UDC 004.5
DOI: 10.20535/SRIT.2308-8893.2024.2.02

MULTI-FACTOR FORECASTING OF STATISTICAL TRENDS FOR DATA SCIENCE PROBLEMS

O. PYSARCHUK, T. ANDREIEVA, O. GRINENKO, D. BARAN

Abstract. The article deals with the processes of multi-factor forecasting of statistical trends for Data Science problems. Most of the classic approaches to data processing consist of studying the consequences of phenomena rather than the factors of their appearance. At the same time, the factors affecting the behavior of the investigated process are assumed to be random and are not investigated. The article discusses the approach to forecasting the parameters of the trend of statistical time series, which consists of the study of factors that lead to changes in the dynamics of the studied process. This approach potentially has better indicators of adequacy, accuracy, and efficiency in obtaining final solutions than classical approaches. The implementation of this approach is shown using an example of the analysis of exchange rate changes. The obtained results show the practicality of considering multifactoriality in forecasting tasks.

Keywords: Data Science, multi-factor forecasting, statistical trends, currency rate forecasting.

INTRODUCTION

The development of information technologies has led to their implementation in many areas. One of the leading directions is the prediction of the indicators behavior of a certain controlled event. The examples of that can be: forecasting fluctuations in currency markets; control of changes in economic performance indicators of trading companies; forecasting the development of the epidemiological situation; forecasting parameters of the technical state of equipment of production lines, aviation systems, etc. All the listed applied tasks have the technological unity of Data Science stages: data acquisition (measurement); their accumulation (storage); data processing for the purpose of obtaining information about the models and behavior of the researched process (processing, forecasting); extraction of knowledge and its manipulation [1; 2]. Currently, the focus of Data Science issues is not on accumulation (measurement, storage), but on data processing with the aim of extracting from them adequate, accurate and operational information and knowledge. These processes in applied aspects of information technologies (IT) take place in the field of Big Data arrays and are manifested in the development of Back-End components of distributed ERP / CRM software systems with intellectual properties.

The key requirement of consumers for the final IT product is high quality indicators of the source information, which are manifested in strict requirements for the adequacy, accuracy and efficiency of the final solutions. It is possible to implement this only in the direction of applying effective mathematical models for processing Big Data arrays.

The experience shows that most classical approaches to data processing, regardless of their classes, directions of improvement and effective implementation to applied software systems, show their limitations [3–5]. They consist in the study of the consequences of phenomena, and not the factors of their appearance. For example, determining the trend and forecasting changes in the exchange rate based on the results of a retrospective analysis of their behavior. At the same time, the factors affecting the exchange rate are assumed to be random and are not investigated.

Therefore, there is a need to implement R&D processes for the development of mathematical support for modern ERP / CRM software systems capable of meeting the high demands of consumers regarding the adequacy, accuracy and efficiency of final solutions.

The article will consider an approach to predicting parameters of the trend of statistical time series, which potentially has better indicators of adequacy, accuracy and efficiency of obtaining final solutions, compared to classical approaches.

Analysis of existing approaches. In its formulated form, we have the classic task of applied statistical analysis / statistical learning: to build a mathematical model based on a statistical sample of data, that ensures the determination of predictive values for the process being studied [1–5]. The key hypothesis in this is the assumption of the random nature of the factors that affect the stochastic fluctuations of each discrete dimension and, accordingly, determine the behavior of the studied process outside the observation interval. As a rule, this happens due to the complex and sometimes unknown nature of cause-and-effect relationships, which determine the actual appearance of stochastic deviations and the development of the situation in the future. Overcoming this a priori uncertainty is classically implemented through assumptions about the general appearance of the trend model and the determination of its variables using complex algorithms, but the principle hypothesis of randomness remains unchanged. That is, the primary stochastic formalization of the problem has certain limitations in the accuracy of the final result, which are determined by data processing methods.

Formulation of the problem. Therefore, the task of improving the methods of statistical analysis / training in the direction of a detailed description and study of factors that lead to the essence of the change in consequences – the dynamics of the researched process – is urgent. The article examines the processes for multifactor forecasting of statistical trends for Data Science tasks. This is implemented in the applied field of economic analysis of exchange rate changes. The transition in statistical education from the analysis of consequences to factors requires the implementation of a complex of R&D processes: the formation of an informational model of factors that influence the change in currency rates; the establishment of indicators (indicators describing change) of factors and criteria; the measurement of indicators; and the statistical processing of indexes / indicators (determination of statistical characteristics, construction of a trend line, forecasting).

Thus, the goal of the article is the implementation of a complex of R&D processes for multifactor forecasting of statistical trends for Data Science tasks using the example of currency exchange analysis.

AN OVERVIEW OF THE MAIN MATERIAL

1. *To form the infographic model of factors that influence on the change of the currency exchange rates*. The ratio of the dollar (USD) to the hryvnia (UAH) was chosen as the exchange rate (hereinafter referred to as the exchange rate). On the basis of the cognitive analysis of primary sources [6–13] and the practice of currency trading, the factors affecting the exchange rate were determined.

