于卓熙　著

相依误差下部分线性模型的统计推断

清华大学出版社

于卓熙 著

相依误差下部分线性模型的统计推断

清华大学出版社
北京

内 容 简 介

全书共分 7 章，主要讨论了 NA 序列、m-相依序列、MA(∞)序列、ARCH(1) 序列误差下部分线性模型的经验似然统计推断，讨论了具有核实数据的鞅差序列误差下部分线性模型的估计和检验问题。这些是目前关于部分线性模型研究的热点问题，对于部分线性模型理论研究的发展起到了一定的推动作用。本书最后一部分是对自回归模型的自回归函数构造了半参数估计量，这种估计量的构造方法为从事估计量构造研究的学者提供一定的借鉴和参考。

本书内容充实、研究深入、论证严谨、写作思路清晰，适合概率论与数理统计专业及相关专业的师生学习，也适合从事概率统计研究的学者和统计工作者参考阅读。

图书在版编目(CIP)数据

相依误差下部分线性模型的统计推断 / 于卓熙 著. —北京：清华大学出版社，2014

ISBN 978-7-302-36224-1

Ⅰ. Ⅰ相… Ⅱ. Ⅰ于… Ⅲ. Ⅰ线性模型—统计推断—研究 Ⅳ. Ⅰ O212.1

中国版本图书馆 CIP 数据核字(2014)第 076544 号

责任编辑：王燊娉　易银荣
封面设计：牛艳敏
版式设计：方加青
责任校对：曹　阳
责任印制：刘海龙

出版发行：清华大学出版社
　　网　　址：http://www.tup.com.cn，http://www.wqbook.com
　　地　　址：北京清华大学学研大厦 A 座　　邮　　编：100084
　　社 总 机：010-62770175　　邮　　购：010-62786544
　　投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
　　质 量 反 馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：三河市金元印装有限公司
经　　销：全国新华书店
开　　本：169mm×230mm　　印　张：8　　字　数：114 千字
版　　次：2014 年 6 月第 1 版　　印　次：2014 年 6 月第 1 次印刷
定　　价：58.00 元

产品编号：058444-01

前　言

部分线性模型是20世纪80年代以来发展起来的一种重要的统计模型，最初是在研究天气变化和供电需求之间的关系时引入的。这个模型自问世以来得到了广泛的重视和研究，并且在工业、农业、经济、医药、金融等领域获得了广泛应用。对于部分线性模型的研究主要是指，在误差是独立同分布的情形下，构造未知参数和未知函数的估计，讨论估计的大样本性质并作假设检验。然而，该模型在实际应用中，误差是独立同分布的假设并不总是适合的，特别是一些序列相关的经济数据，总是表现出误差的相依性。近年来，具有相依误差的部分线性模型的研究，正引起广大统计工作者的浓厚兴趣。

本书是作者几年来科研成果的总结。全书共分7章，主要讨论了NA序列、m-相依序列、MA（∞）序列、ARCH（1）序列误差下部分线性模型的经验似然统计推断，讨论了具有核实数据的鞅差序列误差下部分线性模型的估计和检验问题。这些正是关于部分线性模型研究的热点问题，对于部分线性模型理论研究的发展起到了一定的推动作用。本书最后一部分是对自回归模型的自回归函数构造了半参数估计量，这种估计量的构造方法为从事估计量构造研究的学者提供一定的借鉴与参考。具体内容如下：

（1）针对误差是NA序列的部分线性模型，利用NA序列的性质已经取得的研究成果，构造部分线性模型中回归参数的经验似然比检验统计量，并讨论了统计量在参数取真值时是渐近地服从卡方分布的，相应地构造参数的置信区间。

（2）针对误差是 m-相依序列的部分线性模型，利用分组经验似然在处理相依数据的优势及误差序列的特殊相依关系，采用分组经验似然的方法对模型进行了上述相应的研究。

（3）针对误差是 MA（∞）序列的情形，运用对线性过程的多项式分解方法及对 MA（∞）过程的截断方法，研究 MA（∞）序列误差下模型的经验似然统计推断。

（4）针对误差是 ARCH（1）序列的情形，应用鞅的中心极限定理及鞅差序列的不等式，利用满足一定条件下 ARCH（1）序列的平稳性及四阶矩存在，进行 ARCH（1）序列误差下所构造的经验似然比统计量性质的研究，从而得出相应结论。

（5）针对具有核实数据的部分线性模型，并且在其误差项是鞅差序列的情形下，利用核实数据，估计真实值，构造模型参数和非参数部分的最小二乘估计量，并且证明估计量的相合性。通过模拟计算来说明对于鞅差序列误差下具有核实数据的部分线性模型，我们所构造的估计量的优良性。

