ISSN 2181-2551
E-ISSN 2181-2561

2023
3(11)

IQTISODIYOT VA TURIZM
xalqaro ilmiy va innovatsion jurnali

ЭКОНОМИКА И ТУРИЗМ
международный научно-инновационный журнал

ECONOMICS AND TOURISM
international scientific and innovative journal

ISSN 2181-2551
9 772181 255890

E-ISSN 2181-2561
9 772181 256001

MUNDARIJA

UDC 338

Khidirova Gavkhar Rustamovna,
doctoral student at Bukhara State University.

IMPROVING FORECASTING OF DOMESTIC TOURISM IN THE BUKHARA REGION BY THE ARIMA MODEL.

Abstract: In this article, it is proven that ARIMA data evaluation can be used to forecast the number of tourists, tourism revenue, the role of tourism in the economy, and other tourism indicators in the medium and long term for domestic demand in the region. The main goal of forecasting is to determine the future size of indicators based on existing data. In order to evaluate the main determinants of the Bukhara region (population income, population size, and structure), it was analyzed using the ARIMA method. It is predicted that the number of permanent domestic tourists in the region will exceed 3 million by 2027.

Key words: determinant, ARIMA method, "Stata" program, autocorrelation, hypothesis, econometrics, Maximum Likelihood Estimation method, Stationary, Dickey-Fuller test, model ARI, Portmanteau test.

Аннотация: ушбу мақолада ARIMA услуби маълумотларини баҳолаш минтақада ички талаб учун бугунги кунда туристлар сони, туризм тушуми, туризмни иқтисодиётдаги ўрни ва бошқа туризм кўрсаткичларини прогнозларини ўрта ва узоқ муддатда амалга ошириш мумкинлиги исботланган. Прогнозлаштиришнинг асосий мақсади мавжуд маълумотларига таянган ҳолда кўрсаткичларни келажакдаги ҳажмини аниқлаштириш. Бухоро ҳудудининг асосий детерминантларини (аҳоли даромадлари, аҳоли сони ва таркиби) баҳолаш учун ARIMA услуби ёрдамида таҳлил қилинди, вилоятда 2027 йилга бориб доимий ички туристлар сони 3 млн дан ортиқ бўлишига прогноз қилинган.

Калит сўзлар: детерменант, ARIMA услуби, “Stata” дастури, автокорреляция, гипотеза, эконометрика, Maximum likelihood estimation методи, Стационар, Dickey-Fuller тести, модел ARI, Portmanteau тести,

Аннотация: в данной статье доказывается, что оценка данных ARIMA может быть использована для прогнозирования количества туристов, доходов от туризма, роли туризма в экономике и других показателей туризма в среднесрочной и долгосрочной перспективе для внутреннего спроса в регионе. Основная цель прогнозирования – определение будущих размеров показателей, на основе существующих данных. Для того чтобы оценить основные определяющие факторы Бухарской области (доходы населения, численность и состав населения) с помощью метода ARIMA было проанализировано, что количество постоянных внутренних туристов в регионе к 2027 году прогнозируется на уровне более 3 миллионов.