Table 1. An infographic model of factors that influence the change in currency rates

N	Factor group	N	Factor in the group	Indicator	Data source, frequency of measurement
1	Sale/ purchase of foreign currency	1	Volume of sale/purchase of foreign currency	The official exchange rate of the hryvnia against the US dollar	The official website of the NBU[10], daily
				Saldo of transactions of the natural person on the sale/purchase of foreign currency	
				Saldo of NBU interventions	The official website of the NBU[10], weekly
2	Export of goods	1	Volume of the main Ukrainian export goods	Wheat export volume	Website of the Ministry of Agrarian Policy and Food of Ukraine[8], daily
				Barley export volume	
				Rye export volume	
				Corn export volume	
		2	Export prices for the main agricultural products of Ukraine	Wheat export price	Fenix Agro company website: fenix-agro.com; weekly
				Barley export price	
				Corn export price	
		3	Export prices for the main metal products of Ukraine	Hot-rolled steel export price	Information and analytical resource about industry: gmk.center, daily
				Armature export	
				Scrap steel export price	
				Iron ore raw materials export	
3	Import of goods	1	Global prices for the main imported goods	Oil global price	The website of the Ministry of Finance [12], daily
				Natural gas global price	
4	Foreign investments	1	Participation of non-residents in trading in hryvnia bonds of the domestic state loan	The volume of hryvnia government bonds in circulation at nominal and amortized cost with non-residents	The official website of the NBU[10], daily
				The amount of funds involved in the state budget for placement of domestic government bonds	Website of the Ministry of Finance of Ukraine [9], weekly
5	Interest rates	1	The level of inter-est rates on the interbank market	Interest rates on deposits in the national currency	The official website of the NBU[10], daily
				Interest rates on deposits in US dollars	
		2	Interest rates on deposits	NBU Key Policy Rate	The website of the Ministry of Finance [12], daily
				Ukrainian Overnight I ndex Average (UONIA)	
6	Stock Market	1	Stock indices of Ukraine	UX index	The website of the Ministry of Finance [12], daily
		2	World stock indices	Dollar index	Investing.com, daily

2. To set the indicators (parameters which describes the change) of factors and criteria is implemented as a result of the transformation of Table 1, based on the essence of a specific factor.

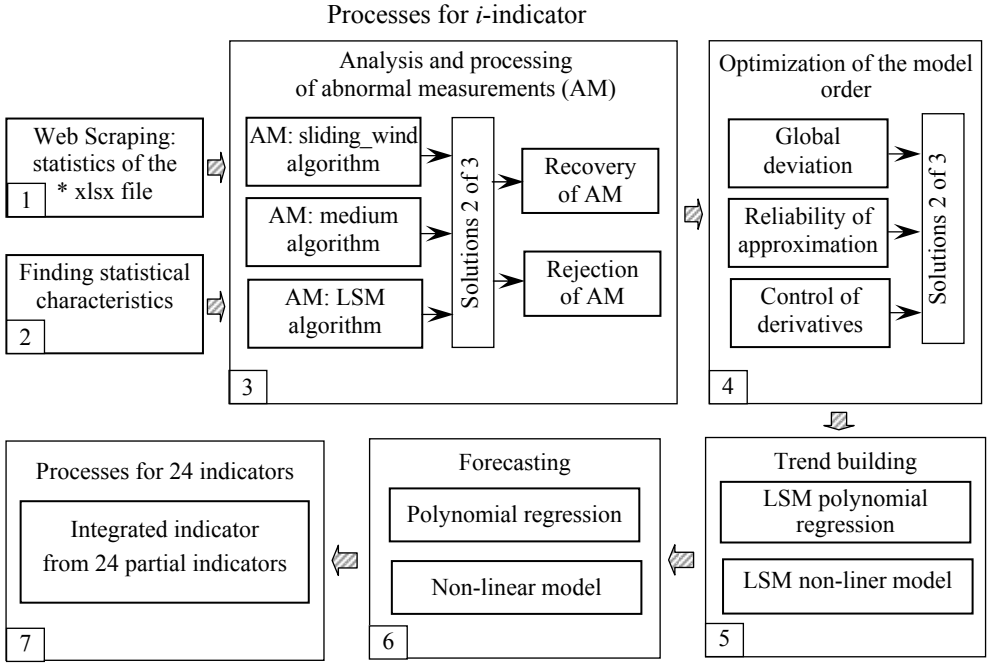
Table 2. Indicators / parameters that describe the change of the factors and criteria

№	Group of factors	№	Indicators	Denotation	Criterion
1	Sale/ purchase of foreign currency	1	The official exchange rate of the hryvnia against the US dollar	ψ	$\psi \rightarrow \min$
		2	Saldo of transactions of the natural person on the sale/ purchase of foreign currency	φ	$\varphi \rightarrow \max$
		3	Saldo of NBU interventions	χ	$\chi \rightarrow \min$
2	Export of goods	4	Wheat export volume	E_{VW}	$E_{VW} \rightarrow \max$
		5	Barley export volume	E_{VB}	$E_{VB} \rightarrow \max$
		6	Rye export volume	E_{VR}	$E_{VR} \rightarrow \max$
		7	Corn export volume	E_{VC}	$E_{VC} \rightarrow \max$
		8	Wheat export price	E_{PW}	$E_{PW} \rightarrow \max$
		9	Barley export price	E_{PB}	$E_{PB} \rightarrow \max$
		10	Corn export price	E_{PC}	$E_{PC} \rightarrow \max$
		11	Hot-rolled steel export price	E_{PS}	$E_{PS} \rightarrow \max$
		12	Armature export	E_{PA}	$E_{PA} \rightarrow \max$
		13	Scrap steel export price	E_{PJ}	$E_{PJ} \rightarrow \max$
		14	Iron ore raw materials export	E_{P0}	$E_{P0} \rightarrow \max$
3	Import of goods	15	Oil global price	I_{POIL}	$I_{POIL} \rightarrow \min$
		16	Natural gas global price	I_{PGAS}	$I_{PGAS} \rightarrow \min$
4	Foreign investments	17	The volume of hryvnia government bonds in circulation at nominal and amortized cost with non-residents	INV_V	$INV_V \rightarrow \max$
		18	The amount of funds involved in the state budget for placement of domestic government bonds	INV_M	$INV_M \rightarrow \max$
5	Interest rates	19	Interest rates on deposits in the national currency	R_{DG}	$R_{DG} \rightarrow \max$
		20	Interest rates on deposits in US dollars	R_{DD}	$R_{DD} \rightarrow \min$
		21	NBU Key Policy Rate	P	$P \rightarrow \max$
		22	Ukrainian Overnight Index Average (UONIA)	UONIA	UONIA$\rightarrow$max
6	Stock Market	23	UX index	UX	$UX \rightarrow \max$
		24	Dollar index	DX	$DX \rightarrow \min$

