

# An Analysis of the Correlation between GDP and the Primary Industry, Secondary Industry and Tertiary Industry in China

Zhongyang Liu

Shandong Qufu Normal University, Qufu Shandong  
Email: liuzy1202@163.com

Received: Mar. 4<sup>th</sup>, 2020; accepted: Mar. 19<sup>th</sup>, 2020; published: Mar. 26<sup>th</sup>, 2020

---

## Abstract

According to the data about primary, secondary and tertiary industry in 1990-2012 from big data platform, this paper sets up a simple linear regression model ( $\hat{y} = -59594.98 + 11.083x_1$ ) between GDP and the adding value of the primary industry by using the statistical software R and SPSS. Then the test of Correlation Coefficient shows that the model has heteroscedasticity and autocorrelation which were eliminated by "one variable weighted least square estimation" and FDM. So the model is improved. Then this paper uses the GDP and the first, the second, the third industry to establish the multiple linear regression model. Due to the significance of the model is not ideal, this paper uses the stepwise regression method to screen the independent variables and finally establishes the multiple linear regression model ( $\hat{y} = 5377.346 + 1.216x_2 + 0.976x_3$ ) between GDP and the secondary industry and tertiary industry. Finally, based on the established model, we can use the future changes in the primary, secondary and tertiary industries to reasonably forecast the GDP growth rate.

## Keywords

Simple Linear Regression, Multiple Linear Regression, Autocorrelation, Stepwise Regression Method

---

# 我国GDP与第一、二、三产业的相关关系分析

刘仲阳

山东省曲阜师范大学, 山东 曲阜  
Email: liuzy1202@163.com

收稿日期: 2020年3月4日; 录用日期: 2020年3月19日; 发布日期: 2020年3月26日

## 摘要

文章通过大数据平台搜集到的1990~2012年GDP与第一、二、三产业的增加值,充分利用统计软件R与SPSS,建立了GDP与第一产业增加值的一元线性回归模型  $\hat{y} = -59594.98 + 11.083x_1$ 。之后通过相关系数检验得出模型存在异方差性与自相关性,通过一元加权最小二乘估计消除了异方差性,同时通过差分法消除了误差项自相关性,从而对模型进行了改进。又通过GDP与第一、二、三产业建立了多元线性回归模型。由于模型的显著性水平不理想,文章利用逐步回归法对自变量进行了筛选,最终建立了GDP与第二、三产业的线性回归模型  $\hat{y} = 5377.346 + 1.216x_2 + 0.976x_3$ 。最终通过所建立的模型,可以通过未来第一二三产业的变化来对GDP增速进行合理的预测。

## 关键词

一元线性回归, 多元线性回归, 差分法, 逐步回归法

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

众所周知,国内生产总值(GDP)是指按市场价格计算的一个国家(或地区)所有常住单位在一定时期内生产活动的最终成果,常被公认为衡量国家经济状况的最佳指标。国内生产总值 GDP 是核算体系中一个重要的综合性统计指标,也是我国新国民经济核算体系中的核心指标,它反映了一国(或地区)的经济实力和市场规模。

而第一产业是指通过人类劳动直接从自然界取得产品的部门。我国规定专指农、林、牧、渔业。第二产业是指对第一产业和本产业提供的产品(原料)进行加工的部门,我国规定专指采矿业,制造业,电力、燃气及水的生产和供应业,建筑业。第三产业是指对消费者提供最终服务和对生产者(包括三个产业的生产者)提供中间服务的部门。指除第一、二产业以外的其他行业。这三种产业深深影响着 GDP 的增速,因此弄清楚四者之间的关系,对于我国今后的经济增长起着有力的推动作用。

## 2. 一元回归模型的建立

### 2.1. 普通最小二乘估计

#### 2.1.1. 普通最小二乘回归方程求解

根据相关文献[1]及生活经验,我们容易得出,第一产业增加值  $x_1$  对 GDP  $y$  有着较为显著的影响,我们通过大数据平台[2]获得了 1990~2012 年国内 GDP、第一产业增加值的数据。

