

累加多层统计模型的建立 及其在经济上的应用研究

刘殿国 著

阶层一: $Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_{\cdot j}) + r_{ij}$

阶层二: $\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}, \beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j}$

其中,

$$Y_{ij} = x_{ij}^{(0)}(i), X_{ij} = \frac{1}{2}[x_{ij}^{(0)}(i-1) + x_{ij}^{(0)}(i)]$$

$$E(u_{0j}) = E(u_{1j}) = 0$$

$$\text{var}(u_{0j}) = \sigma_{u_0}^2, \text{var}(u_{1j}) = \sigma_{u_1}^2, \text{cov}(u_{0j}, u_{1j}) = \sigma_{u_0 u_1}$$

吉林大学出版社

累加多层统计模型的建立及 其在经济上的应用研究

刘殿国 著

吉林大学出版社

图书在版编目(CIP)数据

累加多层统计模型的建立及其在经济上的应用研究/刘殿国著. —长春: 吉林大学出版社, 1905.7

ISBN 978-7-5601-4448-1

I. 累… II. 刘… III. 统计模型-研究 IV. C81

中国版本图书馆 CIP 数据核字(2009)第 069213 号

书 名: 累加多层统计模型的建立及其在经济上的应用研究
作 者: 刘殿国 著

责任编辑、责任校对: 邵宇彤
吉林大学出版社出版、发行
开本: 787×1092 毫米 1/16
印张: 13.25 字数: 300 千字
ISBN 978-7-5601-4448-1

封面设计: 创意广告
长春市东文印刷厂 印刷
2009 年 6 月 第 1 版
2009 年 6 月 第 1 次印刷
定价: 25.00 元

版权所有 翻印必究
社址: 长春市明德路 421 号 邮编: 130021
发行部电话: 0431-88499826
网址: <http://www.jlup.com.cn>
E-mail: jlup@mail.jlu.edu.cn

内容摘要

本书首先在综述了多层统计模型的研究现状,介绍了灰色系统的建模思想和模型的基础上,指出了用多层统计模型和灰色系统模型难以分析具有层结构的小样本数据或数据模型的具体形式难以确定的情形。其次,为了解决具有层结构的小样本数据或数据模型的具体形式难以确定的问题,把灰色系统的建模思想方法和多层统计模型的思想方法的精华结合起来,建立了系列累加多层统计模型。再次,对新模型的参数估计方法,残差估计方法,残差新模型,新旧数据间关系的新模型,多重共线性的克服,在异常方面的预测等问题进行了研究。最后,应用系列累加多层统计模型到香蕉组织绩效以及泛珠三角经济合作区的经济增长研究中,从实际应用的结果可知,模型的分析与预测的结果符合实际,应该说该模型在探索具有层结构的小样本数据方面迈出了一步。

与传统统计模型相比,多层统计模型具有几个优点:①在数据分析中,多层统计模型能得到回归系数更有效的估计;②多层统计模型通过利用集聚信息,能提供正确的标准差,回归系数的置信区间和显著性检验;③通过层结构间的协方差的估计值,能够弄清不同单元之间的差别形成的原因;④当要研究的对象的样本很少时,可以用层结构的特点,把它看成是某一整体的一个随机样本以及利用整个样本数据可得到的信息,对所研究的个体作估计时,可以获得更高的精度;⑤由于模型中不止一个残差项,故存在着非零的单元内相关(intra-unit correlation),这意味着传统的估计方法如多元回归中常用的普通最小二乘法(OLS)是不适用的,多层统计模型正好适用此类情形。由于多层统计模型与传统统计模型相比有着不同的特点,并在对复杂情形分析和预测方面有着明显的优势,因而多层统计模型的理论研究有了长足的发展。模型的变量由简单的单解释变量和单响应变量向着多解释变量和多响应变量方向发展;方差结构由简单的常数型向复杂的函数型方向发展;参数的估计方法也是由一般广义最小二乘法向着数据应满足一定分布的较复杂的估计方法方向发展。其在教育、地理、孩子身高的增长和健康等方面的应用均取得了较好的效果,并且应用的范围逐渐扩大。但它难以解决具有层结构的小样本数据或数据模型的具体形式难以确定的问题。

