

СЕРІЯ «Економіка»

UDC 004.432

[https://doi.org/10.52058/2786-6025-2024-6\(34\)-237-254](https://doi.org/10.52058/2786-6025-2024-6(34)-237-254)

Chupilko Tetjana Anatolyivna PhD in engineering, associate professor, associate professor, University of Customs and Finance, St. Volodymyr Vernadskyi, 2/4, Dnipro, 49000, tel.: (096) 779-84-48, <https://orcid.org/0000-0002-3469-3154>

Ulianovska Yulija Viktorivna PhD in engineering, associate professor, associate professor, University of Customs and Finance, St. Volodymyr Vernadskyi, 2/4, Dnipro, 49000, <https://orcid.org/0000-0001-5945-5251>

Mormul Mykola Fedorovich PhD in engineering, associate professor, associate professor, University of Customs and Finance, St. Volodymyr Vernadskyi, 2/4, Dnipro, 49000, tel.: (098) 487-76-86, <https://orcid.org/0000-0002-8036-3236>

Shchytyov Dmytro Mykolayovych PhD economic sciences, doctoral student, University of Customs and Finance, St. Volodymyr Vernadskyi, 2/4, Dnipro, 49000, tel.: (099) 288-49-41, <https://orcid.org/0000-0003-4306-8016>

Shchytyov Oleksandr Mykolayovych Candidate of Physical and Mathematical Sciences, Associate Professor, Teacher, EC-Lyceum No. 100, Sq. Uspenska, 1, Dnipro, 49000, tel.: (066) 336-29-81, <https://orcid.org/0000-0002-1435-2918>

Chupilko Oleksandr Serhiyovych postgraduate, University of Customs and Finance, St. Volodymyr Vernadskyi, 2/4, Dnipro, 49000, tel.: (096) 779-80-98, <https://orcid.org/0009-0004-4950-6771>

MODELING OF SECURITY INDICATORS OF THE NATIONAL ECONOMY AND DATA PROCESSING USING PYTHON

Abstract. The article considers various aspects of effective data processing, outlines the stages involved and features of each stage. Problems arising during data processing and effective means of their modeling and forecasting are analyzed.

The use of several packages NumPy, Pandas, Matplotlib, SciPy, Statsmodels, Scikit-learn is considered. An example of how Python can be used for customs-

related tasks is given, taking into account the country's economic security indicators. The authors built a series of regression models to analyze the contribution of customs revenues from import and export duties to the state budget of Ukraine. A calculation program has been developed using the above packages. At the same time, the Python programming language was used, using NumPy, Statsmodels, Matplotlib and Xlrd libraries. Preliminary data preparation was carried out in Excel. The obtained results are analyzed. The article demonstrates the possibility of software data processing. Complete statistics could not be obtained from the official website to enable analysis of large sets of similar data. However, problems encountered when modeling based on the aggregated data provided will be reproduced similarly. In addition, the lists contain a warning that more data should be available for an accurate calculation. If the volume of data is insufficient, it is necessary to use adjusted estimates, which is also known from the theory of mathematical statistics. The possibility of using the results of modeling financial and economic indicators for making management decisions is indicated. It is possible to build various single-factor and multi-factor: linear and non-linear models, and thus get a complete picture of what processes are taking place in the industry, in particular, at customs, and identify positive and negative phenomena. Based on the results of modeling, it is possible to obtain a scientifically based forecast and make appropriate management decisions.

Keywords: Python, modeling, forecasting, data processing, regression model.

Чупілко Тетяна Анатоліївна кандидат технічних наук, доцент, доцент, Університет митної справи та фінансів, вул. Володимира Вернадського, 2/4, м. Дніпро, 49000, тел.: (096) 779-84-48, <https://orcid.org/0000-0002-3469-3154>

Ульяновська Юлія Вікторівна кандидат технічних наук, доцент, доцент (завідувач кафедри), Університет митної справи та фінансів, 2/4, вул. Володимира Вернадського, м. Дніпро, Україна, 49000, <https://orcid.org/0000-0001-5945-5251>

Мормуль Микола Федорович кандидат технічних наук, доцент, Університет митної справи та фінансів, вул. Володимира Вернадського, 2/4, м. Дніпро, 49000, тел.: (098) 487-76-86, <https://orcid.org/0000-0002-8036-3236>

Щитов Дмитро Миколайович кандидат економічних наук, докторант, Університет митної справи та фінансів, вул. Володимира Вернадського, 2/4, м. Дніпро, 49000, тел.: (099) 288-49-41, <https://orcid.org/0000-0003-4306-8016>

Щитов Олександр Миколайович кандидат фізико-математичних наук, доцент, викладач, НВК-Ліцей № 100, пл. Успенська, 1, м. Дніпро, Україна, 49000, тел.: (066) 336-29-81, <https://orcid.org/0000-0002-1435-2918>

Чупілко Олександр Сергійович аспірант, Університет митної справи та фінансів, вул. Володимира Вернадського, 2/4, м. Дніпро, 49000, тел.: (096) 779-80-98, <https://orcid.org/0009-0004-4950-6771>

МОДЕЛЮВАННЯ ПОКАЗНИКІВ БЕЗПЕКИ НАЦІОНАЛЬНОЇ ЕКОНОМІКИ ТА ОБРОБКА ДАНИХ ЗА ДОПОМОГОЮ PYTHON

Анотація. У статті розглядаються різні аспекти ефективної обробки даних, окреслюються задіяні етапи та особливості кожного етапу. Проаналізовано проблеми, що виникають при обробці даних, та ефективні засоби їх моделювання та прогнозування.

