Modelling of Air Passenger Transportation in Russia

Olga P. Sushko
Moscow State Technical University of Civil Aviation, Moscow, Russia. o.sushko@mstuca.aero.

ABSTRACT

The use of economic and mathematical methods of forecasting the results of activities of civil aviation organisations, and in particular assessment of the volume of air passenger traffic is quite relevant due to the importance of operational planning of air transport processes, development of strategic directions, technological and technical renewal of air enterprises.

The objective of the study is to plan the traffic flow of air passengers using a regression model, considering the results of multifactorial selection of determinants, particularly distinguishing fundamental macro indicators are distinguished, as well as significant indicators of the aviation market.

The study of passenger air transportation was carried out using methods of system analysis, methods of mathematical statistics and econometrics. Modelling of the process of passenger transportation has identified the main determinants that positively or negatively affect the dynamics of air passenger traffic. The multiple regression of the study of the processes of connectivity and synchronicity of changes in development of passenger traffic and selected macro indicators in a generalised form is the sum of vectors of influencing variables adjusted for the calculated coefficients.

Six-, four- and three-factor regression models were developed. The three-factor model turned to be more reliable with values most close to actual data. Nevertheless, while applying regression model for forecasting air traffic it is necessary to consider not only theoretical aspects, data of official forecasts of macro indicators but expert opinions as well.

Keywords: air transport, factor features, air passenger traffic modelling, multiple regression, statistical verification of the model, forecasting of air passenger flow.


The text of the article originally written in Russian is published in the first part of the issue. Текст статьи на русском языке публикуется в первой части данного выпуска.
INTRODUCTION

The significance and degree of development of transport, including air transport, follows in many aspects the development trends of the global and of the entire country’s economy. The modern aviation business, and Russian air business makes no exception, faces a whole set of problems and challenges of social, economic, and technological character, including fundamentally new ones1-2 [1–5]. The search for solutions to new problems in the field of air transportation involves a comprehensive analysis of long-term fundamental factors and identification of new factors that affect the change in the economic performance of both airlines and aviation enterprises, and of the entire industry. For the purposes of modelling short- and long-term forecasting, it is also important to identify the main determinants that positively or negatively affect the dynamics. This confirms the relevance and necessity of applying economic and mathematical methods for forecasting the results of activity of civil aviation, and in particular of the volume of air passenger transportation due to the importance of operational planning of air transportation processes, development of strategic development guidelines, technological and technical renewal of aviation enterprises.

Overview of Research

Many works dedicated to fundamental and applied research refer to modelling of the processes of transportation of goods and passengers, differing in complexity, the possibility of implementation, and the objectivity of interpreting the data obtained [6–8]. They suggest complex models such as neural networks, artificial intelligence models, Markov chains, and others. Certainly, forecasting techniques developed by airlines can be useful. The Boeing company1 uses a passenger flow forecasting model \(Y = a_0X + a_1X_2 + a_2X_3\), where \(X_1\) is GDP index; \(X_2\) – average fare; \(X_3\) – consideration of other factors; \(a_0, a_1, a_2\) – coefficients.

Scientific interest can focus on a debatable discussion of the impact of new factorial realities, such as global changes in aviation market, institutional transformations of aviation infrastructure, improvement of the technical and

\[\ln Y = -3.21 + 1.88\ln X.\]

Aeroflot Airlines, as far as it is known, uses a method for forecasting the volume of air transportation based on the logarithm of the gross product and the difference in income from passenger air transportation:

\[\ln Y = 1.14 + 2.11\ln X - 0.63\ln Z,\]

where \(Y\) is air transportation pass/km;
\(X\) – GDP;
\(Z\) – income from passenger transportation per pass/km.

Airlines’ data are used among other purposes for planning of airport activity, slot distribution among air carriers4.