（6）针对自回归模型，笔者考虑一种包含参数与非参数两种估计方法在内的新估计方法。回归函数的参数估计首先被作为一个粗略的估计。然后这个粗略的估计被非参数乘子进行修正，可以得到非参数估计。利用一个演变的局部 L_2 拟合准则估计修正项，最终可以得到基于 L_2 拟合准则的自回归系数的半参数估计，并且讨论了估计的相合性和渐近正态性。

本书是在国家自然科学基金项目（项目编号：11326103）、吉林省教育厅“十二五”科学技术研究项目（项目编号：2012186）、吉林财经大学 2013 年专著出版资助计划的资助和支持下完成的。值此专著完成之际，诚挚感谢吉林大学数学研究所王德辉教授多年来的指导和帮助。

由于作者水平有限，书中难免有种种考虑不周之处，诚请广大读者批评指正。

于卓熙

2014 年 1 月于长春

目　录

第1章
引　言

1.1 模型介绍

考虑如下的部分线性模型

$$y_i = x_i^{\tau}\beta + g(t_i) + \varepsilon_i; i = 1,2,\cdots,n \tag{1.1.1}$$

其中 $y_i^{'}s$ 是响应变量；$x_i = (x_{i1}, x_{i2}, \cdots, x_{ip})^{\tau}$；$t_i^{'}s$ 是随机点列或非随机固定设计点列；$\beta = (\beta_1, \beta_2, \cdots, \beta_p)^{\tau}$ 是未知参数向量；$g(\cdot)$ 是取值于闭区间 I 上的未知函数；$\varepsilon_i^{'}s$ 是随机误差；上标 τ 表示一个向量或矩阵的转置。

部分线性模型（Partial linear model），又称为半参数回归模型（Semi-parametric regression model），是 20 世纪 80 年代发展起来的一种重要统计模型，最初是由 Engle 等[21]在研究天气变化与供电需求之间的关系时引入的。这个模型自问世以来得到了广泛的重视和研究，并且在工业、农业、经济、医药、金融等领域获得了广泛的应用。

对模型（1.1.1）的研究主要是构造未知参数 β 和未知函数 $g(\cdot)$ 的估计。对于随机误差 $\varepsilon_i^{'}s$ 是独立同分布的情形，最基本的用于估计 β 和 $g(\cdot)$ 的方法是由 Engle[22]，Green[28]，Shiau[71]等等所给出的带惩罚项的最小二乘法。随着研究的进一步深入，多种估计方法都应用于获得模型（1.1.1）的未知量的估计，包括核方法、样条函数法、序列估计、局部线性估计、二阶段估计、M 估计等，文献可见［9］～［11］，［18］，［22］～［23］，［29］，［32］，［59］～［60］，［67］，［69］等。

在实际应用当中，模型（1.1.1）的误差的独立同分布的假设并不总是适合的，特别是一些序列相关的经济数据，经常表现出误差的相依性。近年来，具有相依误差的部分线性模型的研究正引起广大统计工作者的浓厚兴趣。例如，一种情形是用一阶自回归模型即 AR（1）模型对误差部分建模，Engle 等[21]曾用具有 AR（1）误差的模型（1.1.1）来研究气象对用电需求的影响，其中温度和用电需求之间的关系假设是未知的，而其他一些因素（如收入、电价等）和用电需求之间被假设具有线性关系。

Schick[65]~[67]给出了 AR（1）误差下模型（1.1.1）的自相关系数的估计量，并构造了参数 β 和自相关系数的有效估计。Beran 和 Ghosh[5]，Gao 和 Anh[25]讨论了长记忆误差下模型（1.1.1）的估计问题。Aneiros 和 Quintela[1]应用修正的交叉核实方法选择相依误差下模型（1.1.1）的光滑参数。You 和 Chen[87]构造了误差为随机系数自回归序列的模型（1.1.1）的半参数广义最小二乘估计。Sun 等[74]讨论了 MA（1）误差下模型（1.1.1）的参数估计的收敛速度。Baek 和 Liang[4]讨论了 NA 误差下模型（1.1.1）的参数估计的渐近性质。Li[44]，Li 和 Liu[45]在一定的条件下讨论了误差为鞅差序列的模型（1.1.1）的参数估计问题及估计的强相合性和渐近正态性。

