近代线性回归分析方法

胡宏昌　崔恒建　秦永松　李开灿　编著

科学出版社

内 容 简 介

本书主要介绍几类常见线性回归模型的估计方法：最小二乘估计、泛最小二乘估计、刀切估计、极大似然估计、经验似然方法、稳健估计. 在简要介绍这些回归模型估计方法的古老经典结果之后，有选择地介绍了线性回归模型中相关方法研究的最新成果(提出了泛最小二乘估计、刀切广义岭估计、t 型估计等方法，给出并研究了误差为泛函系数自回归(FCA)时间序列的线性回归模型，用经验似然方法研究了变量含误差(EV)、NA 误差的线性回归模型及缺失数据的线性回归模型，等等)，探索线性回归模型研究新的发展方向和科学规律.

本书适合高等院校统计学专业的高年级大学生、研究生、教师及相关科研工作者阅读参考.

图书在版编目(CIP)数据

近代线性回归分析方法/胡宏昌等编著. —北京：科学出版社，2013.1

ISBN 978-7-03-036579-8

Ⅰ.近… Ⅱ.①胡… Ⅲ.①线性回归—回归分析 Ⅳ.O212.1

中国版本图书馆 CIP 数据核字(2013)第 019337 号

责任编辑：曾 莉/责任校对：董艳辉

责任印制：彭 超/封面设计：苏 波

科 学 出 版 社 出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

武汉中科兴业印务有限公司印刷

科学出版社发行 各地新华书店经销

*

开本：B5(720×1000)

2013 年 1 月第 一 版 印张：13

2013 年 1 月第一次印刷 字数：253 000

定价：29.80 元

(如有印装质量问题，我社负责调换)

前　言

线性模型是统计领域一个经久不衰的研究对象. 有关线性模型的研究文献浩如烟海，关于线性回归模型的专著层出不穷. 本书名为《近代线性回归分析方法》，实则难以全面概述该领域的重要方向，如近年来图模型在生物信息等方面广泛的应用——链回归分析方法，又如离散多元分析中的对数线性模型研究的许多线性回归方法，再如目前在心理学、社会学研究中广泛应用的多层线性回归方法等，书中都难以涉及.

本书在叙述了线性回归模型的一些经典成果和方法后，有选择地介绍了近代线性回归模型研究的一些新课题、新方法（提出了泛最小二乘估计、刀切广义岭估计、t 型估计等方法，用拟极大似然、经验似然等方法研究误差为时间序列及各种混合序列的线性回归模型，等等），其目的就是让读者既掌握经典的研究方法，也了解一些前沿的研究课题.

本书以线性回归模型的估计方法为主线，包含了各类主要线性回归模型的主要研究方法，每种模型各有侧重点. 第 1 章由湖北师范学院李开灿教授执笔，第 2 章至第 5 章及附录由湖北师范学院胡宏昌教授执笔，第 6 章由广西师范大学秦永松教授执笔，第 7 章由首都师范大学及湖北省楚天学者崔恒建教授执笔，全书由胡宏昌教授负责统稿.

感谢所有从事线性回归模型及相关工作研究的统计学人，他们的深入研究成果为笔者提供了丰富的研究资料. 感谢为本书提出宝贵建议和给予帮助与支持的教师（徐侃教授、蔡择林教授、潘继斌老师、陈琴老师等）和硕士研究生（饶少林、苏艳云等）.

本书得到了湖北师范学院重点学科（统计学和数学）、专著专项基金的资助，得到了湖北省教育厅《应用数学》重点学科基金的资助，得到了国家自然科学基金的资助（No. 11071022）和湖北省教育厅科学技术研究项目（No. D20112503），在此一并表示感谢！

线性回归模型的研究方法和结果非常丰富，本书只是作者的一孔之见，疏漏与不足在所难免，恳请同行及广大读者批评指正！

编著者

2012 年 9 月 10 日

目　录

第1章 绪 论

2002年5月由美国国家科学基金资助召开的一个研讨会，其总结报告《统计学：二十一世纪的挑战和机遇》中称，统计学对核心领域的研究集中在对统计模型、方法和根据统计学一般原理的相关理论的研究，并建议加强统计核心的研究. 由此可见统计模型在统计学中的重要地位，其研究将具有十分重要的理论与实际意义. 线性回归模型不仅是现代统计学中应用最为广泛的模型之一，而且是其他统计模型的研究或应用基础，主要原因如下[1]：

第一，在现实世界中，许多变量之间具有线性或近似线性的依赖关系.

第二，尽管现实世界本质上是非线性的，但是许多变量经过适当变换后，新变量之间具有线性或近似线性的依赖关系.

第三，线性关系是数学中最基本的关系，比较容易处理.