Thus, in the aviation business, when forecasting passenger air transportation, special models are used based on the dependence on the gross product, but the resulting forecasts may have errors, since only a single factor is considered. Moreover, according to several researchers and analysts involved in the study of the economics of the transport industry, the existing connection between an increase in passenger air transportation by 1 % and an increase in gross domestic product by 1 % is not characteristic for all the countries, since for developing countries, this ratio is higher. Some research based on data for 2008–2021 do not confirm definitely those proportions due to the fact that the dynamics of passenger turnover has some limits of fall/growth, as well as time lags from movement of economic factors till their impact on the rates of air transportation, which is reflected in the amplitude of traffic fluctuations in periods of crisis and periods of economic growth.

A more complex model for studying the connectivity of passenger flow and tariffs, and other factors is based on several indicators:

\[Y = a_0X_1 + X_2 + X_3,\]

where \(X_1\) is GDP index;
\(X_2\) – average fare;
\(X_3\) – consideration of other factors;
\(a_0, a_1, a_2\) – coefficients.

Scientific interest can focus on a debatable discussion of the impact of new factorial realities, such as global changes in aviation market, institutional transformations of aviation infrastructure, improvement of the technical and

---

technological platform of the aviation business [9–16]. But the results obtained have been till now difficult to interpret and time-consuming for the practical implementation for forecasting the volume of air transportation.

However, the importance of forecasting air travel processes based on econometric modelling remains high, thanks first to possibility to obtain more reliable results for practical use as compared in comparison with the conclusions obtained using other methods.

Statement of the Research Problem

The subject of research is the economy of the aviation industry. At this stage of the study, the object of the study is the dynamics of passenger air transportation and economic indicators as the main regressors.

The objective of this stage of the study is presented as identification of the main factors affecting air passenger traffic and improving the efficiency of air passenger traffic modelling based on multiple regression. Multiple regression modelling allows us to include many factors with establishment of the connectedness of the series and identification of the individual influence of each factor indicator on the simulated series of air transportation, followed by the calculation of the total impact.

The novelty of the study of the processes of air transportation of passengers is associated with development of a multiple regression model for the time series of air passengers and economic indicators, considering in full the information on the correlation between the data of the simulated series and the time series of factor indicators. Also, the resulting model allows monitoring the correlation of time series and the ability to rank factors to ensure greater reliability of the results.

The study of passenger air transportation was carried out using the methods of system analysis, methods of mathematical statistics and econometrics. Development of the experimental base of the study comprised analysis of information and statistics resources of international organisations, state bodies of statistical observation (Federal State Statistics Service, EMISS⁷), information and analytic websites (Statista⁸), including those specialising in aviation topics (Aviastat⁷, Aviaport and Aviapro⁹), as well as data of research publications, other statistical and analytical materials [12–16].

RESULTS

Construction of Fundamentals of the Model and Algorithm

To build a multifactorial model for forecasting air passenger flow, an important task is to select significant indicators as factor variables and determine the number of factors. The selection of indicators that affect the dynamics of air passenger flow for inclusion in the developed regression model was carried out based on the results of a qualitative and quantitative analysis.

The market mechanism for formation of supply and demand depends on the price and many non-price factors, among which the main ones are the number and income of the population. Accordingly, these factors are included in the model. Other factors were ranked according to the results of a survey of researchers involved in the study of transportation processes, specialists, and representatives of the aviation industry. Nevertheless, it is impossible to consider and include all factors in the regression, but this is not so critical, since the regression calculates not only the influence of the included regressors, but also the influence of factors not included, but at the same time associated with the main explanatory variables. It should also be noted that oversaturation of the model with factor variables can lead to statistical insignificance of the parameters according to Student’s t-test, which will be further shown analysing six- and more factor model.

Further, the expediency of including factors in the model was carried out based on statistical methods. The factors were analysed to determine whether they have a correlation dependence on the resulting factor. The research studied the synchronism of changes in development of the main indicator of development of air transportation (passenger traffic) and a number of macro-indicators: gross domestic product, gross domestic product per capita, population size, average wages and average income of the population, the exchange rate of the national currency to

dollar, inflation and unemployment, oil price changes, airfare prices, etc. (Pic. 1) [14–16].