Розглядається використання кількох пакетів NumPy, Pandas, Matplotlib, SciPy, Statsmodels, Scikit-learn. Наведено приклад того, як Python можна використовувати для завдань, пов'язаних з митницею, з урахуванням показників економічної безпеки країни. Авторами побудовано серію регресійних моделей для аналізу внеску митних надходжень від ввізних та вивізних мит до державного бюджету України. Розроблено розрахункову програму з використанням вищенаведених пакетів. При цьому використовувалась мова програмування Python, з використанням бібліотек NumPy, Statsmodels, Matplotlib і Xlrd. Попередню підготовку даних проводили в Excel. Отримані результати проаналізовані. У статті продемонстровано можливість програмної обробки даних. Повні статистичні дані не вдалося отримати з офіційного веб-сайту, щоб уможливити аналіз великих наборів схожих даних. Однак проблеми, які виникли під час моделювання на основі наданих агрегованих даних, відтворюватимуться аналогічно. Крім того, списки містять попередження про те, що для точного розрахунку має бути більший обсяг даних. Якщо обсяг даних недостатній, необхідно використовувати скориговані оцінки, що також відомо з теорії математичної статистики. Вказано на можливість використання результатів моделювання фінансово-економічних показників для прийняття управлінських рішень. Можна побудувати різні однофакторні та багатфакторні: лінійні та нелінійні моделі, і таким чином отримати повну картину того, які процеси відбуваються в галузі, зокрема, на митниці, та виявити позитивні та негативні явища. За результатами моделювання можна отримати науково обґрунтований прогноз і прийняти відповідні управлінські рішення.

Ключові слова: Python, моделювання, прогнозування, обробка даних, регресійна модель.

Introduction. In today's world, companies, enterprises, and institutions manage large volumes of data. Technologies enabling efficient handling of vast amounts of information are becoming increasingly popular. The approach to data

processing depends primarily on the data type, its intended use, and the enterprise's or institution's capabilities in organizing data collection, systematization, and analysis, which are often constrained. Few enterprises can invest significant resources in developing these technologies for their specific needs. Large companies have the resources for business analytics development, while most data processing techniques can handle both large and small datasets, depending on the company's interests and needs. Companies typically work with existing databases rather than continuously generating new data.

Key analysis techniques include cluster and factor analysis, modeling and forecasting using econometric and optimization methods, outlier detection, artificial intelligence, network graphs, and machine learning. Some methods have been around for a long time, while others are more recent. When chosen correctly, various techniques and technologies can effectively evaluate situations across different fields and support informed management decisions. Therefore, understanding which technique suits a specific problem, how these techniques are employed, and their application for modeling indicators is crucial.

Various technologies for studying data utilize similar basic mathematical tools, found in both standard data processing packages and specialized libraries and modules in popular programming languages like Python, as well as in widely-used software products like MS Office and the R language, which is prominent in statistical data analysis.

In their paper, the authors reference the works of foreign researchers [1], [2], which discuss the use of Python and specific software packages for data analysis. Published articles [3] – [5] address issues related to the challenges of this investigation. Numerous papers by both Ukrainian and foreign scientists have been published, discussing access to big data, its implementation in statistics, and its actual benefits. Despite substantial scientific achievements in general data-related issues, there is a notable lack of research focusing on the application of tools for effective data processing using programming languages, especially in the field of customs.

Topicality. Recently, the Python programming language, along with numerous dynamically updating open-source libraries, has become a highly popular and powerful tool. It enables efficient data processing, modeling, and predicting indicators through the interoperable capability of writing code and utilizing ready-made solutions.

Data processing technologies depend on the type of data and the research objectives. Nevertheless, various technological tools share common challenges-data selection and preparation, as well as professional processing of results. Even with a fully automated data processing system, it is crucial to correctly select, sort, and normalize data, choose an appropriate analysis method, and, after the program processes the data, conduct analysis, interpretation, and prediction.

During data preparation, several issues arise due to different data formats from various sources, limited data access because of its value, confidentiality, and strict regulation. Data might have different units of measurement and levels of aggregation. Enhancing data quality, understanding data interactions, and evaluating distributions and conversions to a standardized format are impossible without fundamental knowledge of relevant mathematical tools.

Another significant challenge is data access. Official statistical data, which form the basis for modeling and predicting financial and economic indicators at a national level, particularly in customs, are often limited and mostly aggregated. Frequently, official websites or statistical collections provide data that are not normalized or are insufficient in quantity to build adequate models. Consequently, modeling tasks are constrained to the statistical data that are officially available.

The success of the modeling process hinges on the quality of the data and the expertise of the analyst.

Tasks.

1. Analyze Problems in Working with Data and Application Tools

✓ **Identify Common Challenges:** discuss typical issues encountered during data processing, such as data selection, data preparation, format inconsistencies, access restrictions, and confidentiality concerns.

✓ **Examine Tools and Techniques:** review various tools and methodologies for effective data processing, including cluster analysis, factor analysis, econometric modeling, optimization methods, outlier detection, artificial intelligence, network graphs, and machine learning.

2. Apply the Most Suitable Tools for Modeling Financial and Economic Indicators

✓ **Selection of Tools:** choose the most appropriate Python libraries (e.g., NumPy, Pandas, Matplotlib, SciPy, Statsmodels, Scikit-learn) for the specific task of modeling customs revenues.

