分类号 O 2 1 2　　　　　　　　　　　　　　　密 级

UDC　　　　　　　　　　　　　　　　　　　　编 号 10486

武 汉 大 学
博 士 学 位 论 文

逐段线性回归中
变点问题的统计推断

研 究 生 姓 名：赵　　华　　玲

指导教师姓名、职称：陈 汉 峰　教 授

学 科 、专 业 名 称：概率论与数理统计

研 究 方 向：变　点　问　题

二〇一一年三月

Doctoral Dissertation

On the Statistical Inference on Change-point Problem in Segmented Linear Regression Models

Candidate: Hualing Zhao

Supervisor: Professor Hanfeng Chen

Wuhan University

March 2011

郑 重 声 明

本人的学位论文是在导师指导下独立撰写并完成的, 学位论文没有剽窃、抄袭、造假等违反学术道德、学术规范和侵权行为, 本人愿意承担由此而产生的法律责任和法律后果, 特此郑重声明.

学位论文作者 (签名):

年 月 日

学位论文使用授权书

本论文作者了解学校关于保存、使用学位论文的管理办法及规定，即学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人愿意授权武汉大学将本学位论文的全部或部分内容编入学校数据库和收录到《中国学位论文全文数据库》进行信息服务，可以采用影印、缩印或扫描等复制手段保存或汇编本学位论文。

□本论文提交　年以后，同意发布。□不同意发布。

注：保密学位论文，在解密后适用于本授权书。

学　　院：

学　　号：

专　　业：

作者签名：　　　　　　　　　　指导老师签名：

　　年　月　日　　　　　　　　　　年　月　日

论 文 创 新 点

1. 本论文细致地研究了Liu and Qian (2010)提出的特设检验统计量的大样本性质，获得了该统计量在零假设下的渐近分布，这一渐近分布与Csörgő and Horváth (1997)中参数似然比检验统计量的渐近分布是一致的。在证明过程中，我们发现用于检验逐段线性回归中的变点的似然方法中的拉格朗日系数与Owen (1990)提出的拉格朗日系数具有十分接近的大样本性质。这一结果对拉格朗日系数的逼近与计算以及认识Liu and Qian (2010)提出的特设经验似然方法是十分有用的。

2. 我们改进了Liu and Qian (2010)提出的特设经验似然方法。新方法的主要想法是改进特设经验似然方法中残差的相依性以及充分利用预报变量的信息有效地进行回归分析。通过线性变换使相依的残差成为不相关的量，然后利用回归分析中的正规方程作为经验似然方程的约束方程。因此，提出的新方法被认为是对特设经验似然方法的有效改进。特别指出改进后的方法在计算上与特设经验似然方法是一样的容易计算，并未从本质上加大计算量。蒙特卡洛模拟结果表明新方法极大地改进了Liu and Qian(2010)的特设经验似然方法在有限样本下的有效性。从对一个实际例子的应用中也说明了改进后的方法更有效。此外，本论文给出了检验统计量在零假设下渐近分布的猜想。更进一步地，我们提出了采用自助法(bootstrap)来逼近改进后的新方法的P-值。

摘要

自20世纪50年代以来有关变点的推断和估计一直是统计界的热点问题之一。我们之所以对变点问题很感兴趣是因为变点不仅在工业质量控制（最早产生变点问题的领域）中有广泛的应用，而且在经济、金融、医学、流行病学、心理学、环境科学、气象学，地质学等领域也得到了极大的应用。从统计的角度，变点τ是指一个过程或序列中，从某个未知的时刻τ起，过程或序列的某个统计性质发生了变化。我们对变点问题的研究，主要集中在两个方面：一是检测是否存在变点，也就是假设检验问题；二是如果存在变点，需要对变点做出统计推断，也就是估计问题。本文针对逐段线性回归模型的变点问题进行了研究，其内容分为以下几个部分：

