The Mathematical Statistics Theory Application on the Price Fluctuation Analysis

Jintao Meng[1], Jing Li[2]

[1] **Zhengzhou Institute of Aeronautical Industry Management**
Zhengzhou, 450015, China

[2] **Zhengzhou Railway Vocational & Technical College**
Zhengzhou, 450052, China

Abstract

Grain price and output fluctuation are the normal state of market economy. It is one of the most important economic researches to understand grain price and output fluctuation law, which provides theory basis for the macroeconomic regulation and control. According to the cobweb model theory, the relationship between citrus production and price is accord with the divergence type of cobweb model .This means that simply relying on market regulation can make fluctuation between production and price bigger, go against citrus production and cultivation, thus, affecting the interests of farmers. It is well-known most farmers are concerned about the future price trend and the probability of price fluctuation. This paper uses mathematical statistics theory to study the citrus price changes, and the corresponding change trend, providing a theoretical basis for majority of farmers to better estimate citrus price change trend.

Keywords: *cobweb theory, citrus price fluctuation, mathematical statistics, confidence interval.*

1. Introduction

Currently, more and more attention to the development of the agricultural economy. In such a good market background, it is the theme that the farmers expand production. Each region drastically increases corresponding capital investment. But the market is ruthless. Like roller coaster, citrus fruit price is elusive, and many large investments only get more debt, which becomes the burden of some farmers.

Combining with China Statistical Yearbook 2009, related data of citrus price and production released by the China Citrus Web in 2009, and the cobweb theory[2,3,4], we derive the relationship between citrus production and market price , see [5, 6]:

$$\begin{cases} D_t = 2850.02 - 527.96 P_t \\ S_t = 2935.9 - 546.72 P_{t-1} \end{cases}$$

Where D_t : the demand function, S_t : the supply function, P_t : the price at t time, P_{t-1} : the price at $t-1$ time.

According to the cobweb model theory, this conforms to the second case of the cobweb model. That is, when $|\mu|>|\beta|$, $\lim_{t\to+\infty} p_t$ does not exist, and is tend to be infinite. This shows that with the passage of time, the change in price range is bigger, and the actual price will be getting further away from the equilibrium one, see [1]. Only relying on market to regulate the citrus production would lead to its price and yield far from the equilibrium point, see [7]. The citrus price fluctuation every year may make a phenomenon of “low price hurting farmers " happen repeatedly.

The citrus farmers are most concerned about citrus price prediction and the corresponding possibility. Combining and using the mathematical statistics, the paper will analyze previous price data, and predict the citrus fruit prices and the corresponding probability to provide more scientific price prediction.

2. Mathematical Statistics Preliminaries

As a discipline born in the turn of 20th century, mathematical statistics is widely used as a branch of mathematics. Based on probability theory, mathematical statistics uses experimental and observational data to study random phenomenon, so as to make reasonable estimation and judgment on the objective law of the research object.

Because a large number of random phenomena will necessarily show its regularity, theoretically, observation of the random phenomena enough times will make the regularity of the research object clearly present. But in reality, people are often unable to observe all the research objects(or called overall), only observe or test some (or called the sample) to obtain limited data.

In mathematical statistics certain common objects are collectively referred to as the overall, and its size and scope is decided by the specific research and investigation purpose. Each member constituting the overall is called individual. The number of individuals in the overall is referred to as the overall capacity. The overall distribution is generally unknown. Sometimes the distribution type (such as common normal distribution, binomial distribution, etc.) can be known, but the specific parameters of the distribution (such as expectation, variance, etc.) are unknown. And the task of mathematical statistics is to statistically deduce the unknown distribution of the overall according to some individual data. In order to judge what distribution the overall obeys or estimate what value the unknown parameters should take, we can extract several individuals from the overall to observe, get some data to study the overall, and judge the distribution of the overall and make reasonable estimate on unknown parameters through statistically analyzing the data. The general method is according to certain principles to extract several individuals from the overall to observe. This process is called sampling.

Obviously, the observation result of each individual is random, which can be regarded as a random variable value. The $i-th$ individual indicators extracted from overall X is regarded as $X_i(i=1,2,\ldots,n)$, then X_i is a random variable; and $x_i(i=1,2,\ldots,n)$ can be used as specific observation value of individual index X_i. The X_1, X_2,... , X_n are sample values, individual number in the sample is called sample capacity (or sample size).

Normally, It is assumed that the sample is the one independent and identically distributed.
The overall and sample are two basic concepts in mathematical statistics. On the one hand, the sample is from the overall, naturally with information of the overall, and based on that, some of the characteristics of the overall (distribution or the parameters of distribution) can be studied. On the other hand, using sample to study the overall can be time-saving (especially for destructive sampling experiment). The problem of deducing distribution of the overall X by a sample X_1, X_2,... , X_n is known as the problem of statistical inference.