本章先介绍回归分析，然后介绍线性回归模型及其应用，最后对线性回归模型的推广进行了简要的总结.

1.1 回归分析

1.1.1 相关关系

在人类活动的各种领域中，常常需要研究某些变量之间的关系. 一般来说，变量间的关系可以分为两类. 一类是变量之间具有严格的确定性的关系，如圆的面积 S 与其半径 r 之间有确定性关系 $S=\pi r^2$. 确定性的关系的特征是，一些变量确定另一些变量的值. 另一类是：变量之间存在着一定的制约关系，但这种关系没有密切到可由一个或一些变量决定另一个或一些变量的程度，如人的身高与体重、收入与智商、树叶的长度与宽度等. 我们知道，人的身高与体重有关，但前者并不决定后者. 我们称这类变量之间的关系为“相关关系”. 大体上来说，相关关系的产生有以下几种情形[2]：

(1) 变量之间本来应当有严格的、确定性的关系，但在测量这些变量时有误差，而误差又是随机性，故测量结果之间就呈现相关关系.

(2) 若干个变量，为简单计就举两个变量 X 和 Y，从其性质来看具有因果关系：前者是因，后者是果. 但是影响 Y 的变量不止 X 一个，有的我们知道，但由于种种原因没有去考虑它；有的甚至还不知道，例如，以 X 表示每亩施肥量，Y 表示每

亩产量，则 X 和 Y 有一定的因果关系，但影响产量 Y 的因素不止 X 一个. 因此 X，Y 之间的关系就不可能是确定的，而必然是相关性的.

(3) 有些变量 X，Y 表面上看有一定的因果关系，但它们本身都受到另一些因素的影响，而后者对 X，Y 的关系是相关性的而非确定性的. 这样，X，Y 的关系只能是相关性的，且彼此无合理的因果关系. 例如，以 X 表示一个人一年的旅游支出，Y 表示他一年吃的支出，调查并作统计分析发现 X，Y 有一定的关系，当一个变量增大(或减小)时，另一个变量随之增大(或减小)，但很难说二者具有直接的因果关系. 其实，这后面有一个因素——收入在起作用，收入多，他就可以多花，于是出去旅游，于是多吃或吃的质量提高，等等.

以上分类，对于数据统计分析的方法及其结果的解释都会有一定的影响. 下面我们说明：相关关系本质上是一种概率性质的关系，更确切地说，概率论的概念和方法是描述相关关系的有用工具. 我们以人的身高 X 和体重 Y 的关系为例来说明之.

知道一个人的身高为 1.69 m，并不能由此推出他的体重是多少. 事实上，如果把某一特定的一大群人中身高为 1.69 m 的那些人全挑选出来，再逐一去量他们的体重，则个个不同，而形成一定的概率分布，记为 $F(Y|X=1.69)$，表示在 $X=1.69$ 的条件下 Y 的条件分布. 一般地，指定 X 的一个值 x，可确定具有这个身高的所有人体重的概率分布 $F(Y|X=x)$. 从上面的分析可知，虽然我们无法由身高 X 唯一确定体重 Y，但可以考察 $F(Y|X=x)$ 随 x 变化的情况，弄清了这个情况，也就了解了二者关系的实质. 事实上，$F(Y|X=x)$ 随 x 变化的情况，刻画了 X 和 Y 之间相关关系的本质. 或更准确地说，研究这个条件分布随 x 变化的情况，提供了研究相关关系的重要手段.

1.1.2 回归分析

如上所述，要全面考察两个变量 X 和 Y 之间的相关关系，就要研究 Y 的条件分布 $F(Y|X=x)$ 随 X 的取值 x 的变化情况，这样做比较复杂，作为一个近似，我们可以考察分布 $F(Y|X=x)$ 的某个有代表性的数值，如其期望值. 这个期望值当然与 x 有关，记为 $f(x)$，称为 Y 对 X 的回归函数，而

$$y=f(x) \tag{1.1}$$

称为 Y 对 X 的回归方程. 对于身高和体重的例子，$f(x)$ 就是具有身高 x 的所有人的平均体重. 回归方程是一个确定性的关系，而原变量 X，Y 之间是相关关系. 回归方程的作用，正在于近似地代替这个相关关系.

在实际问题中，回归函数是不知道的，需要由试验或观察数据去估计. 假如做了 n 次试验，得到 n 组数据 (x_i,y_i)，$i=1,2,\cdots,n$. 根据回归函数的意义，有

$$y=E(Y|X=x)=f(x). \tag{1.2}$$

可以写成

$$y_i=f(x_i)+e_i,\quad i=1,2,\cdots,n, \tag{1.3}$$

其中 $Ee_i=0$. 以后我们称 X 为自变量,Y 为因变量或响应变量.

