

# Application of Seemingly Unrelated Model in the Study of Food Production

Bo Wang\*, He Zhang

School of Math and Physics, China University of Geosciences, Wuhan Hubei  
Email: \*1690339596@qq.com

Received: Jun. 18<sup>th</sup>, 2017; accepted: Jul. 4<sup>th</sup>, 2017; published: Jul. 7<sup>th</sup>, 2017

---

## Abstract

The traditional linear model is usually used in the study of grain yield, in which the error terms are assumed independent of each other, and the parameters are estimated by the least squares method. In order to improve the estimation accuracy, seemingly unrelated model is introduced. Least squares model and seemingly unrelated model were built with the data that the yield, planting area and irrigation area of rice, wheat and corn between 1990 and 2009. The result showed that parameter variance of seemingly unrelated model is smaller than the least square model. The output of the three kinds of grain between 2010 and 2015 was predicted by the two models, and the average prediction error of seemingly unrelated model was smaller. It shows that there is some correlation between the yields of the three grain crops, and the seemingly unrelated model is better than the traditional least square model.

## Keywords

Seemingly Unrelated, Least Square Method, Grain Yield, Parameter Estimation

---

# 似乎不相关模型在粮食产量研究中的应用

王博\*, 张贺

中国地质大学(武汉)数学与物理学院, 湖北 武汉  
Email: \*1690339596@qq.com

收稿日期: 2017年6月18日; 录用日期: 2017年7月4日; 发布日期: 2017年7月7日

---

## 摘要

对粮食产量的研究一般采用传统的线性模型, 假设方程误差项相互独立, 采用最小二乘法进行参数估计。

\*通讯作者。

为提高模型的估计精度, 引入似乎不相关模型, 以1990~2009年稻谷、小麦、玉米的产量、种植面积和有效灌溉面积数据分别建立最小二乘模型和似乎不相关模型。结果显示似乎不相关模型的参数方差一致小于最小二乘模型, 用两种模型分别对2010~2015年的三种粮食产量进行预测, 似乎不相关模型预测的平均误差更小。说明三种粮食的产量之间存在一定的相关性, 似乎不相关模型较传统的最小二乘模型效果更好。

### 关键词

似乎不相关, 最小二乘法, 粮食产量, 参数估计

Copyright © 2017 by authors 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

似乎不相关模型最早由 Zellner 于 1962 年提出[1], 它是联立几个回归方程的特殊线性模型, 允许模型方程的自变量及变量个数不同, 使得模型建立更加方便, 大大增加了其应用范围。同时, 似乎不相关模型考虑了各方程误差项间的相关性, 认为模型不同观测之间互不相关, 而同一观测之间是相关的, 且有相同的方差矩阵, 这种特殊的误差结构使得其参数估计方法和效果不同于普通的最小二乘模型, 也是“似乎不相关”名称的由来。

## 2. 模型理论

### 2.1. 最小二乘模型理论

设  $y$  为因变量,  $x_1, x_2, \dots, x_p$  为自变量, 它们存在如下的线性关系:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + e \tag{1}$$

对(1)式中的  $x_1, x_2, \dots, x_p$  和  $y$  进行  $n$  次独立观测, 得到  $n$  组样本数据  $(x_{i1}, x_{i2}, \dots, x_{ip}; y_i)$ ,  $i = 1, 2, \dots, n$ , 记

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

假定不同误差项之间相互独立, 且  $\mathbf{e} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ , 则有

$$\begin{cases} \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \\ E(\mathbf{e}) = \mathbf{0}, \text{var}(\mathbf{e}) = \sigma^2 \mathbf{I}_n \end{cases} \tag{2}$$

(2)式是常见的多元线性模型, 也叫线性回归模型[2]。对参数  $\boldsymbol{\beta}$  的估计, 常采用最小二乘法, 当矩阵  $\mathbf{X}$  列满秩时,  $\boldsymbol{\beta}$  的最小二乘估计量为