1.2 经验似然简介

1.2.1 文献综述

早在 1975 年，Thomas 和 Grunkemeier[76]在建立截尾数据下生存概率的区间估计时已经萌发了经验似然的思想。1988 年，由 Owen[50]首先提出了经验似然方法，Owen[50]~[51]对此方法的一般性质进行了系统的研究。经验似然的性质类似于 Bootstrap 和随机加权方法，然而所不同的是它在第 i 个样本点 X_i 上赋予的概率是带有选择余地的 p_i，而不是 $1/n$。在一些合理约束下，$p_i(i=1,2,\cdots,n)$ 的变化，描绘了支撑在样本空间上的一个“多元似然函数”。这样达到了在已知部分总体信息的情况下，寻找类似于似然函数形式的初衷。

许多研究成果表明，经验似然除了有类似于 Fisher 提出的（参数）似然法的优良性，这一方法与经典的或现代的统计方法相比较还有许多突出的优点。例如：用经验似然方法构造置信区间除有域保持性、变换不变性

及置信域的形状由数据自行决定等诸多优点以外，还有 Bartlett 纠偏性以及无需构造轴统计量等优点。此外，应用经验似然进行统计推断不需要估计方差，而方差的估计通常是一个不容易的问题。因此，经验似然已成为统计界关注的重要研究方向之一。

最初，Owen[50]处理的是一类非参数统计问题，利用经验似然比构造了非参数置信区间和检验方法。后来，人们逐渐发现经验似然方法不但有效而且适应性强，从而这种有力的统计工具被推广到更广泛的统计模型中。如 Owen[50]~[52]由对总体均值的推断提出经验似然并随后将其应用到线性回归模型的统计推断；Kolaczyk[41]应用经验似然于广义线性模型；Monti[47]发展了时序模型的经验似然；Wang 和 Jing[80]~[81]分别就固定设计和随机设计两种情况发展了部分线性模型的经验似然；Chen 和 Qin[13]发展了非参数回归的经验似然；Zhang[89]~[90]应用经验似然于分位数回归 M－泛函的统计推断；Qin 和 Lawless[59]应用经验似然于估计方程；Chen 和 Qin[13]及 Zhong 和 Rao[91]应用经验似然于抽样调查问题的研究；Kitamura[39]等应用经验似然到经济模型；Chuang 和 Chan[15]发展了自回归模型的经验似然方法；Zhou 和 Liang[95]应用经验似然于两样本问题推断；You 和 Zhou[88]应用经验似然于半参数变系数部分线性回归模型；Xue 和 Zhu[86]应用经验似然于单指标模型；Lu 和 Yu[46]应用经验似然于线性转换模型。更多的细节可以参考 Owen[53]的专著以及王启华[84]的综述文章。

近年来，一些统计学家又将经验似然方法应用到不完全数据的统计分析，发展了所谓的被估计的经验似然，调整经验似然及 Bootstrap 经验似然。如 Cui 等[17]~[18]在假设误差结构是可加的情况下，分别把经验似然推广到了误差结构可加的线性度量误差模型和部分线性度量误差模型。Wang[79],[82]~[83]等进一步把经验似然推广到带核实数据和删失数据的度量误差模型。

1.2.2 经验似然

设 $X_1, X_2, \cdots, X_n \in R^d$ 相互独立，且具有共同的累积分布 F，则 F 的非

参数似然是

$$L(F) = \prod_{i=1}^{n} dF(X_i) = \prod_{i=1}^{n} p_i$$

这里 $p_i = dF(X_i) = P_r(X = X_i)$ ，其中 $i = 1,2,\cdots,n$ 。我们知道 X_1，$X_2,\cdots,X_n$ 的经验累积分布函数 $F_n = n^{-1}\sum_{i=1}^{n}\delta_{X_i}$ ，使上式达到最大，其中 $\delta_x(A) = I[x \in A]$，也就是说 F_n 是 F 的非参数极大似然估计。类似于参数似然比，我们可以定义非参数似然比为 $R(F) \doteq L(F)/L(F_n)$ ，容易证明

$$R(F) = \prod_{i=1}^{n} np_i \tag{1.2.1}$$

假如我们要对某个参数 $\theta = T(F)$ 进行统计推断，这里 $T(F)$ 是分布 F 某个泛函。为了简单起见，我们考虑分布 F 的均值 μ ，为了对 μ 作区间估计或检验，Owen[50]定义了如下的经验似然比函数

$$\Re(\mu) = max\left\{\prod_{i=1}^{n} np_i \,\middle|\, \sum_{i=1}^{n} p_i X_i = \mu, \sum_{i=1}^{n} p_i = 1, p_i \geqslant 0\right\} \tag{1.2.2}$$