1.1.3 回归分析的主要内容

回归分析的主要内容如下[3]:

(1) 从一组数据出发,确定这些变量之间的定量关系;

(2) 对这些关系式的可信程度进行统计检验;

(3) 从影响某个变量的许多变量中,判断哪些变量是显著的,哪些是不显著的;

(4) 利用所求的关系式对"生产过程"进行预报和控制;

(5) 针对预报和控制所提出的特别要求,选择试验点,对试验进行某种设计;

(6) 寻找点数较少,且具有较好统计性质的回归设计方法.

1.2 线性回归模型

1.2.1 经典线性回归模型

首先将模型(1.3)中的自变量推广到 $p-1$ 个,即 $X_1,X_2,\cdots,X_{p-1}$,然后考虑其最简单的情形,即线性函数,得到如下线性回归模型:

$$Y=\beta_0+\beta_1X_1+\cdots+\beta_{p-1}X_{p-1}+e, \tag{1.4}$$

其中 e 为误差项,它表示除了 $X_1,X_2,\cdots,X_{p-1}$ 之外其他因素对 Y 的影响以及试验或测量误差;$\beta_0,\beta_1,\cdots,\beta_{p-1}$ 是待估计的未知参数. 假定因变量 Y 和自变量 $X_1,X_2,\cdots,X_{p-1}$ 的 n 组观测值为 $(y_i;x_{i1},\cdots,x_{i,p-1})$,$i=1,2,\cdots,n$,它们满足

$$y_i=\beta_0+\beta_1x_{i1}+\cdots+\beta_{p-1}x_{i,p-1}+e_i,\quad i=1,2,\cdots,n, \tag{1.5}$$

其中 $\{e_i,i=1,2,\cdots,n\}$ 为误差项,常常假定满足 Gauss-Markov 假设,即

$$E(e_i)=0,\quad \mathrm{Var}(e_i)=\sigma^2,\quad \mathrm{Cov}(e_i,e_j)=0\ (i\neq j). \tag{1.6}$$

应用适当的统计方法可以得到未知参数的估计值 $\hat{\beta}_0,\hat{\beta}_1,\cdots,\hat{\beta}_{p-1}$,将它们代入模型(1.4),并略去误差项得到

$$\hat{Y}=\hat{\beta}_0+\hat{\beta}_1X_1+\cdots+\hat{\beta}_{p-1}X_{p-1}, \tag{1.7}$$

称之为(经验)回归方程.

若用矩阵形式表示,则(1.5)变形为

$$\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}=\begin{pmatrix}1&x_{11}&\cdots&x_{1,p-1}\\1&x_{21}&\cdots&x_{2,p-1}\\\vdots&\vdots&&\vdots\\1&x_{n1}&\cdots&x_{n,p-1}\end{pmatrix}\begin{pmatrix}\beta_0\\\beta_1\\\vdots\\\beta_{p-1}\end{pmatrix}+\begin{pmatrix}e_1\\e_2\\\vdots\\e_n\end{pmatrix}.$$

等价地

$$y=X\beta+e, \tag{1.8}$$

其中 y 是 n 维观测列向量，X 为 $n\times p$ 已知设计矩阵，β 为 p 维未知参数列向量，e 为随机误差列向量. 用矩阵形式可将 Gauss-Markov 假设(1.6)写成

$$E(e)=0,\quad \mathrm{Cov}(e)=\sigma^2 I_n. \tag{1.9}$$

模型(1.8)、(1.9)常常称为经典线性回归模型. 下面给出一些例子，便于了解其应用背景.

例 1.1 儿子的身高 Y 与父母亲的平均身高 X 之间有如下关系：

$$Y=\beta_0+\beta_1 X+e.$$

例 1.2 一般来说，体重 Y(单位：kg)与身高 X(单位：cm)的经验回归方程为

$$\hat{Y}=X-105.$$

例 1.3 研究同一地区土壤内所含可给态磷浓度的情况，得到 18 组数据如表 1.1 所示. 其中，X_1 表示土壤内所含无机磷浓度，X_2 表示土壤内溶于 K_2CO_3 溶液并受溴化物水解的有机磷的浓度，X_3 表示土壤内溶于 K_2CO_3 溶液但不受溴化物水解的有机磷的浓度，Y 表示栽在 20 ℃土壤内的玉米中可给态磷的浓度.