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \tag{3}$$

此估计量为最优线性无偏估计量。

## 2.2. 最小二乘模型理论

设  $Y_j$  为因变量,  $\mathbf{X}_j$  为对应的自变量, 它们满足如下的线性关系:

$$Y_j = \mathbf{X}^{(j)}\boldsymbol{\beta}_j + e_j, \quad 1 \leq j \leq q \quad (4)$$

其中,  $\mathbf{X}^{(j)}$  为  $p_j$  维行向量,  $\boldsymbol{\beta}_j$  为  $p_j$  维列向量,  $e_j$  为第  $j$  个方程的误差项。需要注意的是每个方程中的自变量  $\mathbf{X}^{(j)}$  及其个数  $p_j$  不一定相同。在得到样本观测值  $(\mathbf{X}_i^{(j)}, Y_{ij}, 1 \leq i \leq n)$  后, 则有下面的关系:

$$\begin{cases} Y_{1j} = \mathbf{X}_1^{(j)}\boldsymbol{\beta}_j + e_{1j} \\ Y_{2j} = \mathbf{X}_2^{(j)}\boldsymbol{\beta}_j + e_{2j}, \quad 1 \leq j \leq q \\ \vdots \\ Y_{nj} = \mathbf{X}_n^{(j)}\boldsymbol{\beta}_j + e_{nj} \end{cases} \quad (5)$$

假设误差项均值为 0, 不同观测的误差项互不相关, 且有相同的方差矩阵, 即

$$E(e_{ij}) = 0, \quad \text{var}(e_{i\cdot}) = \sigma^2 \boldsymbol{\Sigma}, \quad \text{cov}(e_{i\cdot}, e_{i'\cdot}) = \mathbf{0}$$

其中  $e_{i\cdot} = (e_{i1}, e_{i2}, \dots, e_{iq})$ 。记

$$\mathbf{Y}_j = \begin{pmatrix} Y_{1j} \\ Y_{2j} \\ \vdots \\ Y_{nj} \end{pmatrix}, \quad \mathbf{X}_j = \begin{pmatrix} \mathbf{X}_1^{(j)} \\ \mathbf{X}_2^{(j)} \\ \vdots \\ \mathbf{X}_n^{(j)} \end{pmatrix}, \quad \mathbf{e}_j = \begin{pmatrix} e_{1j} \\ e_{2j} \\ \vdots \\ e_{nj} \end{pmatrix}$$

则有

$$\begin{cases} \mathbf{Y}_j = \mathbf{X}_j \boldsymbol{\beta}_j + \mathbf{e}_j \\ E(\mathbf{e}_j) = \mathbf{0}, \text{var}(\mathbf{e}_{i\cdot}) = \sigma^2 \boldsymbol{\Sigma}, \text{cov}(\mathbf{e}_{i\cdot}, \mathbf{e}_{i'\cdot}) = \mathbf{0} \\ i = 1, 2, \dots, n \quad i \neq i' \quad j = 1, 2, \dots, q \end{cases} \quad (6)$$

模型(6)称为似乎不相关模型[3]。再记

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_q \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & & & \mathbf{0} \\ & \mathbf{X}_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{X}_q \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_q \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_q \end{pmatrix}$$

则可将(6)写成广义一元线性模型的形式:

$$\begin{cases} \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \\ E(\mathbf{e}) = \mathbf{0}, \text{var}(\mathbf{e}) = \sigma^2 (\boldsymbol{\Sigma} \otimes \mathbf{I}_n) \end{cases} \quad (7)$$

其中, “ $\otimes$ ”表示矩阵的 Kronecker 乘积运算符[4]。根据广义一元线性模型的理论, 当矩阵  $\mathbf{X}$  列满秩时, 似乎不相关模型(6)中参数  $\boldsymbol{\beta}$  的估计为:

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}^T (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n) \mathbf{X})^{-1} \mathbf{X}^T (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n) \mathbf{Y} \quad (8)$$