3. *The indicators in Table 2 were measured* on June 1, 2021. – November 1, 2022 according to the sources and frequency (discreteness) specified in Table 1. The result is a multidimensional Big Data array of a statistical training sample of 24 indicators of 156 values, 5 (weekly monitoring) – of 36 values. Technological

efficiency of further processing processes is ensured by saving the received data segment in the * format. xlsx file.

4. *The statistical processing* of indicators / parameters is implemented in the sequence of classical stages of statistical training: determination of statistical characteristics, construction of a trend line, forecasting. To increase the effectiveness of statistical training, a hierarchy of interconnected alternative and innovative stages is proposed (see the structural diagram in Figure). The structural scheme takes into account the features of multi-factor forecasting of statistical trends for Data Science tasks.



Structural diagram of the multi-factor forecasting process of statistical trends for Data Science tasks

The data obtaining (block 1 of the diagram, Figure) is implemented quickly from external sources using Web Scraping technologies.

Determination of the statistical characteristics of the obtained samples (block 2) is carried out a posteriori in the format of calculation: expected value, dispersion, standard deviation (SD), construction of a histogram of the law of distribution of the obtained samples. At the same time, the presence of a trend line is taken into account, which is removed using the Least Square Method (LSM) with a polynomial regression model [4].

Block 3 is intended for cleaning the statistical sample from anomalies. The use of three algorithms for detecting and cleaning anomalies [15] increases the reliability of the implementation of this process. Depending on the number of anomalies, the strategy of rejecting them is used (up to 10% of anomalous measurements – empirically obtained limits) and the recovery strategy (in other cases).

Optimizing the selection of the order of the trend line model (block 4) [14] is implemented with the control of the values of three indicators, which also increases the reliability of the final decisions.

The global linear deviation of the estimate is one that compares across multiple options:

$$\Delta = \frac{1}{n-1}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|.$$

The accuracy of approximation R^2 (coefficient of determination) varies within 0...1 and should be minimal:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2},$$

where n is a sample size; $\bar{y}_i = \frac{1}{n}\sum_{i=1}^{n} y_i$, y_i is a measured value; $\hat{y}_i$ is LSM of estimating the measured value.

The derivatives of the higher orders are the controls of obtaining small values:

$$y_j^{(p)} = \frac{y_{j+1}^{(p-1)} - y_j^{(p-1)}}{\Delta t}, \quad j = \overline{1...m}, \quad p = \overline{1...n}.$$

Determination of the trend line and forecasting (blocks 5, 6) is carried out using the algorithm of the least squares method (LSM) in classical polynomial [3; 4] or R&D nonlinear forms [4; 5].

For the presented research results, a nonlinear in parameters – transcendental model was chosen

$$f(t,c) = a_0 \cos\omega t + b_0 \sin\omega t,$$

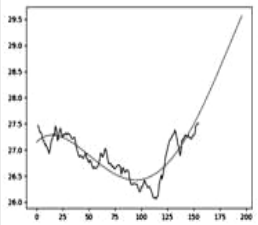
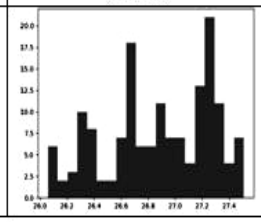
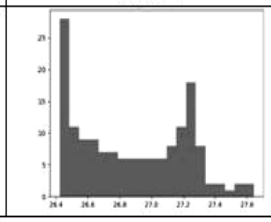
where $c = \{a_0, b_0, \omega\}$ are the unknown parameters of the model. The procedure for determining the parameters of a nonlinear model consistent with the measured values is discussed in detail in [4; 5].

The calculation of the integrated assessment (unit 7) of the effect of factors on the controlled parameter — the exchange rate is carried out according to the scheme of multi-criteria / multi-factor assessment (SCOR) according to the nonlinear scheme of compromises [16]. The data format is a multidimensional discrete set of functions of 24 indicators.

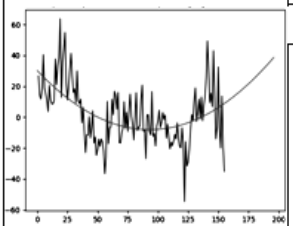
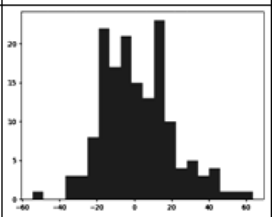
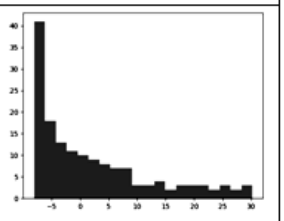
According to the structural diagram of Fig. 1, an alpha version of the computing unit (Backend component) of the ERP system layout was created to support currency trading processes. The software component is implemented in the high-level python programming language with the use of technologies and libraries: Web Scraping, pandas — for obtaining data; numpy — for “raw” programming of data processing algorithms; matplotlib — for visualization of calculation results.