表 1.1 土壤内所含可给态磷浓度的数据

土壤样本	X_1	X_2	X_3	Y	土壤样本	X_1	X_2	X_3	Y
1	0.4	53	158	64	10	12.6	58	112	51
2	0.4	23	163	60	11	10.9	37	111	76
3	3.1	19	37	71	12	23.1	46	114	96
4	0.6	34	157	61	13	23.1	50	134	77
5	4.7	24	59	54	14	21.6	44	73	93
6	1.7	65	123	77	15	23.1	56	168	95
7	9.4	44	46	81	16	1.9	36	143	54
8	10.1	31	117	93	17	26.8	58	202	168
9	11.6	29	173	93	18	29.9	51	124	99

根据这些数据，利用最小二乘法(见第 2 章)得到如下回归方程：

$$\hat{Y}=43.64117+1.78106X_1-0.08209X_2+0.16113X_3.$$

上面我们所讨论的回归模型为线性的. 有些非线性模型可以经过适当变换，使之成为线性模型.

例 1.4 Box-Cox 变换是对回归因变量 Y 作如下变换：

$$Y^{(\lambda)}=\begin{cases}\dfrac{Y^\lambda-1}{\lambda}, & \lambda\neq 0,\\ \ln Y, & \lambda=0,\end{cases} \tag{1.10}$$

其中 λ 是一个待定变换参数(如何确定 λ,参见文献[1]).对不同的 λ,所作的变换自然就不同,所以这是一个变换族.它包括了对数变换($\lambda=0$)、平方根变换($\lambda=1/2$)和倒数变换($\lambda=-1$)等常用变换.对因变量的 n 个观测值 $y_1,y_2,\cdots,y_n$ 应用上述变换,可以得到变换后的向量为

$$y^{(\lambda)}=(y_1^{(\lambda)},y_2^{(\lambda)},\cdots,y_n^{(\lambda)})^{\mathrm{T}},$$

使得 $y^{(\lambda)}$ 满足模型(1.8)和(1.9).因此,Box-Cox 变换是通过参数 λ 的适应选择,达到对原来数据的"综合治理",使其满足一个经典线性回归模型的所有假设条件.

例 1.5 为了对做过某一类型的肝手术病人的生存时间作预报,某医院外科随机地选取了 54 位需要做此类手术的病人为研究对象.手术前对每位病人考察了四个指标:凝血值 X_1,预后指数 X_2,酵素化验值 X_3,肝功能化验值 X_4.手术后跟踪观测各病人的生存时间 Y,得到如表 1.2 所示的 54 组数据[4].

表 1.2 54 位肝手术病人的观测数据

病人编号	生存时间	凝血值	预后指数	酵素化验值	肝功化验值
1	200	6.7	62	81	2.59
2	101	5.1	59	66	1.70
3	204	7.4	57	83	2.16
4	101	6.5	73	41	2.01
5	509	7.8	65	115	4.30
6	80	5.8	38	72	1.42
7	80	5.7	46	63	1.91
8	127	3.7	68	81	2.57
9	202	6.0	67	93	2.50
10	203	3.7	76	94	2.40
11	329	6.3	84	83	4.13
12	65	6.7	51	43	1.86
13	830	5.8	96	114	3.95
14	330	5.8	83	88	3.95
15	168	7.7	62	67	3.40
16	217	7.4	74	68	2.40
17	87	6.0	85	28	2.98
18	34	3.7	51	41	1.55
19	215	7.3	68	74	3.56
20	172	5.6	57	87	3.02

续表

病人编号	生存时间	凝血值	预后指数	酵素化验值	肝功化验值
21	109	5.2	52	76	2.85
22	136	3.4	83	53	1.12
23	70	6.7	26	68	2.10
24	220	5.8	67	86	3.40
25	276	6.3	59	100	2.95
26	144	5.8	61	73	3.50
27	181	5.2	52	86	2.45
28	574	11.2	76	90	5.59
29	72	5.2	54	56	2.71
30	178	5.8	76	59	2.58
31	71	3.2	64	65	0.74
32	58	8.7	45	23	2.52
33	116	5.0	59	73	3.50
34	295	5.8	72	93	3.30
35	115	5.4	58	70	2.64
36	184	5.3	51	99	2.60
37	118	2.6	74	86	2.05
38	120	4.3	8	119	2.85
39	151	4.8	61	76	2.45
40	148	5.4	52	88	1.81
41	95	5.2	49	72	1.84
42	75	3.6	28	99	1.30
43	483	8.8	86	88	6.40
44	153	6.5	56	77	2.85
45	191	3.4	77	93	1.48
46	123	6.5	40	84	3.00
47	311	4.5	73	106	3.05
48	398	4.8	86	101	4.10
49	158	5.1	67	77	2.86
50	310	3.9	82	103	4.55
51	124	6.6	77	46	1.95
52	125	6.4	85	40	1.21
53	198	6.4	59	85	2.33
54	313	8.8	78	72	3.20