灰色系统模型是把杂乱无章的原始数据,用数据累加方法整理成规律较

强的生成数据,以生成数据为基础建模,得到生成数据的预测值,再还原成原始数据。它在实践中已经成为一个重要的模型,但它难以解决具有层结构的小样本数据的问题。

建立的系列累加多层统计模型为:随机系数累加多层统计模型 $AMM(1,1)$ 和 $AMM(1,N)$,改进的随机系数累加多层统计模型(优化的单变量随机系数累加多层统计模型、幂函数的单变量随机系数累加多层统计模型以及对数的单变量随机系数累加多层统计模型),整体模式累加多层统计模型 $AMM(1,1)$ 和 $AMM(1,N)$,改进的整体模式累加多层统计模型(优化的单变量整体模式累加多层统计模型、幂的单变量整体模式累加多层统计模型以及对数的单变量整体模式累加多层统计模型),新陈代謝累加多层统计模型。

在模型的实际应用中,针对能够出现异常值的情形,对收集到的具有层结构的数据,通过数据整理得异常值数据列。以此列为基础,利用 $AMM(1,1)$ 模型进行预测。

给出了系列模型在数据满足不同条件下的参数估计方法:在一般情形下,用一般迭代最小二乘法和广义最小二乘法,收缩估计(Shrinkage Estimates);在数据为随机变量时,用刀切法;在数据的分布已知时,用 $MCMC$ 方法。在得到参数估计后,把所得到的参数估计的分量看作是相应微分方程的系数。从微分方程中解出生成数列的预测值,之后再还原成原始数据的预测值。

具体的随机系数累加多层统计模型为:

阶层一:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{1ij} - \bar{X}_{1j}) + \beta_{2j}(X_{2ij} - \bar{X}_{2j}) + \cdots + \beta_{pj}(X_{pij} - \bar{X}_{pj}) + e_{ij}$$

阶层二:

$$\beta_{0j} = \beta_0 + \mu_{0j}, \beta_{1j} = \beta_1 + \mu_{1j}, \beta_{2j} = \beta_2 + \mu_{2j}, \cdots, \beta_{pj} = \beta_p + \mu_{pj}$$

其中,

$$Y_{ij} = x_{ij}^{(0)}(i), X_{1ij} = \frac{1}{2}[x_{ij}^{(1)}(i-1) + x_{ij}^{(1)}(i)]$$

$$X_{2ij}(i) = x_{ij}^{(2)}(i), \cdots, X_{pij} = x_{ij}^{(p)}(i)$$

$$E(\mu_{1j}) = 0, E(\mu_{2j}) = 0, \cdots, E(\mu_{pj}) = 0$$

$$\text{var}(\mu_{1j}) = \sigma_{\mu 1}^2, \text{var}(\mu_{2j}) = \sigma_{\mu 2}^2, \cdots, \text{var}(\mu_{pj}) = \sigma_{\mu p}^2$$

$$\text{cov}(\mu_{1j}, \mu_{2j}) = \sigma_{\mu 12}, \text{cov}(\mu_{2j}, \mu_{3j}) = \sigma_{\mu 23}, \cdots, \text{cov}(\mu_{kj}, \mu_{rj}) = \sigma_{\mu kr},$$

$$\cdots, \text{cov}(\mu_{p-1j}, \mu_{pj}) = \sigma_{\mu p-1p}.$$

得到 $\hat{\beta}$ 的估计后,把 $\hat{\beta}$ 的分量看作微分方程

$$\frac{dx_{ji}^{(1)}(t)}{dt} = \beta_0 + \beta_1 x_{j1}(t) + \beta_2 x_{j2}^{(1)}(t) + \cdots + \beta_N x_{jN}(t)$$