显然，经验似然比实际上是一种截面非参数似然比函数，它要求 F 在满足约束条件 $\sum_{i=1}^{n} p_i x_i = \mu$ 下使非参数似然比达到极大（在无约束下，极大非参数似然比是1）。而参数 θ（这里是 μ）就通过约束条件引入到这一极大似然比中，从而得到关于参数 θ 的极大截面非参数似然比函数，用这一非参数似然比作假设检验、区间估计或其他统计推断，这一方法就是所谓的经验似然方法。

为了求出式（1.2.2）右边的极大值，我们可以利用 Lagrange 乘数法，求得 p_i 的最优值为

$$p_i = p_i(\mu) = n^{-1}\{1 + \lambda^{\tau}(X_i - \mu)\}^{-1} \tag{1.2.3}$$

其中 λ 为 Largrange 乘子，是如下方程的解

$$\sum_{i=1}^{n} \frac{X_i - \mu}{1 + \lambda^{\tau}(X_i - \mu)} = 0$$

把式（1.2.3）代入式（1.2.2），得到

$$\Re(\mu) = \prod_{i=1}^{n} \{1 + \lambda^{\tau}(X_i - \mu)\}^{-1} \tag{1.2.4}$$

为了能对 μ 进行统计推断，必须求出 $R(\mu)$ 的渐近分布。Owen[49]~[50] 证明了对数经验似然比 $-2\log R(\mu)$，具有类似参数似然比统计量的性质，也具有渐近的 χ^2 分布，这就是经验似然的非参数的 Wilk's 定理。更详细的论述可以参考 Owen[52] 的专著。

1.2.3 经验似然在固定设计部分线性模型中的应用

Owen[52] 将经验似然应用到线性模型进行统计推断，构造了回归参数的经验似然比统计量，并证明了其具有参数的 Wilk's 定理的渐近 χ^2 分布的性质。Wang 和 Jing[80]~[81] 分别将经验似然方法应用于固定设计和随机设计两种情况下的部分线性模型，并得到了相应的结论。

在模型（1.1.1）中，假设 y_i 是反应变量，x_i 是第 i 个 $p \times 1$ 固定设计向量，t_i 是第 i 个固定设计点，β 是未知参数向量，$g(\cdot)$ 是定义在闭区间 $[a,b]$ 上的未知函数，ε_i；$j = 1,2,\cdots,n$ 是 i. i. d. 随机误差列，满足 $E\varepsilon_j = 0$，$0 < \sigma^2 \equiv Var(\varepsilon_j) < \infty$。

用 $W_{nj}(t), j = 1,2,\cdots,n$；表示权函数满足 $0 \leqslant W_{nj}(t) \leqslant 1$，$\sum_{j=1}^{n} W_{nj}(t) = 1$。令

$$\tilde{y}_i = y_i - \sum_{j=1}^{n} W_{nj}(t_i) y_j$$

$$\tilde{x}_i = x_i - \sum_{j=1}^{n} W_{nj}(t_i) x_j$$

$$\tilde{V}_n = n^{-1} \sum_{i=1}^{n} \tilde{x}_i \tilde{x}_i^{\tau}$$

$$Z_i = \tilde{V}_n^{-1/2} \tilde{x}_i (y_i - x_i^{\tau}\beta - g(t_i))$$

$$g_n(t) = \sum_{j=1}^{n} W_{nj}(t)(y_j - x_j^{\tau}\beta)$$

检验 β 是否为参数的真值的问题等价于检验 $EZ_i = 0, 1 \leqslant i \leqslant n$ 是否成

立，参考 Owen[52] 的专著，对此可以应用经验似然方法。令 $p_1, p_2, \cdots, p_n$ 非负，满足 $\sum_{i=1}^{n} p_i = 1$。参数取真值时的经验似然比统计量用 $l_0(\beta) = -2\max_{\sum_{i=1}^{n} p_i \tilde{Z}_i = 0} \prod_{i=1}^{n}(np_i)$ 表示。显然，$l_0(\beta)$ 不仅含有未知参数 β，还含有未知回归函数 $g(\cdot)$。所以，$l_0(\beta)$ 不能直接用于对 β 的统计推断。一个自然的解决问题的办法就是用 $g_n(\cdot)$ 代替未知的 $g(\cdot)$，在 Z_i 当中用 $g_n(\cdot)$ 代替 $g(\cdot)$，可以得到

$$\tilde{Z}_i = \tilde{V}_n^{-1/2} \tilde{x}_i (y_i - x_i^{\tau}\beta - g_n(t_i))$$

