E3S Web of Conferences **548**, 03002 (2024) https://doi.org/10.1051/e3sconf/202454803002
AGRITECH-X 2024

Development of econometric models to forecast indicators of the livestock industry

Ekaterina Kasatkina[1,2*], *Daiana* Vavilova[1], and *Rinat* Faizullin[3]

[1]Kalashnikov Izhevsk State Technical University, 7 Studencheskaya street, Izhevsk, 426069, Russian Federation
[2]Autonomus non-commercial organization "Digital economy of Udmurtia Republic", 21 Lenina street, Izhevsk, 426004, Russian Federation
[3]MIREA – Russian Technological University, 78 Vernadsky Avenue, Moscow, 119454, Russian Federation

Abstract. The article discusses the importance of animal husbandry in ensuring food security and maintaining a high quality of life. In the current study, statistical monthly data on animal husbandry in the Udmurt Republic from 2018 to 2023 is analyzed to create models for forecasting key indicators: the average daily milk yield, the number of cows, and the total volume of milk production. The model of the average daily milk yield takes into account seasonal fluctuations, temperature, and time trends, with an average relative error of just 1.55%. The autoregressive model for predicting the number of cattle with a lag of 12 months has shown high accuracy with an average relative approximation error of 0.19%. The econometric model of total milk production takes into account the average daily milk yield and other factors, demonstrating high accuracy in its forecasts. These results are important to support decision-making on the development of animal husbandry and the agricultural sector in general.

1 Introduction

Livestock farming is one of the economic activities that guarantees the country's food security [1, 2]. It plays an important role in the self-sufficient provision of the population with high-calorie foods (milk, meat, eggs, etc.). The analysis of the indicators of the under-considered agricultural sector is of great importance. This analysis is required for forecasting, assessing prospects, and adjusting livestock development strategies. In addition, analysis and forecasting of development indicators for the livestock industry are necessary in order to solve the problems of the digitalization of agriculture, both for individual countries and for regions. It explains the relevance of this research.

Currently, many papers are devoted to solving the problem of predicting the development indicators of the livestock industry. In particular, researchers from Asia [3-5] solve the problem of forecasting sustainable milk production for Indian regions. They select the most accurate mathematical models, and in their article, they give preference to integrated autoregressive and moving average (ARIMA) models.

* Corresponding author: e.v.trushkova@gmail.com

Similar econometric ARIMA models for predicting dairy farming indicators were proposed by scientists from Montenegro [6]. They analyze and predict the volumes of cow's milk that will be collected and processed into drinking milk and fermented milk products.

The authors' study [7] modeled and forecast milk production in Nigeria using various variants of autoregressive and moving average (ARMA) models. The indicated ARMA models make it possible to identify the current stable trend in the development of milk production and make a forecast only on the basis of production values for previous periods.

We can cite a number of other articles by researchers [8-10], who also use econometric models to solve problems of forecasting agricultural industry indicators. But here, it is important to understand that in order to select the optimal forecasting model, it is necessary to have reliable information in sufficient quantities. And besides, take into account seasonality when solving the problem of forecasting the development indicators of a given industry.

The purpose of this research is to develop the most accurate econometric models for forecasting key indicators of livestock industry development that take into account both seasonal fluctuations of the modeled indicators themselves and the factors influencing them.

The following indicators are considered in the article as key indicators (the designation of the indicator is indicated in parentheses):

- average daily milk yield from one cow, kg (*M*);
- presence of cattle (cows), heads (*C*);
- gross milk production, tons (*Y*).

The construction of econometric models of the dynamics of key indicators is carried out using official statistical data from one of the regions of the Russian Federation, the Udmurt Republic. In terms of many indicators of agricultural development, the Udmurt Republic is a typical region of Russia and is characterized by average Russian values [11]. Therefore, the identified patterns and conclusions obtained during the study can be extrapolated, with some degree of reliability, to the Russian Federation as a whole.

The initial statistical data for constructing econometric models of dynamics (*M*), (*C*) and (*Y*) are the presented indicators broken down by month for the period from January 2018 to December 2023. The total sample size is 72 observations.

