Models for predicting passenger traffic in rail and air transport

A Ya Butyrkin¹, E B Kulikova¹, O N Madyar¹, E I Dmitrieva¹

¹The Russian University of Transport (RUT (MIIT)), 2, Minaevskyi per., 127055, Moscow, Russia

E-mail: a_butyrkin@mail.ru

Abstract. The task of forecasting the value of passenger traffic is traditionally relevant for any type of transport. A huge amount of research and scientific work has been devoted to its solution. There is no universal method. The authors analyzed the main models for forecasting passenger traffic in rail and air transport. The features of various approaches are revealed, the error in the calculations is estimated and the conditions of applicability in practice of various models.

1. Introduction
In modern conditions, the task of forecasting the value of passenger flows is becoming an even more significant and priority area of activity, primarily for air and rail carriers. The solution to this problem allows the passenger complex companies to form an information base for managerial decisions aimed at improving the quality of transport services and fully satisfying consumer demand.

The use of information on future passenger flows in practical activities makes it possible to accurately determine and, if necessary, adjust the transportation capacities and traffic sizes in accordance with demand, improve the quality of rolling stock fleet use, and rationally develop investment and innovation policies.

At the same time, sections of the Transport Strategy of the Russian Federation up to 2030 (approved by order of the Government of the Russian Federation of November 22, 2008 No. 1734-р) have already come into conflict with the real needs of consumers. All this shows, on the one hand, the importance of forecasting, and on the other hand, of the complexity of the process of forecasting passenger flows, since the size of passenger flows and directions depend on the population’s changing needs for movement. In addition, the real ratio of fares, the duration of the trip and the list of services provided by competing modes of transport have a significant impact on real passenger flows.

2. Methods and Results
In the practice of forecasting air and rail transportation, mathematical models are widely used, which are constructed in the form of equations showing the dependence of the predicted indicators on certain parameters (factors). Modern approaches to forecasting are based on such methods as extrapolation of trends, regression analysis, and the method of empirical models.

2.1. Trend extrapolation methods
Extrapolation is a method of scientific forecasting, which is based on the dissemination of past and present trends, patterns, relationships to the future development of the forecasting object.
Extrapolation methods are widely used in practice, since they are methodologically very simple and do not require a large statistical base for calculations. The use of extrapolation methods implies (explicitly or implicitly) two assumptions, which in most cases are satisfied for transport systems:

• The main factors and trends of the past retain their direction in the future, that is, there is a clear trend in the development of the system;
• The phenomenon under study develops evolutionarily, along a smooth trajectory, which can be expressed mathematically and described through the theory of continuous functions.

Extrapolation methods include the moving average method, the method of exponential smoothing, the least squares method.

2.2. The moving average method

The moving average method is one of the well-known methods of smoothing time series. Moving average smoothing is based on the fact that random deviations mutually cancel out in average values. Smoothing the dynamic series of historical values of indicators using a moving average consists in calculating the average level from a certain number of first-order values of the series, then the average level from the same number of values, starting from the second, then starting from the third, etc.

Take the time series consisting of f known historical values (in other words, the length of the time series is f). Then the value of the moving average in general is determined by the formula

$$Y_t = \frac{1}{n} \sum_{i=t-p}^{t} y_i$$

with:

$Y_t$ – being moving average value in period $t$. For $t = 1$, the moving average is not calculated;

$y_i$ - actual (at the appropriate steps - forecast) values of the indicators of the time series;

$n$ - odd number of levels included in the smoothing interval;

$$p = \frac{(n - 1)}{2}.$$

Regarding the value of n, we can say the following: the wider the smoothing interval (greater than n), the smoother the trend. At the same time, requirements for the duration of the initial time series increase noticeably, which creates practical difficulties. For practical use, it is recommended to accept $n = 3$. In this case, formula (1) is simplified and takes the form:

$$Y_t = \frac{1}{3} \sum_{i=t-1}^{t+1} y_i$$

If the original time series consists of f known values, then the values that relate to future periods (forecast $y_{f+1}; y_{f+2}$ etc), are defined by the formula:

$$y_{f+h} = Y_{f+h-2} + \frac{1}{n} (y_{f+h-1} - y_{f+h-2})$$

with:

$y_{f+h}$ - forecast indicator value;

$h = 1, 2, 3$ etc;

$f + h$ – forecast period;

$n$ – odd number of levels included in the smoothing interval (recommended $n =3$);

$Y_{f+h-2}$ – moving average for 2 periods before the forecast;

$y_{f+h-1}; y_{f+h-2}$ – actual (at the appropriate steps - forecast) values of the dynamic range of the determined indicator for the previous period and for two periods preceding the forecast.
We will test the model using the example of passenger turnover of JSC Federal Passenger Company, forecasting the value of this indicator for 2019. Based on the essence of the method, data on passenger turnover until 2015 are not required and do not participate in the calculations. The source data and the calculation results are shown in table 1.