第一章介绍了变点问题的研究背景和意义，给出了有关变点问题的两种模型：分布函数的变点问题和回归模型中的变点问题。

第二章总结了用于研究变点问题已有的方法，主要讨论了参数似然比方法，union-intersection 检验，基于复发回归残差的检验，基于回归残差的检验和信息标准的方法。

第三章介绍了Owen (1988)提出的经验似然方法。由于经验似然既有非参数方法的稳健性又融合了似然方法的有效性和灵活性，所以相对其它经典的统计方法，有许多突出的优点。本章概括了统计泛函和回归模型的经验似然推断。

第四章主要利用经验似然方法研究了逐段线性回归模型中的变点问题。探讨了Liu and Qian (2010)提出的特设检验统计量的大样本性质，得到了其在零假设下的渐近分布，并且这一分布与Csörgő and Horváth (1997)中参数似然比检验统计量的渐近分布是一致的。在本章中提出了一种新的改进特设经验似然的方法。该新方法的主要想法是改进特设经验似然方法中残差的相依性以及充分使用预报变量的信息有效地进行回归分析。通过线性变换使相依的残差成为不相关的量，然后利用回归分析中的正规方程作为经验似然方程的约束方程。因此，提出的新方法被认为是对特设经验似然方法的有效改进。特别指出改进后的方法在计算上与特设经验似然方法是一样的容易计算，并未从本质上加大计算量。蒙特卡洛模拟结果表明新方法极大地改进了Liu and Qian (2010)的特设经验似然方法在有限样本下的有效性。从对一个实际例子的应用中也说明了改进后的方法更有效。此外，本论文给出了检验统

计量在零假设下渐近分布的猜想。更进一步地，我们提出了采用自助法(bootstrap)来逼近改进的新方法的P-值。

第五章分析了美国国家黄石公园喷泉的持续时间和相隔时间的回归关系。通过1980年十月270次数据的散点图可以看出喷泉的持续时间和相隔时间的回归关系可以分为两部分，并且通过R软件得到改进的方法的P-值为0.05比Liu and Qian (2010) 的P-值0.07要小。

最后一章给出了论文的总结与展望。

关键词： 变点；逐段回归模型；似然比方法；经验似然；信息标准；渐近分布；P-值。

Abstract

Since 1950s, the detection and estimation of change-point has been one of the hot topics in statistics. The reasons that we are interested in it are that it has not only important application to industrial quality control, but also can be encountered in many other fields such as economics, finance, medicine, epidemiology, psychology, environment science, meteorology, and geology. From the statistical point of view, a change point refers to an observational time point at which a statistical pattern change occurs during a long-term process or sequence. The research about change-point mainly focus on two aspects : one is to decide if there is any change (often viewed as an hypothesis testing problem), another is to locate the change-point when there is a change present (often viewed as an estimation problem). In this dissertation, we mainly discuss the change-point problem in segmented linear regression models. This dissertation has five chapters that are organized as follows:

In Chapter 1, we introduce the backgrounds and significance of change-point problem and give two models about it, namely, the change-point in distribution functions and the change-point in regression models.

The existing methods used to study the change-point problem are summarized in Chapter 2. It includes parameter likelihood ratio method, the test based on recursive regression residuals and regression residuals, union-intersection test and the method of information criterion.

In Chapter 3, we describe the empirical likelihood method introduced by Owen (1988). Empirical likelihood method combines the reliability of the nonparametric methods with the flexibility and effectiveness of the likelihood approach, so it has many outstanding advantages, compared to other classical methods. This chapter generalizes the empirical likelihood inference about statistical functional and regression model.