✓ **Implementation:** use these tools to create and validate models that predict customs revenues from import and export duties.

✓ **Model Evaluation:** assess the quality of the constructed models using specific criteria to ensure accuracy and reliability of predictions.

✓ Specifically, it uses the example of official customs revenue data related to export and import duties to the state budget of Ukraine.

By fulfilling these tasks, the work aims to demonstrate the practical application of modern data processing tools and techniques in addressing real-world financial and economic challenges, using the case study of Ukraine's customs revenues.

Solving problems.

1. Tools for efficient and fast data processing

For data processing, it is appropriate to use techniques that are most effective for solving a specific type of problem.

Basic Tools: Spreadsheets MS Excel:

- **Accessibility and Simplicity:** Excel is a convenient and accessible tool for initial data analysis. It includes an analysis package that, despite its limitations, is suitable for obtaining preliminary data processing results to understand the nature of the data.

- **Capabilities:** in Excel, you can perform data normalization, correlation and regression analysis, obtain point and interval estimates, and solve linear and nonlinear optimization problems.

- **Limitations:** Excel is not suitable for launching production models, such as artificial intelligence, in a software mode.

Advanced Tools: programming and Specialized Software

Programming Languages:

- **Python:** Python is a powerful tool for data processing due to its extensive libraries like NumPy, Pandas, Matplotlib, SciPy, Statsmodels, and Scikit-learn. These libraries assist in data manipulation, statistical analysis, visualization, and machine learning.

- **R:** R is another popular language for statistical analysis, offering a wide range of packages for data processing, modeling, and graphical representation.

Specialized Software:

- **SPSS (Statistical Package for the Social Sciences):** SPSS is a robust tool for statistical analysis, providing capabilities for regression analysis, factor analysis, cluster analysis, neural network modeling, and more. It also offers excellent options for graphical data representation.

- **SAS (Statistical Analysis System):** SAS is known for its advanced analytics capabilities, including multidimensional analysis, business analytics, data management, and predictive analytics.

Stages of Data Processing

1. Data Collection:

- **Sources:** data can be collected from various sources such as databases, APIs, web scraping, or manual input.

- **Challenges:** ensuring data accuracy, completeness, and relevance.

2. Data Preparation:

- **Cleaning:** handling missing values, outliers, and duplicate records.

- **Transformation:** normalizing data, changing data types, and aggregating data as needed.

- **Integration:** combining data from different sources and ensuring consistency.

3. Data Analysis:

- **Descriptive Statistics:** summarizing data to understand its basic characteristics.

- **Inferential Statistics:** making predictions or inferences about a population based on a sample of data.
- **Advanced Analysis:** using methods such as regression analysis, factor analysis, and cluster analysis to gain deeper insights.

4. Modeling and Prediction:

- **Model Building:** constructing models using methods like regression, decision trees, neural networks, etc.
- **Validation:** assessing model performance using criteria such as accuracy, precision, recall, and F1-score.
- **Forecasting:** using the model to predict new data and provide interval estimates.

5. Data Visualization:

- **Tools:** using tools like Matplotlib, Seaborn (in Python), or ggplot2 (in R) to create graphs and charts that visually represent the data and analysis results.
- **Purpose:** aiding in communicating insights and supporting decision-making processes.

Stages in the Data Processing Workflow

1. Defining the Study's Purpose:

- **Objective Setting:** establish the project's goals, outline the research objectives, and estimate the project's cost.
- **Project Task Preparation:** develop a detailed project plan that aligns with the research purpose.

2. Data Collection and Preparation (Intelligence Analysis):

- **Data Gathering:** collect data from various sources.
- **Normalization:** convert disparate data formats into a uniform format.
- **Handling Incomplete Data:** address issues like incomplete matrices and singularities by choosing appropriate algorithms for gap filling.
- **Outlier Detection and Cleaning:** identify and eliminate significant deviations (outliers) to ensure data quality, as unclean data can lead to inaccurate models.
- **Manual Effort:** this stage often requires meticulous, almost "manual" efforts, along with an intellectual approach and a clear understanding of the study's objectives.

3. Analysis and Data Modeling:

- **Model Selection and Evaluation:** choose an appropriate model and evaluate its parameters.
- **Understanding Data Relationships:** assess how data variables relate to each other, estimate data distributions, and identify outliers.
- **Addressing Statistical Issues:** check for multicollinearity, heteroskedasticity, and autocorrelation, which may necessitate additional data transformations and the application of specialized methods.

- **Statistical and Simple Modeling:** use statistical methods to answer questions about factor relationships, multicollinearity, variable reduction, dependency forms, and model linearity.

- **Model Training:** build various models using a subset of the data, selected randomly from the population. This involves adjusting parameters to select the best model based on criteria such as the least squares method, decision tree methods, or absolute deviation methods.

- **Iterative Process:** model building is iterative, requiring repeated training and parameter adjustment to achieve the best results.

4. Model Adequacy Check and Factor Significance:

- **Quality Evaluation:** assess the model's quality using statistical criteria. Compare the sum of squared residuals and select the parameter set that minimizes this sum.

- **Model Retraining:** if the model quality is unsatisfactory, the model needs to be retrained.

5. Application to Unknown Data (Predictive Modeling):

- **Training Set Selection:** use a "training set" from the same sample for predictive modeling.

- **Forecasting:** apply the model to make predictions on new, unseen data.