THE RESULTS OF THE CALCULATIONS AND THEIR ANALYSIS

1. The official exchange rate of the hryvnia against the US dollar

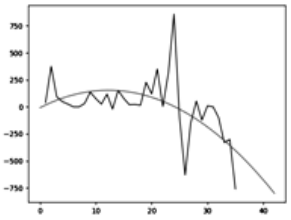
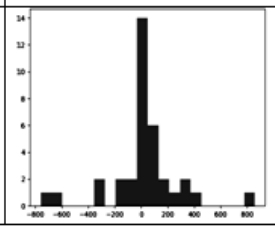
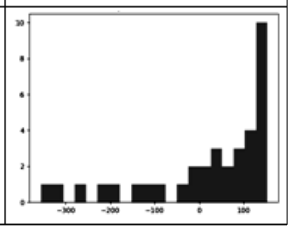
Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 4th order) Coefficient of determination R^2=0,85
	Expected value	26.875997	26.875997
	Dispersion	0.15288	0.11475
	Standard deviation	0.39100	0.33875
	Distribution law		

2. Saldo of operations of physical persons on the sale/purchase of foreign currency

Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 2nd order) Coefficient of determination R^2=0,67
	Expected value	15.352169	7.847362
	Dispersion	137.10981	44.97901
	Standard deviation	11.70939	6.70664
	Distribution law		

The indicator is calculated as the difference between the sale of foreign currency and its purchase.

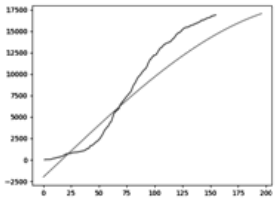
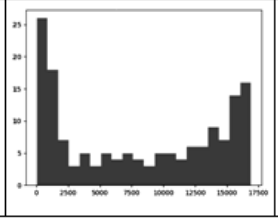
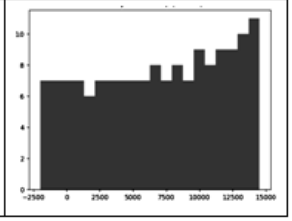
3. Saldo of NBU interventions

Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 2nd order) Coefficient of determination R^2=0,68
	Expected value	163.34	120.4238
	Dispersion	44296.145	6206.705
	Standard deviation	210.4665	78.7826
	Distribution law		

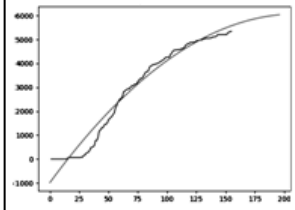
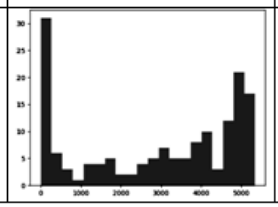
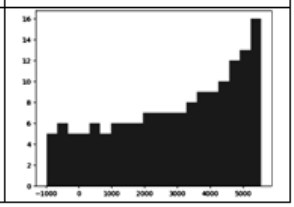
The indicator is calculated as the difference between the purchase of US dollars and their sale.

The volume of the main agricultural products of Ukrainian exports (indicators 4, 5, 6, 7) was calculated as the total volume of exported products, starting from June 1, 2021 (the beginning of the study)).

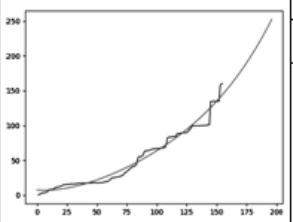
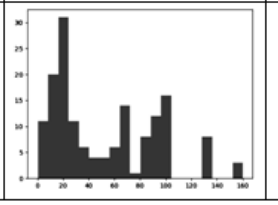
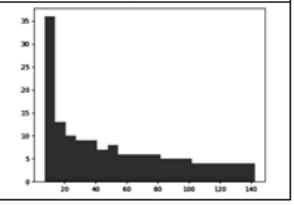
4. Wheat export volume

Chart	Statistical characteristics	Input stream	Estimation (non-linear smoothing) Coefficient of determination R^2=0,91
	Expected value	8045.29677	7103.56007
	Dispersion	37699035.022	20847045.489
	Standard deviation	6139.954	4565.856
	Distribution law		

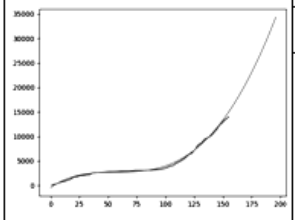
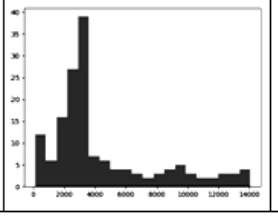
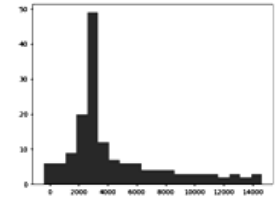
5. Barley export volume

Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 2nd order) Coefficient of determination R^2=0,97
	Expected value	2908.36129	3010.88877
	Dispersion	3804101.947	3088789.380
	Standard deviation	1950.411	1757.495
	Distribution law		