由普通最小二乘法

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (3-1)$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (3-2)$$

(公式里  $x_i$  对应第一产业  $x_1$  的相应值,  $y_i$  对应 GDP 的相应值)。

利用 SPSS 软件[3], 输出结果见表 1:

**Table 1.** The output of ordinary least squares regression to primary sector of the economy and GDP  
**表 1.** 普通最小二乘回归对第一产业增加值和 GDP 的输出结果

| 模型 |      | 系数 <sup>a</sup> |          |       |         | 显著性   |
|----|------|-----------------|----------|-------|---------|-------|
|    |      | 未标准化系数          |          | 标准化系数 | $t$     |       |
|    |      | B               | 标准错误     | Beta  |         |       |
| 1  | (常量) | -59594.980      | 5723.248 |       | -10.413 | 0.000 |
|    | 第一产业 | 11.083          | 0.235    | 0.995 | 47.161  | 0.000 |

<sup>a</sup>. 因变量: GDP。

其中回归系数显著性检验  $P = 8.5634E - 23$ , 远小于显著性水平  $\alpha = 0.05$ , 因此  $y$  与  $x_1$  的线性回归高度显著。

由系数一栏的输出结果可以得出,  $\hat{\beta}_0 = -59594.980$ ,  $\hat{\beta}_1 = 11.083$ , 且二者  $P$  值均小于显著性水平  $\alpha = 0.05$ , 因此二者均高度显著。所以 GDP  $y$  与第一产业  $x_1$  所建立的普通最小二乘回归方程为

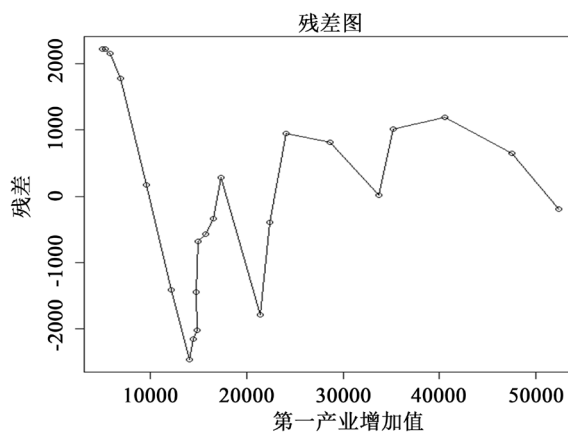
$$\hat{y} = -59594.98 + 11.083x_1 \quad (3-3)$$

这表明第一产业  $x_1$  每增加一个单位, 相应的 GDP 平均增加 11.083 个单位。

### 2.1.2. 残差分析

由上述建立的一元线性回归方程, 利用 SPSS 软件, 计算出每一项所对应的残差  $e_i = y_i - \hat{y}_i$  的值。

以残差为纵坐标, 第一产业增加值为横坐标, 利用 R 软件, 画出图像见图 1:



**Figure 1.** Residuals for different primary sector of the economy  
**图 1.** 不同第一产业增加值的残差图

由以上残差图可以看出, 残差之均匀的散落在直线  $y = 0$  两侧, 但存在一定的异方差性, 说明以上的普通最小二乘法存在一定的局限性。

## 2.2. 一元加权最小二乘估计

### 2.2.1. 异方差性的检验

利用等级相关系数, 通过 SPSS 软件, 所得输出结果见表 2:

**Table 2.** Rank correlation coefficient output  
**表 2.** 等级相关系数输出结果

|          |       | 相关性      |          |          |
|----------|-------|----------|----------|----------|
|          |       |          | 第一产业     | 残差绝对值    |
| 斯皮尔曼 Rho | 第一产业  | 相关系数     | 1.000    | -0.595** |
|          |       | Sig.(双尾) | 0.0      | 0.003    |
|          |       | N        | 23       | 23       |
|          | 残差绝对值 | 相关系数     | -0.595** | 1.000    |
|          |       | Sig.(双尾) | 0.003    | 0.0      |
|          |       | N        | 23       | 23       |

