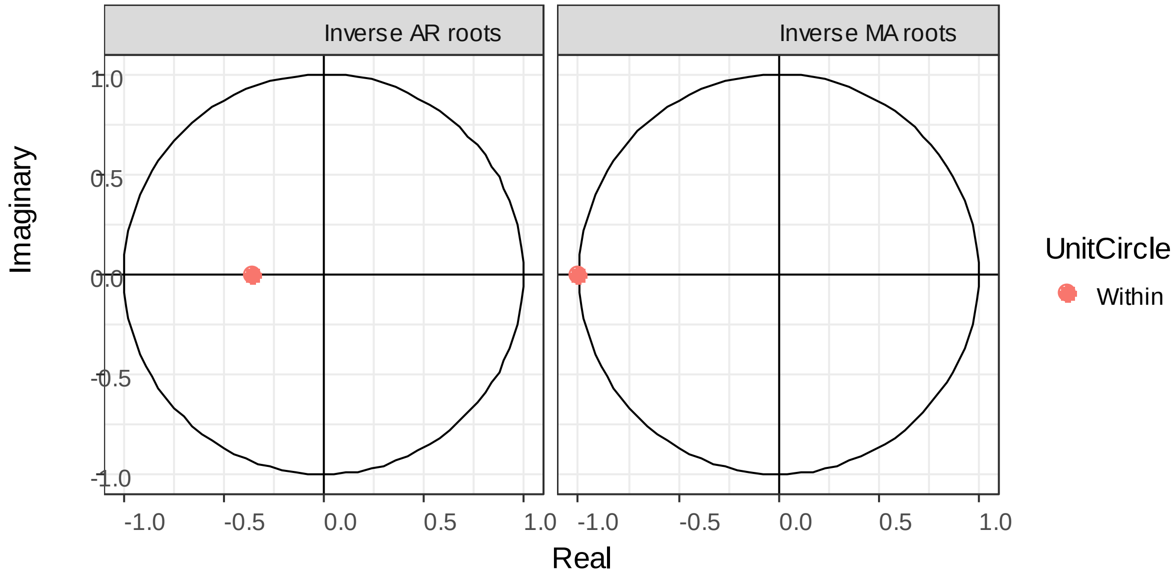

Fig. 5. Display AR and MA

By choosing the best model you can build the predictions presented in the table 2.

Table 2. Prediction of the tourist flow by the ARMA method

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
19	2559804	2030686	3088922	1750588	3369021
20	2399316	1785392	3013240	1460400	3338231
21	2457471	1833267	3081675	1502833	3412109
22	2436398	1810856	3061939	1479714	3393081
23	2444034	1818317	3069751	1487082	3400986

Or, it can be built a schedule of forecasts with intelligent intervals (Fig. 6).

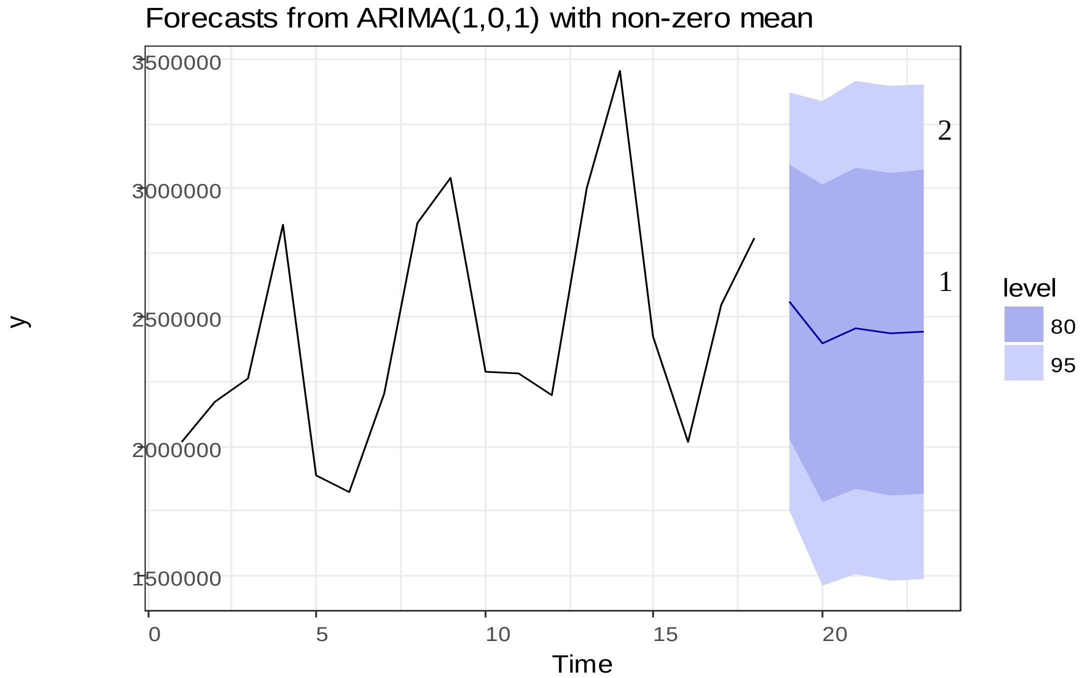

Fig. 6. Forecasting of the tourist flow by the ARMA method

The probability of reaching the predicted value (Fig. 6) in interval 1 is equal to 80% and in interval 2 is equal to 95%. Also, the chart shows that the tourist demand in the next 5 years will continue to decline.

Conclusions. The article investigates:

1. the process of formation of tourist demand and calculated autocorrelation and partial autocorrelation.
2. the behavior of the sample ACF and the partial PACF is evaluated, however, due to lack of data, several standard ARMA (1.1) and ARMA (2.0) competing models have been selected.
3. the model 1 - ARMA (1,1) is selected according to criterion AIC and BIC.
4. for Model 1, the remains of the selected model were checked for the absence of autocorrelation using the Lew Box test.
5. for the selected best model, forecasts for 5 forward periods have been constructed, which shows that tourist demand in the next 5 years will decline.

REFERENCES

1. http://www.ukrstat.gov.ua - Official statistical information of the State Statistics Committee of Ukraine
2. Box J. Analysis of time series. Forecast and management: Vol. 1 (Translated from English by A.A. Levshin) Moscow: Mir, 1974. p. 406
3. Venables W. N., Smith D. M. An Introduction to R [Electronic resource]. Access mode: http://www.e-booksdirectory.com/details.php?ebook=1791
4. Bidyuk P.I. Analysis of time series. / P.I. Bidyuk, V.D. Romanenko O.L. Tymoshchuk - Kyiv: "Polytechnic", NTUU "KPI", 2013. - p. 600
5. Minayev Yu.M., Aponasenko D.V. Intelligent technologies of forecasting of time series on the basis of tensor invariants // Problems of informatization and management: Collection of sciences. Ave - K.: NAU, 2009. - Issue number 2 (26). - p. 105-113
6. Prisenko G.V., Ravikovich E.I. Prediction of Economic Processes: Textbook / Prisenko G.V., Ravikovich E.I. - Kyiv: KNEU, 2005. - p. 382