In Chapter 4, we study the change-point problem in segmented linear regression models via the empirical likelihood method. We study the large sample properties of the ad hoc empirical likelihood detection procedure proposed by Liu and Qian (2010) and derive the limiting null distribution of the corresponding test statistic. The limiting null distribution appears to be the same as that of likelihood ratio test statistics given by Csörgő and Horváth (1997). In the course of development of the large sample theory, it is proved that the Lagrange multi-

plier in the empirical likelihood method setting under the segmented regression models has similar large sample property as in the ordinary cases studied by Owen (1990). This result can greatly help us understand the ad hoc empirical likelihood detection procedure better and is very useful in numerical approximation to the Lagrange multiplier that is critical in use of empirical likelihood methods. In this chapter, we also propose a new detection procedure to improve the efficiency of the ad hoc empirical likelihood detection procedure. The main idea behind the new detection procedure is to transform the highly correlated residuals on which the ad hoc empirical likelihood method is based to uncorrelated quantities and then the norm equation of regression analysis based on the transformed responses is used as the constraints of the empirical likelihood method. The new detection procedure is computationally as easy as the ad hoc empirical likelihood detection procedure and is expected to be a great improvement. Monte Carlo simulation studies are conducted. The simulation results show that the new detection procedure is indeed substantially more efficient with finite samples. Furthermore, the null limiting distribution of the new detecting test statistic is remarked and a bootstrap estimation procedure is proposed to approximate the new test p-value for the purpose of applications in practice.

In Chapter 5, Liu and Qian (2010)'s detection procedure and the new detection procedure we propose are applied to a real-life data set for the eruptions of the Old Faithful geyser in Yellowstone National Park in U.S.A. The same data set was analyzed by Gombay and Horvath (1994). We reanalyze the relationship between the duration time and the interval about the eruptions of the Old Faithful geyser. The scatter plot of 270 eruptions in October 1980 clearly suggests that the geyser eruption has a change point during October 1980. It is indeed turned out that the p-value based on the new detection procedure is 0.05 and the p-value based on Liu and Qian (2010)'s procedure is 0.07, in sharp contrast to the p-value 0.17 based Gombay and Horváth (1994) under normal parametric models.

In the last chapter, we give a summary of the dissertation and outline a future research plan.

Keywords: change-point; segmented linear regression models; parametric likelihood ratio; empirical likelihood; information criterion; the null limiting distribution; P-value.

表 1 符号说明

符号	数学意义
T	向量或矩阵的转置
$\xrightarrow{P}$	以概率收敛
$\xrightarrow{d}$	以分布收敛
$\stackrel{d}{=}$	同分布
$a.e$	几乎处处成立
$O_p(1)$	以概率有界
$o_p(1)$	以概率收敛到0
χ_r^2	自由度为r的卡方分布
$N(\mu, \sigma^2)$	均值为μ，方差为σ^2的正态分布
$E(X)$	随机变量X的数学期望
$Var(X)$	随机变量X的方差
$i.i.d$	独立同分布
dim	向量的维数
R^d	d维欧式空间
C	任意常数
$\propto$	成正比
[ ]	取整函数
$sgn(\ )$	符号函数

目　录

第一章　引言

§1.1　研究背景

自20世纪50年代以来，有关变点问题的统计推断一直是统计界的热点问题之一。从统计的角度看，所谓的变点τ是指在一个过程或序列中，从某个未知的时刻τ起，过程或序列的某个统计性质发生了变化。有关变点问题的统计推断就是根据具体的背景，对变点τ做出估计，并对估计量的性质进行统计分析。

变点问题最初是Page (1954)从工业质量控制中提出来的，人们从生产线上抽检产品以检测产品的质量是否超过其质量控制范围，当产品的质量发生质变，主要是指超出某个可接受的范围时，希望能及时报警，以免出现更多的次品，这个发生质变的时刻就称为变点。

变点问题不仅发生在质量控制中，而且在生物学，医学，流行病学，环境科学，行为学，信号过程，软件工程，经济学中也有大量的应用。比如，在流行病学中，人们希望知道经过一段时间后，传染病在其传染过程中其传染率是否仍然是个常数，如果不是，则希望估计出其发生变化的时刻，以便确定合理的治疗。在心电图心律分析中，变点诊断方法的应用已经构成了模式识别分析的一部分。在股票市场分析中，尽管根据经济理论，股票的价格是服从正态分布，但是现实中会有很多因素会引起股票的波动，使其不再是正态分布，因而投资者最关心的问题就是哪些突发事件会引起股票市场显著的变化。在地质数据分析中，如果从同一地表的不同地理位置采取的样本有明显的不同，则地理学家就想知道哪些地理位置发生了什么变化，是否地下有煤炭、石油或其它矿物质。再比如在1987年美国高速公路的限速从时速55英里提高到65英里时，人们就想知道交通事故的死亡率是否也跟着发生了变化。人们希望准确预报而又难以精确做到的如金融风暴，地震等突发事件都属于变点问题。