In order to deduce the overall from the sample, some appropriate statistic is needed to be constructed, by which unknown overall is got. Such sample statistics is sample function. To construct an excluding overall is a function of unknown parameters for the sample statistics. A function of unknown parameters is constructed as the sample statistics for the sampled sample. Commonly used statistics have:

Sample mean: $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}X_i$.

Sample variance: $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$

The point of origin of Sample (k order):

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^{\ k}, k=1,2,\cdots$$

The sample's original dot pitch (k order):

$$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k, k=2,3,\cdots$$

The commonly used statistical distribution: standard normal distribution, t - distribution, χ^2 - distribution, F - distribution. These distributions have formal tables to look for quantile, or to find the probability value by quantile, so they are commonly used.

There are two ways to estimate the unknown parameter in the overall distribution: one is point estimation, another is called interval estimation: (1) point estimate: if the overall distribution is known, but some parameter in it is unknown, the method of giving reasonable estimate of the unknown parameters from the extracted sample is point estimation, The commonly used point estimation method is distance estimation and maximum likelihood estimation. Distance estimation method is to make the sample distance of the corresponding order number to approximate the overall distance according to the number of unknown parameters of the overall, and solving simultaneous equation can get unknown parameter approximation value. Maximum likelihood estimation method is based on the maximum likelihood function of independently and identically distributed property and structure from the sample, then finds out extreme value point, and lets the extreme value approximate the unknown parameters. (2) the interval estimation - confidence interval: suppose θ is the unknown parameter of the overall distribution , X_1 , X_2,... , X_n are samples taken from the overall X, for a given number $1-\alpha,(0<\alpha<1)$, if the existence of statistics value $\varsigma=\varsigma(X_1,X_2,\cdots,X_n)$, $\xi=\xi(X_1,X_2,\cdots,X_n)$ makes $P\{\varsigma<\theta<\xi\}=1-\alpha$, random interval (ς,ξ) is called $1-\alpha$ bilateral confidence interval of θ, $1-\alpha$ is the confidence degree, and then ς and ξ are respectively the bilateral lower confidence limit and bilateral upper confidence limit of θ.

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 1, January 2013
ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
www.IJCSI.org

The meaning of confidence degree $1-\alpha$: If repeatedly sampled many times in the process of random sample, multiple sample values X_1, X_2,... , X_n can be got. Each corresponding sample value has aconfidence interval (ς,ξ). Each interval contains either a truth value of θ or not. According to Bernoulli large numbers law, When the sampling frequency k is sufficiently large, the frequency of true value θ in the interval is close to the confidence degree $1-\alpha$. That i-s, the number of the intervals containing the true value of θ are about $k(1-\alpha)$, and the number of the intervals not containing the true value θ are about $k\alpha$, e.g. Suppose $1-\alpha=0.95$, repeatedly sample-d 100 times, about 95 intervals contain the true value of θ, and about five interval does not contain the true value of θ.

Confidence degree and estimation precision are a pair of contradiction, the greater the confidence degree, the greater the probability of the true value θ in the confidence interval (ς,ξ), the longer t-he interval (ς,ξ), and the lower the estimate accuracy of unknown parameter θ is. Conversely, the higher the estimate accuracy of the parameter θ , the smaller the length of the confidence interval (ς,ξ), the lower the probability of the true value in (ς,ξ), and the smaller the confidence degree $1-\alpha$ is. The general rule is: in the guarantee of confidence degree, the estimation accuracy should be improved.

The common method of seeking confidence interval is: on the basis of point estimate, appropriate function U containing samples and parameters to be estimated should be constructed. And confidence interval can be derived from the known function U and the given confidence degree.

The general steps: (1) Select Some better estimator $\hat{\theta}$ from unknown parameter θ. (2) Center on $\hat{\theta}$ and construct a function $U=(X_1,X_2,\cdots,X_n,\theta)$ that depends on the sample and parameter θ; And if the distribution of the function is known (independent of θ), the random variable with such nature is called the pivot function. (3) For a given confidence degree $1-\alpha$, determine λ_1 and λ_2, and make $P\{\lambda_1\le U\le\lambda_2\}=1-\alpha$. When λ_1 and λ_2 in $P\{U\le\lambda_1\}=P\{\lambda_2\le U\}=\frac{\alpha}{2}$ are selected, generally, the quantile table can be used. (4) After identical deformation of the inequality $\lambda_1\le U\le\lambda_2$, it turns into $P\{\varsigma<\theta<\xi\}=1-\alpha$, and then (ς,ξ) is the bilateral confidence interval of the confidence degree $1-\alpha$ of θ. In some practical problems, $P\{U\le\lambda_1\}=\alpha$ or λ_1 and λ_2 in $P\{\lambda_2\le U\}=\alpha$ are only needed to meet, after identical deformation of the inequality, it becomes $P\{\varsigma<\theta\}=1-\alpha$ or $P\{\theta<\xi\}=1-\alpha$, so confidence interval similar to $(\varsigma,+\infty)$ or $(-\infty,\xi)$ can be got.

