# Markov Chain Model For the Structure of the Traditional Publishing Industry in China

Eryan Zhang Department of Basic Courses Beijing Institute of Graphic Communication Beijing, China

anpingzhey@yahoo.com.cn

Abstract-In order to analyze the structural change of the traditional publishing industry in China, the prediction technique of Markov Chain is applied to conduct research on the prediction problem of the structure of the traditional publishing industry China, and the estimation model of Markov matrix of state transition probability is provided as well. The conclusion shows that the prediction accuracy of the prediction technique is quite high after the error of simulated result has been tested. The prediction technique is applied to forecast the structure of the traditional publishing industry in China from 2012 to 2016. The application of Markov Chain prediction model in forecasting the structural change trend of the traditional publishing industry in China has greatly expanded the fields where the prediction technique of Markov Chain can be put into practice, and the data about the structural transition of the traditional publishing industry in China can hopefully offer a new perspective on looking into and solving issues to ensure the healthy and sound development of China's changing publishing industry.

Keywords- traditional publishing industry; OLS estimation; Markov Chain; the quadratic programming model; transition probability matrix

## I. INTRODUCTION

The twelfth five-year plan is the key period when publication industry deepens reform and speed up development. In recent years, with the rapid development of national economy in China, the progress of the publishing industry has made an increasing contribution for economic and social development. The publishing of books, periodicals and newspapers is an important part of the cultural industry, which is involved in media industry or publishing industry, even influences the development of great industry in a certain extent. The publication statistical data from 1990-2011 shows that the total number of publishing books, periodicals, newspapers increased year by year in our country, but the proportion of book publishing is in a decreasing trend, while that of the newspaper published is in increasing trend. In order to better analyze the structural changes of publishing books, periodicals, newspaper in our country and predict their development trend, this paper will introduce Markov chain model and linear regression model to analyze and predict the change Xiaofeng Zhu Department of Basic Courses Beijing Institute of Graphic Communication Beijing, China

zhuxiaofeng@bigc.edu.cn

trend of traditional publishing industry structure and total publishing number in our country.

Markov chain prediction method has a very broad prospect of application in social and economic fields, such as the application in product market share [1,2], in grain output prediction[3], in the financial market funds flow [4,5], in the talent flow[6,7], in education evaluation [8] and in industrial structure change [9], etc. Markov chain is a special kind of probability model and its research object is the state and state transition of a operation system. In the publishing industry, many system dynamic processes can be abstracted as state transfer motion issue. For example, dividing the traditional publishing books, periodicals and newspapers published into three different states, affected by the various factors, with the passage of time, all kinds of state flow mutually. One more example, in audio and video publishing industry, sound recording and video products can be considered as two different states, with the passage of time, under the influence of industrial policies, the publishing of recording products may transfer to video products, the publishing of video manufacturing also can transfer to the recording products. All of these can be viewed as the state transition movement issue, and Markov chain forecast method can be used to carry out the analysis and research.

Prediction method has been employed here in the study of structure change in China publishing industry in order to enrich the application of Markov chain prediction method in management. According to the total distribution data of books, periodicals, newspapers published in China from 1990 to 2011, the author combines the least square method with Markov chain and transforms the problem into the least squares estimation problem with the constraint condition [10-12], namely a quadratic programming problem by adding two conditions to the model: the row sum of

transition probability is 1 (  $\sum_{j=1}^n p_{ij} = 1$  ) and nonnegative

( $p_{ij} \ge 0$ , i,  $j = 1, 2, \dots, n$ ). By solving quadratic programming problem, this paper calculates the transition probability valuations of the proportion of published books, periodicals and newspapers in our country, gives their

regression equations, and does the test of significance. In addition, by using the regression equation predict short-term and long-term of publication products, and the error testing of the result based on prediction, it is proved that the model has higher prediction accuracy. This paper is expected greatly to enrich the application of Markov chain prediction method in the publishing industry prediction, and the data about the structural transition of the traditional publishing industry in China will hopefully offer a new perspective in looking into and solving issues to ensure a healthy and sound development of China's changing publishing industry.

### II. THE MARKOV CHAIN MODEL OF TRADITIONAL PUBLISHING INDUSTRY STRUCTURE

Let X be the total number of publishing books, periodicals and newspapers in our country, which can be divided into three kinds of state, let  $s_1$  be books,  $s_2$  be periodicals, and  $s_3$  be newspapers. The variable  $x_i$  respect the i-th component of state  $s_i$  in X, and satisfy  $x_1 + x_2 + x_3 = 1$ .