由于变点问题的广泛应用，用现代统计方法对其进行推断具有非常重要的现实意义。变点问题的统计推断本身也成为数理统计学中一个重要的研究分支。对于变点问题的统计推断困难在于我们感兴趣的变点τ是离散值并且它的值域依赖样本量n。在统计中，对于变点问题

我们主要解决以下几个问题：

1. 检测序列中是否存在变点。
2. 如果存在变点，需要估计出变点的个数和相应的位置。
3. 研究估计量的统计性质，如相合性、收敛速度、精确分布或渐近分布等等。

§1.2　有关变点问题的两种模型

针对变点问题，我们往往有两种提法：一种是经典模型中的变点问题；一种是回归模型中的变点问题。下面我们逐一介绍。

§1.2.1　经典模型中的变点问题

设随机变量（向量）$Y_1, Y_2, \cdots, Y_n$(独立或相依)分别服从分布函数$F_1, F_2, \cdots, F_n$，若存在某一未知整数的$1 \le k < n$，使得分布函数的某些统计性质发生了变化，也即，使得F_k与F_{k+1}有显著的不同，此时称k为序列$Y_1, Y_2, \cdots, Y_n$的一个变点。这是单变点问题。

更一般地，若划分$Y_1, Y_2, \cdots, Y_n$为m组，$\{Y_1, \cdots, Y_{k_1}\}$，$\{Y_{k_1+1}, \cdots, Y_{k_2}\}$，$\cdots$，$\{Y_{k_{m-1}+1}, \cdots, Y_n\}$，每组对应的分布函数为$F_1, F_2, \cdots, F_m$，若$F_i, i = 1, \cdots, m$，各不相同，此时称$k_1, \cdots, k_{m-1}$为序列$Y_1, Y_2, \cdots, Y_n$ 的m个变点。这个问题称为多变点问题。

如果已经检测出存在一个变点k_0，则可以用同样的方法分别检测$F_1, \ldots, F_{k_0}$ 和$F_{k_0+1}, \ldots, F_n$ 是否仍有变点，依次类推，可以检测序列是否存在m个变点，从而多变点问题的检测，可以采用检测单变点问题的方法进行检测，所以以下主要讨论单变点问题的检测。

对于单个变点，假设检验一般的提法是：

$H_0 : F_1 = F_2 = \cdots F_n = F$ ，即不存在变点。

$\leftrightarrow \quad H_1 : F_1 = \cdots = F_k \neq F_{k+1} = F_{k+2} = \cdots = F_n$ ，　存在整数$1 \le k < n$。

特别地，若$F_1, F_2, \ldots, F_n$都属于参数分布族$\{F_\theta, \theta \in R^p\}$，则上述假设检验问题就变为：

$H_0 : \theta_1 = \theta_2 = \cdots = \theta_n = \theta$（未知）。

$\leftrightarrow \quad H_1 : \theta_1 = \cdots = \theta_k \neq \theta_{k+1} = \theta_{k+2} = \cdots = \theta_n$，　存在整数$1 \le k < n$。

§1.2.2 回归模型中的变点问题

对于回归模型中的变点问题，我们有下面模型：

设$(X_1, Y_1), \ldots, (X_n, Y_n)$是独立的随机向量，服从下面的回归模型

$$Y_i = \begin{cases} h_1(X_i; \alpha_1) + \varepsilon_1 & a_0 \le X_i \le \tau_1; \\ h_2(X_i; \alpha_2) + \varepsilon_2 & \tau_1 < X_i \le \tau_2; \\ \vdots & \vdots \\ h_m(X_i; \alpha_m) + \varepsilon_i & \tau_{m-1} < X_i \le a_1. \end{cases}$$