的系数,此微分方程的解为

$$\hat{x}_{j1}^{(1)}(k+1) = \left[x_{j1}^{(0)}(1) + \frac{\beta_0}{\beta_1} + \frac{\beta_2}{\beta_1} x_{j2}^{(0)}(k+1) + \frac{\beta_3}{\beta_1} x_{j3}^{(0)}(k+1) + \cdots + \frac{\beta_N}{\beta_1} x_{jN}^{(0)}(k+1) \right] e^{\beta_1} \\ - \frac{\beta_0}{\beta_1} - \frac{\beta_2}{\beta_1} x_{j2}^{(0)}(k+1) - \frac{\beta_3}{\beta_1} x_{j3}^{(0)}(k+1) - \cdots - \frac{\beta_N}{\beta_1} x_{jN}^{(0)}(k+1)$$

通过式子

$$\hat{x}_{j1}(k+1) = \hat{x}_{j1}(k+1) - \hat{x}_{j1}^{(1)}(k)$$

得到 $x^{(0)}(t)$ 的估计值。

具体的整体累加多层统计模型只须把随机系数累加多层统计模型中的阶层二改为:

阶层二:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + \mu_{0j}, \beta_{1j} = \gamma_{10} + \gamma_{11} W_j + \mu_{1j}, \\ \beta_{2j} = \gamma_{20} + \gamma_{21} W_j + \mu_{2j}, \cdots, \beta_{pj} = \gamma_{p0} + \gamma_{p1} W_j + \mu_{pj}$$

其中, W_j 为层二变量。其它形式与随机系数累加多层统计模型一样。

应用了系列模型到香蕉组织绩效的研究,得到了较为符合实际的研究结果;尤其在只有两个样本数据预测中,各个模型预测的效果都较好,相比较而言,幂的单变量整体模式累加多层统计模型预测效果最好。

应用了系列模型到泛珠三角经济合作区的经济增长研究中,得到了较为符合实际的研究结果;比较了系列随机系数累加多层统计模型在泛珠三角经济合作区内各省的平均 GDP 的预测,对数的单变量随机系数新陈代谢累加多层统计模型与幂的单变量随机系数新陈代谢累加多层统计模型预测效果较好,相比较而言,对数的单变量随机系数新陈代谢累加多层统计模型预测效果更好;比较了系列整体模式累加多层统计模型在泛珠三角经济合作区内各省的 GDP 的预测,得到了预测误差主要受各省的 GDP 与 FDI 影响,当某个省的 GDP 与 FDI 越接近整体 GDP 与 FDI 均值时,预测误差越小。对数单变量整体模式新陈代谢累加多层统计模型对泛珠三角经济合作区内各省的 GDP 进行预测的误差比单变量整体模式新陈代谢累加统计模型和优化的单变量整体模式新陈代谢累加统计模型有较大的改进,甚至比幂的单变量整体模式新陈代谢累加多层统计模型效果更好,应该说,该模型是一个对个体预测的好模型。

对于新模型具体应用时出现的多重共线性的情况,本文初步探讨了它的原因是由于累加算法增加了数据间的相关性造成的。并用岭估计的方法给出了比较有效的克服多重共线性的方法。

关键词: 累加多层统计模型;参数估计;随机系数;整体模式;预测

ABSTRACT

First On account of overviewing the research present condition of multi-level statistical models , introducing the molding thought and method of gray system , pointed out it is difficult to analysis the small sample data of the hierarchical structure or the hard and certain of the concrete form of the data models. Secondly, for resolving to problem of the small sample data of the hierarchical structure or the hard and certain of the concrete form of the data models, combining the thought method of multilevel statistical model with the thought method of gray systems, build up the series accumulated multilevel statistical models. Again, researched parameter estimation, residuals estimation, residuals new model, a new model for relate to of new old data, overcame multicollinearity etc the problem for a new model. End, application the series accumulated multilevel statistical models to banner organize achievement with Pan-Pearl Delta area economic growth , can know from the actual applied result, model of as a result analysis and forecast match physically, should say that that model has the small sample data of the hierarchical structure to exceed in the quest the one step.

