Forecasting potato yield dynamics in the Tashkent region of the Republic of Uzbekistan

V. V. Vakhabov*, and *M. A.* Hidoyatova
Tashkent Institute of Irrigation and Agricultural Mechanization Engineers" National Research University, Tashkent, Uzbekistan

Abstract. Observations of some phenomenon, the nature of which changes over time, generates an ordered sequence, which is called a time series. The statistical regularity of the series of dynamics $\overline{y}_t$ - the average potato yield in the Tashkent region of the Republic of Uzbekistan (based on the materials of the CSO of the Republic of Uzbekistan for 2000-2018 year) was studied by the method of statistical analysis of time series. Point and interval estimates for the average potato yield were built with a 95% guarantee, explicit types of trends were identified and the yield in the Tashkent region was predicted for subsequent years. Using the Darbin-Watson statistical criteria, it was found that the average potato yield in the region has autocorrelation dependence.

1 Introduction

In each area there are phenomena that need to be studied in their development and change over time. For example, to try to predict the future based on knowledge of the past, to manage the process, to describe the characteristic features of a series based on a limited amount of information. When processing time series, the methods rely largely on the methods developed by mathematical statistics for distribution series. To date, statistics has a variety of methods for analyzing time series.

In general, the study of the yield of agricultural processes as a discrete dynamic series and the prediction of their yield based on experimental data play an important role in determining the economic efficiency of farms.

In general, the time series $\{y_t,\ t \in T\}$ consists of four components: trend; fluctuations relative to trend; seasonality effect; random component. The works of Anderson[1], Kendal [2], Lewis [3], Brillinger [4], Chetyrkin [5] and others are devoted to the study and analysis of dynamic series.

2 Results and Discussion

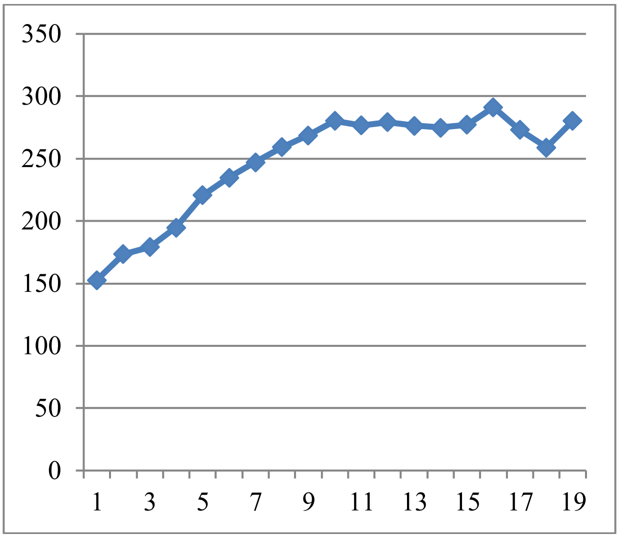
The geometric image of the observed data (Table1) on a rectangular coordinate system gives grounds, in the first approximation, to assume the hypothesis that the trend part of the process under study has a linear dependence $y(t) = a_1 t + a_0$ (see Fig.1).

*Corresponding author: v.vaxobov2019@gmail.com

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

Where the parameters are determined by the least squares method, i.e. based on the observed experimental data, solving the following (1) system of normal equations:

$$\begin{cases} a_0 T + a_1 \sum t = \sum y_t \\ a_0 \sum t + a_1 \sum t^2 = \sum y_1 t \end{cases} \quad (1)$$