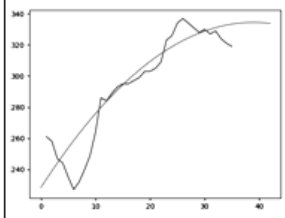
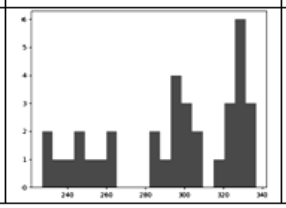
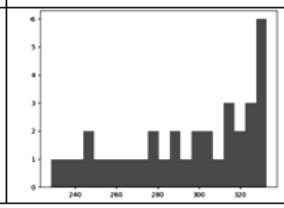
6. Rye export volume

Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 4th order) Coefficient of determination R^2=0,98
	Expected value	51.96645	51.96645
	Dispersion	1650.4386	1605.4167
	Standard deviation	40.6256	40.0676
	Distribution law		

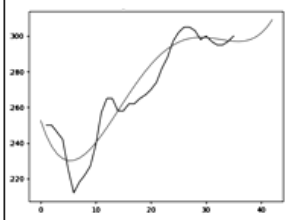
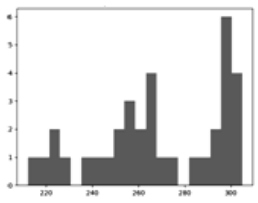
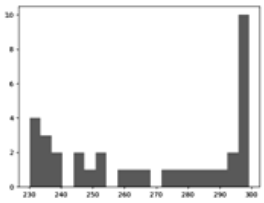
7. Corn export volume

Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 4th order) Coefficient of determination R^2=0,99
	Expected value	4433.81935	4444.55952
	Dispersion	11837263.3093	11682804.9816
	Standard deviation	3440.5324	3418.0118
	Distribution law		

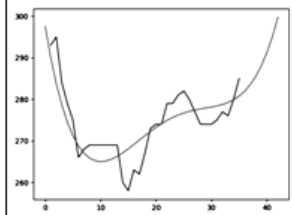
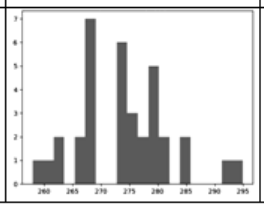
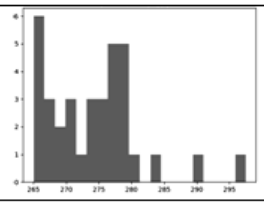
8. The export price of wheat

Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 2nd order) Coefficient of determination R^2=0,90
	Expected value	293.657143	293.657143
	Dispersion	1165.19673	1002.00163
	Standard deviation	34.13498	31.65441
	Distribution law		

9. The export price of barley

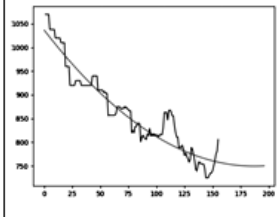
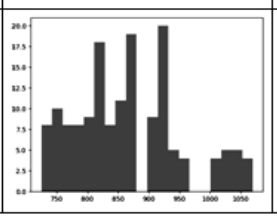
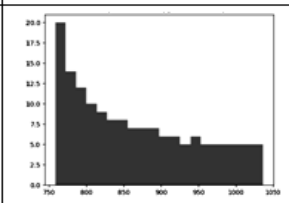
Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 4th order) Coefficient of determination R^2=0,93
	Expected value	268.514286	268.514286
	Dispersion	772.19265	701.16375
	Standard deviation	27.78835	26.47950
	Distribution law		

10. The export price of corn

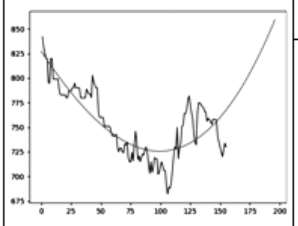
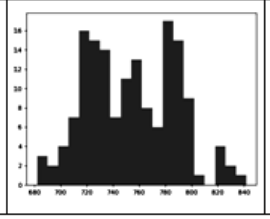
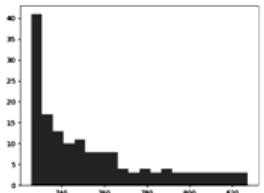
Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 4th order) Coefficient of determination R^2=0,85
	Expected value	274.228571	274.228571
	Dispersion	66.34775	50.32450
	Standard deviation	8.14541	7.09398
	Distribution law		

The export prices for all key agricultural products of Ukraine currently have a positive trend.

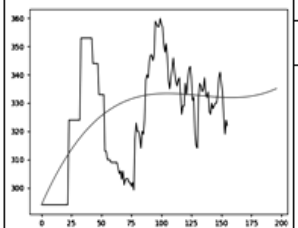
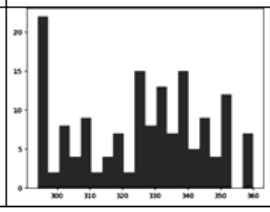
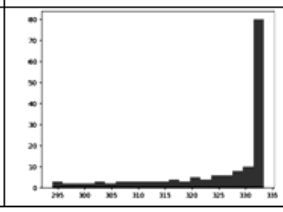
11. Export price of hot rolled steel

Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 4th order) Coefficient of determination R^2=0,93
	Expected value	865.485806	865.485806
	Dispersion	7400.11969	6743.01622
	Standard deviation	86.02395	82.11587
	Distribution law		

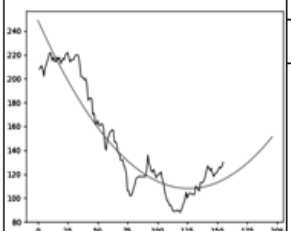
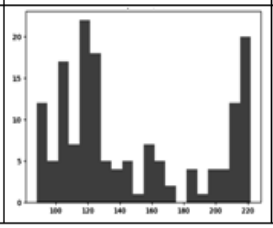
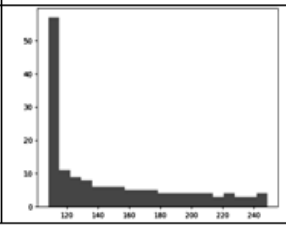
12. The export price of armature

Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 3rd order) Coefficient of determination R^2=0,84
	Expected value	753.474387	753.474387
	Dispersion	1145.28174	785.42695
	Standard deviation	33.84201	28.02547
	Distribution law		

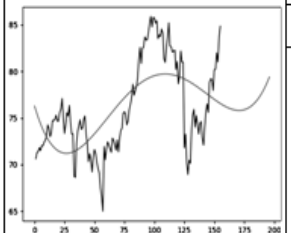
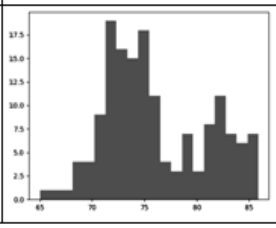
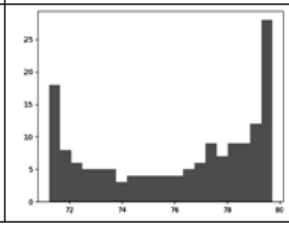
13. Export price of scrap metal

Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 3rd order) Coefficient of determination R^2 = 0,63
	Expected value	325.413419	325.413419
	Dispersion	380.88411	115.86365
	Standard deviation	19.51625	10.76400
	Distribution law		

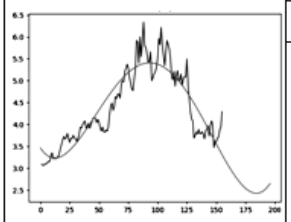
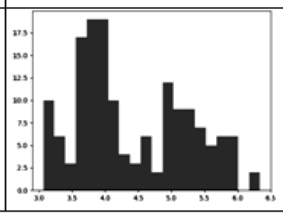
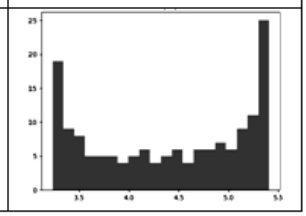
14. Export price for raw iron ore

Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 2nd order) Coefficient of determination R^2=0,92
	Expected value	147.064516	147.064516
	Dispersion	1994.07326	1766.70932
	Standard deviation	44.655048	42.03224
	Distribution law		

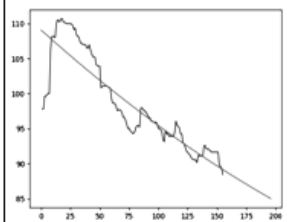
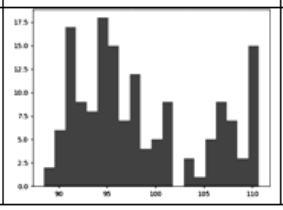
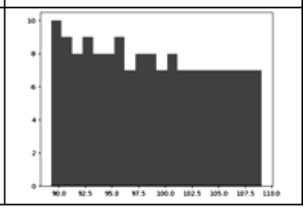
15. The global oil price

Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 4th order) Coefficient of determination R^2 = 0,68
	Expected value	76.080774	76.080774
	Dispersion	24.02192	8.99828
	Standard deviation	4.90122	2.99971
	Distribution law		

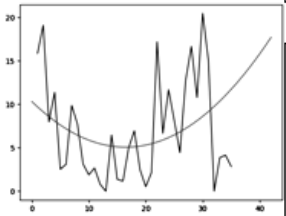
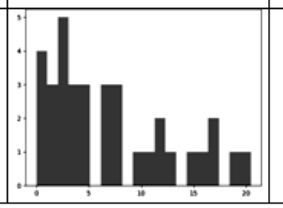
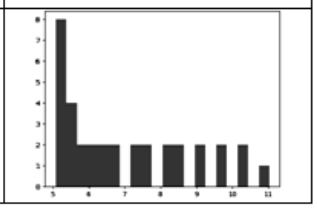
16. The global natural gas price

Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 4th order) Coefficient of determination R^2 = 0,88
	Expected value	4.398355	4.398355
	Dispersion	0.71180	0.58715
	Standard deviation	0.84368	0.76625
	Distribution law		

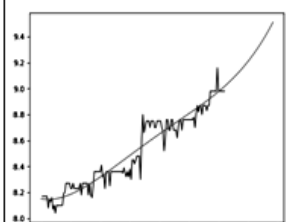
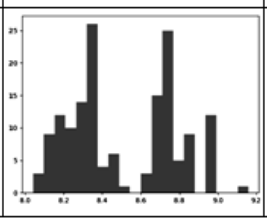
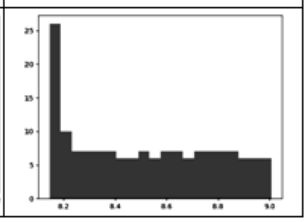
17. The volume of hryvnia government bonds in circulation at nominal and amortized cost with non-residents

Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 2nd order) Coefficient of determination R^2 = 0,87
	Expected value	98.620968	98.620968
	Dispersion	41.82890	33.12315
	Standard deviation	6.46752	5.75527
	Distribution law		