Interpretation of the obtained results of the correlation study showed a different connection and dependence of the data on the dynamics of air passenger flow and the expected factor indicators, which led to the exclusion of several factors. Among the remaining significant factors, only those that have comparable units of measurement should be left since different units of measurement can lead to incomparable coefficients of the equation. Another important procedure for analysing the significance of factor variables as regressors is to bring the values of indicators to the same measurement scale. The rule of including a smaller number of factors compared to the number of observations was also respected. Since data from 2008 to 2021 were selected for the study, 4–5 regressors can be considered optimal.

Based on the preliminary results of the study of many economic factors, significant factor variables were selected to develop the multiple regression equation. The multiple regression of the study of the processes of connectedness and synchronism of changes in development of air transportation (passenger traffic) and pf series of macro indicators in general terms is the sum of the vectors of influencing variables, adjusted for the calculated coefficients, provided that they must be consistent.

Model Formulation

Evaluation of the obtained regression equation for the studied time series [17–19] is performed by compiling a matrix (Table 1) and of an inverse matrix.

The initial regression equation includes six explanatory variables:

\[ Y = -383.08 - 7.496X_1 + 1.093X_2 + 2.293X_3 + 3.687X_4 + 11.498X_5 + 1.091X_6, \quad (1) \]

where \( X_1 \) is average salary;
\( X_2 \) – average income of the population;
\( X_3 \) – population size;
\( X_4 \) – gross product;
\( X_5 \) – air ticket price;
\( X_6 \) – average oil price.

The calculated value of the constant in the multiple regression equation (–383.08) is negative and shows the total impact of unaccounted factors on the result (air passenger flow). And by the difference in multiple determination \( (1 - R^2) \) with the corresponding residual variance, it is possible to determine the share of influence of unaccounted factors.

The coefficients in front of the variables mean a decrease or increase in air passenger flow under the influence of the analysed factor, and the significance of the influence is determined by the value of the coefficient in front of the factor. The air ticket price factor has the greatest influence on the resulting indicator (passenger flow), which shows the maximum coefficient of 11,498 in the obtained regression model. The least influence

![Diagram of the analysis of totality of economic factors](developed by the author)
in this six-factor multiple regression model is exerted by indicators with the minimum values of the coefficients: average income of the population and average oil price.

But before substantiating and excluding insignificant factors from the multiple regression model, it is necessary to calculate the pair correlation coefficients with the compilation of a matrix with a dimension of 14 x 6, taking into account the number of observations 14 and with 6 independent variables and considering the sign (Tables 1 and 2). The significance of determining the pair correlation coefficients is due to the fact that the obtained values show the strength of the correlation of an individual factor on air passenger flow, while the influence of other factors is eliminated. Small values of the correlation coefficients mean that the relationship between factors and air passenger flow is weak, and, accordingly, considering the above feature, such factors can be excluded from the model.

The next procedure in the study was checking the functional relationship of the analysed factors with each other using pair correlation calculations to eliminate intercorrelation. High intercorrelation of factors in the multiple regression model leads to distortion of the results. Therefore, according to the results of pair correlation, collinear variables with a high linear dependence were established in the model (the value of the coefficient is higher than 0.7 modulo) (Table 1). But due to the significance of the impact on air passenger flow of such macro indicators as gross domestic product and population, it is not advisable to exclude them from the regression model, therefore these factors are integrated into a single variable: gross product per capita.

To expand the possibilities of meaningful analysis of the regression model, partial elasticity coefficients were calculated (Table 2).