其中a_0，a_1是已知的常数，$\{X_i\}_{i=1}^n$ i.i.d，且与$\{\epsilon_i\}_{i=1}^n$i.i.d 相互独立。如果$h_j(X_i; \alpha_j) \ne h_{j+1}(X_i; \alpha_{j+1})$，则称$\tau_j$为变点。

当然对于上述模型，我们也可以考虑回归变量为非随机的情况。当回归变量非随机时，有时为了方便，我们往往假设$a_0 \le x_1 < x_2 < \cdots < x_n \le a_1$，此时上面的模型变为：

$$y_i = \begin{cases} h_1(x_i; \alpha_1) + \varepsilon_1, & 1 \le i \le k_1; \\ h_2(x_i; \alpha_2) + \varepsilon_2, & k_1 < i \le k_2; \\ \vdots & \vdots \\ h_m(x_i; \alpha_m) + \varepsilon_m, & k_{m-1} < i \le n. \end{cases}$$

有时，我们也不加区分地统称上面两种模型均为分段回归模型。特别地，当$h_j(x; \alpha_j)$是α_j的线性函数时，即$h_j(x; \alpha_j) = \sum\limits_{k=0}^{q_j} \alpha_{jk} h_{jk}(x)$，其中$\{\alpha_{jk}\}$未知，$h_{jk}(x)$已知，称上面模型为分段线性回归模型。

目前已经有大量文献用来研究分段回归模型中有关变点的假设检验，参数估计以及相关的计算程序，并且分段回归模型在很多领域也得到了广泛的应用，比如生物学(Piegorsch and Bailer (1997))，环境科学(Qian and Ryu (2003), Toms and Lesperance (2003), Piegorsch and Bailer (2005))，医学(Smith and Cook (1980), Cox(1987), Miggco (2003))，流行病学(Pastor amd Guallar (1998))，软件工程(Qian and Yao (2002))，金融学和经济(Chow (1960), Zeileis (2006))。

对于分段回归模型，我们研究最多的是两阶段简单线性回归模型（two-phase linear regression model）。

Y是响应变量，X是预报变量，$\{(x_i, y_i)\}_{i=1}^n$是X, Y的n个独立观测值，满足

$$y_i = (\alpha_0 + \alpha_1 x_i) I_{\{x_i \le \tau\}} + (\beta_0 + \beta_1 x_i) I_{\{x_i > \tau\}} + \epsilon_i, i = 1, \ldots, n. \tag{1.1}$$

其中随机误差$\epsilon_i, i = 1, \ldots, n, i.i.d, \tau$为未知的变点。

设$\{x_{(i)}\}_{i=1}^n$为$\{x_i\}_{i=1}^n$的顺序统计量，若存在未知k满足$x_{(k)} \leq \tau < x_{(k+1)}$，此时称$k$和$\tau$都为变点，从而不失一般性，假设$x_1 \leq x_2 \leq \cdots \leq x_n$，则上述模型变为

$$y_i = (\alpha_0 + \alpha_1 x_i) I_{\{i \leq k\}} + (\beta_0 + \beta_1 x_i) I_{\{i > k\}} + \epsilon_i, i = 1, \ldots, n. \tag{1.2}$$

我们称模型(1.1), (1.2)均为两阶段简单线性回归模型。

Bhattacharya (1990, 1994)对两阶段线性模型给出了如下分类：

当模型(1.2)在k处连续，即$\alpha_0 + \alpha_1 x_k = \beta_0 + \beta_1 x_k$时，称模型(1.2)为逐段（segmented）或两阶段连续线性模型，变点k也称为连接点(joint-point)；反之，称模型(1.2)为两阶段不连续线性模型。

在两阶段不连续线性模型中，若$\beta_0 + \beta_1 x_k - (\alpha_0 + \alpha_1 x_k)$是一不依赖于n的常数，称其为固定跳(fixed-jump)；若当$n \to \infty$ 时，$\beta_0 + \beta_1 x_k - (\alpha_0 + \alpha_1 x_k) \to 0$，称其为近邻跳（contiguous-jump）.

