

The Analysis of Profit Trend by Using the Least Square Method

Xianqing Zhu, Chaoying Zhuo, Yuqi Teng, Xiaojie Yu, Shuping Chen*

College of Mathematics, Xiamen University of Technology, Xiamen Fujian
Email: *shupingchen@126.com

Received: Jun. 28th, 2020; accepted: Jul. 9th, 2020; published: Jul. 16th, 2020

Abstract

Zhejiang Wanma Co., Ltd., as a well-known enterprise in cable industry, has developed rapidly in recent years. In this paper, the Newton interpolation method and polynomial curve fitting method are used respectively to establish the corresponding mathematical model by using more than 10 groups of annual gross profit data of the company. By comparing the errors of the two models, it is found that the mathematical model of polynomial curve fitting is better. Then, the total profit data of the company in 2018 and 2019 are verified in order to further test the validity of the model, and the results show that the fitting effect is good. Finally, the total annual profit of the company in the next two years is forecasted through the above model.

Keywords

Polynomial Curve Fitting, Newton Interpolation, Profit Growth, Error

基于最小二乘法的利润变动趋势分析

朱娴卿, 卓超英, 滕钰琪, 余晓洁, 陈淑萍*

厦门理工学院, 应用数学学院, 福建 厦门
Email: *shupingchen@126.com

收稿日期: 2020年6月28日; 录用日期: 2020年7月9日; 发布日期: 2020年7月16日

摘要

浙江万马股份有限公司作为电缆行业知名企业, 近年来发展迅速, 本文以该公司10余组年利润总额数据为样本, 分别利用牛顿插值与多项式曲线拟合建立相应的数学模型, 并通过比较两种模型各自预测值与

*通讯作者。

实际值之间的误差, 得出多项式曲线拟合构造的数学模型相对更优的结论。接着, 为了进一步检验该模型的有效性, 本文对该公司2018年和2019年的年利润总额数据进行了验证, 结果显示拟合效果良好。最后, 本文利用上述模型对该公司未来2年的年利润总额进行了预测。

关键词

多项式拟合, 牛顿插值, 利润增长, 误差

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

利润是反映一个企业盈利能力的重要指标, 其增减反映着企业的发展状况。浙江万马股份有限公司, 作为电缆行业知名上市公司, 在近年国家基础设施建设大量投入、房地产市场大发展、民生需求逐年提升的背景下, 该公司的经营规模实现了逐年提高。公司在电缆业务稳健发展的前提下, 大力发展新能源业务, 2019年实现营业收入97.5亿元, 是2011年的2.7倍。研究与分析该公司利润增长情况, 不仅有利于预测其市场运营态势与发展趋势, 而且对于该企业寻求最佳的盈利途径及规划目标利润具有积极作用。

曲线拟合在人口预测、疾病感染及工业生产等方面具有非常重要的作用。2008年, 从志鹏在文献[1]中利用傅里叶级数的数学模型来拟合故障电压与电流, 并用残差、残差平方和等误差分析方法检验拟合效果, 确定最佳的拟合方程。郭飞等在文献[2]中, 利用傅里叶级数拟合偏差曲线, 实现了对顶管施工中轴线控制措施的实时反馈, 这将对此类工程的施工具有一定的参考意义。2017年, 刘琴琴在文献[3]中, 利用改进后的B样条曲线拟合算法解决了模型构建时因追求拟合精度而违背毒性物质对发光细菌抑制性作用等问题。

2010年, 陈光在文献[4]中利用MATLAB计算软件包对某水厂的投矾系统进行了分析与研究, 并根据大量历史数据建立了原水水质参数与实际投矾量之间的拟合曲线, 并建立了一个新型的智能化自适应水处理投矾系统, 该系统为曲线拟合在水厂投矾方面的应用做了有益的探索。2015年, 盛立程等在文献[5]中, 根据某地区历年数据, 构建了基于最小二乘法的多项式拟合预测模型, 仿真结果表明, 该方法拟合效果良好, 且具有良好的预测功能, 反映了多项式拟合对电力系统用电量的预测有较大的应用价值。昌霞等在文献[6]中收集了昆明市西山区山洪预警监测区域内13个监测点实测得到的降雨量数据, 利用曲线拟合模型, 对降雨数据进行不同的曲线拟合分析, 结果表明幂函数的拟合效果最好。成高飞等在文献[7]中通过多项式拟合, 得到通风机的风压风量方程, 并与实测数据进行对比, 得出该拟合多项式能够适用于矿井实际风压风量的计算。