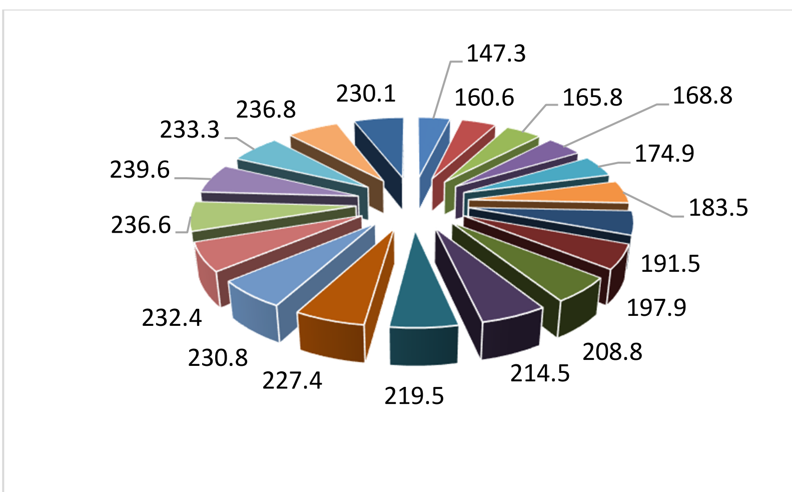


Fig. 1. Polygon distribution

Using the calculation according to Table1, we have

$$\sum Y_t = 3900.1\,, \quad a_0 = \frac{1}{T}\sum Y_t = \frac{3900.1}{19} = 205.67, \quad a_1 = \frac{2943.8}{570} = 5.17.$$

From here, we find the equation of the linear trend (trend) harvest- potato news:

$$y(t) = 5.17t + 205.67 \quad (2)$$

Using statistical criteria ([1-5]), it was found that in equation (2) $y(t) = a_1 t + a_0$ the main hypothesis $H_0: a_1 = 0$ is rejected and the alternative hypothesis $H_1: a_1 \neq 0$ is accepted significance level α=0.05.

From these equations (2) substituting the value $t = 3$ we find the expected forecast potato yield in the region in 2021 will be on average $Y(3) = 221.18\, c/ha$.

For further research, it is necessary to calculate the following finite differences. Denote

$$\Delta Y_t = Y_{t+1} - Y_t\,,\ \ \Delta^2 Y_t = \Delta Y_{t+1} - \Delta Y_{t'},\ \ \Delta^3 Y_t = \Delta^2 Y_{t+1} - \Delta^2 Y_t$$

Table 1. To calculate data for determining the trend of a time series

t	Year	Y_t c/ha	t	t^2	$Y_t \cdot t$	$Y_t \cdot t^2$
1	2000	147.3	-9	81	-1325.7	21697.29
2	2001	160.6	-8	64	-1284.8	25792.36
3	2002	165.8	-7	49	-1160.6	27489.64
4	2003	168.8	-6	36	-1012.8	28493.44
5	2004	174.9	-5	25	-874.5	30590.01
6	2005	183.5	-4	16	-734	33672.25
7	2006	191.5	-3	9	-574.5	36672.25
8	2007	197.9	-2	4	-395.8	39164.41
9	2008	208.8	-1	1	-208.8	43597.44
10	2009	214.5	0	0	0	46010.25
11	2010	219.5	1	1	219.5	48180.25
12	2011	227.4	2	4	454.8	51710.76
13	2012	230.8	3	9	692.4	53268.64
14	2013	232.4	4	16	929.6	54009.76
15	2014	236.6	5	25	1183	55979.56
16	2015	239.6	6	36	1437.6	57408.16
17	2016	233.3	7	49	1633.1	54428.89
18	2017	236.8	8	64	1894.4	56074.24
19	2018	230.1	9	81	2070.9	52946.01
The amount		3900.1	0	570	2943.8	817185.6

Table 2. To the calculation of data to determine the final differences

Year sobserved.	$Y\_(t)$ c/ha	Y_t^2	ΔY_t	ΔY_t^2	$\Delta^2 Y_t$	$\Delta^2 Y_t^2$
2000	147.3	21697.3				
2001	160.6	25792.4	13.3	176.89		
2002	165.8	27489.6	5.2	27.04	-8.1	65.61
2003	168.8	28493.4	3	9	-2.2	4.84
2004	174.9	30590	6.1	37.21	3.1	9.61
2005	183.5	33672.3	8.6	73.96	2.5	6.25
2006	191.5	36672.3	8	64	-0.6	0.36
2007	197.9	39164.4	6.4	40.96	-1.6	2.56
2008	208.8	43597.4	10.9	118.81	4.5	20.25
2009	214.5	46010.3	5.7	32.49	-5.2	27.04
2010	219.5	48180.3	5	25	-0.7	0.49
2011	227.4	51710.8	7.9	62,41	2.9	8.41
2012	230.8	53268.6	3.4	11.56	-4.5	20.25
2013	232.4	54009.8	1.6	2.56	-1.8	3.24
2014	236.6	55979.6	4.2	17.64	2.6	6.76
2015	239.6	57408.2	3	9	-1.2	1.44
2016	233.3	54428.9	-6.3	39.69	-9.3	86.49
2017	236.8	56074.2	3.5	12,25	9.8	96.04
2018	230.1	52946	-6.7	44.89	-10.2	104.04
The amount	3900.1	817185.8	82.8	805.36	-20	463.68

