**Keywords**: demand estimation; railway freight transportation; volume and turnover; ARIMA; forecast quality assessment

**Madiyar SULTANBEK[1], Nazdana ADILOVA[2], Aleksander SŁADKOWSKI[3], Arnur KARIBAYEV[4]**

# ESTIMATING THE DEMAND FOR RAILWAY FREIGHT TRANSPORTATION: A CASE STUDY IN KAZAKHSTAN

**Summary.** This article focuses on the critical importance of demand estimates for effective planning and decision-making in the railway freight transportation industry. Various departments within transportation companies, including marketing, production, distribution, and finance departments, heavily rely on accurate demand forecasts to make informed decisions. Forecasting demand is a crucial aspect of managing business processes, and the methods for doing this can vary across different industries. The ultimate goal remains consistent—to obtain precise predictions of future demand by analyzing historical data and current environmental factors. In the context of transportation services, accurate demand forecasting is essential for successful operational planning and management of functional areas such as transportation operations, marketing, and finance. The current case study specifically examines the National Company Kazakhstan Temir Zholy (KTZ), a transport and logistics holding engaged in rail transportation in Kazakhstan. KTZ's main sources of income are related to freight transportation. The volume of cargo transportation (in tons) and the freight turnover play a significant role in assessing demand and forecasting future revenues from freight traffic. Different techniques for demand forecasting are explored, including qualitative and quantitative methods. Qualitative methods rely on judgments and opinions, while quantitative methods utilize historical data or identify causal relationships between variables. Overall, the present study highlights the critical role of demand forecasting in the railway freight transportation industry and its impact on efficient planning and decision-making processes.

## 1. INTRODUCTION

The accuracy of demand estimates holds utmost importance in any company, as it significantly influences effective planning and decision-making processes. Multiple departments, including marketing, production, distribution, and finance departments, rely on short-to-long-term forecasts to make well-informed choices [1]. The precision and quality of these forecasts are of paramount significance to ensuring that business operations are adequately planned and coordinated.

Demand forecasting plays a crucial role in managing business processes across various industries. Though forecasting methods may vary in complexity from one business to another, their ultimate goal remains consistent: to predict future demand for products or services with reasonable precision.

[1] Al-Farabi KazNU; Al-Farabi av. 71, 050040 Almaty, Kazakhstan; e-mail: cultanbek_madiyar@live.kaznu.kz; orcid.org/ 0000-0002-4408-3552
[2] Al-Farabi KazNU; Al-Farabi av. 71, 050040 Almaty, Kazakhstan; e-mail: adilova.nazdana@kaznu.kz; orcid.org/0000-0002-8184-6993
[3] Silesian University of Technology; Krasińskiego 8, 40-019 Katowice, Poland; e-mail: Aleksander.Sladkowski@polsl.pl; orcid.org/0000-0002-1041-4309
[4] JSV "National company "Kazakhstan Temir Zholy"; Kunayev 6, 010000 Astana, Kazakhstan; e-mail: karibayev_ar@railways.kz; orcid.org/0009-0006-7296-3502
* Corresponding author. E-mail: sultanbek.madiyar@gmail.com

Historical data and current environmental factors, such as political, social, and economic conditions, are leveraged to achieve this objective [2].

Predicting future demand for transportation services is a critical success factor for companies operating in the transportation industry. Such predictions provide fundamental information for planning and managing functional areas like transportation operation planning, marketing, and finance [3].

This paper focuses on a national joint stock company called Kazakhstan Temir Zholy (KTZ), a transport and logistics holding engaged in rail transportation. KTZ's corporate portfolio includes several subsidiaries and structural companies operating in various segments, making it the country's largest employer, with over 115,000 employees. The company's main sources of income are freight and passenger transportation. In Kazakhstan, the volume of applications for freight railway transport nearly matches the volume of traffic, making the demand for railway carrier services dependent on the volume of cargo transportation, in tons, and the distance transported, in kilometers, known as freight turnover. The tariff freight turnover, calculated by multiplying the tariff distance between loading and unloading points by the volume of traffic, forms the basis for calculating future revenues from freight traffic.

The literature presents different forecasting techniques, categorized into qualitative and quantitative approaches. Qualitative methods, like executive opinions, the Delphi technique, sales force polling, and customer services, generate forecasts based on judgments or opinions. On the other hand, quantitative methods include historical data forecasts (e.g., the naive method, trend analysis, time series analysis, and Holt's and Winter's models) and associative forecasts, which identify causal relationships between variables through regression methods. Additionally, this paper introduces mixed models that integrate both qualitative and quantitative approaches.

In summary, this research provides valuable insights into the significance of demand forecasting in business decision-making and emphasizes its critical role in the success of transportation companies. It offers a comprehensive guide for researchers and practitioners, outlining various forecasting methods and their application in the context of the transportation industry. The structured format of this paper facilitates a better understanding and implementation of the discussed concepts.

The primary purpose of this research is to address the significance of demand estimates in effective planning and decision-making within companies. Accurate demand forecasts are crucial for various departments, such as marketing, production, distribution, and finance, to make informed decisions. This paper aims to emphasize the critical role of demand forecasting in business operations and the need for reasonably precise predictions of the future demand for products or services.