For each period t, the division of the total quantity X is according to  $X(t) = (x_1(t), x_2(t), x_3(t))^T$ . Let Y be the structure circle of X, and its area is 1. Dye Y as the division of X, the occupied area of state  $s_i$  in Y is  $y_i = \frac{x_i}{X}$ , here it is the proportion of  $s_i$  occupying in

 $X, y_1 + y_2 + y_3 = 1.$ 

For each period t, the publishing system's structure vector is  $Y(t) = (y_1(t), y_2(t), y_3(t))^T$ . Consider the homogeneous Markov process on Y. Let state  $s_i$  be in period t - 1, the transition probability when t is in  $s_j$  be  $p_{ii}$ , (i, j = 1, 2, 3). So it is

$$y_t(j) = \sum_{i=1}^{3} y_{t-1}(i) p_{ij} , t = 0, 1, 2, \cdots, m.$$
 (1)

Introduce the random error variable  $\mathcal{E}_{i}(j)$  in (1), and obtain the multiple linear regression model about transition probability  $p_{ij}$ .

$$y_{t}(j) = \sum_{i=1}^{n} y_{t-1}(i) p_{ij} + \mathcal{E}_{t}(j), \quad j = 1, 2, 3.$$
 (2)

Add row sum and condition of nonnegative to the transition

probability in Markov chain, that is 
$$\sum_{j=1}^{n} p_{ij} = 1$$
 and  $p_{ij} \ge 0$ ,

i = 1, 2, 3, and obtain Markov chain model of X's structure transition is

$$\begin{cases} y_{t}(j) = \sum_{i=1}^{n} y_{t-1}(i) p_{ij} + \mathcal{E}_{t}(j), \quad j = 1, 2, 3, \\ s.t. \quad \sum_{j=1}^{n} p_{ij} = 1, \quad i = 1, 2, 3, \\ p_{ij} \ge 0, \quad i, j = 1, 2, 3 \end{cases}$$
(3)

Thus the parameter estimation problem transforms into the least square estimation with constraint condition [13-15]. The least square estimation of Markov chain model's transition probability can be completed by using the time sample data of each state proportion in total quantity and solving model (3).

# III. DATA ACQUISITION AND SPECIFIC GRAVITY CALCULATION OF PUBLISHED BOOKS, PERIODICALS AND NEWSPAPERS

According to the statistical data about quantity of published books, periodicals and newspapers (total printed sheet unit: billion printed sheet) from 1990 to 2011 provided by Chinese statistics yearbook (2002-2011) and China's publication network, the proportion data of the three states is calculated. (See Table 1).

TABLE 1 THE PROPORTION OF PUBLISHED QUANTITY OF BOOI	κs,
PERIODICALS AND NEWSPAPERS FROM 1990 TO 2011	

Year	Books' proportion (%)	Periodicals' proportion(%)	Newspapers' proportion(%)
1990	51.4	10.7	37.9
1991	50.1	10.4	39.5
1992	50.6	10.3	39.1
1993	48.2	10.8	41.0
1994	44.5	10.1	45.3
1995	44.2	9.5	46.3
1996	42.6	9.0	48.4
1997	43.9	8.3	47.8
1998	40.6	8.2	51.3
1999	37.6	8.0	54.4
2000	34.8	8.6	56.6
2001	29.5	7.8	62.7
2002	28.1	7.0	64.9
2003	28.0	6.5	65.5
2004	25.6	6.0	68.4
2005	22.2	5.3	72.6
2006	22.1	5.6	72.3

Year	Books' proportion (%)	Periodicals' proportion(%)	Newspapers' proportion(%)
2007	22.2	5.9	71.9
2008	20.7	6.7	72.5
2009	21.2	6.0	72.9
2010	20.9	6.2	72.9
2011	20.7	6.1	73.3

# IV. CALCULATION OF TRANSFER PROBABILITY AND REGRESSION EQUATION

By using Mathematica4.0 to substitute the proportion data of first 21 years in Table 1 to model (3) and calculation, the transition probability matrix's least squares estimate of published books, periodicals and newspapers' structure movement changes can be followed :

 $(\hat{p}_{ij}) = \begin{pmatrix} 0.8761 & 0.0675 & 0.0507 \\ 0.4298 & 0.5956 & 0.0045 \\ 0.0000 & 0.0065 & 0.9935 \end{pmatrix}$ (4)

According to the multiple regression analysis theory, the unbiased estimation of Markov chain's variance is  $\hat{\sigma}^2 = 0.000036$ ,  $\hat{\sigma} = 0.006$ , let inspection level  $\sigma = 0.05$ , because  $\hat{\sigma} = 0.006 < 0.05$ , it can be considered that the industrial model is accordance with the Markov chain model [16-22]. Furthermore, three regression equations can be obtained according to (4):

The regression equation of published books is  $\hat{y}_t(1) = 0.8761 y_{t-1}(1) + 0.4298 y_{t-1}(2) + 0.000 y_{t-1}(3)$ .