According to Table2, we calculate

$$v_k = \frac{\sum_{t=k}^{T}(\Delta^k Y_t)^2}{(T-t)C_{2k}^k}$$

the coefficients of variation of the order differences and establish that their value$V_1 \approx V_2 \approx V_3$. Therefore, first-order finite differces eliminate the linear trend.

Let 's check the presence of autocorrelation in a number of potato yield dynamics using the Darbin - Watson criterion:

$$d = \frac{\sum_{t=1}^{T-1}(Y_{t+1}-Y_t)^2}{\sum_{t=1}^{T-1}Y_t^2} \quad (3)$$

Calculating $d_{observations} = 0.001$ by the formula (3), we compare them with$d_{krit} = 1.08$value. Since$d_{observations} = 0 < d_{krit} = 1.08$, therefore, potato yields have an autocorrelation dependence, where $Y_t = \rho Y_{t-1} + \varepsilon t$, $\rho = \mathrm{Cov}(Yt, Yt+1) = \mathrm{M}[(Y_t - \overline{y}_t)(Y_{t+1} - \overline{y}_t)]$. To calculate the autocorrelation coefficient, we use the formula (4) ([1]-[5]):
where:

$$R_L = \frac{\sum_{t=1}^{N-L} Y_t Y_{t+L} - \frac{\sum_{t=1}^{N-L} Y_t \sum_{t=L+1}^{N} Y_t}{N-L}}{\sqrt{\left[\sum_{t=1}^{N-L} Y_t^2 - \frac{\left(\sum_{t=1}^{N-L} Y_t\right)^2}{N-L}\right]\left[\sum_{t=L+1}^{N} Y_t^2 - \frac{\left(\sum_{t=L+1}^{N} Y_t\right)^2}{N-L}\right]}} \quad (4)$$

According to Table-3, using formula (4), we determine the value of the autocorrelation coefficients R_l: R_1, R_2 ,R_3,R_4,R_5, the difference between their values from zero, gives reason that there is a significant autocorrelation dependence between potato yields. Consequently, the yield of potatoes in the Tashkent region this year depends on the yield of previous years.

Table 3. To calculate data for determining autocorrelation indicators

Year observation	Y_t	$Y_t \cdot Y_{t+1}$	$Y_t \cdot Y_{t+2}$	$Y_t \cdot Y_{t+3}$	$Y_t \cdot Y_{t+4}$	$Y_t \cdot Y_{t+5}$
2000	147.3					
2001	160.6	23656.4				
2002	165.8	26627.5	24422.3			
2003	168.8	27987	27109.3	24864.2		
2004	174.9	29523.1	28998.4	28088.9	25762.8	
2005	183.5	32094.2	30974.8	30424.3	29470.1	27029.6
2006	191.5	35140.3	33493.4	32325.2	31750.7	30754.9
2007	197.9	37897.9	36314.7	34612.7	33405.5	32811.8
2008	208.8	41321.5	39985.2	38314.8	36519.1	35245.4
2009	214.5	44787.6	42449.6	41076.8	39360.8	37516.1
2010	219.5	47082.8	45831.6	43439.1	42034.3	40278.3
2011	227.4	49914.3	48777.3	47481.1	45002.5	43547.1
2012	230.8	52483.9	50660.6	49506.6	48191	45675.3
2013	232.4	53637.9	52847.8	51011.8	49849.8	48525.1
2014	236.6	54985.8	54607.3	53802.8	51933.7	50750.7
2015	239.6	56689.4	55683	55299.7	54485	52592.2
2016	233.3	55898.7	55198.8	54218.9	53845.6	53052.4
2017	236.8	55245.4	56737.3	56026.9	55032.3	54653.4
2018	230.1	54487.7	53682.3	55132	54441.7	53475.2
	3900.1	779461	737774	695626	651085	605908

Based on sample data, using the x7.2019 software package and Excel computers, numerical characteristics y_t-for the average potato yield of the region are calculated (Table 4):