2017年, 杨伟在文献[8]中通过牛顿插值法重构信号, 得到其误差是可以估计的, 并通过模拟实验发现, 通过牛顿插值法恢复出的信号比香农定理更加准确。2019年, 戴仁辉等在文献[9]中利用牛顿插值方法对湖泊沉积软土的压实特性进行分析, 通过常规实验表明, 该方法对不同击数下软土的含水率和干密度具有较好的拟合效果。

目前, 利用多项式拟合与插值研究上市公司利润增长效应的文献还不是很多。本文将利用曲线拟合和Newton插值方法对浙江万马股份有限公司最近几年的年利润总额进行整理与分析, 得到了最优的数学

模型, 并对该公司未来两年的年利润总额进行了预测。

2. 基本概念

2.1. 曲线拟合的基本理论

在实际工作中, 不是所有变量都存在线性关系。曲线拟合是指选择适当的曲线类型来拟合观测数据, 并对拟合数据与观测数据, 分析变量之间的关系。

文献[10]描述了曲线拟合的基本原理, 此算法的本质是用光滑的曲线描述离散的点, 反映出离散点之间的变化规律。多项式拟合又称为最小二乘法拟合, 即对于给定的 $n+1$ 个数据点 (x_i, y_i) , 构造如下形式的拟合函数:

$$\varphi(x) = c_m x^m + c_{m-1} x^{m-1} + \cdots + c_1 x + c_0, \quad (1)$$

其中 c_i 表示此多项式的系数, m 表示多项式的次数。当 $m=1$ 时, 称之为线性拟合或直线拟合。由极值理论可知, 参数 c_1, c_2, \dots, c_m 需要满足如下方程:

$$\begin{bmatrix} n+1 & \sum_{i=0}^n x_i & \cdots & \sum_{i=0}^n x_i^m \\ \sum_{i=0}^n x_i & \sum_{i=0}^n x_i^2 & \cdots & \sum_{i=0}^n x_i^{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=0}^n x_i^m & \sum_{i=0}^n x_i^{m+1} & \cdots & \sum_{i=0}^n x_i^{2m} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_m \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^n y_i \\ \sum_{i=0}^n y_i x_i \\ \vdots \\ \sum_{i=0}^n y_i x_i^m \end{bmatrix}, \quad (2)$$

公式(2)称为求最小二乘拟合多项式的法方程。

2.2. 牛顿插值多项式的基本理论

函数 $f(x)$ 在区间 $[a, b]$ 的 $n+1$ 个相异的插值节点 x_0, x_1, \dots, x_n 处的函数值为 $f(x_i) (i=0, 1, 2, \dots, n)$, 则称

$$f[x_0, x_1, \dots, x_{k-1}, x_k] = \frac{f[x_1, x_2, \dots, x_k] - f[x_0, x_1, \dots, x_{k-1}]}{x_k - x_0} \quad (3)$$

为 $f(x)$ 关于插值节点 x_0, x_1, \dots, x_k 的 k 阶差商, 其中

$$\begin{aligned} f[x_k] &= f(x_k) \\ f[x_k, x_{k+1}] &= \frac{f[x_{k+1}] - f[x_k]}{x_{k+1} - x_k} \end{aligned} \quad (4)$$

依此类推可以得到各阶差商, 且次数不超过 n 次的牛顿插值多项式为:

$$P(x) = \sum_{i=0}^n f[x_0 \cdots x_i] (x - x_0) \cdots (x - x_{i-1}). \quad (5)$$

2.3. 曲线拟合的优劣评判标准

一组数据的拟合曲线不是惟一的, 所以针对拟合结果, 还需要进行误差分析, 找到拟合效果最佳的拟合曲线。误差分析有以下几种情况:

1) 相对误差(Relative Error)

$$\text{相对误差} = \frac{\text{拟合值} - \text{真实值}}{\text{真实值}} \times 100\% . \quad (6)$$

相对误差越小, 说明拟合值越接近真实值, 即拟合效果越好。

2) 残差平方和(SSE)

残差平方和是用连续曲线近似描述平面上离散点组, 其公式为:

$$\text{SSE} = \sum_{i=1}^n w_i (y_i - \hat{y}_i)^2 , \quad (7)$$

其中 \hat{y}_i 为估计值, w_i 是相应的权。一组数据 SSE 越小, 则拟合效果越好, 数据预测越成功。

3) 确定系数(R-square)

确定系数表示一个随机变量与多个随机变量关系的数字特征, 其公式为:

$$\text{R-square} = \frac{\sum_{i=1}^n w_i (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n w_i (y_i - \bar{y}_i)^2} = 1 - \frac{\sum_{i=1}^n w_i (y_i - \hat{y}_i)^2}{\sum_{i=1}^n w_i (y_i - \bar{y}_i)^2} . \quad (8)$$

确定系数的取值范围为 $[0,1]$, 当 R-square 越接近 1, 则拟合效果越好, 反之, 其拟合的参考价值越低。

4) 均方根误差(RMSE)

均方根误差又称标准误差, 是拟合值与真实值偏差的平方和与观测值次数 n 比值的平方根, 计算公式为:

$$\text{均方根误差} = \sqrt{\frac{\text{SSE}}{n}} = \sqrt{\frac{\sum_{i=1}^n w_i (y_i - \hat{y}_i)^2}{n}} . \quad (9)$$

均方根误差对一组测量值中的极值误差特别敏感, 能反映一个数据集的离散程度, 故可很好地反映出测量的精度。

3. 数据的收集和处理

3.1. 公司概况

浙江万马股份有限公司, 起源于 1992 年创建的浙江万马电缆厂, 是国家大型企业浙江万马集团的核心企业。该公司在电缆业务稳健发展的前提下, 大力发展新能源业务。在新能源产业下有万马新能源、爱充网和万马新能源投资三家全资子公司, 2014~2015 年间连续 5 次中标国网充电设备公开采集的企业, 是我国国内较早从事电动汽车充电设备研发、制造和充电站智能运营管理的高科技企业, 更是在 2016 年成为了 G20 杭州峰会指定充电桩品牌。如今, 万马新能源已研发出新型环保材料制造的充电设备、智能化充电设备、多功能一体化交直流充电设备等多项产品, 是智能充电桩方面的领军企业。

3.2. 浙江万马股份有限公司利润总额数据的收集和处理

如表 1 所示, 本文以该公司 2006 年至 2017 年的年利润总额数据作为拟合基础数据, 2018 年和 2019 年的年利润总额数据用于与预测所得的数据进行对比, 以此来验证拟合曲线的拟合度。

Table 1. The total profit of Zhejiang Wanma Co., Ltd. over the years
表 1. 浙江万马股份有限公司历年利润总额数据

| 时间 | 真实值(单位: 亿元) |
|--------|-------------|
| 2006 年 | 0.640165 |
| 2007 年 | 0.750346 |
| 2008 年 | 0.854761 |
| 2009 年 | 1.005452 |
| 2010 年 | 1.113813 |
| 2011 年 | 1.200822 |
| 2012 年 | 1.270236 |
| 2013 年 | 1.366534 |
| 2014 年 | 1.114711 |
| 2015 年 | 1.298535 |
| 2016 年 | 1.252103 |
| 2017 年 | 1.149223 |
| 2018 年 | 1.176168 |
| 2019 年 | 1.464943 |

为了更直观地了解该公司利润总额的变化趋势, 我们结合数据, 做出其各年度利润总额的散点图(见图 1), 以年份为横坐标, 其中 2006 年为第一年记为 1, 以年利润总额为纵坐标。