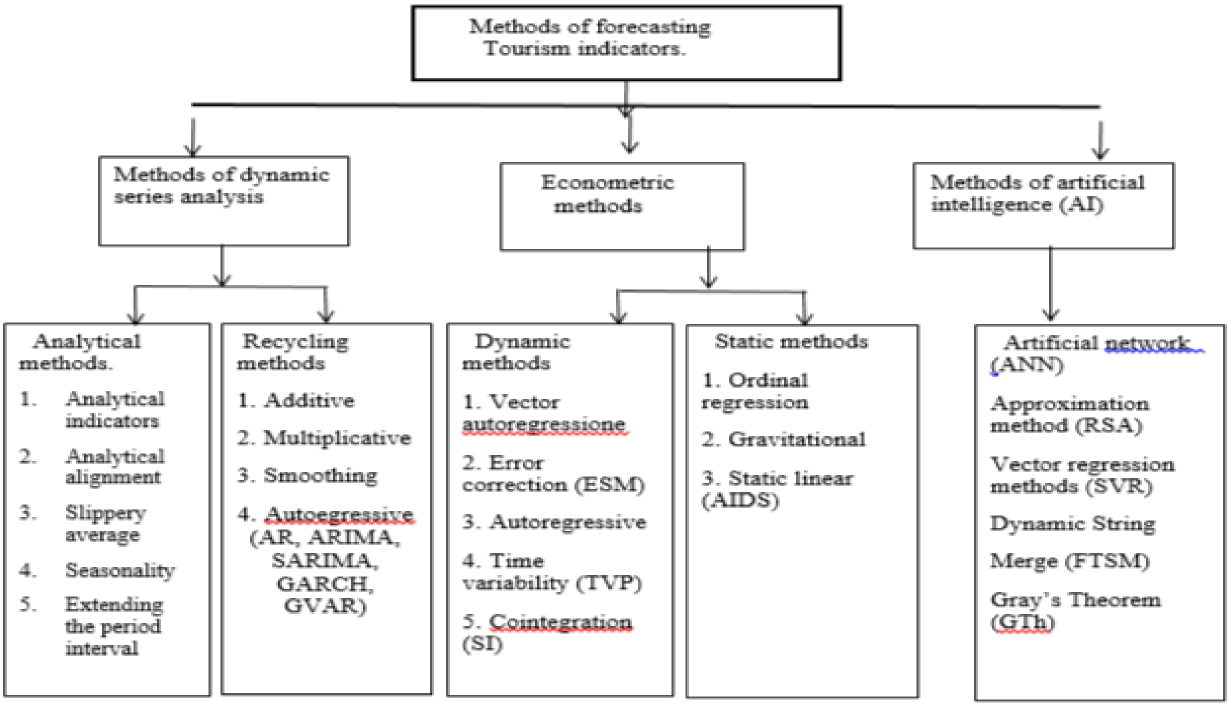
Ключевые слова: детерминант, метод ARIMA, программа «Stata», автокорреляция, гипотеза, эконометрика, метод оценки максимального правдоподобия, стационарный метод, тест Дикки-Фуллера, модель ARI, тест Портманто.

Introduction: It is known that in many countries domestic tourism dominates international flows in terms of volume and economic contribution. However, researchers have recently begun to focus on this phenomenon and its economic impact, as well as its potential to reduce inequality in less developed regions of the world. It follows that research on the determinants of tourism choice across national borders is still rare and focuses mainly on the influence of economic variables.

Today, it has been proven that it is possible to forecast the number of tourists, tourism revenue, the role of tourism in the economy and other tourism indicators in the medium and long term, based on more than 150 methods (Fig. 1). The main goal of forecasting is to determine the future size (number, amount) of indicators based on existing data (real).

The results of our research showed that the conducted scientific research works in the field of tourism, applying statistical methods and modeling processes, were mainly carried out in the cross-section of years, and not enough attention was paid to modeling taking into account the seasonality of the flow of tourists.

Body: This paper aims to contribute to the research field of tourism flow forecasting. Dynamic series analysis methods used in forecasting tourism indicators are widely used. They, in turn, are divided into two groups, the first group includes analytical methods (methods of analytical indicators, analytical smoothing, determining seasonality, extending the period interval, studying slippage), and the second group includes processing methods (additive, multiplicative, autoregression and smoothing). Among the above methods, the most effective method is the ARIMA method, which is expressed in the form of a sum of trend, seasonality, cyclical and random events that reflect the characteristics of a dynamic series. The number of domestic tourism tourists was forecasted using the ARIMA method by estimating the main determinants of Bukhara region based on the regional dataset.



Picture 1. Methods of forecasting tourism indicators.[1]

ARIMA is performed using the Maximum Likelihood method rather than the Least Squares method. The maximum likelihood method was first used in scientific research by (Karl Friedrich Gauss, Pierre Simon Laplace, Thorvard Tillar)[2]. However, this method was widely used between 1912 and 1922. Ronal Fisher, considered one of the founders of econometrics, explained the advantages of this method, as well as its differences from least squares, in his research[3].

Maximum likelihood estimation method indicators are found through the log likelihood function.

The log liklihood function looks like this:

$$P(y_i = 1)^{y_2} P(y_i = 0)^{1-y_i}$$

When $y_i = 1$ *than* $P(y_i = 1)^1 P(y_i = 0)^{1-1} = P(y_i = 1)$

When $y_i = 0$ than $P(y_i = 1)^0 P(y_i = 0)^{1-0} = P(y_i = 0)$

[1] author's development

[2] Edgeworth, Francis Y. (Sep 1908). "On the probable errors of frequency-constants". Journal of the Royal Statistical Society. 71 (3): 499–512. doi:10.2307/2339293. JSTOR 2339293.