3. Data Analysis

Based on the cobweb model theory, price change of citrus is accord with divergent situation of the cobweb model. Further analysis is needed to get the price change trends on the basis of the original price information, and the possible ranges of values, providing a theoretical analysis for the next planting plans.

To analyze the pricing trend of citrus, we have a statistics of the wholesale prices in some area (Unit: 50 Yuan per 50kg). Due to the bigger price change in individual time periods, it is obviously not appropriate to analyze the daily price. We have 30 different prices for statistical analysis, and divide the prices into eight price intervals. Although some errors mean values of grouped sample and variance of the approximate calculation are used in mathematical statistics, after the analysis of the practical problems, the errors are in the acceptable range.

Table 1: the statistics data

Price change interval	Statistical number n_i	Price mean value x_i	$n_i x_i$
[120,130)	1	125	125
[130,140)	3	135	405
[140,150)	6	145	870
[150,160)	14	155	2170
[160,170)	4	165	660
[170,180)	1	175	175
[180,190)	0	185	0
[190, 200)	1	195	195
total	30		4600

price mean of the drawn-out sample can be easily calculated:

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{k}n_i x_i=\frac{4600}{30}\approx 153.33$$

Similarly, the variance approximation value of the sample can be calculated:

$$S^2 \approx \frac{1}{n-1}[\sum_{i=1}^{k} n_i x_i^2 - \frac{1}{n}(\sum_{i=1}^{k} n_i x_i)^2] \approx 172.99.$$

Suppose price fluctuation of citrus accords with most widely normal distribution in nature, if $X \sim N(\mu, \sigma^2)$, among which μ and σ^2 are unknown, the above sample sampling is made to construct pivot function:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

If the confidence degree $1-\alpha = 0.95$, then $T = \frac{153.33-\mu}{13.15/\sqrt{30}} \sim t(29)$, we can obtain the confidence interval [148.41, 158.25] on the confidence degree 0.95.

If the confidence degree is $1-\alpha = 0.99$, then $T = \frac{153.33-\mu}{13.15/\sqrt{30}} \sim t(29)$, we can obtain the confidence interval [146.71, 159.95] on the confidence degree 0.99.

If the confidence degree is $1-\alpha = 0.8$, then $T = \frac{153.33-\mu}{13.15/\sqrt{30}} \sim t(29)$, we can obtain the confidence interval [150.19, 156.47] on the confidence degree 0.8.

According to different requirements, a different confidence degree can be chosen, so as to find the desired confidence interval.

4. Conclusion

Combining with the price information of citrus in some area, the paper used mathematical statistics to predict the price trend. It also can predict more agricultural prices and the corresponding probability, make farmers of economic crops predict prices of agricultural products more purposefully and scientifically and adopt active planting ways to reduce risk and get more income.

References

[1] Richard H Dai. et al. Chaos economics, Shanghai Translation Publishing House, 1996.

[2] Li Zhongmin and Zhang Shiying, "Nonlinear cobweb model of dynamic analysis". Quantity Economic Research, No. 2, 1997, pp. 45-51.

[3] Huang Zelin, "Dynamic Analysis on nonlinear and Disequilibrium Cobweb Model ". Mathematics in Practice and Theory, Vol. 34, No. 3, 2004, pp. 40-45.

[4] Wang Jun and Yang Fuchun, " Some Sufficient and Ecessary Conditions of Convergent Cobweb Model ". Mathematics In Economics, Vol. 23, No. 4, 2006, pp. 364-369.

[5] He Jin and Qi Chunjie, "Empirical Study on Price Formation and Profit Distribution of Citrus in China". Journal of Northwest Agriculture and Forestry University (Social Sciences Edition), Vol. 9, No. 6, 2009, pp. 36-43.

[6] Lu Xiaoxu and Zhang Jie, "Based the cobweb model theory citrus production and price fluctuations analysis". The rural economy, No. 8, 2010, pp. 60-62.

[7] Shao Lu and Sheng Yajun, "Based on the the Cobweb model China's grain price determination mechanism analysis", Agricultural Economics, Vol. 22, No. 9, 2011, pp. 110-111.

Jintao Meng received the Master degree from Zhengzhou University in 2008, Operations research and control theory professional. Lecturer of Zhengzhou Institute of Aeronautical Industry Management.

Jing Li graduated from Zhengzhou University in 2004, Operations research and control theory professional. Lecturer of Zhengzhou Railway Vocational & Technical College.