**Table 1.** Forecast of passenger turnover of JSC Federal Passenger Company by the moving average method (Source: website https://fpc.ru/static/public/ru?STRUCTURE_ID=253).

| Year     | Passenger turnover, bln. passenger-km.* | Moving average $Y_t$, bln. passenger-km. |
|----------|-----------------------------------------|------------------------------------------|
| 2015     | 86.1                                    | -                                        |
| 2016     | 89.5                                    | 87.5                                     |
| 2017     | 87.0                                    | 89.4                                     |
| 2018     | 91.6                                    | -                                        |
| 2019 forecast |                                     | **90.9**                                |

According to preliminary data, the passenger turnover at JSC Federal Passenger Company increased by about 3.5% and amounted to about 94.8 billion passenger-km. Thus, the forecast error is about 4%. This method of model verification (when the predicted value is known) is called backtesting.

Such a big error is explained by a “fracture” in the dynamic range of passenger turnover in 2017, which contradicts the initial premise of the method that there is a stable trend in the development of the system.

Note that the moving averages method is suitable only for the short- and medium-term forecasting period (1-3 years) and is not applicable for long-term forecasts due to the accumulation of calculation errors.

2.3. *The method of exponential smoothing*

The method of exponential smoothing as well as the moving averages method is effective only for short- and medium-term forecasts and is acceptable only for one period ahead. The essence of the method is the calculation of exponentially weighted averages for all historical values of the time series. The basic formula of the exponential smoothing method is as follows:

$$U_{t+1} = a \cdot y_t + (1 - a) \cdot U_t$$

with:

- $U_t$ – with being exponentially weighted average;
- $y_t$ – actual values of the dynamic range indicator;
- $a$ – the smoothing parameter, which in most practical calculations is determined by the following formula:

$$a = \frac{2}{f + 1}$$

with $f$ being the length of historical time series.

As $U_1$, it is proposed to take the arithmetic mean value of the indicators of the time series of length $f$.

As with the previous method, we will test the model using the passenger turnover of JSC Federal Passenger Company as an example, forecasting the value of this indicator for 2019 (row length $f = 8$). The initial data and the calculation results are shown in table 2.
Table 2. Forecast of passenger turnover of JSC Federal Passenger Company by the method of exponential smoothing (Source: website https://fpc.ru/static/public/ru?STRUCTURE_ID=253).

| Year     | Passenger turnover. bln. passenger-km.* | Exponentially weighted average $U_t$, bln. passenger-km |
|----------|------------------------------------------|--------------------------------------------------------|
| 2011     | 111.4                                     | 97.7                                                   |
| 2012     | 114.0                                     | 100.7                                                  |
| 2013     | 107.0                                     | 103.5                                                  |
| 2014     | 94.6                                      | 104.2                                                  |
| 2015     | 86.1                                      | 102.1                                                  |
| 2016     | 89.5                                      | 98.5                                                   |
| 2017     | 87.0                                      | 96.6                                                   |
| 2018     | 91.6                                      | 94.1                                                   |
| 2019 (actual) | 94.8                                 | -                                                      |
| 2019 (forecast) | 93.6                                  | 93.6                                                   |

Graphically, the indicators of actual, forecast passenger turnover and exponentially weighted average are shown in Figure 1.

Figure 1. Indicators of the actual passenger turnover and the exponentially weighted average for the period from 2011 to 2018 and their forecast values for 2019.

The forecast obtained by the method of exponential smoothing differs by a significantly smaller deviation from the actual value in comparison with the forecast by the method of moving averages. Both forecasts give values below the actual passenger turnover received. In our opinion, the main reason for the error is that mathematical models based only on historical data are not capable of taking into account management efforts to form an effective traffic network, dynamically adjust tariffs and other marketing initiatives aimed at attracting passengers without additional adjustment.

The essence of the least squares method is to minimize the sum of the quadratic deviations between the actually known and calculated values. The calculated values are found by the selected equation. Smoothing time series by the least squares method is used to reflect the patterns of development of the studied indicator. In the analytical expression of the trend, time is considered as an independent variable, and the values of the time series act as a function of this independent variable. To correctly establish the type of curve, in other words, the type of analytical dependence of the time series on the time is one of the most difficult tasks in this method.
2.4. Regression analysis models
Regression analysis, in contrast to extrapolation type models, is based not on the presence of a predicted parameter versus time, but on the identification of the dependence of the predicted indicator on factors that have a significant impact on the predicted process. Such models are most widely applicable for long-term forecasts.

The main difficulty for this method lies in the search and testing of factors most related to the predicted indicator in terms of economy.

Studies show that passenger traffic indicators, primarily passenger traffic, are significantly dependent on macroeconomic factors: the dynamics of a country's GDP, per capita GDP, consumer price index, average cash income of the population, and a number of other factors.