Additionally, this research focuses on the transportation industry, particularly on the success of transportation companies. Predicting future demand for transportation services is highlighted as a crucial factor in planning and managing functional areas such as transportation operation planning, marketing, and finance.

This paper makes contributions by providing insights into different forecasting methods used in businesses. It covers qualitative methods, which rely on judgments and opinions to generate forecasts, and quantitative methods, including historical data forecasts and associative forecasts that identify causal relationships between variables. This paper also discusses mixed or combined models that integrate both qualitative and quantitative approaches to enhance forecasting accuracy.

Furthermore, this paper presents a structured format, including an introduction, literature review, case study presentation, detailed account of the materials and methods (including the models used and the experimental process followed), findings, discussion, and conclusion. This paper serves as a comprehensive guide for researchers and practitioners interested in demand forecasting in the transportation industry and other business sectors.

## 2. LITERATURE REVIEW

Several research centers have conducted studies on constructing models to describe the demand for rail services, including [3-7]. However, most of these studies are limited to the analysis of cities. Relatively few models have been developed to assess the functioning of large national rail networks, such as those in Sweden and India, which were examined in previous studies [8-10] that reviewed the application of data envelopment analysis (DEA) in the transport sector, investigating the inputs and outputs used in 69 DEA models reported in the literature. Article [11] proposed various models for forecasting demand in the regular passenger transport industry.

Demand analyses and forecasts are crucial for developing transport policies, but demand data are not always available due to a lack of appropriate mathematical models for generating forecasts. Therefore, it is essential to analyze the railway systems of various countries to select appropriate methods for forecasting transport performance. The objective of this study is to identify the parameters of a mathematical model of rail cargo transport performance based on historical data to reliably forecast future demand. In this paper, we investigate Kazakhstan's national railway system, propose several models dedicated to this type of empirical data, establish selection criteria, identify the best model, and assess its accuracy and effectiveness.

## 3. MATERIALS AND METHODS

KTZ decided to conduct pilot research on the freight transportation volume and freight turnover demand estimations using specialized software that processes and analyzes data sets and compares the quality of the estimation results with the Marketing and Tariff Policy Department (MTPD) experts' estimations (marketing experts' opinions and linear extrapolation).

The experiment is divided into several stages:

1. The monthly historical data on the rail freight volume and turnover from 2012 to 2016 for each nomenclature of unified tariff and statistical nomenclature of goods and 13 aggregated nomenclatures of goods and for all types of communication (export, import, transit, and domestic transportation) were loaded from the KTZ systems into a specialized program for analysis, data science, and forecasting.
2. Macroeconomic indicators (predictors) that potentially correlate with the historical volumes of transportation or freight turnover were found and loaded into a specialized program (260 indicators in the appropriate format and periodicity of data were collected). It is vital that all predictors are uploaded in appropriate granularity to assess the correlation level of the historical freight volume and freight turnover over a five-year period on a monthly basis with predictors.
3. A model was created and tested on test data for 2012-2016. The model automatically generated a monthly forecast for 2017 and then compared these forecasts with the estimations made by MTPD experts in 2016 for 2017. A separate comparison was made for each aggregated nomenclature of cargo and each type of communication (internal, export, import, and transit).
4. The assessment of the quality of the forecast was carried out according to mean absolute percentage error (MAPE) or mean absolute error (MAE) because these are the most common methods of assessment used in forecasting and checking the quality of demand estimation models. Formulas for calculating MAPE and MAE are presented below, where Z(t) is the actual value of the time series and X(t) is the forecast value. MAE is applied if the actual value of the indicator is zero. We compared the forecast with the fact and derived the MAPE/MAE indicator both for the manual forecast made by MTPD experts and for the forecast made by specialized software.

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\frac{|Z(t)-X(t)|}{Z(t)} * 100\% \quad (1)$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|Z(t)-X(t)| \quad (2)$$

IBM Statistical Package for the Social Sciences (SPSS) Modeler, a visual data science and machine learning solution, was chosen as specialized software to be used for data analysis and demand estimation because this product was ranked first in the category of data science platforms in the Gartner ranking in 2017.

Since many statistical and mathematical forecasting methods are available in SPSS, an additional task was to choose the best method for demand estimation. As a result, forecasting was done using three approaches: the autoregressive integrated moving average (ARIMA) model, which is amongst the most generally used to forecast transport demand [13, 3]; the neural net (Nnet) model; and autofitting, comprising neural network methods, C&R tree, Chi-square automatic interaction detection (CHAID) model, linear regression and support vector mechanism). The best model was chosen subsequently.

### 3.1. Models used in the design study

**Neural networks**. Neural networks are models that simplify the functioning of the nervous system in living beings. They consist of basic units called neurons, which are typically organized into layers. These networks employ a simplified model of how the human brain processes information, using interconnected processing elements that represent neurons in an abstract manner.

A neural network is trained by examining records; for each record, the network generates a prediction. If the prediction is incorrect, the weights are adjusted. This process is repeated multiple times, and the accuracy of the predictions gradually improves until a stopping criterion is met.

When a neural network is first initialized, all weights are random, and the responses to input signals are usually meaningless. However, through repeated presentations of examples with known output values, the network learns and adjusts its weights based on the comparison between its response and the known response. As the network learns, its responses become more accurate in reproducing the known output. Once the network is trained, it can be used to predict future observations whose outcomes are unknown.