利用 Box-Cox 变换对因变量 Y 作变换：$\tilde{Y}=\lg Y$. 对变换后的数据用穷举法或逐步回归法得到如下最优回归模型(由 SAS 软件中的 Proc Reg 过程计算得到)：

$$\tilde{Y}=\lg\hat{Y}=0.48362+0.06923X_1+0.00929X_2+0.00952X_3,$$

从而有

$$\hat{Y}=\exp\{0.48362+0.06923X_1+0.00929X_2+0.00952X_3\}.$$

例 1.6 在经济学中，著名的 Cobb-Douglas 生产函数为

$$Q_t=aL_t^bK_t^c,$$

其中 Q_t，L_t 和 K_t 分别为 t 年的产值、劳动投入量和资金投入量；a，b 和 c 为未知参数. 经过对数变换，并令 $y_t=\ln Q_t$，$x_{t1}=\ln L_t$，$x_{t2}=\ln K_t$，$\beta_0=\ln a$，$\beta_1=b$，$\beta_2=c$，再加上误差项，便可得到线性模型

$$y_t=\beta_0+\beta_1x_{t1}+\beta_2x_{t2}+e_t,\quad t=1,2,\cdots,T.$$

1.2.2 经典线性回归模型的推广

尽管对经典线性回归模型的推广有很多种形式，但笔者认为主要有三个方面：其一，从误差结构上进行推广；其二，从模型结构上进行推广；其三，从变量的限制条件上进行推广. 当然，这样的分类不是绝对的，在很多情况下，不仅几个方面相互交叉，而且与完全数据和不完全数据的线性回归模型相互交叉，这也是现代统计学研究的热点.

1. 误差结构上的推广

情形 1 误差项具有不等方差，即异方差性. 这样的线性回归模型称为具有异方差误差的线性回归模型，这方面的研究成果参见文献[5-15]等.

情形 2 $\mathrm{Cov}(e_i,e_j)=0$，$i\neq j$ 不成立，即误差项存在自相关性. 假定误差 $\{e_i,i=1,2,\cdots,n\}$ 为鞅差[16]、各种混合序列[17,18](如 $\alpha,\beta,\rho,\varphi,\psi,\tilde{\rho},\tilde{\varphi}$，负相关序列等)以及它们的线性组合，甚至假定误差为自回归(AR)、滑动平均(MA)、自回归滑动平均(ARMA)、自回归求和滑动平均(ARIMA)、自回归条件异方差(ARCH)、广义自回归条件异方差(GARCH)等时间序列[17]，还可以假定误差为分形时间序列. 尽管这种情形下的线性回归模型的研究结果很多[19-40]，但还有非常大的研究空间.

2. 模型结构上的推广

情形 1 涉及多个回归方程，具有如下形式：

$$y_{ti}=\beta_{0i}+\beta_{1i}x_{ti}+e_{ti},\quad i=1,2,\cdots,M;t=1,2,\cdots,n. \tag{1.11}$$

记 $y_i=(y_{1i},y_{2i},\cdots,y_{ni})^{\mathrm{T}}$，$\beta_i=(\beta_{0i},\beta_{1i})^{\mathrm{T}}$，$e_i=(e_{1i},e_{2i},\cdots,e_{ni})^{\mathrm{T}}$，$i=1,2,\cdots,M$，

$X_i=\begin{pmatrix}1 & x_{1i}\\ 1 & x_{2i}\\ \vdots & \vdots\\ 1 & x_{ni}\end{pmatrix}$，则(1.11)式可写成

$$\begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_M\end{pmatrix}=\begin{pmatrix}X_1 & & & 0\\ & X_2 & & \\ & & \ddots & \\ 0 & & & X_M\end{pmatrix}\begin{pmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_M\end{pmatrix}+\begin{pmatrix}e_1\\ e_2\\ \vdots\\ e_M\end{pmatrix},\tag{1.12}$$

其误差项假定为

$$Ee_i=0,\qquad E(e_ie_j^{\mathrm{T}})=\sigma_{ij}I_n,\quad i,j=1,2,\cdots,M.\tag{1.13}$$

称满足(1.11)、(1.13)的回归方程为半相依回归模型[41-47].

情形 2 时变系数线性模型：

$$Y(t)=X^{\mathrm{T}}(t)\beta(t)+\varepsilon(t),\tag{1.14}$$

其中 t 为时间指标；$X(\cdot),\beta(\cdot),\varepsilon(\cdot),Y(\cdot)$ 分别为 $p\times1$ 维协变量过程、时变回归系数、残差过程和响应过程，或 Y,X 和 ε 与时间 t 无关(有关结果参见文献[48-51]).

