

A New Method for Determining Model Parameters of the k -Order Trend Curve

Kai Zuo^{1*}, Hu Li¹, Yulin Zhang¹, Wenqing Wu²

¹School of Mathematics, Chengdu Normal University, Chengdu Sichuan

²School of Science, Southwest University of Science and Technology, Mianyang Sichuan

Email: *zuokaihelen@163.com

Received: Jul. 20th, 2019; accepted: Aug. 7th, 2019; published: Aug. 14th, 2019

Abstract

In the fields of economy, engineering, energy and environment, non-stationary series caused by deterministic factors can be fitted and predicted by k -order trend curve (or named polynomial function). Inspired by the method of solving the parameters of the modified exponential curve model, this paper proposes a method to determine the parameters of the k -order trend curve model. Compared with the least square method, the modelling process of this new method is simpler and easier. Further, two examples in the textbook show that the proposed method has good accuracy and can be used in data modelling, fitting and prediction.

Keywords

k -Order Trend Curve, Non-Stationary Sequence, Piecewise Summation, Gross Domestic Product, Metal Cutting Machine Tools

一种确定 k 阶趋势曲线参数的新方法

左凯^{1*}, 李虎¹, 张玉林¹, 吴文青²

¹成都师范学院数学学院, 四川 成都

²西南科技大学理学院, 四川 绵阳

Email: *zuokaihelen@163.com

收稿日期: 2019年7月20日; 录用日期: 2019年8月7日; 发布日期: 2019年8月14日

摘要

在经济、工程、能源、环境领域中, 由确定性因素导致的非平稳序列, 往往可以用 k 阶趋势曲线(或称

*通讯作者。

为多项式函数)对数据进行拟合、预测。本文在修正的指数曲线模型参数的求解方法的启发下,提出了一种确定 k 阶趋势曲线模型参数的方法。相比最小二乘法,本方法的求解过程更为简单、易行。通过教材上的两个具体实例验证了本文提出的方法具有很好的精确度,能够用于对数据的建模、拟合和预测中。

关键词

k 阶趋势曲线, 非平稳序列, 分段求和, 国内生产总值, 金属切削机床

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在统计学的教材[1][2]或者时间序列教材[3]中提到,如果现象的长期趋势随着时间的推移按照线性变化,则可以对时间序列拟合趋势直线。如果现象的趋势随着时间的推移出现非线性趋势,则需要配合适当的趋势曲线。趋势曲线的形式又很多,有抛物线型、指数曲线型、修正指数曲线型、龚珀茨曲线型、逻辑斯蒂曲线型等。当现象的变化形态不是按照某种固定的形态变化,而是有升有降,在变化过程中可能有几个拐点。这时就需要拟合 k 阶趋势曲线,即多项式函数。当只有一个拐点时,可以拟合二阶曲线,即抛物线。当有两个拐点时,需要拟合三阶曲线。当有 $k-1$ 个拐点时,需要拟合 k 阶曲线。

段树乔和段方婕[4]提出了带有电源供给能力约束条件的回归函数筛选及其综合方法,并建立了电网公司售电量的多项式回归预测模型及其检验方法,并通过一个具体实例详细说明了该方法的应用。温鲜等人[5]以柳州市房地产价格为例,建立了柳州市房地产价格的多项式回归模型,并进行实证分析、检验模型精度,并预测了未来三年柳州市的房地产价格。于娟和姜潮[6]采用多项式回归模型和灰色预测模型对辽宁省省医院药品比例进行分析预测。徐颖和黄素珍[7]利用百度搜索数据,以江苏省南京市为例建立了新建住宅销售价格指数的多元线性回归、完全二次多项式回归和逐步回归模型。张磊等人[8]将灰色多项式回归组合模型应用在维修器材需求方面。在模型中,作者利用最小二乘法得到了灰色多项式组合预测模型的各项系数,并分别运用灰色二项式、三项式、四项式回归组合模型对数据进行建模分析。计算结果表面,灰色多项式回归组合模型可以大幅度提高数据的拟合精度,并且随着拟合项数的增加,拟合精度逐渐提高。