The regression equation of published periodicals is

 $\hat{y}_{t}(2) = 0.0675 y_{t-1}(1) + 0.5956 y_{t-1}(2) + 0.0065 y_{t-1}(3).$ 

The regression equation of published newspapers is

 $\hat{y}_t(3) = 0.507 y_{t-1}(1) + 0.0045 y_{t-1}(2) + 0.9935 y_{t-1}(3) \,.$ 

### V. MARKOV CHAIN MODEL FITTED DEGREE ANALYSIS AND PREDICTION ACCURACY EVALUATION

The fitting values of published books, periodicals, newspapers proportion data can be calculated by (4) and three regression equations, and comprised with the actual value, calculate the fitting error. (See Table 2).

TABLE 2 THE FITTING ERROR OF PUBLISHED BOOKS, PERIODICALS AND

Year	Books prediction error	Periodicals prediction error	Newspapers prediction error
1991	0.0218	0.0051	-0.0272
1992	-0.0056	0.0095	-0.0042
1993	-0.0230	0.0019	0.0206
1994	0.0085	0.0017	-0.0107
1995	-0.0023	0.0007	0.0013

Year	Books	Periodicals	Newspapers
1996	0.0270		-0.0246
1770	0.0270	0.0027	0.0240
1997	-0.0146	-0.0004	0.0150
1998	-0.0145	0.0010	0.0134
1999	-0.0161	0.0092	0.0066
2000	-0.0470	0.0000	0.0464
2001	-0.0111	-0.0009	0.0113
2002	0.0039	0.0005	-0.0049
2003	-0.0175	-0.0016	0.0188
2004	-0.0284	-0.0051	0.0332
2005	0.0043	0.0051	-0.0097
2006	0.0041	0.0063	-0.0108
2007	-0.0124	0.0124	-0.0005
2008	0.0011	0.0008	-0.0027
2009	-0.0018	0.0070	-0.0058
2010	-0.0033	0.0062	-0.0035
2011	-0.0007	0.0060	-0.0053

The calculation about the fitted error and fitted precision measure indexes from 1991 to 2011 according to Table 2 helps to get the fitted degree of single regression equation. (See Table 3).

TABLE 3 THE FITTING ACCURACY TABLE OF TRADITIONAL PUBLISHING INDUSTRY STRUCTURE BASED ON MARKOV CHAIN

	Books	Periodicals	newspapers
MSE	0.000295	2.8049E-05	0.000305
MAPE	4.0641%	5.663%	2.3463%

According to statistical prediction, if average absolute percentage error MAPE is less than 10%, the model prediction accuracy will be higher. The data above shows that both the MAPE of books prediction model and newspapers forecast model are less than 5%, the MAPE of periodicals prediction model is less than 6%, so the prediction precision of this model is high.

### VI. CONCLUSIONS

1) According to (4), it can be concluded that  $p_{11} \ge p_{12} \ge p_{13}$ ,  $p_{22} \ge p_{21} \ge p_{23}$ ,  $p_{33} \ge p_{32} \ge p_{13}$ . It indicates that the internal keeping in transfer motion process of the three state structure, the probability of books, periodicals and newspapers industry is greater than their transferring externally, which explains that in the internal structure transfer motion process of traditional publishing industry, the "continuity" priority rule of industry still exists. What's more, the internal keeping probability of newspaper industry is 99.35%, while the probability of transferring to books and periodicals is only 0.00% and 0.65%. And the internal keeping probability of transferring to books and

newspapers is 42.98% and 0.45% respectively, so the larger impact is on the periodical industry for sure.

2) By using the transition probability matrix given by this paper and three regression equations, the change of traditional published industrial structure in our country in the next few years can be predicted. (See Table 4).

Year	Books (%)	Periodicals (%)	Newspapers (%)
2012	20.76	5.53	73.85
2013	20.56	5.17	74.45
2014	20.24	4.95	75.05
2015	19.86	4.80	75.61
2016	19.46	4.69	76.15

TABLE 4 THE PUBLISHING INDUSTRY STRUCTURE PREDICTION TABLE IN CHINA BASED ON MARKOV CHAIN MODLE

(The base period of the prediction above is the year of 2011)

Markov chain prediction method on structure transfer in Chinese traditional publishing industry given by this article can hopefully offer a new perspective in relevant research of cultural publishing industry. Independently, the transition probability matrixes gained in this paper represent the industry structure's transfer mode in that period; dynamically, they are all links constituting the whole transfer chain of industrial structure. From the perspective of dynamic development of books, periodicals and newspapers publishing industry, they are the three aspects which are interdependent, mutual influenced, and closely linked in inner. On condition of certain total publishing quantity, the number of books, periodicals, newspapers published is of reciprocal relationship. They supplement one another in the form of reciprocal causation, keeping a certain proportional relation.