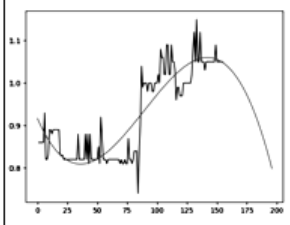
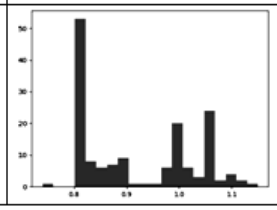
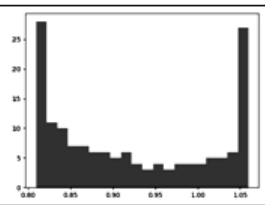
18. The amount of funds involved in the state budget for placement of domestic government bonds

Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 2nd order) Coefficient of determination R^2 = 0,52
	Expected value	7.054516	7.054516
	Dispersion	34.23411	3.17102
	Standard deviation	5.85099	1.78073
	Distribution law		

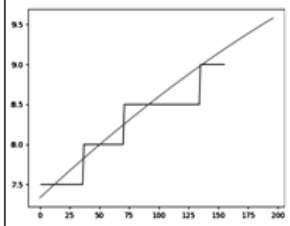
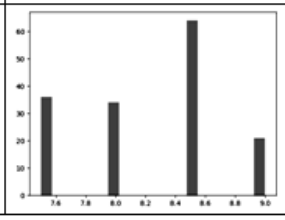
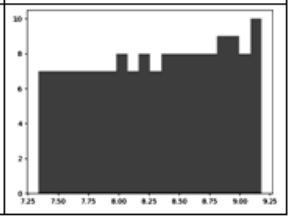
19. The interest rates on deposits in the national currency

Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 4th order) Coefficient of determination R^2 = 0,85
	Expected value	8.511806	8.511806
	Dispersion	0.07945	0.07362
	Standard deviation	0.28187	0.27133
	Distribution law		

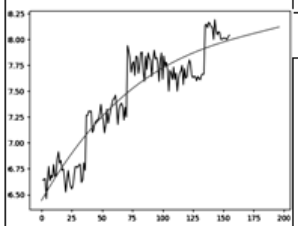
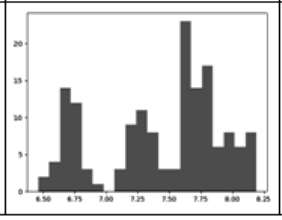
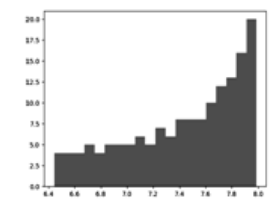
20. The interest rates on deposits in the USA dollars

Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 3rd order) Coefficient of determination R^2 = 0,78
	Expected value	0.923097	0.923097
	Dispersion	0.01054	0.00845
	Standard deviation	0.10268	0.09192
	Distribution law		

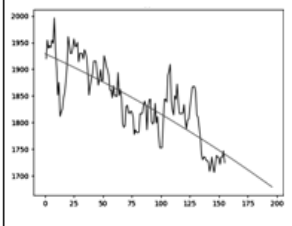
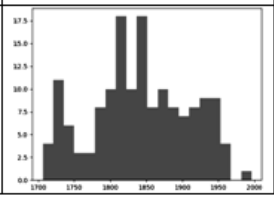
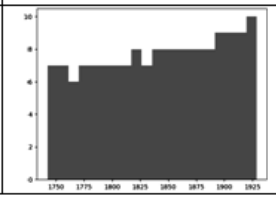
21. NBU Key Policy Rate

Chart	Statistical characteristics	Input stream	Estimation (non-linear smoothing) Coefficient of determination R^2 =0,90
	Expected value	8.225806	8.304335
	Dispersion	0.24579	0.28948
	Standard deviation	0.49577	0.53804
	Distribution law		

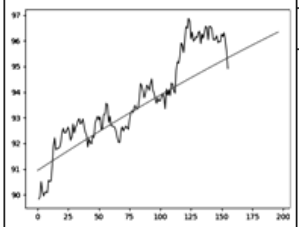
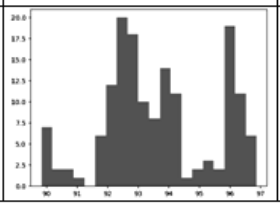
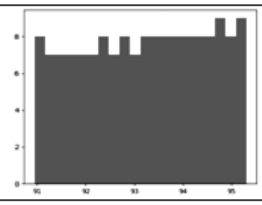
22. Ukrainian Overnight Index Average (UONIA.

Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 3rd order) Coefficient of determination R^2 = 0,90
	Expected value	7.432942	7.432942
	Dispersion	0.22817	0.19639
	Standard deviation	0.47768	0.44316
	Distribution law		

23. UX index

Chart	Statistical characteristics	Input stream	Estimation (polynomial of the 2nd order) Coefficient of determination R^2 = 0,79
	Expected value	1841.581226	1841.581226
	Dispersion	4564.20713	2951.12485
	Standard deviation	67.55892	54.32426
	Distribution law		

24. The dollar index

Chart	Statistical characteristics	Input stream	Estimation (non-linear smoothing) Coefficient of determination $R^2 = 0,83$
	Expected value	93.655419	93.204709
	Dispersion	3.17770	1.61243
	Standard deviation	1.78261	1.26981
	Distribution law		

CONCLUSIONS

The real data obtained and processed allow us to identify useful features. Statistical properties: parameters 1, 2, 3, 10, 11, 12, 23 (see histograms of distribution laws) are characterized by a normal distribution law, the others have combinatorial laws. This demonstrates the decomposition of the factors influencing the exchange rate into unitary and combinatorial components. Inherent natural presence of anomalous values of controlled parameters. The trend of the studied indicators is non-linear, and the dynamics of change may be conflicting according to the minimax analysis; that is, the improvement of certain indicators may be accompanied by the deterioration of others. So, once the infological model is formed, multifactorial consideration of the forecasting problem is appropriate. Further research will include the formation of an integrated indicator from partial factors and a comparison of its dynamics with the dominant effect, the exchange rate. At the same time, one should expect an increase in the accuracy and adequacy of predictive estimates of the studied parameters.