当方差矩阵  $\boldsymbol{\Sigma}$  未知时, 常采用二步回归法。先计算参数  $\boldsymbol{\beta}$  的最小二乘估计量:  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$ ,

然后估计矩阵  $\Sigma$  的元素:

$$\hat{\sigma}_{ij} = \frac{(Y_i - X_i \hat{\beta}_i)(Y_j - X_j \hat{\beta}_j)}{n - \bar{p}} \quad (9)$$

其中  $\bar{p} = \frac{\sum_{i=1}^q p_i}{q}$ , 记  $\hat{\Sigma} = (\hat{\sigma}_{ij})_{q \times q}$ , 可得  $\beta$  的估计量为:

$$\tilde{\beta} = (X^T (\hat{\Sigma}^{-1} \otimes I_n) X)^{-1} X^T (\hat{\Sigma}^{-1} \otimes I_n) Y \quad (10)$$

称  $\tilde{\beta}$  为似乎不相关估计量, 它是  $\beta$  的无偏估计量, 其方差为

$$\text{var}(\tilde{\beta}) = \bar{\sigma}^2 (X^T (\hat{\Sigma}^{-1} \otimes I_n) X)^{-1} \quad (11)$$

### 3. 实例分析

#### 3.1. 粮食产量的线性模型

粮食产量是关系国家发展的一个重要问题, 各个邻域的专家和学者都十分重视对粮食产量问题的研究[5]。目前, 关于粮食产量常用的模型有线性回归、时间序列、灰色预测、神经网络等[6]。线性模型是最简单实用的一种模型, 但由于粮食产量影响因素较多, 具有很大的随机性和突变性, 往往效果不好。似乎不相关模型是一种改进的线性模型, 将其运用到粮食产量问题中, 具有一定的意义。

影响粮食产量的因素主要有粮食种植面积、化肥施用量、受灾面积、农业机械总动力、灌溉面积、人口密度等[7]。

假设粮食产量与其影响因素满足如下的线性关系:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e \quad (12)$$

其中,  $y$  为粮食产量,  $x_1$  为种植面积,  $x_2$  为灌溉面积。以 1990~2009 共 20 年的数据(数据来源于中国统计年鉴), 分别建立三种农作物的线性回归模型:

$$Y_i = X_i \beta_i + e_i \quad (13)$$

其中,  $i=1, 2, 3$  分别代表稻谷、小麦、玉米三种农作物。考虑到三个方程误差项间存在相关性, 将其联立得到似乎不相关模型

$$\begin{cases} Y = X\beta + e \\ E(e) = \mathbf{0}, \text{var}(e) = \sigma^2 (\Sigma \otimes I_{20}) \end{cases} \quad (14)$$

这里  $Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$ ,  $X = \begin{pmatrix} X_1 & & \\ & X_2 & \\ & & X_3 \end{pmatrix}$ ,  $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$ ,  $e = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}$ 。

#### 3.2. 结果对比与预测

分别计算得到三种农作物的回归方程:

稻谷:  $y = -17529 + 0.7544x_1 + 0.2534x_2$  ( $R^2 = 0.855$ )

小麦:  $y = -20807 + 0.4211x_1 + 0.3770x_2$  ( $R^2 = 0.715$ )

玉米:  $y = -3326 + 0.781x_1 - 0.0704x_2$  ( $R^2 = 0.949$ )

模型的拟合效果较好。

似乎不相关模型中, 参数估计结果为:

$$\beta_1 = (-16938, 0.7419, 0.2493)^T$$

$$\beta_2 = (-17988, 0.3799, 0.3445)^T$$

$$\beta_3 = (-2899.8, 0.8052, -0.0898)^T$$

由于两种模型中参数  $\beta$  的估计均为无偏估计, 有必要比较其方差。方差矩阵中对角线上的元素代表参数  $\beta$  各分量的方差, 表 1 列出两种模型三个方程的参数  $\beta$  各分量的方差。

对比表 1 中两种模型的方差结果, 可以发现似乎不相关模型中参数  $\beta$  各分量的方差都略小于最小二乘模型, 说明似乎不相关模型的参数估计更有效[8]。