\*\*在 0.01 级别(双尾), 相关性显著。

由输出结果可以看出, 相关系数  $r_s = -0.595$ ,  $P$  值=0, 认为绝对残差值  $|e_i|$  与自变量  $x_i$  显著相关, 故误差项存在异方差性。

### 2.2.2. 一元加权最小二乘模型的建立

针对上述存在的异方差性问题, 我们采用加权的一元最小二乘估计来改善模型。下面我们对公式(3-1)、(3-2)进行一定的变换

$$\hat{\beta}_{0w} = \bar{y}_w - \hat{\beta}_{1w} \bar{x}_w \tag{3-4}$$

$$\hat{\beta}_{1w} = \frac{\sum_{i=1}^n w_i (x_i - \bar{x}_w)(y_i - \bar{y}_w)}{\sum_{i=1}^n w_i (x_i - \bar{x}_w)^2} \tag{3-5}$$

其中  $\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}$ , 为自变量的加权平均,  $\bar{y}_w = \frac{\sum w_i y_i}{\sum w_i}$  为因变量的加权平均, 权函数  $w_i = \frac{1}{x_i^m}$ ,  $m$  是待定的未知参数。

下面利用 SPSS 软件进行方程拟合:

- ① 确定幂指数  $m$  的最优值, 输出结果如表 3:

**Table 3.** Weight function output  
**表 3.** 权函数输出结果

|   |        | 对数似然值 <sup>b</sup>    |  |
|---|--------|-----------------------|--|
| 幂 | -2.000 | -249.514              |  |
|   | -1.500 | -248.324 <sup>a</sup> |  |
|   | -1.000 | -248.528              |  |
|   | -0.500 | -250.044              |  |
|   | 0.000  | -252.286              |  |
|   | 0.500  | -254.536              |  |
|   | 1.000  | -256.279              |  |
|   | 1.500  | -257.297              |  |
|   | 2.000  | -257.600              |  |

<sup>a</sup>. 选择了相应的幂进行进一步分析, 这是因为, 它使对数似然函数最大化; <sup>b</sup>. 因变量: GDP, 源变量: 第一产业。

根据以上的输出结果，幂指数最优的取值为

$$m = -1.5 \quad (3-6)$$

## ② 建立模型

根据以上选出的最优幂指数，利用 Excel 软件，首先计算出不同时间段的权重大小。

之后由公式(3-4)、(3-5)，利用 SPSS 软件计算出加权最小二乘估计的方程系数，输出结果见表 4：

**Table 4.** UNARY weighted least square estimation output

**表 4.** 一元加权最小二乘估计输出结果

| 模型 |      | 系数 <sup>a,b</sup> |          |       |        | 显著性   |
|----|------|-------------------|----------|-------|--------|-------|
|    |      | 未标准化系数            |          | 标准化系数 |        |       |
|    |      | B                 | 标准错误     | Beta  | t      |       |
| 1  | (常量) | -39785.149        | 5095.595 |       | -7.808 | 0.000 |
|    | 第一产业 | 10.002            | 0.375    | 0.986 | 26.643 | 0.000 |

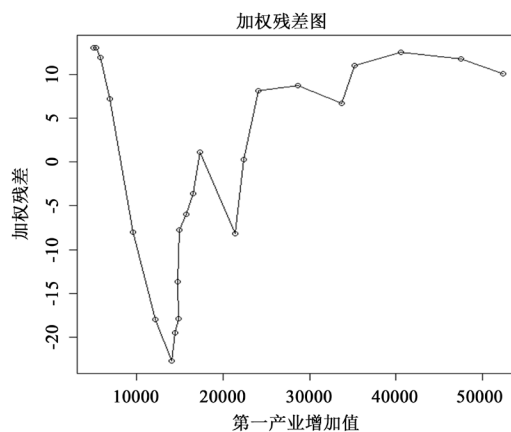
<sup>a</sup>. 因变量：GDP；<sup>b</sup>. 加权最小平方回归 - 按权重加权。