**ARIMA**. The ARIMA technique enables the development of autoregressive integrated moving average models that can be used to refine time series simulations. ARIMA models offer more advanced techniques for modeling trend and seasonal factors than exponential smoothing models, and they have the added benefit of allowing predictor variables to be incorporated. By specifying the autoregressive, differential, and moving average order, as well as their seasonal counterparts, the ARIMA approach can be used to fine-tune a model. However, determining the rational values for these components through trial and error can be a time-consuming process.

### 3.2. Description of Experimental Progress

Historical rail transportation volume and freight turnover data for 13 aggregated cargo nomenclatures over a five-year period (2012 to 2016), presented on a monthly scale, were loaded into SPSS. Combinations were selected for each aggregated nomenclature of goods: unified tariff and statistical nomenclature of goods cargo type code, country of origin, and country of destination. Then, macroeconomic indicators—such as the volume of production of coal, oil, ore, electricity, etc.; the volumes of exported/imported goods; prices for various types of raw materials; exchange rates against the local currency—from countries that have strong trade relations with Kazakhstan were found and loaded into the system. All macro indicators were loaded into SPSS in an appropriate granularity of monthly format for the same period as the historical data on rail freight transportation. Special tools in SPSS were used to analyze the correlation between macro indicators (predictors) and historical data and estimate the influence of predictors on historical data. Then, the model was trained on the training sample and tested on the test data from 2012-2016, and a forecast was formed by month for 2017 for all 13 cargo nomenclatures by rail transportation volume and freight turnover using three different methods (ARIMA, Nnet, and autofitting), the best of which was the ARIMA forecast. The best forecast generated by SPSS (ARIMA) was compared with the 2017 actual freight transportation performance and the MTPD estimates generated using methods described in the Introduction section of this paper.

## 4. CASE STUDY

The current planning process in KTZ lacks modern tools. Previously, researchers [12] employed various techniques, such as the seasonal naive model, exponential smoothing model, exponential smoothing state-space model with Box-Cox transformation, autoregressive moving average errors, trigonometric trends, and the seasonal components model, to forecast demand in Polish Railways. The researchers concluded that the ARIMA method exhibits the least error. Consequently, this study introduces the initial implementation of the ARIMA model to produce a forecast for rail freight transportation in the case of KTZ. The current process of rail freight demand estimation in KTZ completely depends on a person—an expert in the field of freight transportation marketing who makes estimations using MS Excel. KTZ's MTPD is responsible for freight demand estimations in KTZ. The MTPD uses the following methods for freight demand estimation:

1) Expert estimates based on an assessment of the current moment and development prospects. The MTPD experts analyze several years' worth of historical data on transportation, studying the factors that have influenced freight transportation in the past. They also use forecasts of major shippers (if available) and opinions of leading experts in different industries related to transported cargo types.
2) Extrapolation of the distribution of past trends for the future period. Extrapolation is used for prospective calculations of transportation of consignors who are not included in the surveyed group.

Thus, the main scientific method used by MTPD staff to make demand estimations is extrapolation. The other techniques are expert and depend on the judgment and experience of the MTPD expert. Extrapolation [3], as it is known in mathematics and statistics, is a special type of approximation in which the function is approximated outside a given interval rather than between given values. In other words, extrapolation is an approximate determination of the values of a function f(x) in points x lying outside the interval $[x_0, x_n]$ by its values in points $x_0<x_1<...<x_n$. In a more general sense, extrapolation is the transfer of conclusions made about some part of an object or phenomenon to the whole set of related objects or phenomena, as well as to some other part of them. The method of linear extrapolation is most often used.

However, this method, when used by MTPD experts, has significant drawbacks. Namely, extrapolation does not consider changes in the external environment or the impact of external factors on demand estimations. For example, changes in the exchange rate of the national currency to foreign currencies can have a strong impact on the volume and geography of transportation, but the extrapolation method does not consider this.

Based on the above, the following conclusions can be drawn:

1) MTPD experts are the key links in the process at all stages of forming freight demand estimations. In the scientific literature, this method of forecasting is called an expert method. "Expert" in Latin means "experienced." The demand estimations made by an expert or team of experts are based on their professional, scientific, and practical experience and opinions. Expert techniques are typically utilized under the following circumstances: when the subject of investigation is exceedingly uncomplicated or, conversely, in scenarios involving the immense intricacy of the subject of assessment, its innovativeness, uncertainty regarding the establishment of vital characteristics, inadequate data comprehensiveness, or the incapability of fully mathematically formalizing the process of addressing the given problem. The main principle underlying the methods of individual expert evaluations is the maximum possibility of using the individual abilities of the expert. Since MTPD experts have access to a vast amount of available digital historical data on transportation, the expert method of forecasting, as follows from the previous narrative, is not rational.
2) MTPD experts spend most of their time on operations like downloading data from KTZ systems; uploading data to personal computers; generating summary tables; preparing data; generating reports, graphs, and tables; preparing paper questionnaires for shippers; and manually processing survey results. That is, most of the MTPD expert's working time is spent on routine operations.
3) Large data sets from various KTZ systems are processed in MS Excel, whose capabilities for processing large data sets are severely limited. For example, MS Excel is unable to create tables with more than 1,048,576 rows and 16,384 columns. The analysis of thousands of cargo type codes from the unified tariff and statistical nomenclature of goods by hundreds of stations of departure and destination and by hundreds of shippers may create the need for tables with tens of millions of rows and columns. In addition, MS Excel has a limited number of available libraries for forecasting, so MTPD experts use only the linear extrapolation method since it is available in MS Excel. The lack of the technical capability of the MTPD experts to perform a more detailed analysis of the input

historical data and the inability to use many methods of mathematical or statistical analysis other than linear extrapolation leads to simplifications, averaging, and, consequently, a deterioration in the quality of demand estimations.