情形 3 矩阵变量线性回归模型[52]：

$$Y_i=\Theta X_i+U_i,\quad i=1,2,\cdots,n,\tag{1.15}$$

其中 Y_i 为 $p\times r$ 观测矩阵，X_i 为 $q\times r$ 维因变量，U_i 为 $p\times r$ 维随机误差，Θ 为 $p\times q$ 维待估参数.

情形 4 分层分位回归模型[53]. 为了方便起见，我们以具有两层数据的模型为例进行说明. 假定 (X,W,Y) 的一组独立同分布观测值 $(x_i,w_i,y_i),i=1,2,\cdots,n$，其中 y_i 是实数响应变量的值，x_i 为已知的 $1\times d$ 维第一层预测值向量，w_i 为已知的 $d\times f$ 维第二层预测矩阵，满足第一层模型

$$y_i=x_i\beta_i+\varepsilon_i,\quad \varepsilon_i\sim N(0,\sigma^2),\tag{1.16}$$

其中 β_i 是未知的 $d\times1$ 维第一层系数向量，ε_i 是独立同分布的不可观测的随机效应变量. 在第二层模型上，第一层模型中的系数成了输出结果

$$\beta_i=w_i\gamma+u_i,\quad u_i\sim N(0,T),\tag{1.17}$$

其中 γ 为 $f\times1$ 维固定效应向量，u_i 为 $d\times1$ 维第二层随机效应向量，假定它们与 w_i 和 ε_i 独立. 将(1.17)式代入(1.16)式得

$$y_i=x_iw_i\gamma+x_iu_i+\varepsilon_i,\quad \varepsilon_i\sim N(0,\sigma^2),\ u_i\sim N(0,T).\tag{1.18}$$

若在给定 $X=x$ 和 $W=w$ 的条件下响应变量 Y 的条件分布为 $F(y\mid x,w)$，则 Y 的 τ 分位数为

$$q_\tau(x,w)=\inf\{t\in\mathrm{R}:F(t\mid x,w)\geqslant\tau\}=xw\gamma+(xTx^{\mathrm{T}}+\sigma^2)^{1/2}\Phi^{-1}(\tau),\tag{1.19}$$

其中 $0<\tau<1$，$\Phi(\cdot)$为标准正态分布函数. 模型(1.18)和(1.19)一起定义为分层分位回归模型.

另外，还有很多非线性模型可以认为是线性模型的推广，在此略.

3. 变量限制条件上的推广

上述几种情况都是针对模型的误差结构或模型的结构而言的，可视为经典回归模型的推广. 我们还可以考虑对自变量及因变量的限制条件上推广模型.

情形 1 由于人为的或仪器的或测量手段的原因，自变量和因变量的观测值难于精确获得，而是含有随机误差. 以一元回归模型为例，可表示为

$$\begin{cases} y_i=\beta_0+\beta_1 x_i+e_i \\ X_i=x_i+\epsilon_i \end{cases} \tag{1.20}$$

其中 x_i 不能被直接观测到，而只能观测到 X_i，e_i 和 ϵ_i 是随机误差. 统计上称这类模型为误差变量模型(Errors-in-Variables Model)，这方面的研究成果可参见文献[54-73].

情形 2 在模型(1.5)中，若记 $\mu_i=Ey_i$，则

$$\mu_i=\beta_0+\beta_1 x_{1i}+\cdots+\beta_{p-1}x_{i,p-1}.$$

如果存在一个严格增的可微函数 g，使得

$$\begin{cases} g(\mu_i)=\beta_0+\beta_1 x_{1i}+\cdots+\beta_{p-1}x_{i,p-1}, \\ \mu_i=Ey_i, \end{cases} \quad i=1,2,\cdots,n \tag{1.21}$$

则模型(1.21)也是模型(1.5)的一种推广，统计学上称这类模型为广义线性模型，其研究成果非常丰富，可参见文献[74-93]. 如果 $g(p)=\ln[p/(1-p)]$，$0<p<1$，则模型(1.21)为应用很广泛的一类模型——Logistic 模型.

1.2.3 近代线性回归模型的研究方法

研究线性回归模型的方法较多，归纳起来，主要有如下几种：最小二乘估计法、为了克服设计矩阵复共线性问题而提出的岭估计及其他有偏估计方法、刀切估计方法、(拟)极大似然估计方法、经验似然估计方法，稳健估计方法等. 在以后的几章里，我们将从一定的角度介绍这些估计方法，以期读者对线性回归模型的研究方法有一个初步的认识.

参考文献

[1] 王松桂，陈敏，陈立萍. 线性统计模型：线性回归与方差分析[M]. 北京：高等教育出版社，1999.

