Application of Chebyshev Polynomial in Predicting the Grain Yield——A Case of Grain Yield in Jilin Province

ZHANG Hong-qin[1,2*], **GAO Lai-bin**[1]

1. College of Information and Technology, Jilin Agricultural University, Changchun 130118, China; 2. School of Mathematics, Jilin University, Changchun 130021, China

Abstract On the basis of introducing the fundamental principles of the least square methods, the Chebyshev polynomial data fitting method is given, by using this method, the grain yield of Jilin Province from 1952 to 2008 is analyzed. The results show that when analyzing the research data of agricultural economy, the least square method of the Chebyshev polynomials is a good choice; by establishing the prediction model of the least square method of Chebyshev polynomials, we get the results that the grain yield in Jilin Province from 2009 to 2015 is 29.004 millon, 29.836 million, 30.681 million, 31.540 million, 32.412 million, 33.298 million, 34.197 million ton; the annual average growth rate of grain yield from 2009 to 2015 is 2.78%, lower than the annual growth rate of 7.12% from 2000 to 2008.

Key words The least square method, Normal equation, Chebyshev polynomial, China

In the research and production practice of agricultural economy, the data obtained from observation, prediction and calculation should be analyzed and processed, and then the conclusions concerning the problems can be found out. The polynomial least square method is an effective method, because the accuracy and reliability of the least square method relies on the condition number of the coefficient matrix of its normal equation. We adopt the Chebyshev polynomial as basis function, so as to make the small condition number of coefficient matrix of the normal equation of the least square method, and then the accuracy and reliability of the least square method are improved obviously. As a granary industry, Jilin Province is an important grain production base in China, and its grain production has the great significance to the food safety of China. Therefore, accurate prediction of the yield of grain in Jilin Province has great significance to the agricultural development of Jilin Province and the whole country. The research takes the data of grain yield from 1952 to 2008 as the base, adopts the least square fitting of Chebyshev polynomial and predicts the grain yield of Jilin Province from 2009 to 2015.

1 The fundamental principles of the least square method

1.1 The linear least square fitting of general polynomial[1]

we use $P_n(x) = span(1, x, \cdots, x^n)$ to represent the linear space spaned by $1, x, \cdots, x^n$, then to any $p_n(x) \in P_n(x)$, there are $a_0, a_1, \cdots, a_n$, which makes the following equation

$$p_n(x) = a_0 + a_1 x + \cdots + a_n x^n$$

Supposing there is a group of data $\{(x_i, y_i)\}_{i=1}^{N}$, in which the are different to each other, then finding the coefficients $a_0, a_1, \cdots, a_n$ which make the minimum value of the following equation:

$$E_2(a_0, a_1, \cdots, a_n) = \sum_{i=1}^{N} |y_i - p_n(x_i)|^2 = \sum_{i=1}^{N} [\sum_{j=0}^{N} a_j x_i^j - y_i]^2$$

The problem is called the least square, the functions $1, x, \cdots, x^n$ are the radix of the linear least square, $E_2(a_0, a_1, \cdots, a_n)$ is called the residual sum of squares. By mathematical analysis that the least square problems must have a solution and the solution is the linear equation.

$$SA = \begin{bmatrix} s_0 & s_1 & \cdots & s_n \\ s_1 & s_2 & \cdots & s_{n+1} \\ \vdots & \vdots & & \vdots \\ s_n & s_{n+1} & \cdots & s_{2n} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_n \end{bmatrix}, \ A = S^{-1} U$$

In the formula, $s_k = \sum_{i=1}^{N} x_i^k$, $u_k = \sum y_i x_i^k$. The linear equation is called the normal equation of the least square. So the accuracy of the linear least square relies on the accuracy of normal equation. It can be seen form the numerical analysis that when the condition number of the coefficient matrix of the linear equation is very big[2], the solution of the linear equation is unreliable. For example:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ 15 \\ 24 \end{bmatrix}$$

The genuine solution of the equation is (1, 1, 1), by the solution obtained by using the Matlab is (0, 3, 0), the condition number of the coefficient matrix of the linear equation is 1.55×10^{18}. It can be seen from the numerical analysis that, when $\{x_i\}_{i=1}^{N}$ is the consecutive natural numbers, the condition number of the coefficient matrix of the normal equation of general polynomial linear least square is big correspondingly.