从上面的研究文献中可以看出,确定模型系数的方法基本上采用的都是最小二乘法。它通过最小化误差的平方和寻找数据的最佳函数匹配,使得计算得到的数据与实际数据之间的误差的平方和最小。但是,在求解修正的指数曲线的系数时,教材并没有使用最小二乘法,而是采用的分段求和的思想,即三和法求解模型中参数的具体表达式。受到此模型参数求解思路的启发,我们将分段求和的思想引入到 k 阶趋势曲线的参数求解中,得到了模型系数的表达式。最后,以教材中的两个具体实例,进一步分析了该方法的可行性和有效性。

本文的结构安排如下:第2节给出三和法的具体计算过程,以便引入到 k 阶趋势曲线中。第3节讨论了分段求和方法在 k 阶趋势曲线模型参数的具体求解。第4节给出了两个具体的实例,以说明该方法的可行性和有效性。第5节给出了本文的结论。

2. 修正的指数曲线

由文献[1] [2] [3]可知, 如果现象的趋势为: 初期增长迅速, 随后增长率逐渐减低, 最终以一个常数为增长极限, 则可以采用修正的指数曲线进行拟合。在经典指数曲线的基础上增加一个常数 k , 即得到修正指数曲线的方程,

$$Y_t = ab^t + k, \quad (1)$$

其中, a 、 b 、 k 为未知参数, $a \neq 0$, $b \in (0,1) \cup (1,\infty)$, $k \in (0,\infty)$ 。

参数 a 、 b 、 k 估计的基本思想是三和法: 把整个时间序列分成相等的三个数组, 每个组有 m 项, 根据趋势值 Y_t 的三个局部总和分别等于原数列观察值 Y_t 的三个局部总和来确定三个参数。具体为: 设观察值的三个局部总和分别为 S_1 、 S_2 、 S_3 , 得

$$S_1 = \sum_{t=1}^m Y_t, \quad S_2 = \sum_{t=m+1}^{2m} Y_t, \quad S_3 = \sum_{t=2m+1}^{3m} Y_t. \quad (2)$$

由三和法得到如下方程

$$\begin{cases} S_1 = mk + a + ab + ab^2 + \cdots + ab^{m-1} \\ S_2 = mk + ab^m + ab^{m+1} + \cdots + ab^{2m-1} \\ S_3 = mk + ab^{2m} + ab^{2m+1} + \cdots + ab^{3m-1} \end{cases}. \quad (3)$$

通过对方程(3)求解, 得到

$$\begin{cases} b = \left(\frac{S_3 - S_2}{S_2 - S_1} \right)^{\frac{1}{m}} \\ a = (S_2 - S_1) \frac{b-1}{(b^m - 1)^2} \\ k = \frac{1}{m} \left(S_1 - \frac{a(b^m - 1)}{b-1} \right) \end{cases}. \quad (4)$$

3. K 阶趋势曲线

由文献[1]的 13 章可知, k 阶趋势曲线函数的一般形式为

$$Y_t = a_k t^k + a_{k-1} t^{k-1} + \cdots + a_1 t + a_0, \quad (5)$$

其中 $a_k \neq 0$ 。教材中提到参数 a_k, a_{k-1}, \dots, a_0 可以按照多元回归分析中的最小二乘法具体求得, 有兴趣读者可以参看相关资料。这里我们采用分段求和的思想给出参数 a_k, a_{k-1}, \dots, a_0 的具体表达式。

分段求和的思想是: 把整个用于建模的时间序列分成相等的 $(k+1)$ 个数组, 每个组有 m 项, 即用于建模的数据量为 $(k+1)m$ 。根据趋势值 Y_t 的 $(k+1)$ 个局部总和分别等于原序列观察值 Y_t 的 $(k+1)$ 个局部总和来确定参数 a_k, a_{k-1}, \dots, a_0 。具体为: 设观察值的 $(k+1)$ 个局部总和分别为 S_1, S_2, \dots, S_{k+1} , 得