There is a logical conclusion—it is necessary to partly automate the process of demand estimations based on modern software and, thus, accelerate the process of obtaining and loading data, analyzing large sets of data using a variety of methods, not to replace marketing experts with the program but to increase the productivity of marketing experts. It is necessary for marketing experts to spend more time interpreting analyses rather than compiling statistics. This requires the use of special software products to prepare and analyze data sets due to the high performance of database management systems and built-in libraries of algorithms, in which the computing and processing of data sets occur in a matter of seconds (much faster than when they are computed and processed manually).

## 5. RESULTS

The main results of this research work and experiment are presented in graphical form below. Fig. 1 includes three lines that show a comparison of freight traffic for all nomenclatures of goods and all types of cargo: the forecast of experts of the MTPD KTZ (green line), the forecast using ARIMA (blue line), and the actual volume of traffic in 2017 (red line). Fig. 1 clearly shows that the blue line of the ARIMA monthly forecast for the volume of freight transportation and the red line of the actual volume of freight transportation practically coincide starting from the third month of 2017. At the same time, the green line differs significantly from the fact. The value of the MAPE indicator for the forecast of experts of the MTPD KTZ was 9.2%, while for ARIMA, it was 2.0%, which indicates a significant excess of the quality of the ARIMA forecast over the expert forecast.

As can be seen in Fig. 2, the forecast for total coal transportation volume, which is the main nomenclature of cargo type transported by KTZ (the share of coal in freight transportation volume exceeds 40%), was much better predicted by the ARIMA model.

is the strong correlation level of transportation volumes with the macro indicators (predictors) found during this research, as well as the high seasonality of coal transportation. The “predictor screening” feature in IBM SPSS Modeler allows characteristics to be selected, helping to identify the fields most important in predicting certain outputs. From a set of hundreds or even thousands of predictors, the “feature selection” node ranks and selects the predictors that are most important, and it helps provide a faster and more efficient model that uses fewer predictor types, runs faster, and is easier to interpret.

One of the reasons the ARIMA model’s estimation for freight transportation volumes is so accurate

Fig. 3 shows that the ARIMA predictive model gives much better results compared to even the neural network model. However, the neural network model (Nnet in Fig. 3) requires more fine-tuning and has the potential for improvement.

Using the selected ARIMA method, in 2021, a forecast was formed for the volume of traffic in thousands of tons for 2022 based on five years’ worth of historical data. The figures below show the results of comparing the forecast with the fact for 2022 for all types of traffic communication and for all total nomenclatures of goods.

As can be seen in Fig. 4, which shows a comparison of the forecast for 2022 in the export direction with the actual traffic volumes, the first three months of 2022 were forecasted quite well. The forecast and fact lines practically coincide, although the forecast for 2022 was formed in June-July of the previous year in accordance with KTZ corporate procedures. However, after April 2022, export traffic fell sharply and deviated from the forecast values, which affected the value of the MAPE indicator. According to railway industry experts, a sharp decline in exports could be due to political reasons, including the imposition of sanctions against Russian companies, which, as a result, lost many sales markets and, accordingly, began to purchase fewer raw materials and components. However, in the last quarter of 2022, the forecast and actual traffic showed a similar trend line. Overall, the MAPE score for the export forecast was 12.4%, below the target threshold of 10%. Performing an extra investigation into the underlying reasons and potentially enhancing the model is essential.