Have the different characteristics in comparing multilevel statistical model with tradition statistical model: ① in data analyze, multilevel statistical model can obtain efficient estimates of regression coefficients. ② The multilevel statistical model can provides correct standard error, confidence intervals and significance tests, by using the clustering information. ③ The passes the hierarchical structure estimate of variance value , can make clear the reason of the difference formation between different unit; ④ When the sample of the object is few, we can obtain more precision by using the characteristics of the hierarchical structure and regarding the individual as a sample from a population and using the information available from the whole sample data when making estimates for any one individual. ⑤ The existence of a non-zero intra- unit correlation, resulting from the presence of more

than one residual term in the model, means that traditional procedures, such as OLS which are used for example in multiple regression, are inapplicable. multilevel statistical model to apply this kind of situation at the right moment. As a result multilevel statistical model theories research, pressing to be explain by the simple explanatory variables, respond variables to change to many explanatory variables, many respond variables direction development; variance structure from the simple constant type to a direction of complicated function development; The method of the parameter estimate is form GLS to the more complicated estimate of method direction development that be should satisfy to certainly distribute It is in educate, the growth of the geography, child's height, health etc. of application all obtained to compare good of result, and applied scope gradual extension. But it is difficult to analysis the small sample data of the hierarchical structure or the hard and certain of the concrete form of the data models.

Gray system is chase miscellaneous and disorderly original data, use accumulate method sorting to become the stronger and generating data of regulation, take generating data as the foundation to set up the mold, get the estimate value of the generating data, then the revivification becomes the original data. It has already become an important model in the fulfillment, But it is difficult to analysis the small sample data of the hierarchical structure or the hard and certain of the concrete form of the data models.

The series accumulated multilevel statistical models of the establishment is: The accumulated multilevel statistical models of random coefficient $AMM(1,1)$ and $AMM(1,N)$, The improvement accumulated multilevel statistical models of random coefficient (accumulated multilevel statistical models of optimistic random coefficient, accumulated multilevel statistical models of power random coefficient and accumulated multilevel statistical models of logarithmic random coefficient), Accumulated multilevel statistical models of whole model $AMM(1,1)$ and $AMM(1,N)$, The improvement accumulated multilevel statistical models of whole model (accumulated multilevel statistical models of optimistic whole model, accumulated multilevel statistical models of power whole model and accumulated multilevel statistical models of logarithmic whole model), The metabolism accumulated multilevel statistical models.

In the actual application of the model, aim at the situation that can appear the excrescent value, have the data of the hierarchical characteristic towards collect of, pass the data sorting to get with the excrescent value data row. List as the foundation with this, make use of AMM(1,1) model to carry on the forecast.

Give the parameter estimate the method of series model under data satisfy different the condition: Under the condition of general, use the Iterative Least Squares and Generalized Least Squares, the Shrinkage Estimates; When data is random variable, use jackknife; At the data distribute to have already know, with the MCMC method. After get the parameter estimate of, see it is coefficient that corresponds differential equation. Solve the forecast value of generating data from the differential equation, after again restore the forecast value of the original data.

An ordinary accumulated multilevel statistical model of random coefficient is:

level 1:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{1ij} - \bar{X}_{1j}) + \beta_{2j}(X_{2ij} - \bar{X}_{2j}) + \cdots + \beta_{pj}(X_{pij} - \bar{X}_{pj}) + e_{ij}$$

level 2:

$$\beta_{0j} = \beta_0 + \mu_{0j}, \beta_{1j} = \beta_1 + \mu_{1j}, \beta_{2j} = \beta_2 + \mu_{2j}, \cdots, \beta_{pj} = \beta_p + \mu_{pj}$$

here,

$$Y_{ij} = x_{1j}^{(0)}(i), X_{1ij} = \frac{1}{2}[x_{1j}^{(1)}(i-1) + x_{1j}^{(1)}(i)]$$

$$X_{2ij}(i) = x_{2j}^{(1)}(i), \cdots, X_{pij} = x_{pj}^{(1)}(i)$$

$$E(\mu_{1j}) = 0, E(\mu_{2j}) = 0, \cdots, E(\mu_{pj}) = 0$$

$$\text{var}(\mu_{1j}) = \sigma_{\mu 1}^2, \text{var}(\mu_{2j}) = \sigma_{\mu 2}^2, \cdots, \text{var}(\mu_{pj}) = \sigma_{\mu p}^2$$