$$S_i = \sum_{t=(i-1)m+1}^{im} Y_t, \quad i=1, 2, \dots, k+1. \quad (6)$$

于是, 得到如下方程

$$\begin{cases} S_1 = a_k \sum_{t=1}^m t^k + a_{k-1} \sum_{t=1}^m t^{k-1} + \cdots + a_1 \sum_{t=1}^m t + a_0 m \\ S_2 = a_k \sum_{t=m+1}^{2m} t^k + a_{k-1} \sum_{t=m+1}^{2m} t^{k-1} + \cdots + a_1 \sum_{t=m+1}^{2m} t + a_0 m \\ \vdots \\ S_{k+1} = a_k \sum_{t=km+1}^{(k+1)m} t^k + a_{k-1} \sum_{t=km+1}^{(k+1)m} t^{k-1} + \cdots + a_1 \sum_{t=km+1}^{(k+1)m} t + a_0 m \end{cases} \quad (7)$$

在表达式(7)的计算中,我们要对形如 $\sum_{t=(i-1)m+1}^{im} t^p$ 的表达式进行求解。很显然给出它的一般形式是很困难的,有兴趣的可以去看 Faulhaber 数或者 Harmonic 数关于如何求解的问题。将表达式(7)写出矩阵的形式为:

$$\begin{pmatrix} \sum_{t=1}^m t^k & \sum_{t=1}^m t^{k-1} & \sum_{t=1}^m t^{k-2} & \cdots & \sum_{t=1}^m t^2 & \sum_{t=1}^m t & m \\ \sum_{t=m+1}^{2m} t^k & \sum_{t=m+1}^{2m} t^{k-1} & \sum_{t=m+1}^{2m} t^{k-2} & \cdots & \sum_{t=m+1}^{2m} t^2 & \sum_{t=m+1}^{2m} t & m \\ \sum_{t=2m+1}^{3m} t^k & \sum_{t=2m+1}^{3m} t^{k-1} & \sum_{t=2m+1}^{3m} t^{k-2} & \cdots & \sum_{t=2m+1}^{3m} t^2 & \sum_{t=2m+1}^{3m} t & m \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \sum_{t=(k-2)m+1}^{(k-1)m} t^k & \sum_{t=(k-2)m+1}^{(k-1)m} t^{k-1} & \sum_{t=(k-2)m+1}^{(k-1)m} t^{k-2} & \cdots & \sum_{t=(k-2)m+1}^{(k-1)m} t^2 & \sum_{t=(k-2)m+1}^{(k-1)m} t & m \\ \sum_{t=(k-1)m+1}^{km} t^k & \sum_{t=(k-1)m+1}^{km} t^{k-1} & \sum_{t=(k-1)m+1}^{km} t^{k-2} & \cdots & \sum_{t=(k-1)m+1}^{km} t^2 & \sum_{t=(k-1)m+1}^{km} t & m \\ \sum_{t=km+1}^{(k+1)m} t^k & \sum_{t=km+1}^{(k+1)m} t^{k-1} & \sum_{t=km+1}^{(k+1)m} t^{k-2} & \cdots & \sum_{t=km+1}^{(k+1)m} t^2 & \sum_{t=km+1}^{(k+1)m} t & m \end{pmatrix} \begin{pmatrix} a_k \\ a_{k-1} \\ a_{k-2} \\ \vdots \\ a_2 \\ a_1 \\ a_0 \end{pmatrix} = \begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ \vdots \\ S_{k-1} \\ S_k \\ S_{k+1} \end{pmatrix} \quad (8)$$

通过求解矩阵方程(8),即可得到参数 a_k, a_{k-1}, \dots, a_0 的值,然后代入 k 阶曲线方程即可对数据进行拟合、预测。

- 当 $k=1$ 时,求解线性趋势曲线 $Y_t = a_1 t + a_0$ 参数的表达式。首先,有

$$\begin{cases} S_1 = a_1 \sum_{t=1}^m t + a_0 m, \\ S_2 = a_1 \sum_{t=m+1}^{2m} t + a_0 m. \end{cases} \quad (9)$$

通过对方程(9)的求解得到,

$$\begin{cases} a_1 = \frac{S_2 - S_1}{m^2}, \\ a_0 = \frac{S_1}{m} - \frac{(m+1)(S_2 - S_1)}{2m^2}. \end{cases} \quad (10)$$