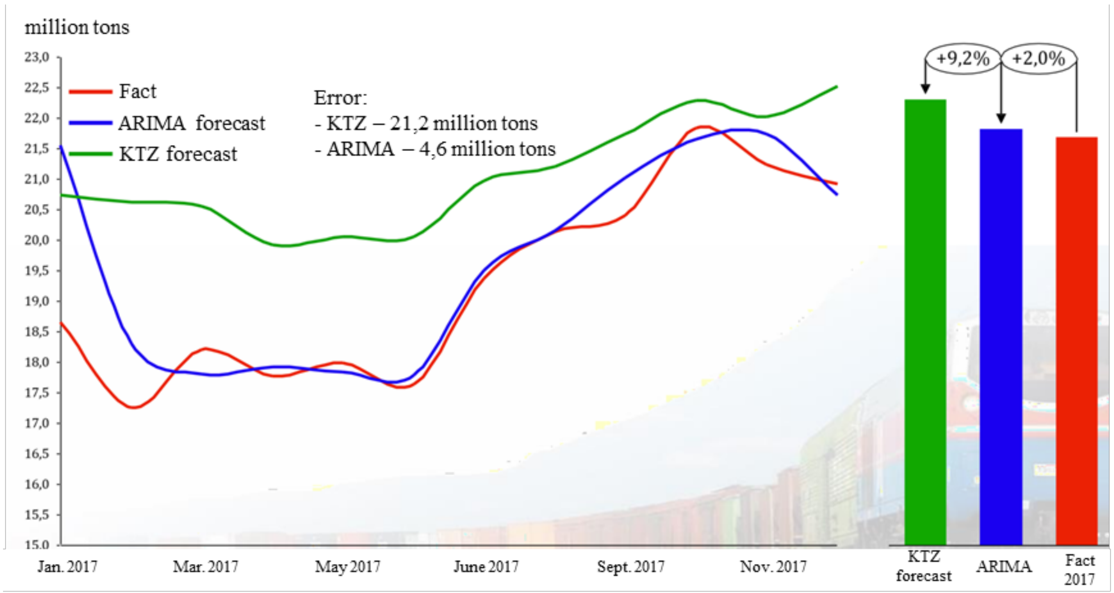


Fig. 1. Cumulative estimation of rail freight volume for all cargo types' nomenclature and communication types compared to the 2017 fact and MTPD KTZ experts' estimations

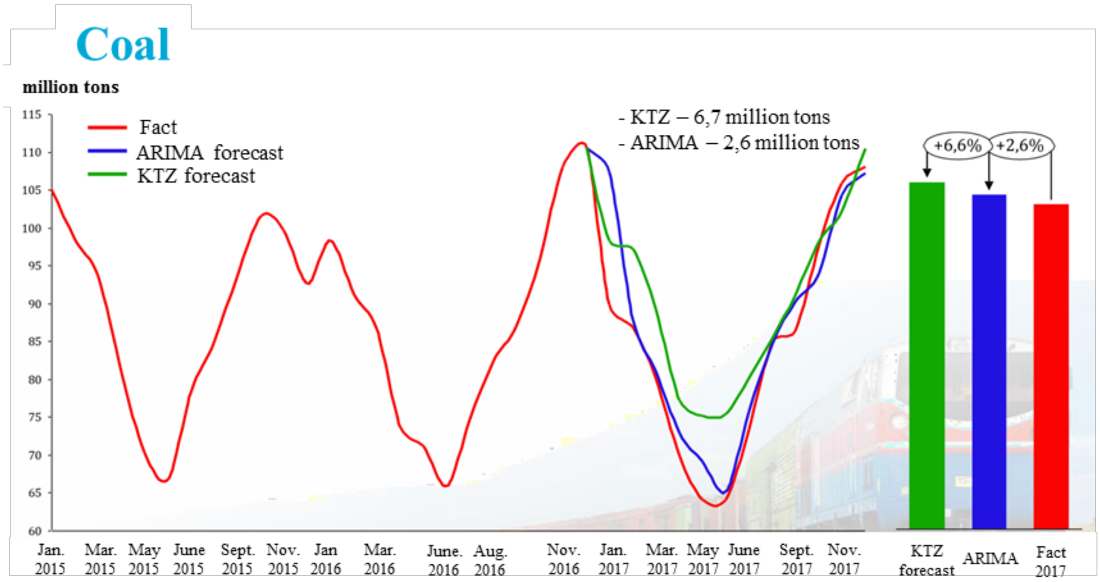


Fig. 2. Comparison of the actual freight transportation volume in 2017 with MTPD experts' estimations and the ARIMA model's forecast of the total coal transportation volume in all types of communication (internal, export, import, and transit)

Fig. 5 compares the forecast of imports for all types of goods in 2022 with the actual volumes of imported freight traffic. Significant fluctuations in the volume of imports are seen in the first four months of 2022, followed by more stable monthly traffic volumes in the subsequent months of 2022. In the period from Month 5 to Month 9, the forecast line and the actual value line are very close to each other, which affected the final value of the MAPE indicator of 5.3%, which fits into the target range of forecast quality. Abnormal values of imports at the beginning of 2022 could also be dictated by political and, as a result, foreign economic factors related to Kazakhstan's proximity to Russia. Given the sharp fluctuations at the beginning of the year, the import-forecasting model requires adjustment.

Fig. 6 shows the forecast and fact lines for the transit direction of cargo transportation. Again, a dramatic drop and a complete divergence from the forecast are seen in transit in the first four months of 2022. Transit traffic is usually the most difficult to predict, as it involves the movement of goods that originate outside domestic territory and are sent to foreign customers, often located far beyond the national borders of Kazakhstan. Considering that the main transit through Kazakhstan traditionally passed further through the territory of Russia, with which Kazakhstan has the longest land border in the world, the influence of political and economic factors on the volume of transit traffic at the beginning of 2022 is obvious and difficult to predict. However, due to the sharp growth and the further relative stabilization of transit traffic volumes in the second half of 2022, the MAPE indicator for the whole year amounted to 8.1%. However, given what happened, the transit forecasting model also requires further study and fine-tuning.