Each of these stages requires specific tools and expertise to ensure efficient and accurate data processing, leading to reliable and actionable insights.

The described approach is used for modeling and forecasting tasks in machine learning. For instance, Python's Scikit-learn library offers a variety of algorithms suitable for these tasks.

Machine learning is becoming increasingly popular and promising among data analysts (data scientists). The market for machine learning has been expanding rapidly, with its value surpassing \$1 billion in 2016 and projected to grow to \$39.98 billion by 2025. Currently, 60% of companies worldwide utilize machine learning.

Machine learning can address various tasks, including modeling and forecasting indicators based on one or more factors, as well as optimization tasks. These tasks employ both traditional econometric analysis methods, such as single-factor and multi-factor models based on the method of least squares, and non-traditional methods like decision trees with numerous adjustable parameters, offering flexibility in parameter modeling.

Neural networks are also gaining popularity in the field. In modeling, the concept of risk is often employed, with its quantitative characteristics calculated based on the numerical traits of discrete and continuous random variables.

Over the past decade, Python has emerged as a pivotal programming language in data science, machine learning, and general-purpose software development in both academic and industrial settings. Enhanced libraries for Python have made it a strong competitor in solving data processing application challenges.

Many modern environments still rely on a common set of legacy libraries written in FORTRAN and C, which include implementations of algorithms for linear algebra, optimization, integration, and more. Consequently, many companies use Python as a “glue” language to integrate these long-established programs with new data processing applications.

2. Python packages for working with data.

NumPy (Numerical Python) is a fundamental package for scientific calculations in Python. It serves as the foundation for many other libraries. Key features of NumPy include:

- Efficient creation and manipulation of multidimensional arrays.
- Functions for performing calculations on array elements and mathematical operations involving multiple arrays.
- Methods for reading from and writing to disk data sets in array form.
- Linear algebra operations, Fourier transform and random number generation.
- Integration with code written in C, C++, or Fortran.

NumPy significantly speeds up working with arrays, making data storage and manipulation more efficient compared to Python’s built-in data structures. Many Python computing tools use NumPy arrays as their underlying data structure or integrate with NumPy.

Pandas provides data structures and functions that simplify and accelerate working with structured data. The main Pandas objects are:

- *DataFrame*: a two-dimensional table with labeled rows and columns.
- *Series*: a one-dimensional array object with labels.

Pandas combines the high performance of NumPy’s array tools with the flexible data manipulation capabilities of spreadsheets and relational databases (e.g., SQL). As data manipulation, preparation, and cleaning are crucial in data analysis, Pandas is one of the primary tools used.

Main capabilities of Pandas include:

- Advanced indexing tools for reshaping data sets, forming slices, performing aggregation, and selecting subsets.
- Data structures with labeled axes that support automatic or explicit data alignment, eliminating common errors when working with unaligned data from different sources.
- Built-in functionality for handling time series data.
- Support for both time series and other types of data within the same structures.
- Arithmetic operations on objects treated as numerical data.
- Flexible handling of missing data (imputation).
- Data integration and support for connection and other relational operations available in popular databases (e.g., SQL-based).

These packages together form a robust toolkit for data analysis, enabling efficient data manipulation, analysis, and visualization.

Pandas offers many features available in the R language or through additional R packages. The name “pandas” comes from “panel data”, a term used in econometrics for multidimensional structured datasets, and “Python data analysis”. Pandas simplifies working with structured data through powerful data manipulation capabilities.

Matplotlib is the most popular Python tool for creating graphs and visualizing two-dimensional data. It is suitable for creating publication-quality graphs and integrates well with other parts of the Python ecosystem. Although other packages offer visualization capabilities, *Matplotlib* remains the most widely used.

SciPy is a collection of packages for solving various standard computing problems. Key submodules include:

- `scipy.integrate`: routines for numerical integration and solving differential equations.
- `scipy.linalg`: subroutines for linear algebra and matrix expansion, complementing `numpy.linalg`.
- `scipy.optimize`: algorithms for optimizing functions (finding extrema) and root finding.
- `scipy.signal`: tools for signal processing.
- `scipy.sparse`: algorithms for working with sparse matrices and solving sparse linear systems.
- `scipy.special`: a wrapper around the SPECFUN Fortran library, implementing many standard mathematical functions, including the gamma function.
- `scipy.stats`: standard continuous and discrete probability distributions, statistical tests, and descriptive statistics.

Scikit-learn is the primary Python machine learning library, offering tools for various models:

- classification: support vector machines, nearest neighbors, random forests, logistic regression, etc;
- regression: lasso, ridge regression, etc;
- clustering: k-means, spectral clustering, etc;
- dimensionality reduction: principal component analysis, feature selection, matrix factorization, etc;
- model selection: grid search, cross-validation, metrics;
- preprocessing: feature selection, normalization.

Scikit-learn focuses on forecasting and prediction, making it a core toolkit for machine learning in Python.

Statsmodels is a statistical analysis package focusing on classical (frequentist) statistics and econometrics, complementing *Scikit-learn*. It includes submodules for:

- regression models: linear regression, generalized linear models, mixed-effects models, etc.
- ANOVA (Analysis of Variance).
- time series analysis: AR, ARMA, ARIMA, VAR, and other models.
- non-parametric methods: kernel density or regression estimation.
- visualization of statistical modeling results.