$$\text{cov}(\mu_{1j}, \mu_{2j}) = \sigma_{\mu 12}, \text{cov}(\mu_{2j}, \mu_{3j}) = \sigma_{\mu 23}, \cdots, \text{cov}(\mu_{kj}, \mu_{rj}) = \sigma_{\mu kr},$$

$$\cdots, \text{cov}(\mu_{p-1j}, \mu_{pj}) = \sigma_{\mu p-1p}.$$

After obtained $\hat{\beta}_{pj}$ estimate value, see $\hat{\beta}_{pj}$ is coefficient of differential equation.

$$\frac{dx_{j1}^{(1)}(t)}{dt} = \beta_0 + \beta_1 x_{j1}(t) + \beta_2 x_{j2}^{(1)}(t) + \cdots + \beta_N x_{jN}(t)$$

Solution of the differential equation . is

$$\hat{x}_{j1}^{(1)}(k+1) = \left[x_{j1}^{(0)}(1) + \frac{\beta_0}{\beta_1} + \frac{\beta_2}{\beta_1} x_{j2}^{(1)}(k+1) + \frac{\beta_3}{\beta_1} x_{j3}^{(1)}(k+1) + \cdots + \frac{\beta_N}{\beta_1} x_{jN}^{(1)}(k+1) \right] e^{\beta_1 k}$$

$$-\frac{\beta_1}{\beta_1} - \frac{\beta_2}{\beta_1} x_{j2}^{(1)}(k+1) - \frac{\beta_3}{\beta_1} x_{j3}^{(1)}(k+1) - \dots - \frac{\beta_N}{\beta_1} x_N^{(1)}(k+1)$$

using formula

$$\hat{x}_{j1}(k+1) = \hat{x}_{j1}(k+1) - \hat{x}_{j1}^{(1)}(k)$$

obtained $x^{(0)}(t)$ estimate value.

Only change level 2 formula of an ordinary accumulated multilevel statistical model of random coefficient for an ordinary accumulated multilevel statistical model of whole model:

level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + \mu_{0j}, \beta_{1j} = \gamma_{10} + \gamma_{11} W_j + \mu_{1j},$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21} W_j + \mu_{2j}, \dots, \beta_{pj} = \gamma_{p0} + \gamma_{p1} W_j + \mu_{pj}$$

here, W_j is level 2 variable, other formula is the same with an ordinary accumulated multilevel statistical model of random coefficient.

Applying the series accumulated multilevel statistical models to banner organize achievement, getting match the actual research result; Particularly to only have two samples data forecast, the result of each model forecast is all better, comparing but talking, the accumulated multilevel statistical models of power whole model forecast result best.

Applying the series accumulated multilevel statistical models to Pan-Pearl Delta area economic growth, getting to match the actual research result; Compared the forecast of the average GDP of each province about the series accumulated multilevel statistical models of random coefficient in Pan-Pearl Delta area economic cooperate area, the metabolism accumulated multilevel statistical models of logarithmic random coefficient and the metabolism accumulated multilevel statistical models of power random coefficient forecast result better, compare but talk, the metabolism accumulated multilevel statistical models of logarithmic random coefficient forecast result more better; Compared the forecast of the average GDP of each province about the series accumulated multilevel statistical models of whole model in Pan-Pearl Delta area economic cooperate area, forecast error mainly is influenced by the GDP and the FDI of each province, GDP and FDI that be a certain province near to the whole GDP and FDI to all be worth more, forecasting error more small. Forecast error of the GDP of each province have bigger improvement in Pan-Pearl Delta area economic cooperate area for using the me-

tabolism accumulated multilevel statistical models of logarithmic whole model than using the metabolism accumulated multilevel statistical models of whole model and the metabolism accumulated multilevel statistical models of optimistic whole model, even compare the accumulated multilevel statistical models of power whole model the result is better, should say, that model is a predicts to the individual of good model.