[3] Wilks, S.S. (1938). "The large-sample distribution of the likelihood ratio for testing composite hypotheses". Annals of Mathematical Statistics. 9: 60–62. doi:10.1214/aoms/1177732360

An overview of the log likelihood function:

$$\sum_{i}^{n} = 1(^{y_i} * \log_P(y_i = 1) + (1 - y_i) * \log_P(y_i = 0))$$

If we replace P(y=1) with G(xß) log likelihood function also changy:

$$\sum_{i}^{n} = 1^{(y_i} * log(G(x\beta)) + (1 - y_i) * log(1 - G(x\beta))$$

The coefficient β is found by maximizing this equotion. Thus based on the method of maximum likelihood estimotion, to find the coefficient β the following equotion represents the situation when it has a maximum.

$$max \sum_{i}^{n} = 1(y_i * logP(y_i = 1) + (1 - y_i) * logP(y_i = 0))$$

Based on the maximum likelihood estimotion method the statistical significance of the β coefficient is checked with Wald test, score test, likelihood ratio test.

Statistical significance of the coefficient based on maximum likelihood estimation method is checked by Wald test, score test, likelihood ratio test.

At this point, we will briefly touch on the ARIMA method. This method was introduced to science in the 1970s by scientists George Box & Jillian Jenkins, who were awarded the Nobel Prize for their invention. Trend, seasonality, randomness, cycle components in time series are taken into account in forecasting using ARIMA method. The essence of the ARIMA method is reflected in its name: AR - Autoregression, I - Integrated, MA - Moving Average.

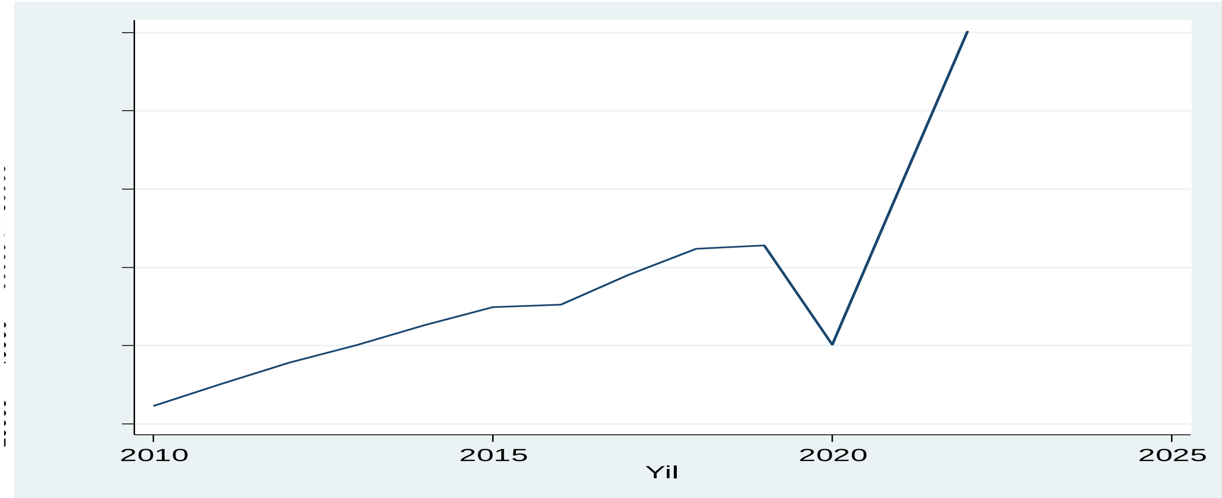
It is known that economic processes expressed in the form of time series are forecasted in 2 different ways: univariate as a one-dimensional time series, and multivariate modeling in the study of connected time series. A simple time series (univariate) is based on the change of only one indicator over time. A multivariate time series is forecasted by a model built on the basis of other indicators influencing the selected indicator. Only simple time series are forecasted by the ARIMA method. ARIMA forecasting has 3 steps:

Identification- IDENTIFICATION;

Estimation- ESTIMATION;

Forecasting- FORECASTING.