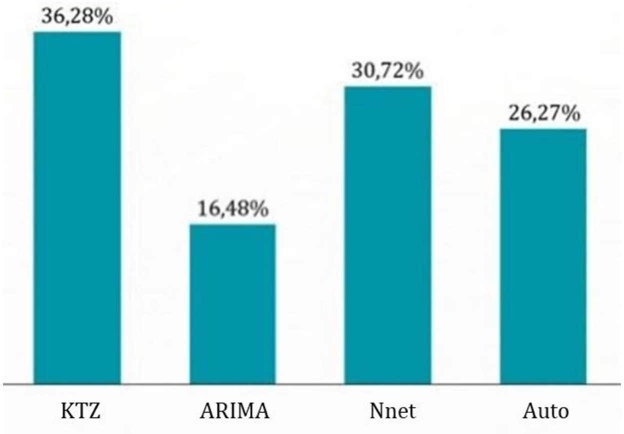


Fig. 3. The indicator of the MAPE for all nomenclatures of goods, all types of forecasts (volume and cargo turnover), and all types of communication using various forecasting methods available in IBM SPSS

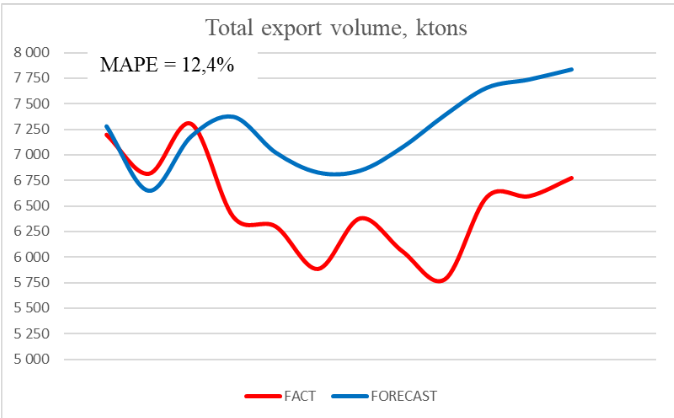


Fig. 4. Total 2022 export volume for all types of cargo nomenclature

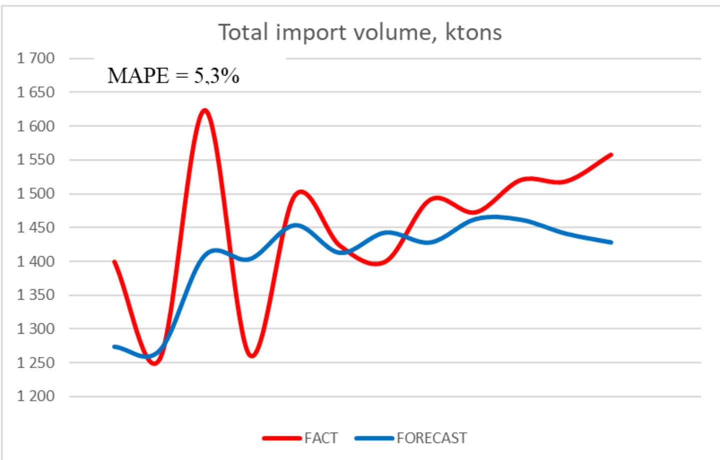


Fig. 5. Total 2022 import volume for all types of cargo nomenclature

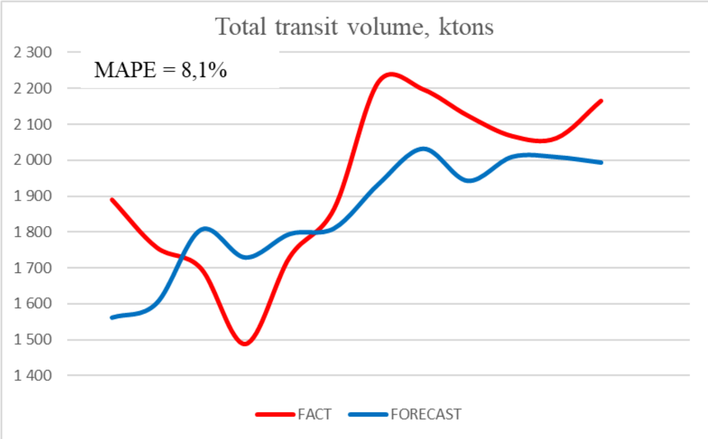


Fig. 6. Total 2022 transit volume for all types of cargo nomenclature

Fig. 7 compares the forecast of internal cargo traffic within the country's borders with the actual volumes of goods transported. The figure shows that the first eight months of the forecast and fact lines are very close to each other, and the discrepancies are visible only in the final four months of 2022. Obviously, the model predicted future traffic volumes very well, although there is a tendency to somewhat overestimate the volume of traffic. The value of the MAPE indicator was 6.1%.