### Table 1
Matrix of paired correlation coefficients of the analysed indicators [developed by the author]

<table>
<thead>
<tr>
<th>Passenger flow</th>
<th>Average salary</th>
<th>Average income</th>
<th>Size of the population of the Russian Federation</th>
<th>GDP</th>
<th>Air ticket price</th>
<th>Oil price</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>x_1</td>
<td>x_2</td>
<td>x_3</td>
<td>x_4</td>
<td>x_5</td>
<td>x_6</td>
</tr>
<tr>
<td>y_1</td>
<td>0.7033</td>
<td>0.8387</td>
<td>0.7646</td>
<td>0.8403</td>
<td>-0.1154</td>
<td>0.6834</td>
</tr>
<tr>
<td>x_1</td>
<td>0.7033</td>
<td>0.9791</td>
<td>0.8568</td>
<td>0.9871</td>
<td>0.07357</td>
<td>0.5612</td>
</tr>
<tr>
<td>x_2</td>
<td>0.8387</td>
<td>0.9791</td>
<td>0.891</td>
<td>0.9871</td>
<td>-0.0051</td>
<td>0.5791</td>
</tr>
<tr>
<td>x_3</td>
<td>0.7646</td>
<td>0.8568</td>
<td>0.891</td>
<td>0.849</td>
<td>0.1705</td>
<td>0.3353</td>
</tr>
<tr>
<td>x_4</td>
<td>0.8403</td>
<td>0.9871</td>
<td>0.849</td>
<td>1</td>
<td>-0.04295</td>
<td>0.6392</td>
</tr>
<tr>
<td>x_5</td>
<td>-0.1154</td>
<td>-0.00518</td>
<td>0.1705</td>
<td>-0.04295</td>
<td>1</td>
<td>-0.2984</td>
</tr>
<tr>
<td>x_6</td>
<td>0.6834</td>
<td>0.5612</td>
<td>0.5791</td>
<td>0.3353</td>
<td>0.6392</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table 2
Partial coefficients of elasticity of factorial variables in a six-factor model [developed by the author]

<table>
<thead>
<tr>
<th>Partial coefficient of elasticity</th>
<th>Coefficient value</th>
<th>Value interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>E_1</td>
<td>3.05</td>
<td>Average salary significantly affects the effective feature Y</td>
</tr>
<tr>
<td>E_2</td>
<td>1.36</td>
<td>Average income affects the effective feature Y moderately</td>
</tr>
<tr>
<td>E_3</td>
<td>3.95</td>
<td>Number of the population significantly affects the effective feature Y</td>
</tr>
<tr>
<td>E_4</td>
<td>3.50</td>
<td>GDP significantly affects the effective feature Y</td>
</tr>
<tr>
<td>E_5</td>
<td>1.75</td>
<td>Air ticket price significantly affects the effective feature Y</td>
</tr>
<tr>
<td>E_6</td>
<td>0.05</td>
<td>Oil price affects the effective feature Y insignificantly</td>
</tr>
</tbody>
</table>
After calculating the partial coefficients of elasticity, the factors that have the least and insignificant influence were excluded.

After this procedure, a new multiple regression model was developed with three explanatory variables:

\[ Y = -80,510 - 7,445X_1 + 6,159X_2 + 14,689X_3, \]  

where \( X_1 \) – average salary; \( X_2 \) – gross product per capita; \( X_3 \) – air ticket price.

Statistical analysis of the regression equation was subjected to a standard test for the significance of the equation and its coefficients, the study of absolute and relative approximation errors. Statistical analysis of the resulting regression equation showed a small average approximation error of 9.49\% and an estimate of the root mean square deviation equal to 10.3\%.

Part of the errors in the resulting modelling equation can be explained by the difference in the units of measurement of indicators, scale and dimension.

The statistical significance of the regression coefficients for factor variables is confirmed by \( t \)-statistics since they are all less than Student’s tabular \( t \)-test (Table 3). Comparison of the \( p \)-value of all variable factors for which it is less than the significance level we adopted (0.05), showed that the obtained regressor coefficients are significant, and in some cases they are significantly less, respectively, the significance of the alternative hypothesis and the rejection of the null hypothesis are higher.