- 当 $k=2$ 时,求解线性趋势曲线 $Y_t = a_2 t^2 + a_1 t + a_0$ 参数的表达式。首先,有

$$\begin{cases} S_1 = a_2 \sum_{t=1}^m t^2 + a_1 \sum_{t=1}^m t + a_0 m, \\ S_2 = a_2 \sum_{t=m+1}^{2m} t^2 + a_1 \sum_{t=m+1}^{2m} t + a_0 m, \\ S_3 = a_2 \sum_{t=2m+1}^{3m} t^2 + a_1 \sum_{t=2m+1}^{3m} t + a_0 m. \end{cases} \quad (11)$$

通过对方程(11)的求解得到,

$$\begin{aligned} a_2 &= \frac{(6S_1 - 6S_3)m(3m+1) - (6S_1 - 6S_2)m(5m+1) - 6m(m+1)(S_2 - S_3)}{2m^3(-12m^2 + m + 1)}, \\ a_1 &= \frac{2m(S_1 - S_2)(38m^2 + 15m + 1) - 2m(S_1 - S_3)(14m^2 + 9m + 1)}{2m^3(-12m^2 + m + 1)} \\ &\quad + \frac{4m(S_2 - S_3)(m+1)\left(m + \frac{1}{2}\right)}{2m^3(-12m^2 + m + 1)}, \\ a_0 &= \frac{-m[S_1(3m+1) - S_2(m+1)](38m^2 + 15m + 1)}{2m^4(-12m^2 + m + 1)} \\ &\quad + \frac{m[S_1(5m+1) - S_3(m+1)](14m^2 + 9m + 1)}{2m^4(-12m^2 + m + 1)} \\ &\quad - \frac{2m^2[S_2(5m+1) - S_3(3m+1)](m+1)\left(m + \frac{1}{2}\right)}{2m^4(-12m^2 + m + 1)}. \end{aligned}$$

4. 数值验证

首先, 我们给出衡量模型精度的表达式, 根据计算值与实际值确定绝对百分误差(APE)和平均绝对百分误差(MAPE)如下。

$$\text{APE} = \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100\%, t = 1, 2, \dots, n, \quad (12)$$

$$\text{MAPE} = \frac{1}{v-l+1} \sum_{t=l}^v \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100\%, v \leq n, \quad (13)$$

在表达式(13)中, 当 $l=2, v=(k+1)m$, MAPE 为拟合误差, 记为 MAPE_{fit} ; 当 $l=(k+1)m+1, v=n$, MAPE 为预测误差, 记为 $\text{MAPE}_{\text{fore}}$; 当 $l=2, v=n$, MAPE 为总误差, 记为 $\text{MAPE}_{\text{total}}$, 表达式中 Y_t 为计算得到的值, $(k+1)m$ 为用于建模的个数, n 为原始序列的总个数。

例 1: 本例中的数据来源于文献[2]第 167 页表 6~7 的数据, 作者在书中利用 Excel 分析工具对 2005 年至 2015 年我国国内生产总值数据建立了线性回归模型。作者们给出的线性回归方程的表达式为

$$Y_t^{\text{book}} = 52305.1982t + 112731.7927. \quad (14)$$

在本文中, 我们对上述数据采用我们提出的方法也进行了研究, 其中建模的数据 $2m=10$, 经过计算得到方程的表达式为

$$Y_t = 53933.9440t + 104911.7880。 \quad (15)$$

计算得到的相关数据见表 1 和图 1。从表 1 可以看出，我们的结果与作者的结果非常的接近，其中书上的建模误差，拟合误差和预测误差分别为 2.4021%，1.6818%和 2.3300%。我们的建模误差，拟合误差和预测误差分别为 1.8471%，3.1738%和 1.9798%。