Table 4. Numerical characteristics y_t-for the average potato yield of the region

Selective characteristics	Estimates sample characteristics
Average potato yield $\bar{y}_T cen/h$	205.67
Variance	923.24
The mean square deviation σ_T	30.38
Coefficientofvariation v (%)	14.77%
AsymmetryA_ς	-0.54
ExcessE_{K_ς}	1,19
Error average value$\bar{y}_T$,m_y	$m_y = \frac{\sigma_y}{\sqrt{n}} = 6.97$
Limit error m'_y	$m'_y = tm_y = 2.11 \cdot 6.97 = 14.71$
Error mean square deviationσ_T	$m_\sigma = \frac{\sigma}{\sqrt{2n}} = \frac{30.38}{6.16} = 4.93$
Interval estimates (95%) $\bar{y}_T \pm tm_y$	$\bar{y}_T \pm tm_y = 205.67 \pm 14.71, (190.96; 220.38)$c/ha
Statistical hypothesis testing $H_0 : P(\bar{y}_T < x) = \Phi_{a,\sigma}(x)$	95% guarantees hypothesisH_0 accepted

3 Conclusions

Based on the above statistical analyses of the dynamics average potato yield in the Tashkent region of the Republic of Uzbekistan with a reliability of γ= 0.95, the following conclusions can be drawn:

1) Point and interval statistical estimates for their sample characteristics are constructed;

2) The explicit type of trend is determined and its linearity is established;

3) The Darbin-Watson criterion established that autocorrelations in the considered series of dynamics have linear trends;

4) Using statistical criteria, it was found that average potato crop of the region forms a non-stationary time series.

References

1. Kendal M., Stuart A. “Multivariate statistical analysis and time series” M. “Nauka”, 1976, 736 p.
2. Lewis C.D. “Methods of forecasting economic indicators”, M. “FiS”, 1986, 136 p.
3. Ya.Ya.-F.Vaynu “Correlations of series of dynamics”, M. “Statistics”, 1977
4. Ayvazyan S.A. "Statistical study of dependencies". M., "Metallurgy", 1968, 227 p.
5. Achilov M.U., Fayziev A.A. “The analysis of dynamics of fruits and berry froductivity grown in Uzbekistan”, EPRA International journal of Research and Development. Volume: 4. Issue: 8. August 2019, 5-9 p.
6. Fayziev A.A., Vakhobov V. “Forecasting the dynamics of cotton yields in the Fergana region”, Tashkent Institute of Irrigation and Agricultural Mechanization. Journal “Irrigation and Melioration”, No. 6. 2020, 182 - 188 pages.

7. Abdullayev, A., Hidoyatova, M. Exact method to solve finite difference equations of linear heat transfer problems (2021) AIP Conference Proceedings, 2402, № 070021. DOI: 10.1063/5.0071430
8. Abdullayev, A., Kholturayev, K., Safarbayeva, N. Exact method to solve of linear heat transfer problems. E3S Web of Conferences (2021), 264, 02059
9. https://doi.org/10.1051/e3sconf/202126402059
10. Abdullayev, A., Zhuvanov, K., Ruzmetov, K. A generalized solution of a modified Cauchy problem of class R_2 for a hyperbolic equation of the second kind. Journal of Physics: Conference Series, (2021), 1889(2), 022121 DOI:10.1088/1742-6596/1889/2/022121
11. Yuldashev, T.K., Islomov, B.I., Abdullaev, A.A. On Solvability of a Poincare–Tricomi Type Problem for an Elliptic–Hyperbolic Equation of the Second Kind. Lobachevskii Journal of Mathematics, (2021), 42(3), pp. 663–675 DOI: 10.1134/S1995080221030239
12. Abdullayev, A.A., Ergashev, T.G. Poincare-tricomi problem for the equation of a mixed elliptico-hyperbolic type of second kind. Vestnik Tomskogo Gosudarstvennogo Universiteta, Matematikai Mekhanika, (2020), (65), pp. 5–21/ DOI 10.17223/19988621/65/1
13. Abdullaev, A., Hidoyatova, M. Innovative distance learning technologies. Journal of Critical Reviews, 2020, 7(11), pp. 337–339. DOI: 10.31838/jcr.07.11.57