1.2 The linear least square of Chebyshev polynomial[3]

The expression of the Chebyshev polynomial is that:

$$T_k(t) = \cos(k \arccos t) \quad (k \geqslant 0)$$

In the interval $[-1, 1]$, the Chebyshev polynomial can be obtained by the following recurrence relations:

$$T_{k+1}(t) = 2xT_k(t) - T_{k-1}(t) \quad k = 1, 2, \cdots, n$$

Received: November 1, 2010 Accepted: November 13, 2010
Supported by the Projects of Jilin Provincial Science & Technology Department (20080644).

* Corresponding author. E-mail: zhanghq_@126.com

$T_0(t)=1 \quad T_1(t)=t$

By using transformation, the Chebyshev polynomial in the interval of $[a, b]$ can be obtained:

$$t=\frac{2x-(a+b)}{b-a},\ x\in[a, b],\ t\in[-1, 1]$$

$$\tilde{T}k(x)=T_k(t)=T_k\frac{2x-(a+b)}{b-a}$$

Giving a group of data $\{(x_i, y_i)\}_{i=1}^{N}$, supposing $a=x_1<x_2<\cdots x_N=b$, selecting $\tilde{T}_k(x)$, $k=0, 1, \cdots, n$ as the basis function, solving c_0, c_1, $\cdots$, c_n making minimum value of $\varphi(x)=c_0+c_1\tilde{T}_1(x)+\cdots c_n\tilde{T}(x)$, $E_2(c_0, c_1, \cdots, c_n)=\sum_{i=1}^{N}|y_i-\varphi(x_i)|^2=\sum_{i=1}^{N}[\sum_{j=1}^{n}c_j\tilde{T}_j(x_i)-y_i]^2$, that is: $\tilde{T}_j=[\tilde{T}_j(x_1), \cdots \tilde{T}_j(x_N)]^T$, $j=0, 1, \cdots, n$, $y=[y_1, \cdots, y_N]^T$, $C=[c_0, \cdots, c_n]^T$. Then its normal equation of the least square is:

$$GC=b$$

In the formula, $G=[(\tilde{T}_j, \tilde{T}_k)]$, $(\tilde{T}_j, \tilde{T}_k)$ refers to the inner product of $\tilde{T}_j$ and $\tilde{T}_k$; $b=[(y, \tilde{T}_0), \cdots, (y, \tilde{T}_n)]^T$. The condition number of the coefficient matrix of this normal equation is very tiny, so the fitting accuracy and practicability is high.

2 The prediction of the grain yield of Jilin Province

2.1 Fitting function We conducts the least square fitting on the data of grain yield of Jilin Province from 1952 to 2008 and the grain yield of it from 2009 to 2015 can be predicted. The grain yield of Jilin Province from 1952 to 2008 is as Table 1.

Table 1 Grain yield of Jilin Province from 1952 to 2008 $\times 10^4$ t

Year	Yield	Year	Yield	Year	Yield	Year	Yield
1952	613.20	1967	647.74	1981	921.91	1995	1 992.40
1953	561.45	1968	622.15	1982	1 000.04	1996	2 326.60
1954	530.95	1969	498.70	1983	1 477.98	1997	1 808.30
1955	556.53	1970	738.80	1984	1 634.46	1998	2 506.00
1956	493.64	1971	713.05	1985	1 225.26	1999	2 305.60
1957	429.35	1972	556.99	1986	1 397.71	2000	1 638.00
1958	528.84	1973	783.00	1987	1 675.81	2001	1 953.40
1959	526.60	1974	858.15	1988	1 693.25	2002	2 214.80
1960	394.70	1975	906.50	1989	1 351.29	2003	2 259.60
1961	398.55	1976	755.50	1990	2 046.52	2004	2 510.00
1962	437.16	1977	728.35	1991	1 898.87	2005	2 581.21
1963	501.67	1978	914.70	1992	1 840.31	2006	2 720.00
1964	491.80	1979	903.34	1993	1 900.90	2007	2 453.80
1965	525.10	1980	859.60	1994	2 015.70	2008	2 840.00
1966	597.60						

By using the Matlab mathematic software, the fitting result of the polynomial least square is as follows:

$$p(x)=0.682\,3x^2-2.659\,2\times10^3x+2.591\,2\times10^6$$

The fitting result of Chebyshev polynomial least square is:

$$a=1\,952,\ b=2\,008,\ t=\frac{2x-(a+b)}{b-a},$$

$$\varphi(x)=1\,350.6+1\,200.6t+267.5(2t^2-1)$$

It can be seen from Fig.1 that the effect of the general polynomial least square is bad, but the fitting result of the Chebyshev polynomial least square is more accurate. By calculating, we get the condition number of the coefficient matrix of the linear least square of general polynomial is $4.888\,8\times10^{21}$, but that of the Chebyshev polynomial is just 3.432 6.