The dependence of the volume of passenger traffic on a complex of factors of national economic importance was pointed out by many Russian economists during the planned economy [1]. They noted that the medium-term and long-term estimates of the parameters of passenger transportation should be based on the study of the correlation dependence of transport mobility on the level of welfare of the population. It was proposed to use national income, the level of real and monetary incomes of the population as indicators of well-being. An almost linear functional relationship (correlation coefficient of more than 0.98) between the mobility of the population and real per capita incomes has been established; population mobility and national income; mobility and growth in cash income.

Later papers confirm the accuracy of the stated provisions in the market environment. In particular, the work [2] confirmed an almost linear relationship between annual GDP per capita and passenger turnover in rail transport. A review of the work on this topic and calculations of the influence of macroeconomic factors on the indicators of passenger transportation by rail over a long time horizon are given in [3].

Studies of the dependence of the quantitative characteristics of the work of the aviation transport industry, primarily the number of passengers carried, on macroeconomic indicators were carried out in a number of works [4], [5], [6], [7], [8]. All the works show a close relationship between the passenger turnover of air transport and GDP, which persists for a long time. Figure 2 shows the graphs of changes in passenger turnover of JSC Federal Passenger Company and GDP by years.

![Figure 2. Graphs of changes in passenger turnover of JSC Federal Passenger Company and GDP by years.](image-url)
Figure 3 shows a one-factor regression dependence of changes in passenger turnover of JSC Federal Passenger Company on changes in GDP.

![Graph showing regression analysis](image)

**Figure 3.** One-factor regression dependence of changes in passenger turnover of JSC Federal Passenger Company on changes in GDP.

As can be seen from the graphs, the relationship between passenger traffic and GDP is well traced at sufficiently large time horizons and is of a true indicative nature. The forms of the graphs of changes in GDP and passenger turnover almost coincide. At the same time, in the context of one or two years, uneven dynamics of changes in demand for passenger transportation and GDP can be observed, which determines a high variance in the estimate of passenger turnover based on this dependence. This is due to some delay in the formation of demand for passenger transportation depending on changes in the macroeconomic situation in the country and the impact of other factors affecting the transport mobility of the population. This explains the relatively low correlation coefficient of 0.26. A detailed analysis of this phenomenon is presented in [3].

Regression models are the main tool for long-term forecasting of passenger traffic at home and abroad, especially in the field of aviation. The main forecasting technique in this type of transport is the technique proposed by the International Civil Aviation Organization ICAO, which can be represented as a multiplicative form for describing the functional dependence of passenger turnover on conjuncture-forming factors (formula (6), the so-called multiplicative regression model [9]:

$$Y = a \cdot X_1^b \cdot X_2^c$$

(6)

with:
- $Y$ – being passenger turnover
- $X_1, X_2$ – independent variables reflecting the influence of various factors on air travel;
- $a, b, c$ – constant coefficients characterizing the elasticity of demand in relation to the corresponding variable.

ICAO 8991 documents note that there can be more than two independent variables. However, in world forecasting practice, when compiling forecasts as a whole for a route network, as a rule, only two variables are used, moreover, one always acts as a macroeconomic indicator (usually GDP), and the other as an indicator characterizing the cost of transportation (usually a profitable rate per passenger-kilometer).
Since regression analysis models are most often used for indicative forecasting for a long-term period (up to 20 years or more), an important observation should be made. The historically correct and currently operating close correlation between passenger traffic and macroeconomic indicators is likely to weaken in the long run (on the horizon for more than 10 years). This is explained, in particular, by the approach of the transportation market to saturation.

For the first time, this problem was clearly formulated by the Swedish economist L. Tornqvist, proposing to use the analytical dependencies (Tornqvist functions) of demand $Q$ and income $C$ for three groups of goods and services: “basic necessities”, “second necessities” and “luxury goods” (Figure 4). For essential goods, the function (curve 1) has a convex form, leaves the zero point, and asymptotically approaches the upper limit of market demand $Q_1$ (1), which characterizes the level of saturation of this product. For the goods of the second group, the function (curve 2) also has a convex form and approaches with the growth of incomes to the upper limit of consumption of goods of this group $Q_2$ (2). Demand for a group of goods of the second group begins with a certain income $C_2$ (2), after which it becomes possible to purchase goods of this group. The approximation of consumption to the limit values indicates the saturation of the market. For luxury goods, the volume of consumption does not have an upper limit and manifests itself after reaching incomes at the level of $C_3$ (3).

According to the Tornquist theory, passenger transportation should be attributed to the goods of the second group. For this type of service, there is a threshold income level $C_2$ (2), upon reaching which the population begins to use the services of air and rail transport. The upper limit of consumption $Q_2$ (2) indicates the existence of a maximum for trips per year, and as a consequence - the attenuation of demand for transportation.