### ACKNOWLEDGEMENTS

This study is supported by Institute Level Projects Funded by Beijing Institute of Graphic Communication (Eb-2013-28), Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality (PHR201107145), Science Project of Beijing Municipal Education Commission (KM201210015011).

#### REFERENCES

[1] Li Yunxia. The forecast to the market share and development of Markov model, J. Statistics and Decision, Vol.12 (2009) pp.180-181. (In Chinese)

[2] Tang Xiaowo, The estimation of transition probability of Markov chains in market forecasting, J. Journal of UEST of China, Vol.23 (1994) pp.646-648. (In Chinese)

[3] He Fuli, Hu Yong, Chen Chun. Forecast of Grain Output in China with Grey Markov Forecasting Model, J. Mathematics in Practice and Theory. Vol.33 (2003), No.12, pp.46-47. (In Chinese)

[4] Li Xingping, Zhang Qing, Wu Shiyong. Markov Modeling and Its Application in the Stock Market, J. Journal of Kunming University of Science and Technology(Science and Technology), Vol.34 (2009), No.6, pp.116-118. (In Chinese) [5] Li Haitao.Using the Markov Chain Prediction Method to Predict the Stock Price, J. Statistics and Decision, Vol.5(2002), pp.25-26. (In Chinese)

[6] Li Yi, Huo Kaijie. Enterprise Human Resources Flow Analysis by Markov Chain Model, J. Science of Management, Vol.1(2008), pp.108-109. (In Chinese)

[7] Xia Xiufang, Fang Yuanyuan. The Application of Markov Forecasting Pattern on Human Resources, J. Journal of Qingdao Institute of Architecture and Engineering, Vol.2(2001), pp.75-76. (In Chinese)

[8] Dong Jianguo, Hui Shurong, Chen Zhongwei. Application of Markov Chain Model to Mathematics Education Evaluation, J. Journal of Shenyang Agricultural University (Social Sciences Edition), Vol.12(2010), No.4, pp.468-471. (In Chinese)

[9] Li Jinghua. The Markov Chains for Industrial Structure Transition in China, J. Mathematics in Practice and Theory, Vol.31(2001), No.2, pp.156-161. (In Chinese)

[10] Sheng Zhou, Probability Theory and Mathematical Statistics. Beijing, Higher Education Press, 1989.

[11] Altman E. Denumerable constrained Markov decision problems and finite approximations, J. Math. Operes, Vol. 19(1994), pp.169-191.

[12] Hu Guanghua, Zhang Sheng. On Discounted Semi-markov Decision Processes with A Constraint, J. Acta Mathematicae Applicatae Sinica, Vol. 20(1997), No.2, pp.187-195. (In Chinese)

[13] S.M Lauce. Random Process . Beijing, China Statistical Press, 1997, pp.114-211.

[14] Zhang Dewen and Wei Fuxuan. Exposition and argumentation on constrained least square method, J. Chinese Journal of Computational Mechanics, Vol.17. (2000), No.4, pp.398-404. (In Chinese)

[15] Jiang Tongsong. A note on exposition and argumentation on constrained least squares method, J. Chinese Journal of Computational Mechanics, Vol.20(2003),No.1, pp.119-126. (In Chinese)

[16] Yuan Yaxiang, Sun Wenyu.Theory and Methods of Optimization. Beijing, Science Press, 1997.

[17] Shi Guangyan,Dong Jiali. Optimization method, M. Beijing, Higher Education Press, 1999.

[18] Xiao Shutie. Random mathematics. Beijing, Higher Education Press, 2000.

[19] Tan Yingshi. Applied mathematics. Guangzhou, The South China Science and Technology University Press , 1994.

[20] Yi Danhui. Statistical forecast. Beijing, China Renmin University Press, 1990.

[21] Shen Dingxuan. Random process. Wuchang, Central China university of science and technology Press, 1990.

[22] Hang Bin, Li Jinghua. Least squares estimation of chain transfer probability of Markov chain, J. Mathematicae Applicatae Sinica, Vol. 23(1993), No.1, pp.29-35. (In Chinese)