Statsmodels emphasizes statistical inference, providing uncertainty estimates and p-values of parameters. It works seamlessly with NumPy and Pandas. Python also has libraries for convenient and quick data reading from various formats such as spreadsheets, databases, CSV files, and more. These libraries facilitate efficient data import and export, streamlining the data analysis workflow.

3. The application of Python to model customs revenues of Ukraine.

For this research, data from official statistics [6] was used. It's important to note that publicly available information is very limited. The consolidated data used for this task includes the total volume of receipts from customs authorities to the state budget of Ukraine, as well as receipts from import and export duties. This data is shown in Table 1.

Table 1.

Raw data for modeling receipts to the state budget of Ukraine

Years	Income to the state budget of Ukraine Y, UAH	Income to the state budget of Ukraine by import duty X1, UAH	Income to the state budget of Ukraine by export duty, X2, UAH
2015	5,34694E+11	1,7422E+10	2,4500E+08
2016	6,16219E+11	2,0004E+10	3,7000E+08
2017	6,98405E+11	2,2257E+10	6,4300E+08
2018	8,33615E+11	2,3301E+10	5,1600E+08
2019	8,79833E+11	2,2778E+10	2,3000E+08
2020	8,77603E+11	2,1538E+10	2,5700E+08
2021	10,84E+11	1,320E+10	1,4E+08
2022	13,89E+11	1,343E+10	1,2E+08
2023	30,75E+11	1,034E+10	3,2E+08

According to the data in Table 1, to analyze the relationships between import and export duties and the total incomes from customs authorities, we will use econometric modeling techniques. Specifically, we will:

1. Evaluate the Presence and Closeness of Relationships
2. Determine the Form and Type of the Model
3. Estimate Regression Parameters
4. Assess Model Adequacy

5. Evaluate Statistical Significance of Parameters
6. Check for Autocorrelation
7. Predict Values with Point and Interval Estimations
8. Construct Regression Confidence Intervals

We will use Python and the following libraries for our analysis:

- NumPy for numerical operations.
- Pandas for data manipulation.
- Statsmodels for statistical modeling.
- Matplotlib for data visualization.
- Xlrd for reading data from Excel files.

Step-by-Step Analysis

1. Data Collection and Preparation

- Read the data from an Excel file.
- Clean and normalize the data to ensure consistency.

2. Exploratory Data Analysis (EDA)

- Plot the data to understand initial patterns and distributions.

3. Model Selection and Training

- Use linear regression to model the relationships.
- Evaluate the models using the least squares method.

4. Model Evaluation

- Assess the model's adequacy using the Fisher test.
- Evaluate the statistical significance of the parameters using the Student's

t-test.

5. Check for Autocorrelation: perform tests to check for the presence of autocorrelation in residuals.

6. Prediction and Confidence Intervals: predict future values and construct confidence intervals for the predictions.

In this section, we provide the results of the software execution in Python and graphical visualizations to represent the constructed models. We will analyze how the incomes to the state budget from customs authorities depend on the income from import duties.

The analysis of the main simulation results gave the following:

1. Dependent Variable: Total customs revenue (y).
2. Independent Variable: Import duty revenue.
3. Regression Equation: the regression equation derived from the model is shown in Fig. 1 (visual representation not included here).
4. Correlation Coefficient: 0.95, this indicates a strong positive correlation between import duty revenue and total customs revenue.
5. Adjusted Coefficient of Determination (0.885): this means that 88.5% of the variance in total customs revenue can be explained by the variance in import duty revenue.

6. F-statistic:

- Calculated F-value: 55.03
- Critical F-value: 5.98 (for the given degrees of freedom and significance level of 0.05)
- The calculated F-value significantly exceeds the critical value, confirming that the model is statistically adequate.

7. t-statistics:

- Slope: 7.42
- Intercept: -3.25
- Both t-values indicate that the corresponding parameters are statistically significant, as they exceed the critical t-value of 2.45 with a confidence probability of 0.975.

8. Statistical Significance:

- The regression parameters (slope and intercept) are statistically significant, indicating that they contribute meaningfully to the model.

The high correlation coefficient of 0.95 suggests a strong relationship between import duty revenue and total customs revenue. The adjusted of 0.885 indicates that the model explains a significant portion of the variance in total customs revenue. The F- and t-statistics both support the adequacy and significance of the model. Therefore, this OLS model is reliable and can be used to predict total customs revenue based on import duty revenue.

Confidence Intervals of Regression Parameters

1. Slope:

- Confidence Interval: (39.14, 77.67)
- This range indicates that the true value of the slope parameter, with 97.5% confidence, lies between 39.14 and 77.67.

2. Intercept:

- Confidence Interval: (-8.75E+11, -1.23E+11)
- This range indicates that the true value of the intercept parameter, with 97.5% confidence, lies between -8.75E+11 and -1.23E+11.

Autocorrelation Check

- Durbin–Watson Statistic:
- This statistic tests for the presence of autocorrelation in the residuals of the regression model.
- The Durbin–Watson statistic for this model indicates the absence of autocorrelation, meaning the residuals are not correlated with each other.

Covariance Matrix: the covariance matrix of the regression parameters is correctly specified, indicating that the estimates of the regression coefficients are reliable.

Model Utility for Prediction

• The model can be utilized to predict the indicator, which in this context is the total customs revenue.

• Forecast Estimates:

• Both point and interval estimates for future values can be determined using this model.

Coefficient of Elasticity

• Elasticity Coefficient: 1.69

• This indicates that the total customs revenue is elastic with respect to the import duty revenue.