After downloading the data into Stata, our research begins with the initial IDENTIFICATION step of the ARIMA method. At the detection stage, the values of "p", "d", "q" are determined. For this, the time series is checked for stationarity, in which the period of the integration process is found, which is expressed by the value "d".



Picture 2. Determining stationarity graphically

It is known that there are 3 different ways to check the stationarity of a time series, consisting of the graphical method, the autocorrelation method, and the test method.

By graphically checking the time series for stationarity in the "Stata" program, the presence or absence of elements of trend, randomness, and seasonality in the time series is shown.

It can be seen from the above graph that the time series has elements of trend, cyclicity and randomness. However, in order to be completely sure of the stationarity of the time series, it is desirable to evaluate it by means of a test method. One such method is the Dickey-Fullar test. The Dickey-Fuller test, like other methods, requires a p-value < 0.05. Based on the p-value, the following hypotheses are accepted or rejected:

H0: the time series belongs to the stationary type

H1: the time series is not stationary.

If p-value < 0.05, the main hypothesis is accepted, the alternative hypothesis is rejected, on the contrary, if p-value > 0.05, the alternative hypothesis is accepted, the main hypothesis is rejected. If a time series is found to be non-stationary, it can be converted to a stationary representation. this requires differentiating the time series. In econometrics, this is called an integration process.

When the time series is proved to be stationary at any stage of the integration process, this period represents the value d.

If we check the time series using the Dickey-Fuller test in the Stata program, it was found that our time series is non-stationary at the initial stage.

```
Dickey-Fuller test for unit root                   Number of obs   =        10

                               ----------- Interpolated Dickey-Fuller ----------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)              0.274            -3.750            -3.000            -2.630
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.9761

.
```

Picture 3. Dickey-Fuller test.

According to the results of the Dickey-Fuller test, p-value = 0.9761, which does not satisfy the condition of p-value < 0.05, so it is necessary to perform the Dickey-Fuller test again by differentiating the time series.

When the time series is re-examined by the Dickey-Fuller test, we see that it is stationary:

```
. dfuller d.TBL

Dickey-Fuller test for unit root                   Number of obs   =         9

                               ----------- Interpolated Dickey-Fuller ----------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.870            -3.750            -3.000            -2.630
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0489
```

Picture 4. Dickey-Fuller test.

According to the results of the repeated Dickey-Fuller test above, p-value = 0.0489, which fully satisfies the condition of p-value < 0.05, that is, the integration process is stopped here.

Therefore, the d value required for the identification step of the ARIMA method is equal to 1: d =1.

The meaning of p and q values is that they indicate the presence of autocorrelation of the time series and its residuals. If there is no autocorrelation, it is taken as p=0 or q=0, if there is autocorrelation, its order is taken as p= [1;2] or q= [1;2]. It should be remembered that if the time series is converted to a stationary form in the differentiated state, the p and q values are found on the same transformed time series.

First, to determine whether the variables in the time series are related or not, that is, to determine the value of q, we give the necessary command in the "Stata" program and get the following graph.

There is no autocorrelation in the time series because all lags lie on the surface of the confidence interval. Based on this, we can set the value of q as 0: q=0.

A special feature of the ARIMA method is that it selects types (families) of forecasting based on the determined p, d, q values. In particular, since q=0, the model representation is expressed as ARI.

There is an autocorrelation of order 2 between the residuals. This is because 2 of the lags lie outside the confidence interval surface. This means that the indicator in the current state of the residuals depends on the change in the indicator of 2 periods (years) before it. Therefore, the number 2 is accepted as the value of p: p=2.

Thus, we write down the combinations of p, d, q values necessary for the ARIMA method in the following sequence:

ARIMA = (p,d,q) = (2, 1, 0)

ARIMA = (p,d,q) = (1, 1, 0)

ARIMA = (p,d,q) = (0, 1, 0)

As ARIMA = (p,d,q) = (0, 1, 2).

In these combinations, only the value of d does not change, the value of p and q varies between 0 and 2.

At the Estimation stage of the ARIMA method, the significance (quality) of the models selected above is evaluated according to 5 different criteria and the best one is selected.

According to the evaluation results, among the selected models, ARIMA (0,1,0) is optimal compared to other models, because this model fully meets all criteria requirements.