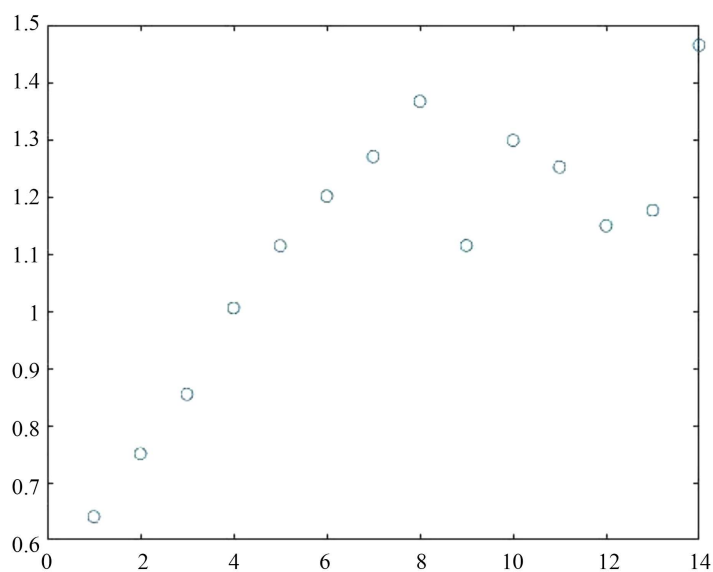


Figure 1. Scatter chart of total profit of Zhejiang Wanma Co., Ltd. over the years

图 1. 浙江万马股份有限公司历年利润总额数据散点图

借助拟合工具箱 *cftool* 对基础数据进行不同阶数的多项式拟合, 以此找到拟合精度最高的拟合多项式, 各阶数多项式的误差分析如表 2 所示。

Table 2. Root mean square error of polynomial curve fitting
表 2. 各阶拟合曲线函数的均方根误差

| 阶数(n) | 均方根误差(RMSE) | 确定系数(R-square) | 残差平方和(SSE) |
|-----------|-------------------|-------------------|-------------------|
| 1 | 0.135886026480219 | 0.647096972483080 | 0.203115134118411 |
| 2 | 0.062277046336955 | 0.924953727860717 | 0.043193263990705 |
| 3 | 0.062071188349112 | 0.926364726303119 | 0.042381156653780 |
| 4 | 0.059430196532827 | 0.932497459467521 | 0.038851430859234 |
| 5 | 0.057988152035917 | 0.935733548151986 | 0.036988883541947 |
| 6 | 0.053051501014020 | 0.946210050363395 | 0.030959079358246 |
| 7 | 0.051495462270489 | 0.949319171292473 | 0.029169608978965 |

通过比较数据可得, 阶数越高, 该公司的利润总额数据拟合精度越高, 故当多项式的最高次 $n = 4, 5, 6, 7$ 时, 所得到的拟合效果可能最佳, 图 2 为基础数据与 4 至 7 阶拟合曲线的对比图。

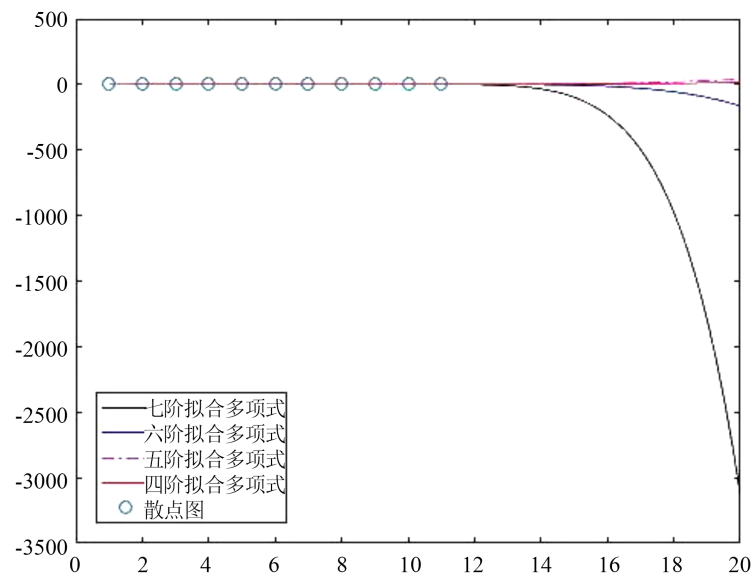


Figure 2. Curve fitting for each order
图 2. 各阶拟合基础数据曲线

为验证拟合多项式的拟合度, 我们对该公司 2018 年和 2019 年的年利润总额进行预测, 可以得到各阶多项式所对应的误差(见表 3):