The authors proposed econometric forecast models that, in addition to information for previous time periods, take into account the influence of intermediate factors of production. The econometric modeling approach proposed in this article is original and new. It can be used to solve the problem of forecasting agricultural indicators in other regions and countries.

2 Forecasting model for average daily milk yield per cow

When constructing an econometric model, it is taken into account that the average daily milk yield per cow is influenced by factors such as seasonality, average monthly air temperature, and an established trend over time (trend). Taking into account these assumptions, the mathematical model for forecasting the average daily milk yield per cow (M) looks like this:

$$\begin{gathered}\widetilde{M} = a + bt + d_1 m_1 + d_2 m_2 + d_3 m_3 + d_4 m_4 + d_5 m_5 + d_6 m_6 + \\ + d_7 m_7 + d_8 m_8 + d_9 m_9 + d_{10} m_{10} + d_{11} m_{11} + cT,\end{gathered} \tag{1}$$

where t – the number of the time period (month number in order); $m_1,...,m_{11}$ – the dummy variables corresponding to months of the year $(m_1 - \text{September},...,m_{11} - \text{November})$, the dummy variable for December (m_{12}) is not included in the model due to multicollinearity [12]; T – the average monthly air temperature.

A model of type (1) is a multiple linear regression with dummy variables. The unknown parameters of model (1) are estimated using the classical least squares method [13].

To obtain a significant model, a method of sequential exclusion of variables is used, which consists in constructing a model with an exhaustive number of factors at the first stage, and at the next stages of econometric modeling, it is assumed that factors that are insignificant according to the student criterion are excluded from the model [14].

The result of a step-by-step test of the significance of the parameters of the econometric model (1) using the sequential elimination method is presented in Table 1.

Table 1. Construction of model (1) using the method of sequential exclusion of factors.

Model factor (parameter)	Stage 1		Stage 2		Stage 3	
	Parameter value	Student's t-statistic	Parameter value	Student's t-statistic	Parameter value	Student's t-statistic
Constant (a)	15.528*	44.93	15.40*	69.55	15.379*	76.27
Time (b)	0.088*	18.86	0.088*	19.10	0.088*	19.40
January (d_1)	0.434	1.51	0.480*	1.80	0.505*	2.06
February (d_2)	0.986*	3.44	1.041*	4.00	1.065*	4.47
March (d_3)	1.371*	4.32	1.471	6.15	1.493*	6.82
April (d_4)	1.774*	4.09	1.943*	7.48	1.962	8.02
May (d_5)	2.282*	3.78	2.524*	7.48	2.540*	7.79
June (d_6)	2.893	4.33	3.163*	8.47	3.177*	8.75
July (d_7)	1.719*	2.37	2.011*	4.95	2.025*	5.10
August (d_8)	1.345*	1.98	1.615*	4.16	1.629*	4.31
September (d_9)	0.344	0.64	0.552*	1.75	0.570*	1.88
October (d_{10})	-0.22	-0.49	–	–	–	–
November (d_{11})	-0.163	-0.49	-0.066	-0.25	–	–
Average monthly temperature (c)	-0.05*	-2.20	-0.059*	-3.88	-0.059*	-3.93

* The model parameter is significant at the significance level α = 0,05.

Table 1 shows that at the first stage of the method of sequential exclusion of factors, the average daily milk yield per cow in October differs slightly from the average daily milk yield in December. Therefore, at the second stage, the variable corresponding to the month of October is excluded from the model. At the same time, at the second stage, the average daily milk yield per cow differs slightly in November from the average daily milk yield in December and October, therefore, at the third stage, the variable corresponding to November is excluded from the model. All parameters of the model obtained at stage 3 are significant.

Next, we evaluate the significance of the econometric model as a whole using the Fisher criterion [15], having previously calculated the coefficient of determination of the model:

$$R^2 = 1 - \frac{\sum_{t=1}^{n}\left(M_t - \widetilde{M}_t\right)^2}{\sum_{t=1}^{n}\left(M_t - \overline{M}\right)^2}, \tag{2}$$

where R^2 – the coefficient of determination; M_t , $\widetilde{M}_t$, $\overline{M}$ – actual, model and average value of the average daily milk yield from 1 cow in the t-th month, respectively; n – the number of observations.