这样我们可以定义一个估计的经验似然比

$$L(\beta) = -2 \max_{\sum_{i=1}^{n} p_i \tilde{Z}_i = 0} \prod_{i=1}^{n} (np_i)$$

应用 Lagrange 乘子法，p_i 有最优值

$$p_i = n^{-1}(1 + \lambda^{\tau}\tilde{Z}i)^{-1}$$

这里 λ 是如下方程的解

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{Z}_i}{1 + \lambda^{\tau}\tilde{Z}i} = 0$$

所以，得到经验似然比

$$l(\beta) = 2\sum_{i=1}^{n} \log(1 + \lambda^{\tau}\tilde{Z}i) \tag{1.2.5}$$

Wang 和 Jing[76] 证明了在一定的条件下：如果 β_0 是 β 的真值，则 $l(\beta) \xrightarrow{D} x_p^2$，这里"$\xrightarrow{D}$"表示随机变量列依分布收敛。

1.2.4　分组经验似然

Kitamura[38] 应用分组经验似然的方法来处理强平稳的相依数据。分组经验似然的方法主要是将 bootstrap 的分组思想用于经验似然。对于平稳的混合时间序列 $\{X_t, t = 1, 2, \cdots, T\}$，以长度 M >1 进行分组，以 L 表示相邻

两组始点间的距离，则 $n = \left[\frac{T-M}{L}+1\right]$ 为组的个数，其中 $[z]$ 表示对实数 z 取整。令

$$B_i = (X_{(i-1)L+1}, \cdots, X_{(i-1)L+M}); i = 1,2,\cdots,n$$

若参数 θ 满足 $E(m(X_t,\theta)) = 0$，定义分组估计方程

$$b(B_i,\theta) = \frac{1}{M}\sum_{j=1}^{M} m(X_{(i-1)L+j},\theta)$$

显然有 $E(b(B_i,\theta)) = 0$。如果当 $T \to \infty, L = M \to \infty$时，在一些假设下，$b(B_i,\theta)$ 间的相依关系可以忽略。这样，对于参数 θ 的分组经验似然统计推断就基于下面的检验似然比

$$\Re(\theta) = sup\{\prod_{i=1}^{n} w_i \mid w_i > 0, \sum_{i=1}^{n} w_i = 1, \sum_{i=1}^{n} w_i b(B_i,\theta) = 0\}$$

Kitamura[37] 在一定的假设条件下，其中包括 $M \to \theta, MT^{-1/2} \to 0$，证明了当 $T \to \infty$时，有

$$-2\left(\frac{T}{nM}\right)\log\Re(\theta_0) \xrightarrow{D} \chi_p^2,$$

这里 p 是参数 θ 的维数，θ_0 是参数 θ 的真值。

1.3 本书结构

第 2 章：针对误差是 NA 序列的部分线性模型，利用 NA 序列的性质已经取得的研究成果，构造部分线性模型中回归参数的经验似然比检验统计量，并讨论了统计量在参数取真值时是渐近地服从卡方分布的，相应地构造参数的置信区间。

第 3 章：针对误差是 m-相依序列的部分线性模型，利用分组经验似然处理相依数据的优势及 m-相依误差序列的特殊的相依关系，分组经验似然

的方法对模型进行了上述相应的研究。

第4章：针对误差是MA（∞）序列的情形，运用对线性过程的多项式分解的方法及对MA（∞）过程截断的方法，研究MA（∞）序列误差下模型的经验似然统计推断。

第5章：针对误差是ARCH（1）序列的情形，应用鞅的中心极限定理及鞅差序列的不等式，利用满足一定条件下ARCH（1）序列的平稳性及四阶矩存在，进行ARCH（1）误差下所构造的经验似然比统计量性质的研究，而得出相应结论。

第6章：针对具有核实数据的部分线性模型，并且在其误差项是鞅差序列的情形下，利用核实数据，估计真实值，构造模型参数和非参数部分的最小二乘估计量，并且证明估计量的相合性。通过模拟计算来说明对于鞅差序列误差下具有核实数据的部分线性模型，我们所构造的估计量的优良性。

第7章：针对自回归模型，我们考虑一种包含参数与非参数两种估计方法在内的新的估计方法。回归函数的参数估计首先被作为一个粗略的估计。然后这个粗略的估计被非参数乘子进行修正，可以得到非参数估计。利用一个演变的局部L_2拟合准则估计修正项，最终可以得到基于L_2拟合准则的自回归系数的半参数估计，并且讨论了估计的相合性和渐近正态性。