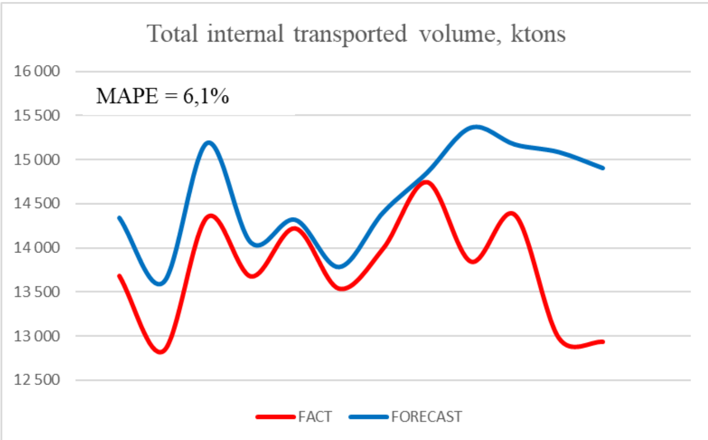


Fig. 7. Total 2022 internal volume for all types of cargo nomenclature

Fig. 8 shows the total volume of freight transportation in all directions and for all types of cargo. Considering that the volume of domestic traffic significantly exceeds the volume of freight transportation in the export, import, and transit directions, the forecast and actual lines are very similar to those in Fig. 7. Therefore, the MAPE indicator is 6.1%, as it was in the domestic freight transportation forecast.

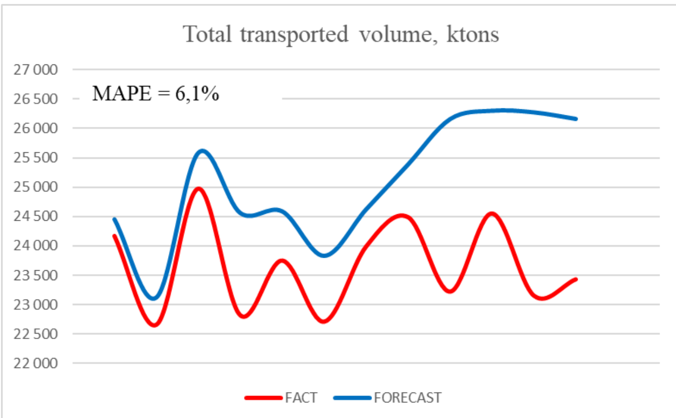


Fig. 8. Total 2022 transported volume for all types of communication and all types of cargo nomenclature

In general, it can be seen that in domestic transportation and in exports, the model as a whole somewhat overestimates the actual values. Meanwhile, it provides underestimates for imports and transit. These features of the models also need to be studied and refined.

## 6. DISCUSSION

The research presented in this paper focuses on demand forecasting in the transportation industry, with a particular emphasis on the case study of joint stock company KTZ, a transport and logistics holding engaged in rail transportation. This study explores the significance of demand estimates in effective planning and decision-making within transportation companies, especially for planning transportation operations, marketing, and finance.

The present research uses historical data and macroeconomic indicators to develop forecasting models for rail freight transportation. It employs various forecasting techniques, including qualitative methods, such as expert opinions and extrapolation, and quantitative methods, such as ARIMA and neural network models. The results of the experiment demonstrate that the ARIMA model outperforms other forecasting methods, including the expert-based estimation method currently used by MTPD experts in KTZ.

One of the key findings of the present study is that the ARIMA model provides more accurate demand estimations compared to the manual expert-based approach. The ARIMA model benefits from its ability to consider changes in the external environment and the impact of external factors, leading to improved forecasting accuracy. The correlation between transportation volumes and macro indicators enhances the ARIMA model's predictive power.

This paper identifies the limitations of the current manual approach used by MTPD experts, including its heavy reliance on routine operations and the limited data processing capabilities of tools like MS Excel. The proposed solution is to automate part of the demand estimation process using modern software to increase the productivity of marketing experts. This automated approach would enable faster data analysis, support the use of various forecasting methods, and reduce the time spent on routine tasks, ultimately improving the quality of demand estimations.

The present research also acknowledges some areas that require further refinement and study. For instance, the ARIMA model tends to slightly overestimate the demand for domestic transportation and exports while underestimating the demand for imports and transit. These tendencies may be influenced by specific factors, and additional research could help fine-tune the models to address these variations.

Overall, this study provides valuable insights into the demand forecasting process in the transportation industry, particularly in the context of rail freight transportation. By demonstrating the effectiveness of the ARIMA model, this research highlights the potential benefits of adopting modern software tools to enhance demand estimation accuracy. The findings could serve as a guide for transportation companies, including KTZ, in improving their demand forecasting practices and, in turn, enhancing their planning and decision-making processes. The structured format of this paper facilitates a better understanding and implementation of the discussed concepts by researchers and practitioners in the field.

## 7. CONCLUSIONS

The purpose of the current research was to compare the quality of two estimations of the rail freight transportation volume with the actual performance of KTZ in 2017. For this purpose, a five-year period (from 2012 to 2016 inclusive) was selected, and actual historical data on the volume of transportation and freight turnover were loaded into a specialized software product for data analysis and forecasting—namely, IBM SPSS Modeler. Various macroeconomic indicators for the Republic of Kazakhstan and other countries that are trading partners with Kazakhstan were also loaded into the system. SPSS generated a forecast of transportation volume and freight turnover for all nomenclatures of goods and types of communication for 2017, and the IBM forecast values were compared with the official actual data of KTZ performance in 2017 and the estimations made by MTPD experts in 2016 for 2017.