REFERENCES

1. Foster Provost and Tom Fawcett, *Data Science for Business*. Printed in the United States of America. Published by O'Reilly Media, Inc., 2013, 409 p.
2. David Dietrich, Barry Heller, and Beibei Yang, *Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting.* Indianapolis, Indiana: Data John Wiley & Sons, Inc., 2015, 420 p.
3. Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning Data Mining, Inference, and Prediction;* 2nd edition. Springer, 2020, 768 p.
4. S.V. Kovbasyuk, O.O. Pisarchuk, and M.Yu. Rakushev, *The least squares method and its practical application.* Zhytomyr: NAU, 2008, 228 p.
5. O.O. Pisarchuk, V.P. Kharchenko, *Nonlinear and multicriterial modeling of processes in traffic control systems*. K.: Institute of Gifted Child, 2015, 248 p.
6. F.O. Zhuravka, *Monetary policy in the context of transformational changes in Ukraine's economy: monograph*. Sumy: "Business Perspectives," Ukrainian Academy of Banking of the National Bank of Ukraine, 2008, pp. 63–123.
7. S. Kulitsky, "Dynamics of US dollar exchange in Ukraine in 2019: attempted situational analysis," *Ukraine: Events, Facts, Comments*, no. 11, pp. 35–46, 2019. Available: http://nbuviap.gov.ua/images/ukraine/2019/ukr11.pdf
8. *Ministry of Agrarian Policy and Food of Ukraine* [official site]. Available: https://minagro.gov.ua
9. *Ministry of Finance of Ukraine* [official site]. Available: https://mof.gov.ua
10. *National Bank of Ukraine* [official site]. Available: http://bank.gov.ua

11. I. M. Sysoyeva, "The enterprise's profit forecasting depending on accounting policy methods," *Economics and State*, no. 10, pp. 93–94, 2010. Available: http://nbuv.gov.ua/UJRN/ecde_2010_10_26
12. *Ukrainian Information Portal on Finance and Investments "Minfin.com.ua"*. Available: https://minfin.com.ua
13. A. Khivrenko, *Exchange rate: how it is determined, who influences it and what should be guided*. Available: https://www.epravda.com.ua/publications/2020/09/19/665288/
14. O.O. Pysarchuk, O.V. Korochkin, and D.R. Baran, "Determining the order of a polynomial model for constructing a trend line in Data Science problems," *Problems of Informatization and Management*, 3(71), pp. 35–40, 2022. doi: 10.18372/2073-4751.71.17001.
15. O. Pysarchuk, Yu. Mironov, I. Pysarchuk, and D. Baran, "Algorithms of Statistical Anomalies Clearing for Data Science Applications," *System Research & Information Technologies*, no. 1, pp. 78–84, 2023. doi: 10.20535/SRIT.2308-8893.2023.1.06.
16. O. Pysarchuk, A. Gizun, A. Dudnik, V. Griga, T. Domkiv, and S. Gnatyuk, "Bifurcation Prediction Method for the Emergence and Development Dynamics of Information Conflicts in Cybernetic Space," *Proceedings of the International Workshop on Cyber Hygiene (CybHyg-2019) co-located with 1st International Conference on Cyber Hygiene and Conflict Management in Global Information Networks (CyberConf 2019). Kyiv, Ukraine, November 30, 2019*, pp. 692–709. Available: https://ceur-ws.org/Vol-2654/paper54.pdf

Received 04.09.2023

INFORMATION ON THE ARTICLE

Oleksii O. Pysarchuk, ORCID: 0000-0001-5271-0248, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: PlatinumPA2212@gmail.com

Tetiana V. Andreieva, ORCID: 0009-0009-7033-9054, National Aviation University, Ukraine, e-mail: tetyanaandreieva@gmail.com

Olena O. Grinenko, ORCID: 0000-0001-9673-6626, National Aviation University, Ukraine, e-mail: gsa_ck@ukr.net

Danylo R. Baran, ORCID: 0000-0002-3251-8897, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: danil.baran15@gmail.com

БАГАТОФАКТОРНЕ ПРОГНОЗУВАННЯ СТАТИСТИЧНИХ ТРЕНДІВ ДЛЯ ЗАДАЧ DATA SCIENCE / О.О. Писарчук, Т.В. Андреєва, О.О. Гріненко, Д.Р. Баран

Анотація. Розглянуто процеси багатофакторного прогнозування статистичних трендів для задач Data Science. Більшість класичних підходів до оброблення даних полягають у дослідженні наслідків явищ, а не факторів їх появи. При цьому фактори, що впливають на поведінку досліджуваного процесу, вважаються випадковими та не досліджуються. Розглянуто підхід до прогнозування параметрів тренду статистичних часових рядів, який полягає в дослідженні факторів, що призводять до зміни динаміки досліджуваного процесу. Такий підхід потенційно має кращі показники адекватності, точності і оперативності отримання кінцевих рішень порівняно з класичними підходами. Наведено реалізацію цього підходу на прикладі аналізу зміни курсу валют. Отримані результати розрахунків показують доцільність розгляду багатофакторності у задачах прогнозування.

Ключові слова: Data Science, багатофакторне прогнозування, статистичні тренди, прогнозування курсу валют.