The final stage of the ARIMA method - we develop forecast options for the optimal ARIMA (0,1,0) selected in the Forecasting stage. Nowadays, it has become an integral principle of applied econometrics - evaluation of any model by its residuals is also important in ARIMA forecasting. Based on this, before forecasting, it is advisable to check the residuals of the selected model using different methods. The average of the sum of residuals is equal to 0 (or at least close to 0) and the [AR][I][MA] parameters, i.e. the available eigenvalues in the form of p,d,q are located inside the unit circle are the basis for reliable forecasting using the selected model is considered

To determine the sum of the balance column, the results of the ledger will be displayed:

```
. summ qoldiq4
```

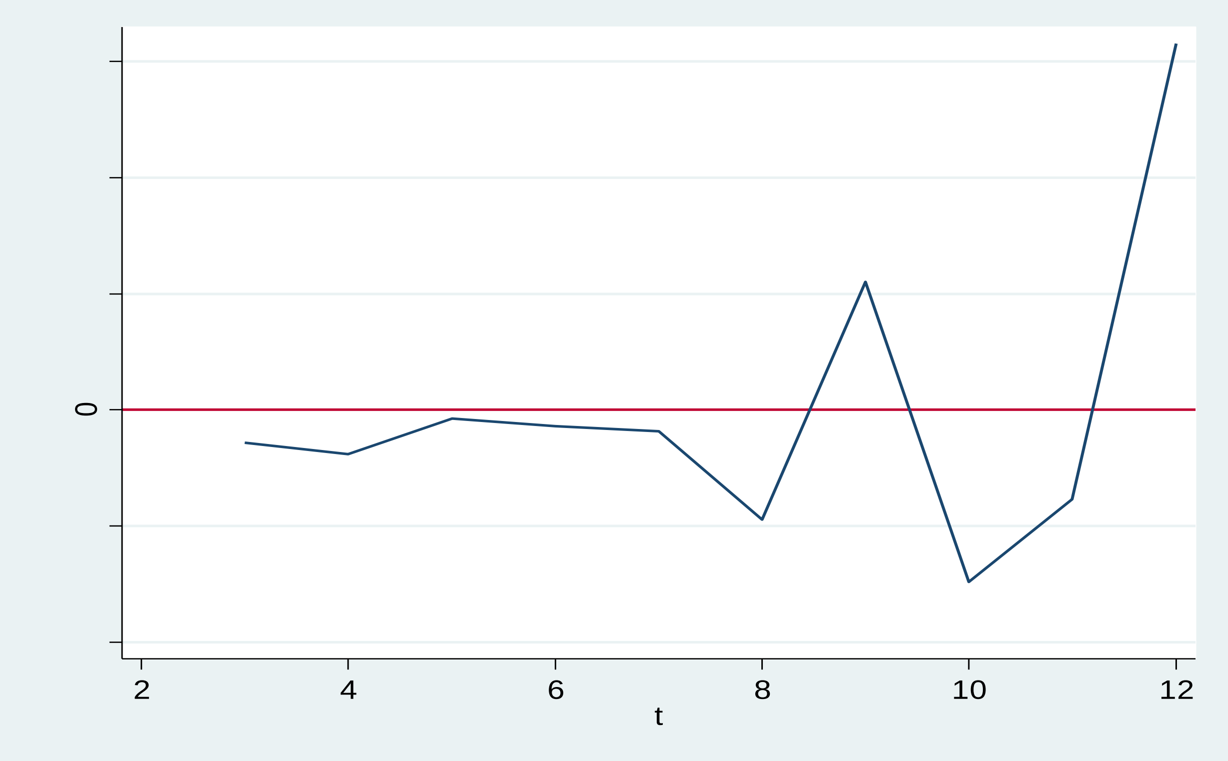
Variable	Obs	Mean	Std. Dev.	Min	Max
qoldiq4	10	.8125	2.59e+07	-2.96e+07	6.30e+07

Pictures 5. Determining the sum.

From the image above, we can see that the average of the sum of the 21 rows of the residual column is 0.000, which when rounded up represents 0, so the condition is satisfied.

Also, in order to graphically express the fact that the residuals are in a stationary state, it is enough to give the necessary command to the sequence of commands, in which the following image is displayed:

It can be seen from Figure 6 that the residuals are almost stationary, that is, during the period, the residuals have changed on average around the line u=0.000.