[2] 傅權，胡蓓华. 基本统计方法教程[M]. 上海：华东师范大学出版社，1989.

[3] 茆诗松，丁元，周纪芗，等. 回归分析及其试验设计[M]. 上海：华东师范大学出版社，1981.

[4] 梅长林，范金城. 数据分析方法[M]. 北京：高等教育出版社，2006.

[5] Cheng T C. Robust diagnostics for the heteroscedastic regression model[J]. Computational Statistics and Data Analysis,2011,55:1845-1866.

[6] Cook R D, Weisberg S. Diagnostics for heteroscedasticity in regression[J]. Biometrika, 1983,70:1-10.

[7] Welsh A H,Carroll R J,Ruppert D. Fitting heteroscedastic regression models[J]. Journal of the American Statistical Association,1994,89:100-116.

[8] Harvey A C. Estimating regression models with multiplicative heteroscedasticity[J]. Econometrika,1976,38:375-386.

[9] Wen M J, Chen S Y, Chen H J. On testing a subset of regression parameters under heteroscedasticity[J]. Computational Statistics and Data Analysis,2007,51:5958-5976.

[10] Muller H,Stadtmuller U. Estimation of heteroscedasticity in regression analysis[J]. The Annals of Statistics,1987,15:610-625.

[11] Dixon S L, McKean J W. Rank-based analysis of the heteroscedastic linear model[J]. Journal of the American Statistical Association ,1996,91:699-712.

[12] Dufour J M, Khalaf L, Bernard J T, et al. Simulation-based finite-sample tests for heteroskedasticity and ARCH effects[J]. Journal of Econometrics,2004,122:317-347.

[13] Özkale M R. A jackknifed ridge estimator in the linear regression model with heteroscedastic or correlated errors[J]. Statistics and Probability Letters,2008,78:3159-3169.

[14] Cheng T C. On simultaneously identifying outliers and heteroscedasticity without specific form[J]. Computational Statistics and Data Analysis,2012,56:2258-2272.

[15] Bianco A,Boente G,Rienzo J. Some results for robust GM-based estimators in heteroscedastic regression models[J]. Journal of Statistical Planning and Inference,2000,89:215-242.

[16] Hall P, Heyde C C. Martingale Limit Theory and its Application[M]. Academic Press, New York,1980.

[17] Fan J Q,Yao Q W. Nonlinear Time Series[M]. Berlin:Springer-Verlag,2005.

[18] 吴群英.混合序列的概率极限理论[M].北京:科学出版社,2006.

[19] Jun F. Moderate Deviations for M-estimators in Linear Models with φ-mixing Errors[J]. Acta Mathematica Sinica,English Series,2012,28(6):1275-1294.

[20] Babu G J. Strong representations for LAD estimators in linear models[J]. Probability Theory and Related Fields,1989,83:547-558.

[21] Wu Q Y. Strong consistency of m estimator in linear model for $\tilde{\rho}$-mixing samples[J]. Acta Mathematica Scientia,2005,25A(1):41-46.

[22] Wu Q Y. Further study strong consistency of M estimator in linear model for $\tilde{\rho}$-mixing Random samples[J]. J Syst Sci Complex,2011,24:969-980.

[23] Hu H C. QML estimators in linear regression models with functional coefficient autoregressive processes[J]. Mathematical Problems in Engineering,2010,Doi:10.1155/2010/ 956907.

[24] Song L, Hu H C, Cheng X S. Hypothesis testing in GLM with FCA[J]. Mathematical Problem in Engineering,2012,Doi:10.1155/2012/862398.

[25] Maller R A. Asymptotics of regressions with stationary and nonstationary residuals[J]. Stochastic Processes and their Applications,2003,105:33-67.

[26] Fuller W A. Introduction to Statistical Time Series (Second Edition)[M]. New York:John Wiley & Sons,1996.

[27] Pere P. Adjusted estimates and Wald statistics for the AR(1) model with constant[J]. Journal of Econometrics,2000,98:335-363.

[28] Yajima Y. On estimation of a regression model with long-memory stationary errors[J]. The Annals of Statistics,1988,16:791-807.

[29] Yajima Y Asymptotic properties of the LSE in a regression model with long-memory stationary errors[J]. The Annals of Statistics,1991,19:158-177.

[30] Koul H L,Mukherjee K. Asymptotics of R-,MD- and LAD estimators in linear regression with long range dependent errors[J]. Probability Theory and Related Fields,1993,95:538-553.

[31] Dahlhaus R. Efficient location and regression estimation for long range dependent regression models[J]. The Annals of Statistics,1995,23:1029-1047.