For the new model applies in a specific way appear multicollinearity circumstance, this text first step probe multicollinearity creating reason, this is because accumulated calculate way to increase the relativity of the data to result. The method that counteracts the ridge estimate overcome multicollinearity.

Keywords: accumulated multilevel statistical models; Parameter Estimate; random coefficient; whole model; forecast.

目 录

第一章 导论	1
1.1 研究的背景、目的与意义	1
1.1.1 研究背景	1
1.1.2 研究目的	2
1.1.3 研究意义	3
1.2 本书的内容及其结构	3
1.3 在应用统计的研究中要以研究方法论为指导	4
1.3.1 集聚与个体	4
1.3.2 事前假设与事后假设	5
1.3.3 真理性的检验与统计检验	6
1.3.4 逻辑分析与统计分析	7
1.3.5 相关分析与因果分析	8
1.4 本书的研究方法	9
第二章 多层的普遍性与多层模型文献综述	10
2.1 多层的普遍性	10
2.1.1 自然现象中的层结构	10
2.1.2 社会现象中的层结构	11
2.1.3 思维现象中的层结构	12
2.2 多层统计模型的文献综述	13
2.2.1 在参数估计方法方面	13
2.2.2 在处理方差分量方面	14
2.2.3 在处理误差方面	14
2.2.4 在处理多个响应变量方面	14
2.2.5 在非线性多层模型方面	15
2.2.6 在处理不同类型的数据方面	15
2.2.7 在处理缺失数据方面	15
2.2.8 在析因分析多层结构方程模型方面	16
2.2.9 多层模型应用越来越广泛	16

第三章 累加多层线性模型的基础理论	17
3.1 一般多层线性模型简介	17
3.1.1 二层数据与单层模型	17
3.1.2 二层模型的形式	20
3.1.3 二层模型的估计	20
3.1.4 残差估计	21
3.2 灰色系统模型简介	23
3.2.1 灰色系统的含义	23
3.2.2 灰色系统的建模思想	24
3.2.3 生成数据	25
3.2.4 灰色模型的分类	26
第四章 多层统计模型在经济增长中的应用研究	31
4.1 经济增长的有关理论	31
4.1.1 古典经济增长理论	31
4.1.2 新古典经济增长理论	32
4.1.3 经济增长因素分析	33
4.1.4 新经济增长理论	34
4.1.5 内生经济增长理论	35
4.1.6 巴罗的增长理论	36
4.2 泛珠三角经济合作区经济增长实证分析	36
4.2.1 泛珠三角经济合作区简介	36
4.2.2 数据介绍	37
4.2.3 数据分析	43
4.2.4 结论与建议	49
第五章 随机系数累加多层统计模型的建立	53
5.1 单变量随机系数的累加多层模型的建立	53
5.1.1 生成数	53
5.1.2 单变量随机系数累加二层统计模型的形式	54
5.1.3 单变量随机系数累加二层统计模型的 方差分量模型的参数估计	56
5.1.4 单变量随机系数累加二层统计模型的参数估计	58
5.1.5 单变量随机系数累加二层统计模型的残差估计	59
5.1.6 单变量随机系数累加二层统计模型的等价形式	60
5.2 多变量随机系数累加多层模型的建立	61