• An elasticity coefficient greater than 1 suggests that a 1% change in import duty revenue leads to a more than 1% change in total customs revenue.

• The growth rate of total customs revenue is slowing down, primarily due to changes in import duties.

The additional analysis reaffirms the robustness of the OLS model. The confidence intervals of the regression parameters suggest that the estimates are reliable. The absence of autocorrelation, as indicated by the Durbin–Watson statistic, further strengthens the model's validity. The correctly specified covariance matrix ensures the reliability of the parameter estimates. With an elasticity coefficient of 1.69, the model indicates that customs revenue is highly responsive to changes in import duty revenue. This model can thus be effectively used for forecasting and policy analysis related to customs revenues.

Then we'll construct and analyze the regression of customs revenues to the export duty budget (Fig. 3). The same OLS model is applied. The dependent variable is y .

The corresponding listing is shown in Fig. 4.

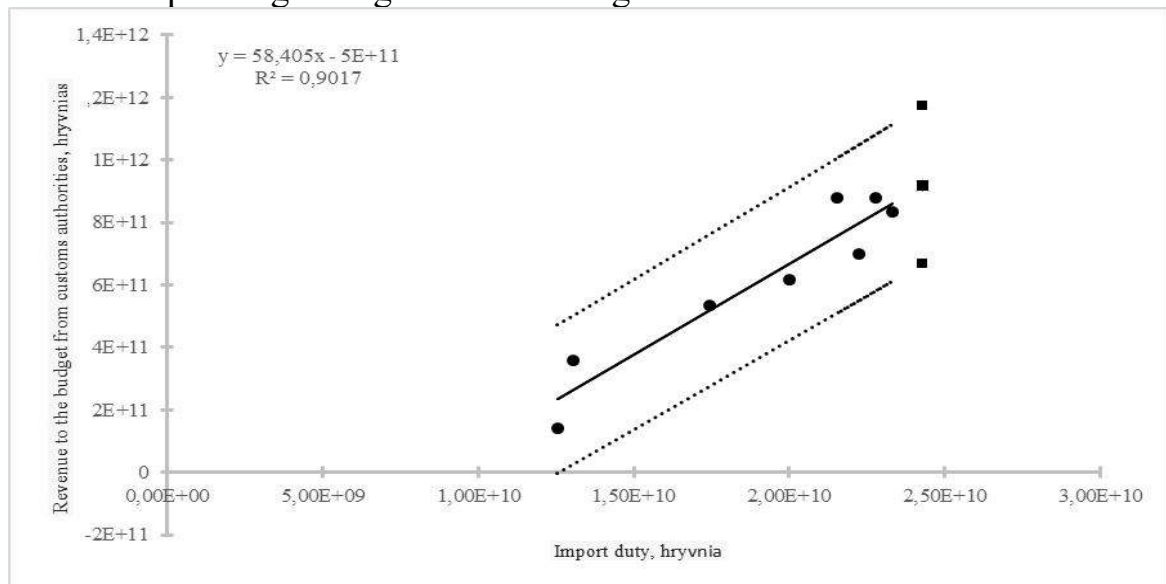


Fig. 1 Regression of incomes to the budget of Ukraine from customs authorities on import duty, constructed with a reliability of 0.95