Hawkins (1980)说明了两阶段回归模型在变点处是否连续对参数的估计和推断都有很大的影响。

对于两阶段回归模型的变点问题的假设检验提法为：

$H_0 : k \geq n$(即不存在变点) $\longleftrightarrow$ $H_1 : 1 \leq k < n$。

或$H_0 : \alpha = \binom{\alpha_0}{\alpha_1} = \beta_0 = \binom{\beta_0}{\beta_1}$ $\longleftrightarrow$ $H_1 : \alpha = \binom{\alpha_0}{\alpha_1} \neq \beta_0 = \binom{\beta_0}{\beta_1}$。

自Quandt (1958)首次利用最大似然比的方法检测两阶段回归模型中的变点以来，至今有不少的文章对两阶段回归模型中的变点问题进行了讨论，比如Hawkins (1989)，Chow (1960)，Sen (1985), Huskova (1993)，Csörgő and Horváth (1997)，Liu, Zou and Zhang (2008)，Liu and Qian (2010)等文献都有讨论，我们将在第二章对其进行详细讨论。

第二章　研究回归模型中变点问题已有的方法

回归分析一直是统计学的一个重要分支，当经典的线性回归模型不足以应对现实生活中的复杂数据时，非线性回归是最好的选择，而非线性回归中应用较广泛的是分段线性回归。由于分段线性回归在经济和其它领域的广泛应用，本章主要介绍用于研究分段线性回归中变点问题已有的方法。

§2.1　似然比检验

两阶段回归模型是由Quandt (1958)最早提出来的，并且利用最大似然法给出了模型中参数的估计。Quandt (1960)用似然比检验对回归模型中是否存在变点做出了检验，并给出了当随机误差服从正态分布时，检验统计量的渐近分布。Csörgő and Horváth (1997)对此做出了更进一步的总结。

Quandt (1958)中假设样本$(x_i, y_i), i = 1, \ldots, n$ 独立，满足下面的回归模型:

$$y_i = \begin{cases} \alpha_0 + \alpha_1 x_i + \epsilon_1 & i \leq k; \\ \beta_0 + \beta_1 x_j + \epsilon_2 & j > k. \end{cases} \tag{2.1}$$

其中$\epsilon_1 \sim N(0, \sigma_1^2)$，与$\epsilon_2 \sim N(0, \sigma_2^2)$独立。
则$(x_i, y_i), i = 1, \ldots, n$的似然函数为

$$(\frac{1}{\sqrt{2\pi\sigma_1^2}})^k (\frac{1}{\sqrt{2\pi\sigma_2^2}})^{n-k} \exp\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{k}(y_i - \alpha_0 - \alpha_1 x_i)^2 - \frac{1}{2\sigma_2^2}\sum_{j=k+1}^{n}(y_j - \beta_0 - \beta_1 x_j)^2\}.$$

取对数后得对数似然为:

$$\begin{aligned} L = & -n\log(\sqrt{2\pi}) - k\log(\sigma_1) - (n-k)\log(\sigma_2) \\ & - \frac{1}{2\sigma_1^2}\sum_{i=1}^{k}(y_i - \alpha_0 - \alpha_1 x_i)^2 - \frac{1}{2\sigma_2^2}\sum_{j=k+1}^{n}(y_j - \beta_0 - \beta_1 x_j)^2. \end{aligned} \tag{2.2}$$

