

Statistical Analysis on the Publishing Indices of Books, Newspapers and Magazines in China

Xiaofeng Zhu¹, Xidan Yang², Shuyang Zhang²

¹⁾ Department of Basic Science, Beijing Institute of Graphic Communication, Beijing, China (zhuxiaofeng@bigc.edu.cn)

²⁾ School of Economics and Management, Beijing Institute of Graphic Communication, Beijing, China

Abstract—First the authors through understand the opportunities and challenges which the publishing industry faced in the development in China, determine printed sheets of the books, newspapers and magazines is one of the important Indices. Then use statistical analysis method of hierarchy analysis and regression analysis to establish statistical models of the publishing indices for publishing industry in China. Last analyze printed sheets of the books, newspapers and magazines in 1987–2011, and forecast publishing indices in 2012–2016. In order to provide useful data for the relevant research departments, and promote development of the publishing industry in China.

Keywords—publishing indices, statistical analysis, regression analysis, book, newspaper, magazine

中国书报刊出版指标的统计分析

朱晓峰¹ 杨喜丹² 张舒扬²

¹⁾ 北京印刷学院基础部, 北京, 中国

²⁾ 北京印刷学院经济管理学院, 北京, 中国

摘 要 首先了解我国出版业发展面临的机遇和挑战, 并分析了影响出版业发展的各项指标。然后利用层次分析、回归分析、指数增长模型分析等多种统计分析方法, 建立了出版业发展指标模型。最后对我国 1987 年—2011 年期间的图书、报纸和期刊总印张进行统计分析, 并对十二五末和十三五初的出版指标进行了预测, 以期对相关研究部门提供有效的分析数据, 促进我国出版业的发展。

关键词 出版业, 统计分析, 回归分析, 图书, 报纸, 期刊

1. 引言

目前我国出版业已经形成产业化、规模化, 发展也呈现出一定的规律性[1]。但是出版业作为一种特殊的产业, 其发展又具有自身的特点。一是发展变化的趋势比较复杂。二是进行量化分析时需要量化的指标多。三是各数据之间通常为非线性关系。因此如何建立适宜数学模型, 通过对我国出版业进行量化分析, 找出其发展过程中的一些规律性信息, 为我国出版业的发展提供有必要、有参考价值的分析数据, 成为出版业发展的一个重要课题。

2. 出版业发展指标的量化分析与预测

出版物的分类有很多种, 其中它分为传统出版物, 包括图书出版、报纸出版和期刊出版; 电子出版物, 包括软磁盘、只读光盘、交互式光盘、照片光盘等[2]。由于电子

出版物的发展年代较短, 数据规律性较弱, 因此这里主要对传统出版物图书、报纸、期刊的指标进行分析研究。为

表1 出版数据统计表

单位: 亿印张

年份	图书	期刊	报纸	年份	图书	期刊	报纸
1987	261.2	72.7	223.6	2000	376.2	100.0	799.8
1988	269.0	71.2	231.3	2001	406.1	100.9	938.9
1989	243.6	50.7	179.4	2002	456.4	106.4	1067.4
1990	232.1	48.1	182.8	2003	462.2	109.1	1235.6
1991	266.1	54.4	205.8	2004	465.6	110.5	1524.8
1992	280.0	62.7	238.8	2005	493.3	125.3	1613.1
1993	282.3	64.2	287.1	2006	512.0	136.9	1658.9
1994	297.2	63.9	310.7	2007	486.5	157.9	1700.8
1995	316.8	67.0	359.6	2008	561.1	158.0	1930.6
1996	360.5	68.1	392.4	2009	565.5	166.2	1969.4
1997	364.0	73.3	459.8	2010	604.7	181.1	2148.0
1998	373.6	79.9	540	2011	634.5	192.7	2272.0
1999	391.4	96.8	636.7				

了研究图书出版发展趋势,我们调研了从1987年到2011年我国图书、期刊和报纸主要出版业指标数据如表1所示。

2.1 图书出版分析预测

运用 Mathematica 软件绘制我国 25 年图书趋势曲线图(图1)。可以观察到,我国图书出版行业的发展呈逐渐上升趋势,没有太大波动性,因此这里笔者采用趋势方程拟合合法对图书出版数据进行分析预测。