The resulting multiple correlation index of 0.929 and the determination coefficient of 0.862 show a good quality of design results compared to the actual ones (Table 4). To rank factor variables by the strength of their influence on the result, a regression model was calculated on a standard scale with standardised coefficient values:

\[ T_y = -3,648X_1 + 4,457X_2 + 0,399X_3. \]  

A stronger influence on air passenger traffic will be exerted by such factors as average salary and gross product per capita. According to the coefficients of separate determination, the share of each factor in the total variation of the effective trait was identified:

\[ d_1^2 = 0.77\times(-3.648) = -2.821; \]
\[ d_2^2 = 0.84\times4.457 = 3.741; \]
\[ d_3^2 = -0.12\times0.399 = -0.046. \]

At the final stage, the parameters of the equation were checked for stability of the relation (Table 5). Checking the obtained model over a long period showed fairly close values with the actual data (Pic. 2), which indicates the quality and reliability of the model obtained.

CONCLUSION
The study has confirmed the synchronism of changes in development of the main indicator for assessing air transportation (passenger flow) and of a number of macro indicators: gross product, gross product per capita, population size, average salary and gross product per capita.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Estimated coefficient values and their statistical significance [developed by the author]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients</td>
<td>Standard error</td>
</tr>
<tr>
<td>Y-intersection</td>
<td>-80,510</td>
</tr>
<tr>
<td>Variable ( X_1 )</td>
<td>-7,4466</td>
</tr>
<tr>
<td>Variable ( X_2 )</td>
<td>6,1588</td>
</tr>
<tr>
<td>Variable ( X_3 )</td>
<td>14,689</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Regression statistics of the equation [developed by the author]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicator</td>
<td>Value</td>
</tr>
<tr>
<td>Multiple R</td>
<td>0.935201</td>
</tr>
<tr>
<td>R-square</td>
<td>0.8746</td>
</tr>
<tr>
<td>Normalised R-square</td>
<td>0.83698</td>
</tr>
<tr>
<td>Standard error</td>
<td>10.29697</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5</th>
<th>The values of parameters of the equations of single-factor dependence and the coefficients of stability of the relation [developed by the author]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor</td>
<td>Parameters of the equations of single-factor dependence</td>
</tr>
<tr>
<td>1</td>
<td>( Y_{x_1} = 128.1 (1 - 0.85231 \times d_{x_1/x_{min}}) )</td>
</tr>
<tr>
<td>2</td>
<td>( Y_{x_2} = 128.1 (1 - 1.09607 \times d_{x_2/x_{min}}) )</td>
</tr>
<tr>
<td>3</td>
<td>( Y_{x_3} = 128.1 (1 - 24.21201 \times d_{x_3/x_{min}}) )</td>
</tr>
</tbody>
</table>
and average income of the population, the national currency exchange rate to dollar, inflation and unemployment rates, changes in oil prices, etc.

The correlation dependences of the time series of passenger air transportation and economic parameters from 2008 to 2021, revealed using different methods, showed a direct and close relationship with GDP, GDP per capita, average income and average salary, population size and exchange rate. For some economic factors, an inverse dependence of different correlation strengths in dynamics over the analysed period was revealed. But since an inclusion of larger number of factors in the regression model causes a decrease in interpretability of the results, and a more complex identification of cause-and-effect relationships, it is necessary to proceed with a high-quality multi-stage selection of factors. Also, considering the intercorrelation of factors and the multicollinear effect on the result, it is advisable to exclude several factors and to replace them with an integrated factor. As a result, six, four- and three-factor regression models have been developed, of which the latter has the highest reliability with values sufficiently close to the actual data.

Further research plan includes the development of an ARIMA autoregressive integrated moving average model, autoregressive and integrated components on analytical platforms (Foresight), using the Python programming language, in statistical programs.

The development of different aspects of implementation of regression model for forecasting future air passenger traffic which is also the further task of the study is associated with the complex process of consideration not only of theoretical aspects, data of official forecasts of macro indicators, the factors influencing their stability and volatility, but of expert opinions as well. The obtained forecast results should be subject to adjustment after discussion with the experts, after renewal of statistical data, planned of forecast indicators of social and economic development.

Moreover, complex transformational processes in modern world suppose the search for new forecast methods as it is witnessed by large number of recent research publications (e.g., [20–25]), that’s why the further scientific discussion on the topic is of great importance.

REFERENCES