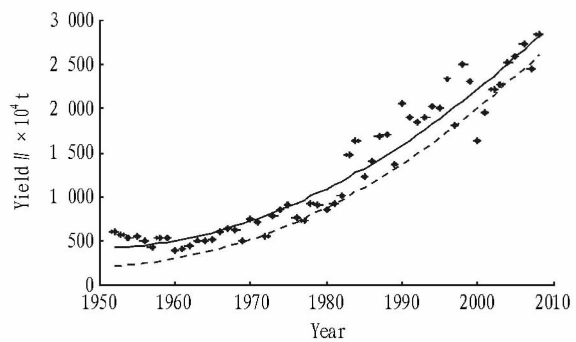

Fig.1 The actual grain yield and its fitting curve of Jilin Province from 1952 to 2008

2.2 Evaluation on the fitting results

2.2.1 Evaluation on the mean square error[4]. The formula used to evaluate the unbiased variance is:

$$MSE=\frac{1}{N}\sum_{i=1}^{N}e_i^2=\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2$$

In the formula, $\hat{y}_i$ is the fitting evaluation value of x_i. When using the unbiased variance to evaluate, the usually followed principle is the smaller the mean square error, the better the method used. The linear least square of general polynomial $MSE=8.233\,8\times10^4$, but the linear least square of the Chebyshev polynomial $MSE=3.815\,3\times10^4$.

2.2.2 Determination of the coefficient[5]. The formula of the determination coefficient is as follows:

$$R^2=1-\frac{\sum(y_i-\hat{y}_i)^2}{\sum(y_i-\bar{y})^2}$$

In the formula, $\bar{y}$ is the average value of y_i. When using the determination coefficient R^2 to evaluate, usually we follow the principle that R^2 closes to 1 is the best method. the linear least square of general polynomial is $R^2=0.853\,8$, but the linear least square Chebyshev polynomial is $R^2=0.932\,2$.

3 Conclusions and discussions

Firstly, when analyzing the data of agricultural economic research, the linear least square of the Chebyshev polynomial is the best method.

Secondly, through establishing the Chebyshev polynomial least square model of the grain yield in Jilin Province, the grain yield in Jilin Province from 2009 to 2015 can be predicted, which is 29.004 million ton, 29.836 million ton, 30.681 million ton, 31.540 million ton, 32.412 million ton, 33.298 million ton and 34.197 million ton respectively. According to the prediction result, the annual growth rate of grain of Jilin Province from 2009 to 2015 is 2.78%, lower than the annualized growth of 7.12% in the years from 2000 to 2008. This growth rate with the tendency from fast to slow is in accordance with the feature of agricultural production and the general rule of development, which provides scientific evidence for Jilin government to make the agricultural development policies.

Thirdly, the process of grain production is rather complex,

(To page 44)

market, increasing saturation of job opportunities, difficult working conditions of migrant workers and the existence of large amount of returning migrant workers. Therefore, most farmers transfer their land temporally.

3 Conclusions and suggestions

Through investigation, the factors such as the non-agricultural job opportunity, education degree and age of rural households, government policy and rural collective work, rural social security system, the condition of land transfer and some other factors all affect the willingness of rural land transfer.

Most farmers say that they will not transfer their land or only transfer small part of their land out as a result of the unhealthy social security system and unstable non-agricultural economic income. Some people have transferred out their land completely because they have completely qualified to free from agricultural production. That is to say, they have stable non-agricultural income. there are also some people who want to transfer in land, for they think agriculture is profitable and they do not have the ability to undertake other non-agricultural production (if they have the ability to undertake non-agricultural production and have stable income and the income is higher than undertaking agricultural production, they will not transfer land). In the village, there are surplus labors and their management skills allow them to expand the operation scale, but the current land situation can not meet the demand, which shows the inharmonious relations between people and land.