Forecasts were compared with actual data for all cargo types and all types of communication in terms of tons of freight transportation volume and ton-kilometers of freight turnover. The comparison revealed that demand estimations using the ARIMA model showed quite comparable or even better results than three months' worth of manual process results obtained from MTPD experts. More accurate results were obtained for the nomenclatures of goods and types of communication for which there were holistic data sets without gaps loaded into SPSS and for which it was possible to find macro indicators correlating with the volume of transportation and freight turnover in the past. Using the aforementioned ARIMA model, this article presents the results of forecasting demand for freight transportation volume in 2022 and compares the forecast values with the fact based on MAPE. The forecast for 2022 for all types of rail traffic communication showed good forecast quality results, with the exception of the estimates for exports. While export results may have been affected by political factors in 2022, a further investigation into the possible reasons for the deviation of the forecasts from an acceptable quality range will be carried out.

Brief conclusions from the present research are as follows:

1. Modern mathematical and statistical models and software should be introduced into the practice of the largest enterprises in Kazakhstan, including the transport industry.
2. The experiment shows the potential for using the above methods in practice for data set analysis and forecasting while saving up to 8,000 man-hours annually among MTPD experts. This reduction in man-hours would allow MTPD experts to focus not only on the manual processing of data but also on their analysis and interpretation.
3. The experiment was conducted on real historical data, and a comparative analysis of the results was performed for all cargo types and types of communication used in KTZ practice together with the MTPD experts and presented to the KTZ's top management. The results of the pilot experiment on the use of specialized forecasting software served as the basis for launching the "Integrated Planning System" project.

The methodology discussed in this article, the results of the scientific experiment, and the recommendations based on the results of the experiment have been put into practice in KTZ.

**References**


1. Punia, S. & Shankar, S. Predictive analytics for demand forecasting: A deep learning-based decision support system. *Knowledge-Based Systems.* 2022. Vol. 258. No. 109956. DOI: https://doi.org/10.1016/j.knosys.2022.109956.
2. Merkuryeva, G. & Valberga, A. & Smirnov, A. Demand forecasting in pharmaceutical supply chains: A case study. *Procedia Computer Science.* 2019. Vol. 149. P. 3-10. DOI: https://doi.org/10.1016/j.procs.2019.01.100.
3. Milenković, M. & Švadlenka, L. & Melichar, V. & Bojović, N. & Avramović, Z. SARIMA modelling approach for railway passenger flow forecasting. *Transport.* 2015. Vol. 33. No. 2. P. 1-8. DOI*:* https://doi:10.3846/16484142.2016.1139623.
4. Roos, J. & Gavin, G. & Bonnevay, S. A dynamic Bayesian network approach to forecast short-term urban rail passenger flows with incomplete data. *Transportation Research Procedia.* 2017. No. 26. P. 53–61. DOI: https://doi.org/10.1016/j.trpro.2017.07.008.
5. Zhang, J. & Chen, F. & Shen, Q. Cluster-based LSTM network for short-term passenger flow forecasting in urban rail transit. *IEEE Access.* 2019. Vol. 7. P. 147653-147671. DOI: https://doi.org/10.1109/ACCESS.2019.2941987.
6. Tang, Q. & Cheng, P. & Li, N. 基于GSO-BPNN方法的城市轨道交通客流短时预测[J]. 2017. 通信领域的技术与经济(1): 1–4. [In Chinese: Tang, Q., Cheng, P., Li, N. Short time forecasting of

passenger flow in urban railway using GSO-BPNN method. *Technology & Economy in Areas of Communications.* 2017]

7. Намиот, Д.Е. & Покусаев, О.Н. & Лазуткина, В.С. О моделях пассажирского потока для городских железных дорог. *International Journal of Open Information Technologies*. 2018. Vol. 6. No. 3. P. 9-14. [In Russian: Namiot, D.E. & Pokusaev, O.N. & Lazutkina, V.S. On passenger flow models for urban railways. *International Journal of Open Information Technologies*].
8. Andersson, M. & Brundell-Freij, K. & Eliasson, J. Validation of aggregate reference forecasts for passenger transport. *Transportation Research Part A: Policy and Practice.* 2017. Vol. 96. P. 101-118. DOI: https://doi.org/10.1016/j.tra.2016.12.008.
9. Prakaulya, V. & Sharma, R. & Singh, U. & Itare, R. Railway passenger forecasting using time series decomposition model. In: *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA).* 20-22 April 2017. Coimbatore, India. P. 554-558.
10. Markovits-Somogyi, R. Measuring efficiency in transport: the state of the art of applying data envelopment analysis. *Transport.* 2011. Vol. 26. No. 1. P. 11-19. DOI: https://doi.org/10.3846/16484142.2011.555500.
11. Banerjee, N. & Morton, A. & Akartunalı, K. Passenger demand forecasting in scheduled transportation. *European Journal of Operational Research.* 2020. Vol. 286. No. 3. P. 797-810. DOI: https://doi.org/10.1016/j.ejor.2019.10.032.
12. Borucka, A. & Mazurkiewicz, D. & Lagowska, E. Mathematical modelling as an element of planning rail transport strategies. *Transport.* 2021. Vol. 36. No. 4. P. 354-363. DOI: https://doi.org/10.3846/transport.2021.16043.
13. Bermúdez, J.D. & Segura, J.V. & Vercher, E. Holt–Winters forecasting: an alternative formulation applied to UK air passenger data. *Journal of Applied Statistics.* 2007. Vol. 34. No. 9. P. 1075-1090.

Received 06.02.2022; accepted in revised form 31.08.2023