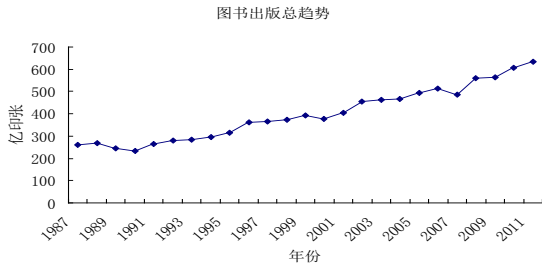


图1 图书出版趋势曲线图

2.1.1 趋势方程拟合法

趋势方程拟合法是通过拟合以时间 t 为解释变量、所考察指标为被解释变量的回归方程来测定研究对象的长期变化趋势。其最大的优点是:不仅可以测定出时间序列中各期的趋势值,而且所拟合的趋势方程还具有延伸外推的功能,可以根据趋势方程直接预测。长期趋势可以分为线性趋势和非线性趋势。趋势形态不同,所拟合的趋势方程也就有不同的形式([3],[4],[5])。

由于时间序列的逐期增长量大致相同、长期趋势可近似的用一条直线来描述,时间序列具有线性趋势,因此可用下列形式的线性趋势方程来描述 $\hat{y}_t = a + bt$ 。式中 \hat{y}_t 为时间序列 y_t 的趋势值; t 为时间(通常取 $t=1,2,\dots,n$); a 为趋势线的截距,表示 $t=0$ 时的趋势值(即既定时间序列长期趋势的初始值); b 为趋势线的斜率,表示当时间 t 每变动一个单位,趋势值的平均变动量。

估计线性趋势方程中参数 a 、 b 的方法通常采用最小二乘法,其计算公式与直线回归方程中参数的计算公式相同,即计算公式为

$$\begin{cases} b = \frac{n \sum ty_t - \sum t \sum y_t}{n \sum t^2 - (\sum t)^2} \\ a = \bar{y} - b\bar{t} \end{cases} \quad (1)$$

2.1.2 趋势方程法建立与求解

从图书的印张数曲线可以看出,图书每年的印张数变化大体近似于一条直线,因此可以拟合线性趋势方程。取

$t=1,2,\dots,25$,根据式(1)可计算出参数的估计值 $a=190.43$, $b=16.003$,即所求的线性趋势方程为 $\hat{y}_t = 190.43 + 16.003t$,并由此得出拟合曲线图(图2)。

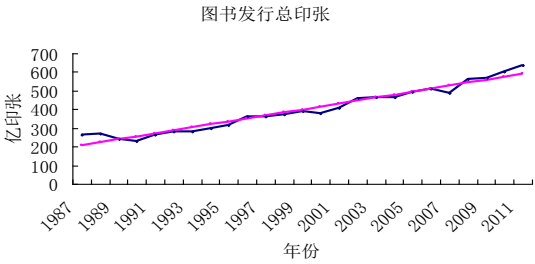


图2 图书拟合曲线图

2.1.3 图书出版分析预测

根据实际数据与趋势值的对比,对趋势方程式进行检验,符合发展趋势。根据此方程可以预测出2012到2016年(十二五末和十三五初的)未来五年图书出版总印张数如表3所示。其中2012年图书出版的实际数据为666.99亿印张,与预测数据相比,二者相差较小,仅有9.06%,说明该模型预测数据比较准确。

从拟合的函数关系看,随着时间的增长,图书的出版总印张数与时间呈线性关系逐年增长。一方面,从微观经济学的角度把商品分为必需品和超必需品。前者是维持日常生活所不可缺少的商品,比如衣食住行等;后者是相对而言日常生活中可有可无的,比如读书,看报,旅游,听音乐等。图书作为超必需品,随着我国国民经济的增长,人们收入增加,消费水平的不断提高,图书的消费量也呈逐年增加的趋势。另一方面,随着全球化的到来,我国施行“引进来,走出去”的政策,近些年引进多种外文图书,翻译出版等受到广大人民的喜爱,致使图书的出版逐年增加。

2.2 期刊出版分析预测

1987—2011年我国期刊出版总印张数如表1所示,运用 Mathematica 软件绘制我国25年期刊趋势曲线图(如图3所示)。观察曲线图可以看出,我国期刊发行行业的发展呈逐渐上升趋势,刚开始成缓慢增长,后来呈较快速度增长。因此采用最小二乘法模型对期刊发行数据进行分析预测。

表2 2012-2016图书出版预测数据

年份	2012	2013	2014	2015	2016
图书总印张 (单位:亿印张)	606.5	622.5	638.5	654.5	670.5

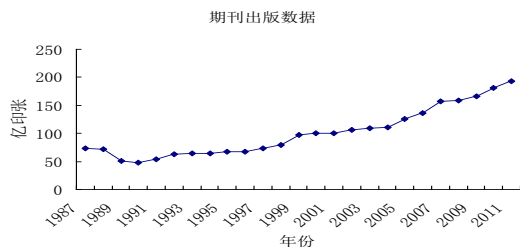


图3 期刊出版趋势曲线图

2.2.1 最小二乘法模型

在科学实验的统计方法研究中,往往要从一组实验数据中寻找出自变量 x 和因变量 y 之间的函数关系 $y=f(x)$ 。由于观测数据往往不够准确,因此并不要求 $y=f(x)$ 经过所有的点,而只要求在给定点上误差按照某种标准达到最小,通常采用欧氏范数作为误差量度的标准。这就是所谓的最小二乘法。