Table 3. Root mean square error of fitting functions
表 3. 5、6、7 阶拟合函数的均方根误差

| 阶数 | 残差平方和(SSE) |
|----|--------------------|
| 4 | 0.060334623441549 |
| 5 | 0.581160115727685 |
| 6 | 8.936825632326258 |
| 7 | 37.757447636630403 |

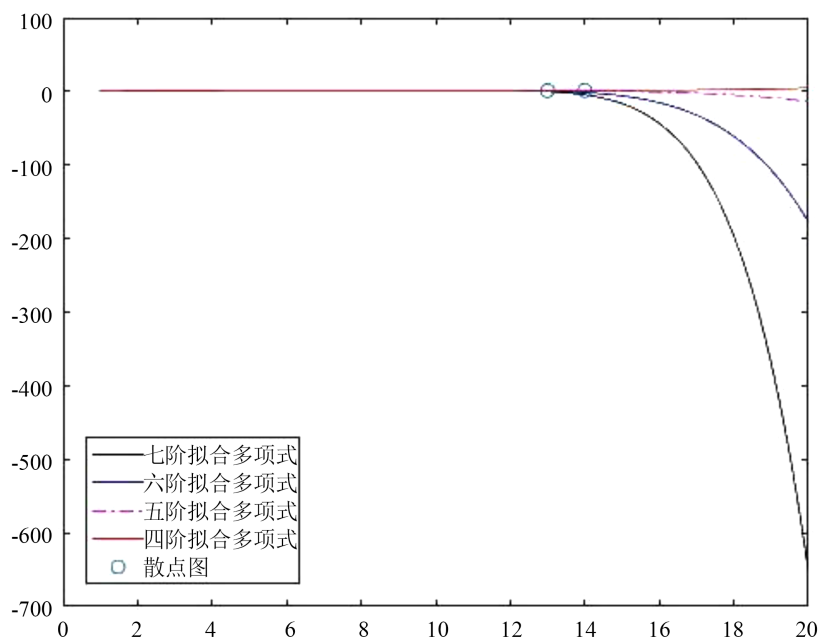


Figure 3. The testing data points and curve fitting for each order
图 3. 各阶拟合曲线与验证数据点

由上述图 3 和表 3 可知，当 $n = 4$ 时均方根误差最小，此时预测结果较佳。则可得到其拟合函数为：

$$f(x) = p_1x^4 + p_2x^3 + p_3x^2 + p_4x + p_5 \tag{10}$$

且 p_i 的值为：

$$\begin{aligned} p_1 &= 0.000193640078671 \\ p_2 &= -0.005299428758741 \\ p_3 &= 0.037509425480770 \\ p_4 &= 0.028027214597900 \\ p_5 &= 0.577940590909094 \end{aligned}$$

接着，本文选择 5 个插值节点，得到如下 4 次牛顿插值多项式：

$$g(x) = -0.000104533x^4 + 0.0007074x^3 - 0.00550247x^2 + 0.153305x + 0.46176 \tag{11}$$

为了比较两种算法的优劣，表 4 分别给出了 2018 年和 2019 年利润总额的误差。

Table 4. The error of total profit for two algorithms in 2018 and 2019

表 4. 2018 年和 2019 年两种算法利润总额的误差

| 时间 | 真实值 | 插值方法的预测值 | 插值方法的误差 | 多项式拟合预测值 | 多项式拟合的误差 |
|--------|----------|-----------|-----------|----------|----------|
| 2018 年 | 1.176168 | 0.0933984 | 1.0827696 | 1.169097 | 0.007071 |
| 2019 年 | 1.464943 | -0.545088 | 2.010031 | 1.219414 | 0.245529 |