[32] Hall P,Lahiri S N,Polzehl J. On bandwidth choice in nonparametric regression with both short and long range dependency errors[J]. The Annals of Statistics,1995,23:1921-1936.

[33] Robinson P M,Hidalgo F J. Time series regression with long-range dependence[J]. The Annals of Statistics,1997,25:77-104.

[34] Iglesias P,Jorquera H,Palma W. Data analysis using regression models with missing observations and long-memory:an application study[J]. Computational Statistics & Data Analysis,2006,50:2028-2043.

[35] Kleiber C. Finite sample efficiency of OLS in linear regression models with long-memory disturbances[J]. Economics Letters,2001,72:131-136.

[36] Koul H L,Surgailis D. Goodness-of-fit testing under long memory[J]. Journal of Statistical Planning and Inference,2010,140:3742-3753.

[37] Zhou Z,Wu W B. On linear models with long memory and heavy-tailed errors[J]. Journal of Multivariate Analysis,2011,102:349-362.

[38] Wu W B. M-estimation of linear models with dependent errors[J]. The Annals of Statistics,2007,35(2):495-521.

[39] Wu R N,Wang Q. Shrinkage estimation for linear regression with ARMA errors[J]. Journal of Statistical Planning and Inference,2012,142:2136-2148.

[40] Bera A K,Zuo X L. Specification test for a linear regression model with ARCH process [J]. Journal of Statistical Planning and Inference,1996,50:283-308.

[41] Zellner A. An efficient method of estimating seemingly unrelated regression equations and test for aggregation bias[J]. Journal of American Statistical Association, 1962, 57: 348- 368.

[42] Srivastava V K,Giles (Eds.) D E A. Seemingly Unrelated Regression Equations Models

[M]. New York: Marcel Dekker, Inc., 1987.

[43] Velu R, Richards J. Seemingly unrelated reduced-rank regression model[J]. Journal of Statistical Planning and Inference, 2008, 138: 2837-2846.

[44] Wang H. Sparse seemingly unrelated regression modelling: Applications in finance and econometrics[J]. Computational Statistics and Data Analysis, 2010, 54: 2866-2877.

[45] Cadavez V A P, Henningsen A. The use of seemingly unrelated regression (SUR) to predict the carcass composition of lambs[J]. Meat Science, 2012.

[46] Zellner A, Ando T. Bayesian and non-Bayesian analysis of the seemingly unrelated regression model with Student-t errors and its application for forecasting[J]. International Journal of Forecasting, 2010, 26: 413-434.

[47] Ma T F, Ye R D. Efficient improved estimation of the parameters in two seemingly unrelated regression models[J]. Journal of Statistical Planning and Inference, 2010, 140: 2749-2754.

[48] Zhou Z, Wu W B. Simultaneous inference of linear models with time varying coefficients [J]. Journal of the Royal Statistical Society B, 2010, 72: 513-531.

[49] Ramsay J, Silverman B W. Functional Data Analysis[M]. New York: Springer, 2005.

[50] Fan J, Zhang W Y. Simultaneous confidence bands and hypothesis testing in varying-coefficient models[J]. Scand. J. Statist., 2000, 27: 715-731.

[51] Honda T. Quantile regression in varying coefficient models[J]. Journal of Statistical Planning and Inference, 2004, 121: 113-125.

[52] Viroli C. On matrix-variate regression analysis[J]. Journal of Multivariate Analysis, 2012.

[53] 田茂再, 陈歌迈. 条件分位中的分层线性回归模型[J]. 中国科学(A 辑), 2006, 36(10): 1103-1118.

[54] Gleser L J. Estimation in a multivariate "errors in variables" regression model: large sample results[J]. The Annals of Statistics, 1981, 9(1): 24-44.

[55] Amemiya Y, Fuller W A. Estimation for the multivariate errors-in-variables model with estimated error covariance matrix[J]. The Annals of Statistics, 1984, 12(2): 497-509.

[56] Deaton A. Panel data from a time series of cross-sections[J]. Journal of Econometrics, 1985, 30: 109-126.

[57] Fuller W A. Measurement error models[M]. New York: Wiley, 1987.

[58] Mittag H J. Estimating parameters in a simple errors-in-variables model: a new approach based on finite sample distribution theory[J]. Stat. Pap., 1989, 30: 133-140.

[59] Cui H J. Asymptotic normality of M-estimates in the EV model[J]. J Syst Sci Math Sci, 1997, 10(3): 225-236.

[60] Cheng C L, Ness J W V. Statistical Regression with Measurement Error[M]. Arnold, London, 1999.

[61] Cui H J, Chen S X. Empirical likelihood confidence region for parameter in the errors-in-variables models[J]. Journal of Multivariate Analysis, 2003, 84: 101-115.