2.2.2 模型建立与求解

观察上述曲线图,呈曲线状态,可以用二次多项式来拟合期刊发行总印张与时间之间的效应图。设拟合二次多项式方程为 $y=a_0 \cdot 1+a_1 \cdot t+a_2 \cdot t^2$, 式中 t 为时间, y 为期刊每年总印张。记

$$R = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_n & t_n^2 \end{bmatrix}, \quad A = [a_0, a_1, a_2]^T, \quad Y = [y_1, y_2, \dots, y_n]^T。$$

式中 $n=25$, $t_i (i=1,2,\dots,25)$ 分别表示时间所对应的数据; $a_i (i=0,1,2)$ 待求拟合曲线系数; y_i 表示 $t_i (i=1,2,\dots,25)$ 时间对应的期刊发行总印张。由公式 $R^T R A = R^T Y$ 即可求得在最小二乘法下的系数 $A = [a_0, a_1, a_2]^T$ 。

运用 Mathematica 软件可直接将上面数据输入,而求得期刊发行总印张关于时间 t 的曲线方程如下,并由此得出拟合曲线图(图4)。

$$y = 64.55 - 2.2756t + 0.2975t^2。$$

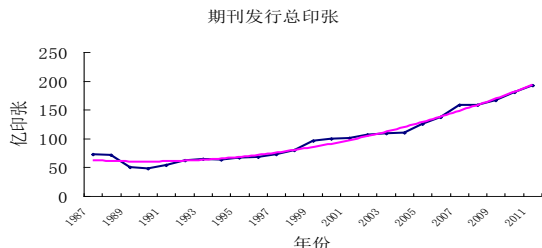


图4 期刊拟合曲线图

表3 2012-2016期刊出版预测数据

年份	2012	2013	2014	2015	2016
期刊总印张 (单位:亿印张)	206.5	220.0	2341	248.8	264.0

2.2.3 期刊出版分析预测

根据实际数据与拟合曲线数据的对比,对拟合方程进行检验,符合该发展趋势。由此推测,十二五末和十三五初的未来五年期刊发行总印张如表3,其中2012年期刊的实际数据为196.01亿印张,预测值与实际数据相比二者相差很小,偏差仅有-5.08%。说明该模型预测数据比较准确。

从拟合的函数关系看,随着时间的增长期刊的出版总印张数与时间呈二次方关系,表明总印张数的增长速度要快于时间的增长速度。一方面,随着现在经济水平的发展,期刊在广大读者中有广泛的影响,已成为人们日常生活中一种必须的文化消费品,给人们传递着来自四面八方的、手机与网络等无法提供的深度信息,为人们的日常生活增加了乐趣。另一方面,虽然近几年电子出版物逐渐发展,但是期刊在某些方面有着不可替代的作用。特别是在近几年,中国期刊市场在改革中有了很大的发展和变化,更是为期刊出版业的发展提供了优越的条件。

2.3 报纸出版分析预测

1987——2011年我国期刊出版总印张数如表1所示,运用 Mathematica 软件绘制我国25年期刊趋势曲线图(图5)。由这个折线图可以看出我国报纸总印张发展趋势从1987年到1997年增长相对较慢,1997年到2005年成加速增长趋势,2006年到现在出现快中有稳定的增长。目前期刊出版的总印张数呈现出指数增长趋势。

2.3.1 模型建立与求解

随着社会的发展和经济的增长,报纸总印张不可能无限制增长,可能会出现最大发行量 K ,因此指数模型与实际报纸总印张(后期的)增长率存在差异。指数模型只适用于限制相对较少的初期发展阶段。

随着经济的增长,报纸行业发展受到外界空间大小、

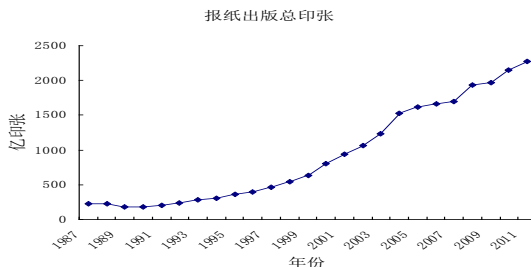


图5 报纸出版趋势曲线图

经济发展水平、减少砍伐树木等的限制，将对报纸发行量增长起到阻滞作用，并随着报纸发行数量的增加，阻滞作用也将越来越大。考虑到这种因素，运用阻滞增长模型，并进行求解。