Table 1. The application of linear trend curve in China's GDP

表 1. 线性趋势曲线在我国国内生产总值中的应用

年份	GDP	Y_t	APE(%)	Y_t^{book}	APE(%)
2005	183867.9	158845.7320	13.6088	165036.9909	10.2415
2006	210871.0	212779.6760	0.9051	217342.1891	3.0688
2007	270232.3	266713.6200	1.3021	269647.3873	0.2164
2008	319515.5	320647.5640	0.3543	321952.5855	0.7627
2009	349081.4	374581.5080	7.3049	374257.7837	7.2122
2010	413030.3	428515.4520	3.7492	426562.9819	3.2764
2011	489300.6	482449.3960	1.4002	478868.1801	2.1321
2012	540367.4	536383.3400	0.7373	531173.3783	1.7014
2013	595244.4	590317.2840	0.8277	583478.5765	1.9766
2014	643974.0	644251.2280	0.0430	635783.7747	1.2718
2015	676708.0	698185.1720	3.1738	688088.9729	1.6818
	MAPE _{fit} (%)		1.8471		2.4021
	MAPE _{fore} (%)		3.1738		1.6818
	MAPE _{total} (%)		1.9798		2.3300

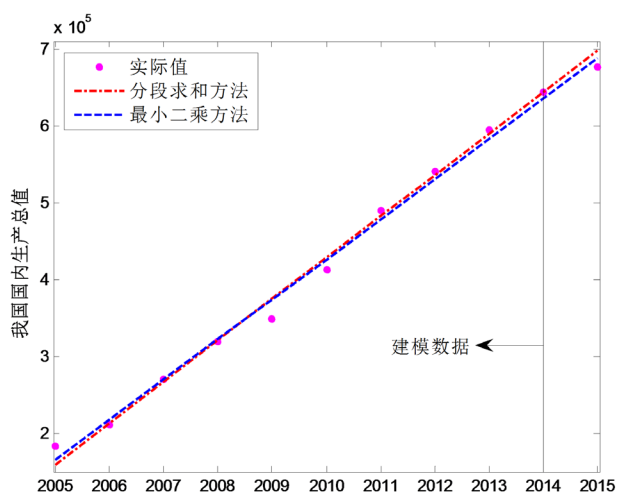


Figure 1. The linear trends of the China's GDP

图 1. 国内生产总值线性趋势图

例 2: 本例中的数据来源于文献[1]第 345 页例 13.11 的数据，作者在书中利用 Excel 分析工具对 1990 年至 2009 年金属切削机床产量建立了四阶趋势曲线模型。作者们给出的四阶趋势曲线方程的表达式为：

$$Y_t^{book} = -0.0078t^4 + 0.3268t^3 - 4.2017t^2 + 19.1011t - 4.0453。 \quad (16)$$

在本文中，我们对上述数据采用我们提出的方法也进行了研究，其中建模的数据 $4m = 20$ ，经过计算得到方程的表达式为

$$Y_t = -0.0084t^4 + 0.3557t^3 - 4.6523t^2 + 21.6510t - 7.6325。 \quad (17)$$

计算得到的相关数据见表 2 和图 2。从表 2 可以看出，我们的结果与作者的结果非常的接近，其中书上的误差为 8.9661%，而我们的误差为 8.4582%。

Table 2. Application of the fourth order trend curve in the output of metal cutting machine
表 2. 四阶趋势曲线在金属切削机床产量中的应用

年份	产量(万台)	Y_t	APE(%)	Y_t^{book}	APE(%)
1990	13.5	9.7135	28.0481	11.1731	17.2363
1991	16.4	19.7713	20.5564	19.8397	20.9738
1992	22.9	24.3717	6.4266	23.6345	3.2074
1993	26.2	25.1435	4.0323	24.0503	8.2050
1994	20.7	23.5133	13.5907	22.3927	8.1773
1995	20.3	20.7051	1.9955	19.7801	2.5611
1996	17.7	17.7410	0.2315	17.1437	3.1429
1997	18.7	15.4407	17.4296	15.2275	18.5695
1998	11.9	14.4216	21.1901	14.5883	22.5908
1999	14.2	15.0991	6.3316	15.5957	9.8289
2000	17.7	17.6860	0.0788	18.4321	4.1362
2001	25.6	22.1932	13.3076	23.0927	9.7941
2002	30.9	28.4292	7.9962	29.3855	4.9013
2003	30.6	36.0001	17.6474	36.9313	20.6905
2004	48.7	44.3100	9.0143	45.1637	7.2614
2005	51.1	52.5607	2.8585	53.3291	4.3622
2006	57.3	59.7516	4.2785	60.4867	5.5614
2007	64.7	64.6800	0.0308	65.5085	1.2496
2008	71.7	65.9410	8.0320	67.0793	6.4445
2009	58.6	61.9273	5.6780	63.6967	8.6974
MAPE _{total} (%)			8.4582		8.9661