The coefficient of determination for the model obtained at the third stage R^2 is 0.951. It shows that 95.1% of the variation in average daily milk yield depends on the variation of factors taken into account in the model. According to the Fisher criterion, the mathematical model is significant, therefore, it can be argued that it can be used to predict the average daily milk yield from one cow in the future.

To assess the adequacy and applicability of the model for forecasting, the average relative error of approximation is calculated [16]:

$$\overline{A} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{M_i - \widetilde{M}_i}{M_i}\right| \cdot 100\%. \tag{3}$$

The average relative error of approximation for the model of average daily milk yield is 1.55%, which indicates the high accuracy and suitability of the model for forecasting.

Figure 1 presents actual data on average daily milk yield per cow and model values for the period of known data, as well as forecast values until December 2024.

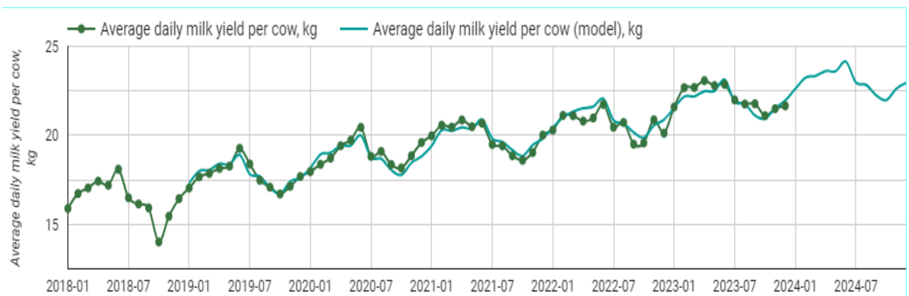

Fig. 1. Dynamics of real, model, and forecast values of average daily milk yield per cow in the Udmurt Republic.

3 Cow population forecast model

When constructing a mathematical model, it is taken into account that the number of cows (C) is influenced by the presence of cows in previous periods of time, therefore, an autoregressive process of the form is used:

$$C_t = \alpha_0 + \beta t + \alpha_1 C_{t-1} + \alpha_2 C_{t-2} + ... + \alpha_p C_{t-p}, \tag{4}$$

where $\alpha_j, j = \overline{0, p}$ –the parameters of the autoregressive model; β –trend component parameter; $C_{t-j}, j = \overline{1, p}$ – values of the number of cows for previous time periods; p – the order of the autoregressive process.

An autocorrelation function is constructed to establish the order of the autoregressive process p, based on the analysis of which the maximum significant time lag is selected (this is the order of the autoregressive process [17]). As a result of the analysis of the autocorrelation function (Fig. 2), it is found that in the cow population forecast model, it is advisable to use a lag of 12, i.e. $p = 12$.

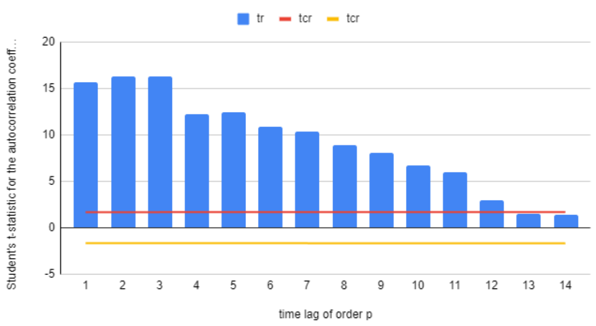

Fig. 2. Autocorrelation function of a time series characterizing the cow population and a graph of student statistics values.

As a result of assessing model (4) using the maximum likelihood method [18], a statistically significant model is obtained, for which the coefficient of determination shows that 93.2% of the variation in the number of cows depends on the variation of factors taken into account in the model. The livestock model is suitable for forecasting cow numbers since the average relative error of the approximation of this model is only 0.19%.