阻滞作用体现在对报纸总印张增长率的影响上，天使得报纸增长率随着报纸总印张的增加而下降。如果设 t 为时间， $x(t)$ 为报纸总印张， r 为报纸总印张固有的年增长率， K 中国报纸最大发行总印张。于是可以得到报纸总印张 $x(t)$ 的微分方程

$$\frac{dx}{dt} = rx \left(1 - \frac{x}{K} \right), \quad x(0) = x_0. \quad (2)$$

其中 rx 体现了报纸总印张自身的增长趋势，因子 $(1-x/K)$ 体现了环境和经济对报纸总印张增长的阻滞作用。显然报纸总印张是两个因子共同作用的结果。

方程 (2) 的解可用分离变量法得到 $x(t) = \frac{K}{1 + Ce^{-rt}}$ ，其中 $C = (K - x_0) / x_0$ 。

用这个模型可以较好的描述一定时间空间内我国报纸总印张的增长状况（根据目前报纸发展状况还未到达 S 的平缓阶段）。

2.3.2 报纸年总印张分析与预测

分析 1987 年到 2011 年的中国报纸总印张（如表 1 所示），我们取 1996 年的报纸总印张为 $x(0) = x_0 = 392.4$ 。假设中国报纸最大可发行量 $K = 2750$ ，则有 $C = 5.62589$ 。由不同的 r 值代入

$$x(t) = \frac{2750}{1 + 5.62589e^{-rt}},$$

再运用 Mathematica 软件可画出曲线如图 6 所示。

通过对比 K 取不同值得到的分析图，发现当 $K=2750$ 时候最符合发展趋势。再对不同 r 值经过一系列的检验分析，当 $r=0.2$ 时，模型预测比较符合实际的增长速度。所以综上分析，由此预测得到的未来五年报刊出版数据如表 4 所示。其中 2012 年报纸的实际出版数据为 2211 亿印张，预测值与实际数据相比二者相差很小，偏差仅有 -7.21%。

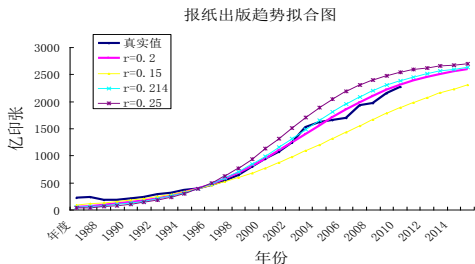


图 6 不同 r 值时的报刊出版数据预测和真实值比较

表4 2012到2016报纸出版预测数据

年份	2012	2013	2014	2015	2016
期刊总印张 (单位:亿印张)	2382.8	2449.7	2505.7	2552.1	2590.2

说明该模型预测数据比较准确。

观察报纸总印张的走势大致可分为三个阶段。从 1987 年到 1997 年增长相对较慢，1997 年到 2005 年成加速增长态势，2006 年到现在出现快中有稳定的增长趋势。一方面，近几年报纸的发展受到电子出版物的冲击发展速度有所减缓。另一方面，像《经济观察报》、《北京青年报》、《中国日报》等这些报纸都有着不可替代的作用。但总的发展还是呈现增长趋势的。

3. 结束语

近 30 年，我国出版业逐步实现了由“引进来”向“走出去”的转变，国际竞争实力日益增加。但是随着信息化时代的到来，我国出版业也面临许多竞争和挑战。笔者从出版业指标统计分析的角度，对出版业数据进行分析。针对不同出版物的发展趋势不同，笔者采用了不同的分析方法和数学模型，对我国 1987 年——2011 年期间的图书、报纸和期刊的总印张进行了分析，并对十二五末十三五初的未来五年数据做出了预测，希望对出版业的未来发展有所帮助。

致谢

本项目由 2013 年北京市本科生科学研究计划项目资助；北京印刷学院校级重点资助项目(Ea201231)；2013 年学科与研究生教育（研究生教育）专项资金资助；2013 年北京印刷学院校级教学改革重点项目资助；北京市教委教师队伍建设资助项目（青年英才计划入选人员）；北京市属高等学校人才强教计划资助项目（PHR201107145）。

参考文献(References)

- [1] Zhensheng Hao, "The report of development for Chinese publishing industry in 2009-2010", China Books Press, 2010.
- [2] Guanyi Wang, Yuhong Hua, "Study on the performance evaluation of Chinese publishing industry", China Financial and Economic Press, 2010.
- [3] The writing group of mathematical model, "Mathematical model", South China University of Technology Press, 2001.
- [4] Wuyi Zeng and Hongye Xiao, "Statistics Introduction", Science Press, 2012.
- [5] Xizu Yan, Zhongmin Song and Chunjia Bi, "Mathematical modeling and experiment", Science Press, 2009.