由图 4 和表 4 可知，多项式的拟合误差较小，即多项式拟合的效果较好。故根据式(10)，可预测该公司 2020 年和 2021 年的利润总额，其结果如表 5。

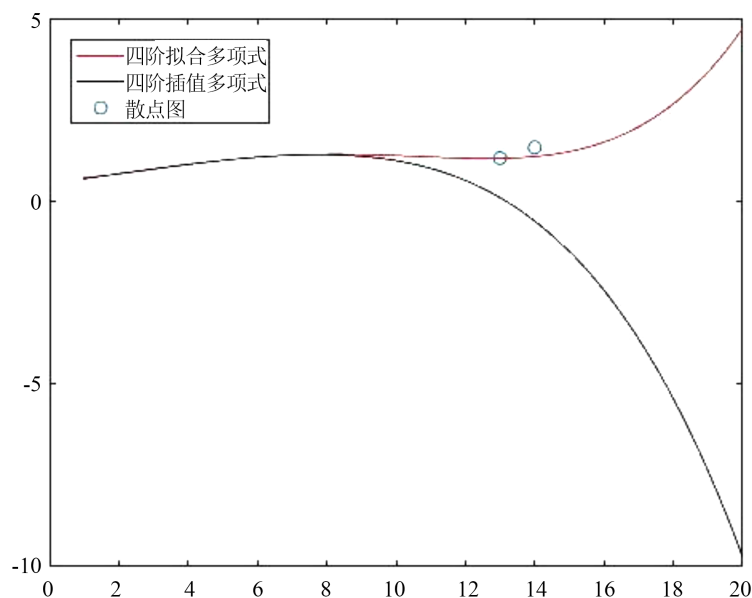


Figure 4. The testing data points and polynomial curve of fourth order
图 4. 4 阶多项式曲线与验证数据点

Table 5. Total profit forecast for 2020 and 2021

表 5. 2020 年和 2021 年的利润总额预测值

| 时间 | 预测值(单位: 亿元) |
|--------|-------------------|
| 2020 年 | 1.355426465034970 |
| 2021 年 | 1.612724947552455 |

4. 总结

本文基于多项式拟合与 Newton 插值方法分别对浙江万马股份有限公司 2006 年至 2017 年的年度利润总额数据进行了分析, 并根据确定系数与残差平方和等相关指标, 得到最优拟合曲线。同时, 为了进一步验证所选模型的优劣, 本文还利用 Newton 插值方法构造了牛顿插值多项式, 理论结果显示本文所选的数学模型是最优的。最后, 本文利用所选的数学模型对公司未来两年的年利润总额进行了预测。其数值结果表明, 公司的年度利润总额呈现出持续上升的趋势。希望本课题的研究能给相关企业利润预测提供理论参考。

参考文献

- [1] 从志鹏. 基于曲线拟合方法的故障测距应用研究[D]: [硕士学位论文]. 济南: 山东大学, 2008.
- [2] 郭飞, 罗川舟. 基于曲线拟合的大直径顶管施工技术研究与应[J]. 中外建筑, 2019(1): 191-192.
- [3] 刘琴琴. 曲线拟合预测模型及算法在水质远程监测系统中的应用研究[D]: [硕士学位论文]. 杭州: 浙江理工大学, 2017.
- [4] 陈光. 基于曲线拟合的智能投矾系统的应用研究[D]: [硕士学位论文]. 广州: 中山大学.
- [5] 盛立程, 曾喆昭, 李莎. 基于代数多项式模型的用电量预测研究[J]. 电力科学与技术学报, 2015(1): 35-38.
- [6] 吕霞, 刘赛娥. 基于曲线拟合的降雨侵蚀力模型初探[J]. 计算机时代, 2019(2): 71.
- [7] 成高飞, 郭媛. 基于多项式曲线拟合方程法的矿用风机风压特性曲线的对比研究[J]. 山西煤炭, 2019, 38(1): 48-50.

- [8] 杨伟. 基于牛顿插值法的模拟信号采样与重构研究[D]: [硕士学位论文]. 兰州: 兰州大学.
- [9] 戴仁辉, 陈士军, 王增亮, 易进翔. 基于牛顿插值法的软土压实特性应用研究[J]. 工程质量, 2019, 37(3): 70-71.
- [10] 裴玉茹, 马庚宇. 数值分析[M]. 第二版. 北京: 机械工业出版社, 2014.