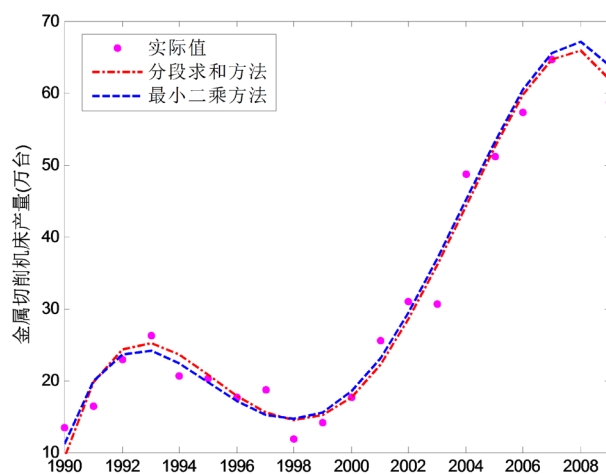


Figure 2. The fourth order trend curve in the output of metal cutting machine
图 2. 金属切削机床产量的四阶趋势曲线图

5. 结论

本文讨论了 k 阶趋势曲线的参数的求解问题,利用分段求和的思想给出了模型中参数的矩阵表达式,并在 $k=1$ 和 $k=2$ 的情形下给出了模型中参数的具体表达式。进一步,将所提出的方法与相应教材上采用的最小二乘法方法的计算结果进行比较,从计算结果来看,两种的模型精度基本上是一致的。但是,从分析的过程中可以明确知道,我们提出的计算方法更为简单、直接。

参考文献

- [1] 贾俊平,何晓群,金勇进. 统计学[M]. 第5版. 北京: 中国人民大学出版社, 2014.
- [2] 宋廷山,王坚,刁艳华,郭思亮. 应用统计学——以 EXCEL 为分析工具[M]. 第2版. 北京: 清华大学出版社, 2008.
- [3] 周永道,王会琦,吕王勇. 时间序列分析及应用[M]. 北京: 高等教育出版社, 2016.
- [4] 段树乔,段方婕. 基于多项式回归函数的电网公司售电量预测方法[J]. 数学的实践与认识, 2012, 42(18): 76-81.
- [5] 温鲜,霍海峰,黄秋和. 基于多项式回归模型的柳州市房产价格预测[J]. 市场研究, 2013(8): 53-54.
- [6] 于嫦,姜潮. 采用多项式回归和灰色模型对辽宁省省医院药品比例的分析预测[J]. 中国卫生经济, 2006, 25(1): 72-75.
- [7] 徐颖,黄素珍. 大数据时代房地产价格预测模型的研究[J]. 中国商论, 2017(4): 134-137.
- [8] 张磊,于战果,李世民. 基于灰色多项式回归组合模型的维修器材需求预测方法研究[J]. 兵器装备工程学报, 2019, 40(1): 179-183.

知网检索的两种方式：

1. 打开知网首页：<http://cnki.net/>，点击页面中“外文资源总库 CNKI SCHOLAR”，跳转至：<http://scholar.cnki.net/new>，搜索框内直接输入文章标题，即可查询；
或点击“高级检索”，下拉列表框选择：[ISSN]，输入期刊 ISSN：2324-7991，即可查询。
2. 通过知网首页 <http://cnki.net/>顶部“旧版入口”进入知网旧版：<http://www.cnki.net/old/>，左侧选择“国际文献总库”进入，搜索框直接输入文章标题，即可查询。

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：aam@hanspub.org