5.2.1 多变量累加二层统计模型的形式	61
5.2.2 多变量累加二层统计模型的残差估计	64
5.2.3 多变量累加二层统计模型的等价形式	65
第六章 随机系数累加多层统计模型在经济中的应用研究	66
6.1 单变量随机系数累加多层统计模型在经济中的应用	66
6.1.1 在香蕉组织绩效分析中的应用	66
6.1.2 在香蕉组织绩效预测中的应用	75
6.1.3 在泛珠三角经济区经济发展预测中的应用	77
6.2 多变量随机系数累加多层统计模型在经济中的应用	80
6.2.1 在香蕉组织绩效分析中的应用	80
6.2.2 在香蕉组织绩效预测中的应用	82
第七章 改进的随机系数的累加多层统计模型建立及其应用	85
7.1 改进的单变量随机系数累加多层统计模型的建立	85
7.1.1 优化的单变量随机系数累加多层统计模型的建立	85
7.1.2 对数的单变量随机系数累加多层统计模型的建立	86
7.1.3 幂的单变量随机系数累加多层统计模型的建立	87
7.2 改进的单变量随机系数累加多层统计模型的应用	87
7.2.1 在香蕉绩效预测上的应用研究	87
7.2.2 在泛珠三角区域经济发展预测上的应用研究	93
第八章 整体模式累加多层统计模型的建立	99
8.1 单变量整体模式累加多层统计模型的建立	100
8.1.1 建模的条件	100
8.1.2 模型的构造	100
8.1.3 参数估计	101
8.1.4 残差整体模式累加多层统计模型	103
8.1.5 整体模式累加多层统计模型 型在应用中的数据整理方法	106
8.2 多变量整体模式累加多层统计模型的建立	114
8.2.1 建模的条件	115
8.2.2 模型的构造	115
8.2.3 参数估计	117
8.2.4 多重共线性的产生原因初探和克服方法	117
8.3 改进的单变量整体模式累加多层统计模型的建立	120
8.3.1 优化的单变量整体模式累加多层统计模型的建立	121

8.3.2	对数的单变量整体模式累加多层统计模型的建立	122
8.3.3	幂的单变量整体模式累加多层统计模型的建立	123
第九章	整体模式累加多层统计模型的应用	124
9.1	单变量整体模式累加多层统计模型的应用	124
9.1.1	香蕉组织的绩效分析	124
9.1.2	香蕉组织的绩效预测	127
9.1.3	泛珠三角区经济发展分析	129
9.1.4	泛珠三角区经济发展预测	132
9.2	多变量整体模式累加模型的应用	136
9.2.1	香蕉组织的绩效分析	136
9.2.2	香蕉组织的绩效预测	138
9.3	改进的单变量整体模式累加模型的应用	140
9.3.1	改进的单变量整体模式的累加模型在香蕉组织绩效预测上的应用	140
9.3.2	改进的单变量整体模式的累加模型在泛三角经济发展预测上的应用	145
第十章	新旧数据取舍累加多层统计模型的建立及其应用	159
10.1	新旧数据取舍累加多层统计模型的建立	159
10.1.1	全数据累加多层统计模型的建立	159
10.1.2	新息累加多层统计模型的建立	160
10.1.3	新陈代谢累加多层统计模型的建立	160
10.2	新陈代谢累加多层统计模型的应用	161
10.2.1	随机系数及改进的新陈代谢累加多层统计模型的应用	161
10.2.2	改进的随机系数新陈代谢累加多层统计模型的应用	163
10.2.3	整体模式及其改进的新陈代谢累加多层统计模型的应用	168
10.2.4	改进的整体模式新陈代谢累加多层统计模型的应用	172
第十一章	结论与展望	185
11.1	研究结论	185
11.2	创新点	187
11.3	研究的局限性和未来的展望	187
11.3.1	研究的局限性	187
11.3.2	未来的展望	187
参考文献		188