Note that there is an alternative point of view, especially regarding the prospects of mobility of the population of the Russian Federation. The fact is that the average Russian makes 4 to 7 times less trips per year than the average resident of Western Europe. Based on this fact, as some experts believe, the saturation of demand for trips in our country will not be reached for at least several decades.
The theory of the limit of saturation of demand has received modern development in the form of the so-called logistic curve [14]. The curve is S-shaped. At the initial time, the slope of this curve increases almost like an exponent, and then horizontally decreases to zero. For large values of t (long forecast horizon), the curve merges with the horizontal line (Figure 5). This is the so-called equilibrium state, to which a promising long-term volume of traffic is striving.

![Logistic curve](image)

**Figure 5.** Logistic curve used to develop a long-term forecast taking into account the saturation limit.

2.5. Methods of empirical models
Specific methods for forecasting passenger traffic include the use of so-called gravity models [14], [15]. Using gravity models, the volume of traffic between large cities and industrial centers is usually estimated. The structure of the calculation formulas is similar to the gravitational formula of the interaction of Newton's forces (hence the name of the models)

\[ Y_{1,2} = k \frac{S_1 \cdot S_2}{W_{1,2}^n} \]  

with:
- \( Y_{1,2} \) – being passenger turnover between centres 1 and 2;
- \( W_{1,2} \) – distance between cities;
- \( S_1, S_2 \) – population of the cities;
- \( k, n \) – empirical constants.

After adjusting the formula, which means selecting the constants \( k \) and \( n \) individually for each pair of cities of interest according to the historical values of passenger traffic, the calculation of the forecast passenger traffic actually comes down to forecasting the population of the respective cities. These values can be found in long-term plans for the development of cities for the medium and long term on the Internet in the public domain, usually on the site of the city administration.

A variation of gravity models is the Bjerkman model, according to which the growth rate of traffic between two cities \( (dY_{1,2}) \) is determined by the product:

\[ dY_{1,2} = dP \cdot dD \cdot dS^2 \]  

with \( dP, dD, dS \) being the growth rates, respectively, of the population, its purchasing power and improving the quality of services provided.
Methodological issues and practical calculations of demand elasticity regarding the quality of transport services deserve a separate detailed consideration and are beyond the scope of this article. In this regard, we can recommend the work [16].

3. Discussion
In the conditions of fierce competition for the client in the transport market, the issue of forecasting the amount of work ahead takes a crucial place, determining the development strategy of the transport company and its effectiveness. Unfortunately, companies often do not pay due attention to the entire spectrum of related issues that determine the depth of the necessary research on the transport market and the accuracy of forecast data. Work related to forecasting should be accompanied by activities to form an analytical base of influencing factors, their constant monitoring, research activities to analyze the impact of each of them on potential customer demand.

4. Conclusions
Based on the analysis, we can conclude that there is no universal method for predicting the amount of passenger traffic in vehicles that allows accurately to calculate the potential traffic for a time period of more than three years and to avoid serious errors. The work related to forecasting should be carried out constantly and systematically, and the models used should be selected based on their principles of expediency and available resources of the company.

References
[1] Pravdin N V, Negrey V Ya 1980 Prediction of passenger flows (M.: Transport) p 223
[2] Butyrkin A Ya, Mikhailov V I 2012 Economics of Railways 10 19
[3] Kagan D Z 2019 World of Transport 15(1) 140
[4] Kosobrev S I 2014 Economics and planning of air transport (M.: Aeroflot Redisdat) p 56
[5] Balashov V, Smirnov A 2013 World of Transport 4 78
[6] Egorova T P On the transport mobility of the population and its role in the strategic planning of transport enterprises www.esa-conference.ru
[7] Gyazova M 2015 Resources, Information, Supply, Competition 91
[8] Yaroshevich N Yu, Dubrovsky V Zh 2014 Izvestiya Ural State Economic University 54
[9] 1985 Air Transport Forecasting Guide (ICAO, Doc 8991 - AT / 722/2)
[10] Bassovsky L E 2009 Forecasting and planning in a market setting: Textbook (M.: Infra) p 259
[11] Vladimirova L P 2008 Forecasting and planning in market conditions (M.: Dashkov and K.) p 307
[12] Chetyrkin E M 1975 Statistical forecasting methods (M.: Statistics) p 184
[13] Slutskiy L N 2006 MBA course in business forecasting (M: Alpina Business Books) p 277
[14] Kotenko A G, Makarova E A 2017 Organization of passenger traffic (M.: Federal State Budgetary Institution Educational and Methodological Center for Education in Railway Transport) p 136
[15] Persianova V A 2019 Economics of passenger transport: a training manual (M.: KNORUS) p 390
[16] Lavrov I M 2014 World of Transport 1 86