4 Forecasting model for gross milk production in the region

When constructing an econometric model of gross milk production in the region, it is taken into account that it is influenced by factors such as the average daily milk yield per cow (*M*), the number of cows in the region (*C*), and the number of days in a month (*D*). The following multiplicative model is proposed for forecasting gross milk production:

$$\widetilde{Y} = C^{\gamma_1}(DM)^{\gamma_2}, \tag{5}$$

where γ_1, γ_2 – the parameters of the linear model, estimated by the least squares method, are obtained as a result of the logarithm of equation (5):

$$\widetilde{y} = \ln \widetilde{Y} = \gamma_1 \ln C + \gamma_2 \ln(DM). \tag{6}$$

As a result of estimating the parameters of model (6), a model of gross milk production is obtained, the coefficient of determination of which is equal to 0.962. Consequently, 96.2% of the variation in gross milk production depends on the variation of factors taken into account in the model. Model (6) for predicting gross milk production is generally statistically significant according to Fisher's test. Its average relative error of approximation is 1.22%, which indicates high accuracy and that the model can be used to predict gross milk production.

Table 2 presents the results of forecasting the average daily milk yield, the number of cows, and the results of forecasting gross milk production in the Udmurt Republic for the short term (until December 2024).

Table 2. Forecast results of the livestock industry key indicators in the Udmurt Republic for the period until December 2024.

Month and year forecast	**Average daily milk yield, kg *M***	**Cows population, heads *C***	**Number of days in a month, days *D***	**Forecast values gross milk production, tons *Y***
01.2024	22.62	122 478	31	84 954
02.2024	23.24	122 425	29	81 463
03.2024	23.34	122 592	31	87 838

04.2024	23.62	122 683	30	85 951
05.2024	23.61	122 854	31	88 981
06.2024	24.15	122 806	30	88 019
07.2024	22.97	123 090	31	86 503
08.2024	22.84	123 198	31	86 016
09.2024	22.25	123 299	30	80 858
10.2024	21.98	123 419	31	82 661
11.2024	22.63	123 473	30	82 358
12.2024	22.96	123 614	31	86 603

The results of the forecast of key indicators of the livestock industry in the Udmurt Republic for the period until January 2025 in Table 2 indicate that in the Udmurt Republic, on average, in the forecast year 2024, the average daily milk yield will be 23.02 kg per cow, and the number of cows will be 122,994 heads. At the same time, gross milk production in total for the forecast year 2024 will amount to 1,022,205 tons.

Fig. 2 presents actual data, model values for the period of known data, as well as forecast values for the period until December 2024 for the number of cows and gross milk production.

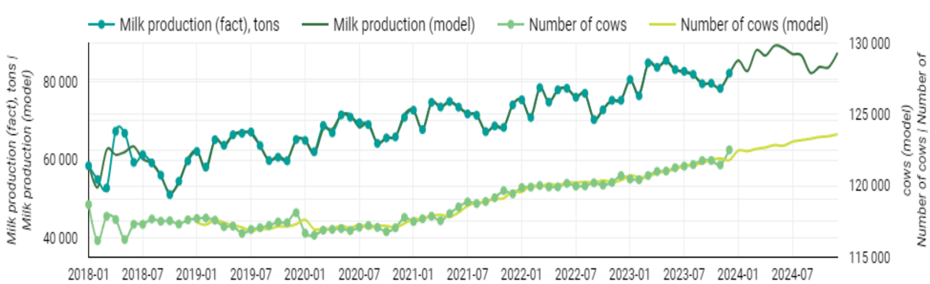

Fig. 3. Dynamics of the real model and forecast values of gross milk production and cows in the Udmurt Republic.

At the end of 2024, it is expected that gross milk production will amount to 1.022 million tons in the Udmurt Republic as a whole (50.8 kg per month per capita). The increase in output compared to 2023 will be 4.41%.

5 Conclusion

The research systematized statistical data on livestock production indicators in the Udmurt Republic for the period 2018-2023. It was required to develop econometric models for forecasting key indicators of the development of the livestock industry: average daily milk yield, number of cows, and gross milk production.

The econometric model of average daily milk yield was constructed taking into account seasonality factors, average monthly air temperature, and established trends over time. The average relative error of the model for forecasting the average daily milk yield was 1.55%, which indicates high accuracy.

The most suitable econometric model for forecasting the number of cows was an autoregressive model with a lag of 12. The constructed model is statistically significant: the average relative error of approximation was 0.19%.