Picture 6. Expression of stationarity.

A graph alone is not enough to be completely sure of stationarity of the residuals, there are many tests other than the Dickey-Fullar test to confirm stationarity that can be used to assess stationarity, including the Portmanteau test, as well as the IM-test, Brosh-Pagan, Shapiro-Wilk, Brosh-Godfrey tests, the p> chi indicator should be greater than 0.05. When determining this value, the following hypotheses are accepted or rejected:

H0: the residuals tend to the property of stationarity;

H1: residuals are not stationary.

After checking for stationarity using the portmanteau test, the following table appears:

```
Portmanteau test for white noise
───────────────────────────────────────────────

 Portmanteau (Q) statistic =      3.3830
 Prob > chi2(3)            =      0.3363
```

Picture 7. Portmanteau test

From picture 7, we can see that according to the results of the Portmanteau test, p > chi 2 (1) = 0.33, which is significantly greater than 0.05, so we can start by rejecting the alternative hypothesis.

Another important precondition for forecasting is the stability condition of the eigenvalues, which requires that the values of the parameters p and q lie within the unit circle. Since the selected optimal model is in the form of ARI, we only examine p values.

Based on the results of the above investigation, we can say that the selected ARI (0.1.0) model has been positively evaluated for its stationarity and stability under the required conditions, and this model can be used as the most optimal (optimal) model for forecasting future growth patterns.

Visualizing how close the initial given empirical data are to each other over a cross-section of years, where both time series are displayed on the same graph:

We can interpret the results of the 5-year forecast calculated according to the selected ARI (0.1.0) model from the pessimistic and optimistic point of view: (Table 1). The number of domestic tourism tourists in the region in 2027 is forecasted to be almost 3.0 million. This forecast shows that the number of tourists will increase almost 20.0 times in the near future

Table 1

Forecast indicators of the number of tourists visiting Bukhara region through ARIMA method[1]

Years	2023	2024	2025	2026	2027
Forecasting for 1000 people	150.0	456.0	485.0	1253.0	2730.0

Our analysis based on the ARIMA (0.1.0) method, its significance was confirmed by a number of tests, particularly Dickey-Fullar, Portmanteau, Brosh-Pagan, Shapiro-Wilk, Brosh-Godfrey tests. According to our hypothesis, the population was divided into five groups, each of them was compared with the predicted values of the correlation coefficient with the number of domestic tourism tourists (the coefficient of 0-9 years old 4.056, 10-19 years old-4.32, 20-29 years old-2.79, 30 -3.6 under the age of 54, 1.79 for the population over the age of 55)

In conclusion, based on the above ideas, the following suggestions have been developed in order to improve the forecasting of the factors determining the internal tourism of Bukhara region in the ARIMA model:

- since tourism has a seasonal nature, it is necessary to use methods that express seasonality in tourism forecasting;
- the results of international practice show that ARIMA is the best method for forecasting tourism indicators;
- in the course of the research, forecasts of the number of domestic tourism tourists in the Bukhara region were determined and proven by the ARIMA (0.1.0) method;
- it has been determined that the number of permanent domestic tourists in Bukhara region will exceed 3.0 million by 2027, and all the factors that divide the country's population into 5 groups are taken into account;
- improving the tourism infrastructure (roads, hotels, service facilities) in Bukhara region in order to put the forecast indicators into practice.

List of used literature

1. Edgeworth, Francis Y. (Sep 1908). "On the probable errors of frequency-constants". Journal of the Royal Statistical Society. 71 (3): 499–512. doi:10.2307/2339293. JSTOR 2339293
2. Wilks, S.S. (1938). "The large-sample distribution of the likelihood ratio for testing composite hypotheses". Annals of Mathematical Statistics. 9: 60–62. doi:10.1214/aoms/1177732360
3. Каморников С.Ф., Каморников С.С. Эконометрика: учеб. Пособие.-М.: Интеграция, 2012.-262с.
4. И.И.Елисеева. Эконометрика: учеб./под ред.-М. Э40 Проспект, 2009.-288с.
5. Carla Massidda, Ivan Etzo. The determinants of Italian domestic tourism: A panel data analysis. Tourism Management Volume 33, Issue 3, June 2012, Pages 603-610

[1] Source - author accounts