第一章 导 论

1.1 研究的背景、目的与意义

1.1.1 研究背景

系统科学方法是 20 世纪科学的重大成果,它的产生不仅极大地改变了世界的科学图景,而且也为人认识世界、改造世界提供了强有力的方法工具。因此,各个学科的研究者都想用系统科学的观点和方法来解释本学科的内容和问题。统计学科也不例外,在上世纪 80 年代中期,许多统计学家就开始探讨怎样把系统科学方法引入到统计模型和层结构分析中。其实,早在上世纪 70 年代初期,Lindley 及 Smith(1972)、Smith(1973)在研究线性模式的贝氏估计(Bayesian estimation of linear models)时即对于具有复杂误差结构的巢状数据的分析有着浓厚的兴趣,但受限于当时估计的瓶颈而无法突破^[80]。其后 Dempster, Laird, 及 Rubin(1977)所推导出的 EM 估计法^[33],则对于 HLM 之估计技术提供了决定性的突破。在应用方面,Strenio, Wong, Bryk(1983)将 HLM 应用于分析成长(growth)问题;Mason, Wong, Entwistle(1983)将 HLM 应用于多层结构的横断研究,Laird, Ware(1982)将 HLM 应用在纵贯研究上^[150]。Goldstein(1987)写出了第一本关于多层统计学的专著^[57]。他们系统地总结了与传统统计模型相比多层统计模型具有几个优点:①在数据分析中,多层统计模型能得到回归系数更有效的估计;②多层统计模型通过利用集聚信息,能提供正确的标准差,回归系数的置信区间和显著性检验;③通过层结构间的协方差的估计值,能够弄清不同单元之间的差别形成的原因;④当要研究的对象的样本很少时,可以用层结构的特点,把其看成是某一整体的一个随机样本以及利用整个样本数据可得到的信息,对所研究的个体作估计时,可以获得更高的精度;⑤由于模型中不止一个残差项,故存在着非零的单元内相关(intra-unit correlation),这意味着传统的估计方法如多元回归中常用的普通最小二乘法(OLS)是不适用的,多层统计模型正好适用此类情形。Draper(1995)将多层统计模型应用到元分析(meta-analysis)中^[150]。之后,多层统计模型按上述的几个方向逐渐丰富和发展着,并已经成为国际统计学研究中一个新兴而重要的领域^[177]。

综合来看,运用多层模型对多层数据进行分析的研究主要表现在以下五个领域。多层模型首先可以广泛用于组织和管理研究;多层模型的第二个应用领域体现在对个体进行追踪、多次观测的发展研究中;多层模型的第三种应用可以视为前两种应用的综合;多层模型的第四个应用领域为做文献综述;多层模型的第五种应用是利用多层的数据来回答单层数据的问题^[178]。

然而,多层统计模型在实际应用中要求数据满足:

(1)有足够的样本(样本量最少要 30 个,且变量与样本的比例是 1:10);

(2)数据间的关系应满足:数据间的关系是线性的;数据间的关系是非线性的,但应大致地知道其曲线的类型;时间序列是平稳的或能转换成平稳的。

那么,对于小样本或数据间关系的具体形式难以确定的情形[如:一层是每公顷上香蕉产量的时间序列数据(产量与时间之间的非线性关系的具体形式难以确定),二层是组织形式的数据,由于不同的组织形式只是最近几年才出现的事物,因而相应的数据较少;年平均降雨量的时间序列、早霜或洪水出现的时区的时间序列等难以确定量之间的具体形式。]我们该怎么办呢?

其实,由邓聚龙(1982)提出的灰色系统模型对这种杂乱无章以及少样本的数据就有所研究^[152]。它通过累加方法将杂乱无章的数据变成规律性较强的生成数据,以生成数据为基础建立灰色系统模型。由于累加处理可以将无规则数据转化成有规则数据,因而可以用较少的数据得到较好的结果。原因是:如果 X_1, X_2, Δ, X_n 是一些随机变量,则 $X_1 + \Delta X_n$ 分布,在一般的情况下,和的极限分布就是正态分布^[149]。在应用中,当分布形态已知时,利用较少的数据就可以得到较好的估计^[170]。灰色系统模型在一些实际问题中已取得了较好的效果,它已经成为一个重要的模型。然而,模型在实际应用中还存在着下列问题:①当原始数据只有两个数据时是难以预测的;②对于具有层结构的数据它没有考虑层结构。

因此,多层统计模型和灰色系统模型都难以解决具有层结构的小样本数据或数据模型的具体形式难以确定的情形。

1.1.2 研究目的

为了解决具有层结构的小样本或模型的具体形式难以确定问题,本书把多层统计模型的思想方法与灰色系统模型的思想方法有机地结合起来,建立了随机系数累加多层统计模型,讨论了随机系数累加多层统计模型的参数和残差的估计进的随机系数累以及相关的理论;建立了改加多层统计模型;建立了整体模式累加多层统计模型以及改进的整体模式累加多层统计模型;建立了新陈代谢累加多层统计模型;并应用上述模型评价了区域经济的发展,预测了只有两年度数据的香蕉组织绩效;为政府各决策部门制订区域经济与香蕉