分别用两种模型对 2010~2015 年的粮食产量进行预测, 结果如表 2 所示。

结果显示, 最小二乘模型中, 三种农作物预测产量相对真实值的平均误差分别为 4.29%、7.38%、3.28%, 而似乎不相关模型的平均相对误差分别为 4.08%、5.95%、3.08%, 误差分别减小了 4.90%、19.38%、6.10%。

#### 4. 总结

本文以粮食产量、种植面积及灌溉面积为变量, 分别对稻谷、小麦、玉米建立独立的最小二乘模型和三个方程联立的似乎不相关模型, 结果表明似乎不相关模型的参数方差小于最小二乘模型, 估计更有效。用两种模型对后续的粮食产量进行预测, 似乎不相关模型的平均相对误差整体上小于最小二乘模型,

Table 1. Variance of the parameters of the two models

表 1. 两种模型参数  $\beta$  各分量方差

	最小二乘模型			似乎不相关模型		
	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_0$	$\beta_1$	$\beta_2$
稻谷	$1.45 \times 10^7$	0.0056	0.0013	$1.36 \times 10^7$	0.0052	0.0012
小麦	$2.35 \times 10^7$	0.0043	0.0039	$2.22 \times 10^7$	0.0040	0.0037
玉米	$6.49 \times 10^6$	0.0101	0.0077	$6.46 \times 10^6$	0.0101	0.0077

Table 2. Prediction results of the two models (Unit: Million tons)

表 2. 两种模型预测结果(单位: 万吨)

时间		2010	2011	2012	2013	2014	2015	平均相对误差
最小二乘模型	稻谷	20298	20775	21040	21421	21690	21956	4.29%
	小麦	12162	12671	12975	13282	13664	14197	7.38%
	玉米	17810	18529	19635	20572	21125	21809	3.28%
似乎不相关模型	稻谷	20270	20739	21000	21374	21639	21901	4.08%
	小麦	12016	12481	12759	13040	13389	13876	5.95%
	玉米	17848	18567	19692	20642	21194	21876	3.08%
实际值	稻谷	19,576.10	20,100.09	20,423.59	20,361.22	20,650.74	20,822.52	
	小麦	11,518.08	11,740.09	12,102.36	12,192.64	12,620.84	13,018.52	
	玉米	17,724.51	19,278.11	20,561.41	21,848.90	21,564.63	22,463.16	

说明前者的预测精度更高。总体来说, 由于似乎不相关模型考虑了方程误差项间的相关性, 参数估计的效果比最小二乘模型更好, 模型优于传统的最小二乘模型。

### 参考文献 (References)

- [1] Zellner, A. (1962) An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association*, 57, 348-368.
- [2] 向东进, 李宏伟, 刘小雅. 实用多元统计分析[M]. 武汉: 中国地质大学出版社, 2005.
- [3] 唐守正. 生物数学模型的统计学基础[M]. 北京: 科学出版社, 2002.
- [4] 许君一, 孙伟, 齐东旭. 矩阵 Kronecker 乘积及其应用[J]. 计算机辅助设计与图形学学报, 2003, 15(4): 377-388.
- [5] 陈秧分, 李先德. 中国粮食产量变化的时空格局与影响因素[J]. 农业工程学报, 2013, 29(20): 1-10.
- [6] 林芳. 灰色神经网络在粮食产量预测中的应用[J]. 计算机仿真, 2012(4): 225-228.
- [7] 顾乐民. 基于最小一乘准则的中国粮食产量与影响因素的相关性分析[J]. 农业工程学报, 2013(11): 1-10.
- [8] 罗文强, 冀雅楠, 王淳越, 宋子健. 多监测点滑坡变形预测的似乎不相关模型研究[J]. 岩石力学与工程学报, 2016(S1): 3051-3056.

#### 期刊投稿者将享受如下服务:

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

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

期刊邮箱: [aam@hanspub.org](mailto:aam@hanspub.org)