An econometric model of gross milk production in the region is constructed taking into account the factors of average daily milk yield, the number of cows, and the number of days in a month. The average relative error of approximation of the described model was 1.22%, which indicates high accuracy and the possibility of use for forecasting gross milk production.

The development of econometric models for the dairy farming industry in the Udmurt Republic made it possible to obtain accurate forecasts of key indicators of the industry, which is important for making informed decisions on the development of the agricultural industry.

The proposed modeling approach can be used to predict livestock performance in other regions and countries, ensuring more efficient use of resources and identifying priority areas for industry development.

References

1. D.K. Kwaghtyo, C.I. Eke, Artificial Intelligence Review **56** (2023). https://www.doi.org/10.1007/s10462-022-10266-6
2. S. Mourtzinis, P.D. Esker, J.E. Specht, Scientific Reports **11** (2021). https://www.doi.org/10.1038/s41598-021-97380-7
3. S.S. Deshmukh, R. Paramasivam, Asian Journal of Dairy and Food Research **35** (2016). https://www.doi.org/10.18805/ajdfr.v35i1.9246
4. S.T. Pavan Kumar, K. Milind, K. Hanamaraddi, L. Biswajit, Indian Journal of Dairy Science **76** (2023). https://www.doi.org/10.33785/IJDS.2023.v76i05.011
5. S. Shankar Vishnu, R. Ajaykumar, S. Ananthakrishnan, A. Aravinthkumar, Asian Journal of Dairy and Food Research **42** (2023). https://www.doi.org/10.18805/ajdfr.DR-2103
6. M. Joksimovic, S. Ivanovic, S. Jankovic, Ekonomika poljoprivrede **67**, 2 (2020). https://www.doi.org/10.5937/ekoPolj2002391J
7. T.K. Samson, IOP Conf. Ser.: Earth Environ. Sci. **1219** (2023). https://www.doi.org/10.1088/1755-1315/1219/1/012026
8. Y.S Amirkhalili, A. Aghsami, F. Jolai, International Journal of Hybrid Information Technology **13** (2020). https://www.doi.org/10.21742/ijhit.2020.13.1.02
9. P. Mishra, Ch. Fatih, H.K. Niranjan, Sh. Tiwari, M. Devi, An. Dubey, Indian Journal of Animal Research **54** (2020). https://www.doi.org/10.18805/ijar.B-3918
10. S. Mohapatra, R. Sendhil, Aj. Singh, A.K. Dixit, R. Malhotra, K. Ponnusamy, Indian Journal of Dairy Science **74** (2021). https://www.doi.org/10.33785/IJDS.2021.v74i02.009
11. K.V. Ketova, E.V.Kasatkina, D.D. Vavilova, Journal of Physics: Conference Series **1889**, 3, (2021). https://www.doi.org/10.1088/1742-6596/1889/3/032029
12. R.V. Faizullin, S. Hering, IOP Conference Series: Materials Science and Engineering **971** (2020). https://www.doi.org/10.1088/1757-899X/971/4/042075
13. C. Zhang, C. Hu, F. Tang et al., Acta Geod Geophys (2024). https://www.doi.org/10.1007/s40328-023-00432-9
14. Y. Ding, L. Wang, Y. Li, D. Li, Computers and Electronics in Agriculture **151** (2018). https://www.doi.org/10.1016/j.compag.2018.06.004
15. A.K. Dubey, A. Kumar, A.K. Sharma, Sustainable Energy Technologies and Assessments **47** (2021). https://www.doi.org/10.1016/j.seta.2021.101474
16. E.V. Kasatkina, D.D. Vavilova, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **13217** (2022). https://www.doi.org/10.1007/978-3-031-16500-9_20

17. A. Hosovsky, J. Pite, M. Adamek, Journal of Building Engineering **34** 101955 (2021). https://www.doi.org/10.1016/j.jobe.2020.101955
18. M.A. Amin, M.A. Hoque, *Comparison of ARIMA and SVM for Short-term Load Forecasting*, in Proceedings of the 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference, IEMECON, 13-15 March 2019, Jaipur, India (2019). https://www.doi.org/10.1109/IEMECONX.2019.8877077