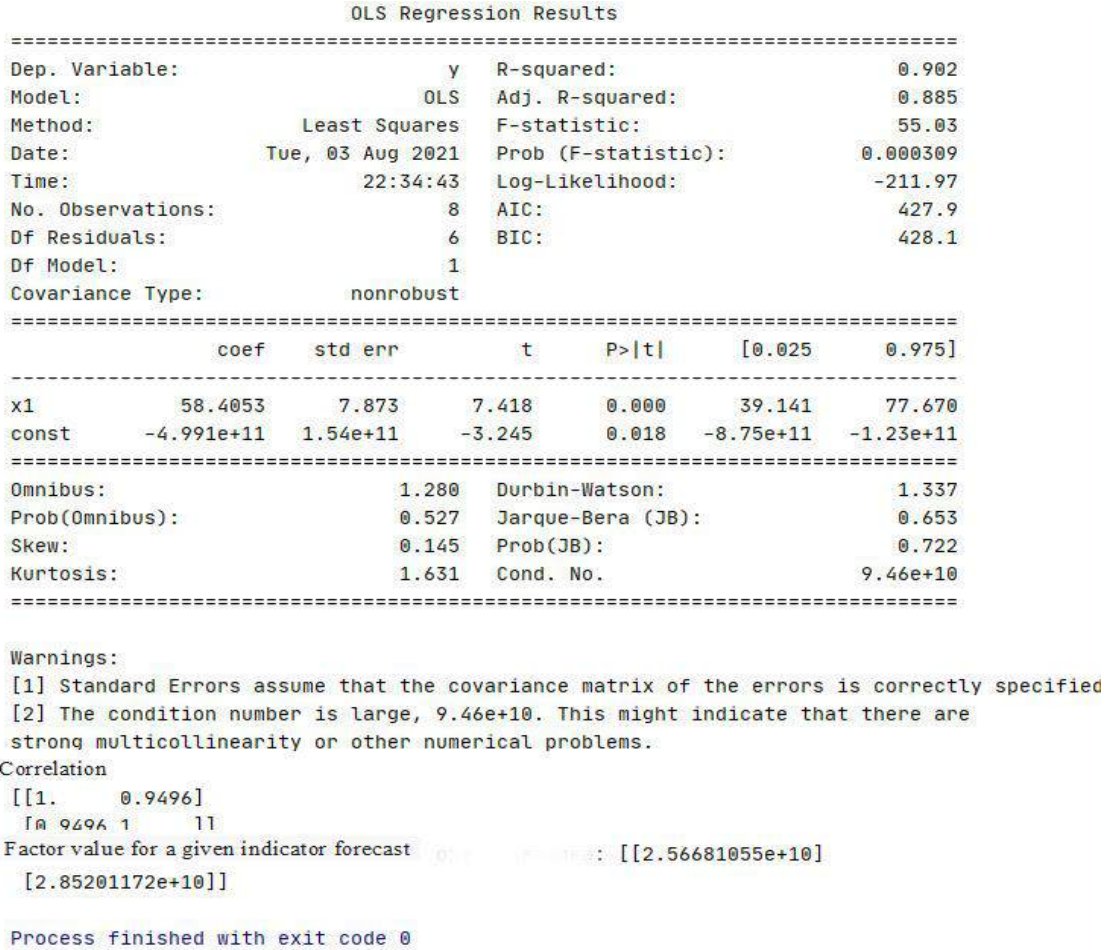


Fig. 2 List of execution of the program (model of dependence of customs revenue to the budget on the import duty)

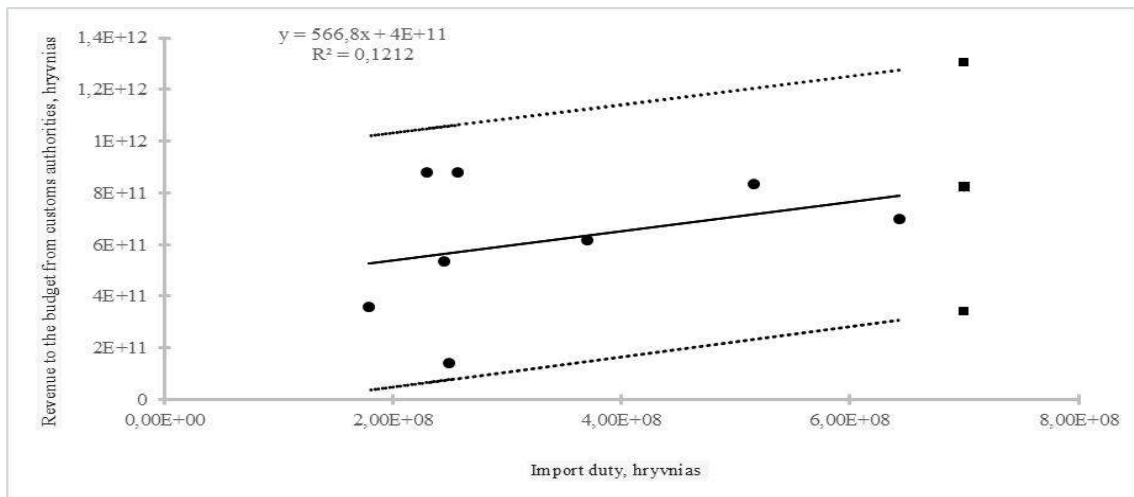


Fig. 3 Regression of incomes to the Ukraine's budget from customs authorities on export duty, confidence interval of the forecast and regression, with a reliability of 0.95

Результат розрахунку:

OLS Regression Results

```

=====
Dep. Variable:          y      R-squared:                0.121
Model:                  OLS    Adj. R-squared:           -0.025
Method:                 Least Squares  F-statistic:              0.8276
Date:                   Tue, 03 Aug 2021  Prob (F-statistic):      0.398
Time:                   22:29:22    Log-Likelihood:          -220.73
No. Observations:      8          AIC:                     445.5
Df Residuals:          6          BIC:                     445.6
Df Model:               1
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
x1	566.7962	623.032	0.910	0.398	-957.708	2091.300
const	4.267e+11	2.3e+11	1.856	0.113	-1.36e+11	9.89e+11

```

=====
Omnibus:                0.138    Durbin-Watson:           0.410
Prob(Omnibus):          0.933    Jarque-Bera (JB):        0.213
Skew:                   -0.204   Prob(JB):                 0.899
Kurtosis:                2.313    Cond. No.                 8.94e+08
=====

```

Warnings:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified
- [2] The condition number is large, 8.94e+08. This might indicate that there are strong multicollinearity or other numerical problems.

Correlation