根据输出结果可以看出，回归方程显著，回归系数  $\bar{\beta}_{0w} = -39785.149$ ,  $\bar{\beta}_{1w} = 10.002$ ，且二者均显著，因此加权之后的回归方程为

$$\hat{y} = -39785.149 + 10.002x_1 \quad (3-7)$$

改进后的回归方程表明，自变量  $x_1$  每增加一个单位，相应的 GDP 平均增加 10.002 个单位。

为了更直观地对比加权前与加权后的残差变化，我们以第一产业增加值  $x_1$  为横轴，以权变换残差  $\sqrt{w_i}e_{iw}$  为纵轴画残差图见图 2：



**Figure 2.** Weighted least square estimation residual diagram

**图 2.** 加权最小二乘估计残差示意图

由图像我们能够看出明显的自相关现象，因此接下来考虑利用迭代的方法消除自相关性，来建立新的一元回归模型。

## 2.3. 迭代法建立模型

### 2.3.1. 样本相关系数的计算

利用 SPSS 软件，计算第一产业增加值与 GDP 之间的相关性，输出结果见表 5：

**Table 5.** Output table of the correlation coefficient between primary sector of the economy and GDP  
**表 5.** 第一产业增加值与 GDP 的相关系数输出表

|          |          | 相关性 |         |         |
|----------|----------|-----|---------|---------|
|          |          |     | 第一产业    | GDP     |
| 第一产业     | 相关系数     |     | 1.000   | 0.999** |
|          | Sig.(双尾) |     | 0.0     | 0.000   |
|          | N        |     | 23      | 23      |
| 斯皮尔曼 Rho | 相关系数     |     | 0.999** | 1.000   |
|          | Sig.(双尾) |     | 0.000   | 0.0     |
|          | N        |     | 23      | 23      |
| GDP      | 相关系数     |     | 0.999** | 1.000   |
|          | Sig.(双尾) |     | 0.000   | 0.0     |
|          | N        |     | 23      | 23      |

\*\*在 0.01 级别(双尾), 相关性显著。

由输出结果能够看出二者之间存在显著的相关性, 相关系数  $\rho = 0.999012$ , 非常接近 1, 因此采用一种特殊的迭代法——差分法, 消除残差自相关性。

### 2.3.2. 利用差分法消除自相关性

首先计算差分  $\Delta y_t = y_t - y_{t-1}$ ,  $\Delta x_t = x_t - x_{t-1}$ 。

然后利用 SPSS 做  $\Delta y$  对  $\Delta x$  过原点的最小二乘回归, 输出结果见表 6:

**Table 6.** The output of the regression model established by the difference method  
**表 6.** 差分法回归模型建立的输出结果

|    |          | 系数 <sup>a,b</sup> |       |       |          |       |
|----|----------|-------------------|-------|-------|----------|-------|
| 模型 |          | 未标准化系数            |       | 标准化系数 | <i>t</i> | 显著性   |
|    |          | B                 | 标准错误  | Beta  |          |       |
| 1  | 差分后的第一产业 | 0.181             | 0.059 | 0.559 | 3.088    | 0.006 |

<sup>a</sup>. 因变量: 差分后的 GDP; <sup>b</sup>. 过原点线性回归。

由输出结果可以看出, 模型显著,  $\hat{\beta}_1 = 0.181$ , 因此差分法建立的模型如下

$$\Delta y = 0.181 \Delta x \quad (3-8)$$

即

$$\hat{y}_t - \hat{y}_{t-1} = 0.181(x_t - x_{t-1}) \quad (3-9)$$

## 3. 多元回归模型的建立

### 3.1. 多元回归模型的普通最小二乘估计

#### 3.1.1. 回归模型的建立

由相关文献[4]我们知道, GDP  $y$  的变化不仅由第一产业  $x_1$  影响着, 同时还受第二产业  $x_2$ 、第三产业  $x_3$  的影响, 因此通过查阅相关资料, 我们能够获得 1990-2012 年间 GDP 以及第一、二、三产业增加值的相关数据。

解释变量有三个, 被解释变量有一个, 根据多元线性回归的回归系数公式

$$\hat{\beta} = (x^T x)^{-1} x^T y \quad (4-1)$$

利用 SPSS 软件去计算, 输出结果见表 7:

**Table 7.** The output of the coefficient calculation of multivariate regression equation  
**表 7.** 多元回归方程系数计算输出结果

|    |      | 系数 <sup>a</sup> |       |       |            |       |
|----|------|-----------------|-------|-------|------------|-------|
| 模型 |      | 未标准化系数          |       | 标准化系数 | t          | 显著性   |
|    |      | B               | 标准错误  | Beta  |            |       |
| 1  | (常量) | 0.008           | 0.038 |       | 0.213      | 0.833 |
|    | 第一产业 | 1.000           | 0.000 | 0.090 | 159741.712 | 0.000 |
|    | 第三产业 | 1.000           | 0.000 | 0.447 | 463696.876 | 0.000 |
|    | 第二产业 | 1.000           | 0.000 | 0.464 | 404016.996 | 0.000 |

a. 因变量: GDP。

由输出结果可以看出, 常数变量的显著性  $P = 0.833$  远远大于显著性水平  $\alpha = 0.05$ , 因此回归方程的拟合效果并不理想。

下面考虑自变量之间的相关性(表 8)。

**Table 8.** Independent variable correlation test output  
**表 8.** 自变量相关性检验输出结果

|          |      | 相关性      |         |         |         |
|----------|------|----------|---------|---------|---------|
|          |      |          | 第一产业    | 第三产业    | 第二产业    |
| 斯皮尔曼 Rho | 第一产业 | 相关系数     | 1.000   | 0.999** | 0.999** |
|          |      | Sig.(双尾) | 0.0     | 0.000   | 0.000   |
|          |      | N        | 23      | 23      | 23      |
|          | 第三产业 | 相关系数     | 0.999** | 1.000   | 1.000** |
|          |      | Sig.(双尾) | 0.000   | 0.0     | 0.0     |
|          |      | N        | 23      | 23      | 23      |
|          | 第二产业 | 相关系数     | 0.999** | 1.000** | 1.000   |
|          |      | Sig.(双尾) | 0.000   | 0.0     | 0.0     |
|          |      | N        | 23      | 23      | 23      |

\*\*在 0.01 级别(双尾), 相关性显著。

由输出结果能够看出, 第一产业与第二产业, 第二产业与第三产业有着高度的自相关性, 因此需要做一定的自变量处理。

### 3.1.3. 逐步回归法[5]选择自变量

取  $\alpha_{\text{entry}} = 0.1$ ,  $\alpha_{\text{removal}} = 0.15$ , 利用 SPSS 软件进行运算, 输出结果见表 9:

**Table 9.** The result of the stepwise regression  
**表 9.** 逐步回归法输出结果

|    |      | 系数 <sup>a</sup> |          |       |         |       |
|----|------|-----------------|----------|-------|---------|-------|
| 模型 |      | 未标准化系数          |          | 标准化系数 | t       | 显著性   |
|    |      | B               | 标准错误     | Beta  |         |       |
| 1  | (常量) | -90.437         | 1313.438 |       | -0.069  | 0.946 |
|    | 第二产业 | 2.155           | 0.013    | 1.000 | 169.267 | 0.000 |

Continued

|   |      |          |         |       |            |       |
|---|------|----------|---------|-------|------------|-------|
|   | (常量) | 5377.346 | 620.561 |       | 8.665      | 0.000 |
| 2 | 第二产业 | 1.216    | 0.074   | 0.564 | 16.407     | 0.000 |
|   | 第三产业 | 0.976    | 0.077   | 0.437 | 12.702     | 0.000 |
|   | (常量) | 0.008    | 0.038   |       | 0.213      | 0.833 |
| 3 | 第二产业 | 1.000    | 0.000   | 0.464 | 404016.951 | 0.000 |
|   | 第三产业 | 1.000    | 0.000   | 0.447 | 463696.824 | 0.000 |
|   | 第一产业 | 1.000    | 0.000   | 0.090 | 159741.694 | 0.000 |