With the acceleration of urbanization, rural households' willingness on land transfer has changed slightly[11]. But at present, the quality of Chinese farmers is low and they lack the competitiveness in the rapidly developed secondary and tertiary industry, so it is hard for them to get stable job and non-agricultural income. Therefore, in the first place, the government should vigorously develop the rural education; improve farmers' quality and cultural level; develop the township enterprises; develop the secondary and tertiary industry and absorb rural surplus labors. In the second place, the government should stabilize the land system owned by the collective; further define the land property relations; further intensify the support on three agricultural problems concerning agriculture, countryside and farmers; develop rural economy; supply capital, technology, agricultural product processing and sales for the farmers throughout the process of agricultural production; provide favorable environment for rural land transfer and ensure and improve the economic interest of rural land transfer. In the third place, the government should regulate the behaviors of rural land transfer; strengthen the management on rural land transfer procedure; protect the legal rights of the two parties of land transfer. Last but not the least, the government should solve the rural social security problems concerning employment, medical care and old-age security well from the reality, and fundamentally solve farmers' worries on land transfer.

References

[1] DU PH, OU MH. A demonstrative study of factors influencing farmers' behavior in farmland transfer—with Jiangsu Province as an example [J]. Scientific and Technological Management of Land and Resources, 2008, 25(1): 53–56. (in Chinese).
[2] ZHAN HP. Empirical prospect of the rural land transfer in China [J]. Journal of Anhui Agricultural Sciences, 2007, 35(24): 7710–7711, 7713. (in Chinese).
[3] CHEN MQ, XIAO HL, LONG Y, *et al.* Research progress on households willingness for arable land circulation and driving force and prospect [J]. Chinese Journal of Agricultural Resources and Regional Planning, 2008, 29(2): 70–73. (in Chinese).
[4] WU C. Study on the actuality, problems and countermeasure of farmland transfer in Guangdong Province—based on the survey of 654 households in 10 cities of Guangdong Province [J]. Southern Rural, 2008(6): 33–36. (in Chinese).
[5] JIAO YL. Positive research of the peasant's land circulating will in traditional farming region in middle Shandong Province [J]. Journal of Shandong Agricultural University: Social Science Edition, 2005 (1): 82–86, 120. (in Chinese).
[6] DU WX, HUANG XJ. Regional difference and influencing factors of farm households' willingness of rural land transmission: a case study of Shanghai, Nanjing, Taizhou and Yangzhou Cities in Yangtze-Delta Region [J]. Resources Science, 2005(6): 90–94. (in Chinese).
[7] WANG JX. Transition and development of village-level economy—based on fieldwork of China and foreign countries [M]. Beijing: China Economic Publishing House, 2005. (in Chinese).
[8] TAO ZG. Village economy in the marketed process in China [M]. Taiyuan: Shanxi Economic Press, 2005: 54–63. (in Chinese).
[9] ZHANG ZX, ZHANG HY. Summarization of research on rural land leasing [J]. Journal of Xinyang Teachers College: Philosophy and Social Sciences Edition, 2004(2): 57–61. (in Chinese).
[10] WANG JX. Agricultural land system and reconstruction of the pattern of interests in the process of modernization [M]. Beijing: China Economic Publishing House, 2005. (in Chinese).
[11] QIAN WR. A positive study on the intensions and behaviors of farmers' households for land circulating in traditional grain growing areas in North Zhejiang [J]. Chinese Rural Economy, 2002(7): 60–69. (in Chinese).

(From page 38)

and during the process many factors may affect the grain production, for example, sown acreage, agricultural policy, agricultural infrastructure, climatic factor, natural disasters and so on. Besides, the influence factors vary according to the time periods. Therefore, there must be some errors if using grain yield in the past to predict future grain yield, and it is the problems of all the prediction models. So we should adjust the prediction model in due time, so as to reduce the error and realize the accurate prediction.

References

[1] ZHOU YS, LIANG XZ. Numerical analysis [M]. Beijing: Higher Education Press, 2008: 127–131. (in Chinese).
[2] LIN CS. Numerical analysis [M]. Beijing: Science Press, 2007: 197–201. (in Chinese).
[3] HEATH MT. Scientific computing [M]. Beijing: Tsinghua University Press, 2005: 59–61. (in Chinese).
[4] MAO SS, CHENG YM. Tutorial of probability and mathematical statistics [M]. Beijing: Higher Education Press, 2004: 418–419. (in Chinese).
[5] DING CF. Application of combination forecasting method in grain production forecast [J]. Research of Agricultural Modernization, 2007(1): 101–103. (in Chinese).