```

1 1 1.      0.3482]
[0.3482 1.      ]

```

```

Factor value for a given indicator forecast      : [[4.44254957e+08]
[4.73616617e+08]]

```

Process finished with exit code 0

Fig. 4 List of program implementation (model of dependence of customs incomes to the budget on export duty)

The correlation coefficient is 0.348. This measures the strength and direction of the linear relationship between two variables. The coefficient of determination of 0.1212 means that approximately 12.12% of the variability in the indicator (dependent variable) can be explained by the factor (independent variable). This is quite low, suggesting that the model does not explain much of the variability in the indicator.

The F-statistics indicates is 0.83, the calculated F-value of 0.83 is much lower than the critical value of 5.98 at the 0.05 significance level. Calculated t-statistic values are 0.91 for the slope and 1.86 for the regression intercept. This tests if the slope of the regression line is significantly different from zero. The critical value of the t-statistic is 2.45 for the given degrees of freedom and significance level (0.05).

The elasticity coefficient based on average indicators for the last four years is 0.34, increasing to 0.5, indicating that the indicator is inelastic to the factor, and the growth rate of budget revenues from customs receipts is accelerating due to export duties.

A very wide confidence interval is due to a large deviation of the original data, and accordingly, a wide confidence interval of the forecast has no practical meaning. The Durbin-Watson statistic indicates no autocorrelation in the model. The covariance matrix is specified correctly.

The model obviously cannot be recommended for use in predicting the indicator.

Therefore, for both models:

- 1) the method of least squares was used, which gives the best approximation of the initial data with the smallest sum of squared errors;
- 2) estimates obtained by the Gauss-Markov theorem are efficient and unbiased;
- 3) the calculation was carried out using the author's Python program, using NumPy, Statsmodels, Matplotlib and Xlrd libraries.

Preliminary data preparation was carried out in Excel.

The article demonstrates the possibility of software-based data processing. Complete statistical data could not be obtained from the official website to enable analysis on large datasets of similar data. However, the problems encountered during modeling based on the provided aggregated data would similarly reproduce. Additionally, the listings include a warning that the amount of data required for accurate calculation must be larger. If the data volume is insufficient, adjusted estimates must be used, which is also known from the theory of mathematical statistics.

It can be noted that with a small amount of data, the statistical analysis provided by Excel could be used, and it would be the most effective solution. However, the purpose of the current work was to apply Python's capabilities for modeling. The same program would provide efficient calculations for large datasets, which is not a problem when programming directly. It can also be noted that data can be stored in databases, and Python has appropriate tools for working with such data. Therefore, we have relevant economic conclusions, which reveal significant issues in revenues from customs authorities and, in particular, from export duties.

It is possible to construct various single-factor and multi-factor linear and nonlinear models to obtain a complete picture of the processes occurring in the industry, particularly in customs, and to identify positive and negative phenomena.

Based on the modeling results, a scientifically grounded forecast can be obtained, and appropriate managerial decisions can be made.

Conclusions.

1. The problems arising in data processing and effective means of data modeling and forecasting have been analyzed.

2. Models analyzing customs revenues to the state budget of Ukraine from import and export duties have been constructed. The Python programming language and packages suitable for the tasks were used.

3. The possibility of using the results of financial and economic indicators modeling for making managerial decisions has been indicated.

References:

1. Makkynny, U. (2023). *Python y analiz dannykh* (per. s angl. A. Slinkina) [Python and Data Analysis]. Moskva: DMK Press, 536 s. [in Russian].

2. Devy, S., Arno, M., Mohamed, A. (2016). *Introducing Data Science: Big Data, Machine Learning, and more, using Python tools*. New-York: dreamtech press india, 336 p. [in English].

3. Chupilko, T. A. (2021). Aktualni problemy vysokoeffektivnoi obrobky danykh. Modeliuvannya pokaznykiv za dopomohoiu movy prohranuvannya Python [Actual problems of highly efficient data processing. Modeling indicators using the Python programming language]. *Aktualni napriamy rozvytku tekhnichnoho ta vyrob-nychoho potentsialu natsionalnoi ekonomiky*. Dnipro: Porohy, 151-163 [in Ukrainian].

4. Chupilko, T. A. (2020). Bazovi instrumentarii u suchasnykh tekhnolohiiakh kompiuternoi biznes-analitiky [Basic tools in modern technologies of computer business analytics]. *Innovatsiini tekhnolohii, modeli uprav-linnia kiberbezpekoiu ITMK-2020*, Dnipro, T.2, 53-54 [in Ukrainian].

5. Chupilko, T. A. (2020). Kompiuterni tekhnolohii ta ekonomiko-matematychni metody v upravlinni biznes-protsesamy na pidpriemstvi [Computer technologies and economic and mathematical methods in the management of business processes at the enterprise]. *Mizhnar. Nauk. Konf. «Innovatsiini tekhnolohii, modeli upravlinnia kiberbezpekoiu ITMK-2020»*, Dnipro, T.1, 26-28 [in Ukrainian].

6. Ministerstvo finansiv Ukrainu [Ministry of Finance of Ukraine] (2023). URL: <http://mof.gov.ua> [in Ukrainian].

Література:

1. Маккинни У. Python и анализ данных (пер. с англ. А. Слинкина). – Москва: DMK Пресс, 2023. – 536 с.

2. Devy S., Arno M., Mohamed A. *Introducing Data Science: Big Data, Machine Learning, and more, using Python tools*. – New-York: dreamtech press india, 2016. – 336 p.

3. Чупілко Т. А. Актуальні проблеми високоєфективної обробки даних. Моделювання показників за допомогою мови програмування Python // Актуальні напрями розвитку технічного та виробничого потенціалу національної економіки. – Дніпро: Пороги, 2021. – С. 151-163.

4. Чупілко Т. А. Базовий інструментарій в сучасних технологіях комп'ютерної бізнес-аналітики // Іноваційні технології, моделі управління кібербезпекою ITMK-2020. – Дніпро, Т.2, 2020. – С. 53-54.

5. Чупілко Т. А. Комп'ютерні технології та економіко-математичні методи в управлінні бізнес-процесами на підприємстві // Міжнар. наук. конф. «Іноваційні технології, моделі управління кібербезпекою ITMK-2020». – Дніпро, Т.1, 2020. – С. 26-28.

6. Міністерство фінансів України. <http://mof.gov.ua>. Data of application: 20 serpnia, 2023.