<sup>a</sup>. 因变量: GDP。

由输出结果可以得出, 逐步回归的最优子集模型为  $y$  与  $x_2$ 、 $x_3$  的模型, 即排除掉了第一产业, 仅建立 GDP 与第二产业、第三产业的回归模型, 相应的方程为

$$\hat{y} = 5377.346 + 1.216x_2 + 0.976x_3 \quad (4-2)$$

以上方程表明当固定第三产业  $x_3$  不变, 第二产业  $x_2$  每增加一个单位, 相应的 GDP 平均增加 1.216 个单位。同时若固定第二产业  $x_2$  不变, 第三产业  $x_3$  每增加一个单位, 相应的 GDP 平均增加 0.976 个单位。

## 3.2. 残差分析

### 3.2.1. 残差计算与绘图

首先通过 Excel 软件计算出各时间段的残差值。

利用 R 软件画出残差值与时间的关系图见图 3:

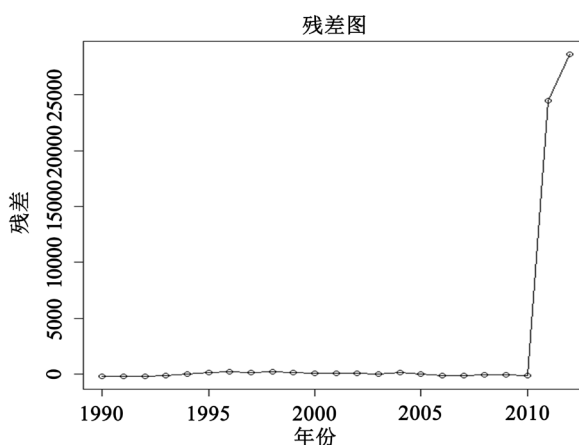


Figure 3. Residuals graph of multiple linear regression

图 3. 多元线性回归残差图

由图像能够看出, 除了 2011、2012 年的残差较大之外, 其余时间段残差值较平稳, 不存在异方差、自相关现象, 因此模型较为合理。

## 4. 评价与总结

本文通过一元最小二乘估计、多元最小二乘估计, 分别建立了 GDP 与第一产业增加值和 GDP 与第二、三产业增加值之间的回归模型, 结果如公式(3-8)、(4-2)。表面上看是两个不同的回归模型, 但由于



第一产业与第二、三产业之间存在较强的相关关系，即第一产业的变化一定会有相应的第二、三产业的变化，因此实质上本文是通过两种不同的途径来建立了两个相似的回归方程。这便于我们通过今后的产业量来估计我国 GDP 的增速，从而做出一些相关的政策调整。

## 参考文献

- [1] 张怡,熊常伟. 基于南充市生产总值(GDP)的灰色预测与因素分析[EB/OL].  
<http://kns.cnki.net/KXReader/Detail?TIMESTAMP=637183386499466250&DBC CODE=CJFQ&TABLEName=CJFD2014&FileName=IGNE201403008&RESULT=1&SIGN=venkkEOhs%2bHgiQwGe53xKddtbr0%3d>, 2019-12-12.
- [2] 中华人民共和国国家统计局. 中国统计年鉴: 2013[M]. 北京: 中国统计出版社, 2013.
- [3] 何晓群, 刘文卿. 应用回归分析(第 5 版)[M]. 北京: 中国人民大学出版社, 年份.
- [4] 孙玲, 张静. 基于多元线性回归的粤西地区经济发展影响因素分析[EB/OL].  
<http://kns.cnki.net/KXReader/Detail?TIMESTAMP=637185147123372500&DBC CODE=CJFO&TABLEName=CJFDLA ST2019&FileName=DAJJ201910016&RESULT=1&SIGN=UGCJi1txT2nNKA45Ms5XROo3p1I%3d>.
- [5] 蔡思美. 逐步回归法的财政收入影响因素分析[EB/OL].  
<http://kns.cnki.net/KXReader/Detail?TIMESTAMP=637185150096966250&DBC CODE=CJFD&TABLEName=CJFDLA ST2019&FileName=KJSZ201914011&RESULT=1&SIGN=XdvbkRAj2erlXnUJ3PQ6hs8jx Fk%3d>.