从而，对任意固定的k, $\alpha_1, \alpha_0, \beta_1, \beta_0, \sigma_1^2, \sigma_2^2$的最大似然估计分别为:

$$\hat{\alpha}_1 = \frac{k\sum_{i=1}^{k} x_i y_i - \sum_{i=1}^{k} x_i \sum_{i=1}^{k} y_i}{k\sum_{i=1}^{k} x_i^2 - (\sum_{i=1}^{k} x_i)^2}, \qquad \hat{\alpha}_0 = \frac{\sum_{i=1}^{k} y_i}{k} - \hat{\alpha}_1 \frac{\sum_{i=1}^{k} x_i}{k},$$

$$\hat{\beta}_1 = \frac{(n-k)\sum_{j=k+1}^{n} x_j y_j - \sum_{j=k+1}^{n} x_j \sum_{j=k+1}^{n} y_j}{(n-k)\sum_{j=k+1}^{n} x_j^2 - (\sum_{j=k+1}^{n} x_j)^2}, \qquad \hat{\beta}_0 = \frac{\sum_{j=k+1}^{n} y_j}{n-k} - \hat{\beta}_1 \frac{\sum_{j=k+1}^{n} x_j}{n-k},$$

$$\hat{\sigma}_1^2 = \frac{\sum_{i=1}^{k} (y_i - \hat{\alpha}_0 - \hat{\alpha}_1 x_i)}{k}, \qquad \hat{\sigma}_2^2 = \frac{\sum_{j=k+1}^{n} (y_j - \hat{\beta}_0 - \hat{\beta}_1 x_j)}{n-k}. \tag{2.3}$$

上面得到$\alpha_1, \alpha_0, \beta_1, \beta_0, \sigma_1^2, \sigma_2^2$的MLE显然也是其最小二乘估计，将$\hat{\alpha}_1, \hat{\alpha}_0, \hat{\beta}_1, \hat{\beta}_0, \hat{\sigma}_1^2, \hat{\sigma}_2^2$ 带入(2.2)式。我们得到关于k的函数：

$$L(k) = -n\log(2\pi) - k\log(\hat{\sigma}_1) - (n-k)\log(\hat{\sigma}_2) - n/2. \tag{2.4}$$

则取k的估计$\hat{k} = \arg\max_k L(k)$。

Quandt (1958)指出在求$\hat{k}$时需要注意以下几个问题：

1. $L(k)$ 中的变量k取整数，因而$L(k)$ 关于k并非连续变化的，从而在求其最大值时不能采用微分的办法。

2. 由于$L(k)$ 可能存在多个极大值点，所以仅仅满足条件$L(k-1) < L(k), L(k+1) < L(k)$ 的k也是不行的。

3. 对所有可能的k求$L(k)$，然后取使得$L(k)$达到最大的那个k。

当然现在通过运用统计软件求$\hat{k}$变得相当容易了。但是Quandt并未对最大似然估计的统计性质比如相合性、渐近分布作出任何分析。

Robison (1964)在假设变点k属于某个区间的条件下，给出了回归系数的区间估计. Feder (1975a)考虑了一类更广的由r阶段组成的回归模型且模型在变点处连续，并且指出运用最大似然估计的渐近理论分析分段回归模型的困难在于：最大似然估计渐近正态性是在假设对数似然满足某些正则条件下得到的，而分段回归模型由于变点k未知，不满足要求的那些正则条件，从而对最大似然估计的统计性质作出分析是有些困难的。后面我们在假设模型满足一定的条件下给出了估计量的相合性和收敛速度。

事实上，在估计参数之前，我们应该先判断是否存在变点，也即需要对假设

$$H_0:\alpha=\begin{pmatrix}\alpha_0\\ \alpha_1\end{pmatrix}=\beta_0=\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix},\ \sigma_1=\sigma_2 \ \longleftrightarrow\ H_1:\alpha=\begin{pmatrix}\alpha_0\\ \alpha_1\end{pmatrix}\neq\beta=\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix},\sigma_1\neq\sigma_2$$

做出检验。

Quandt (1960)首次用似然比检验对上面假设进行了检验。记

$$\hat{\alpha}_{11}=\frac{k\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{k}x_i\sum_{i=1}^{n}y_i}{k\sum_{i=1}^{n}x_i^2-(\sum_{i=1}^{n}x_i)^2},\qquad \hat{\alpha}_{10}=\frac{\sum_{i=1}^{n}y_i}{n}-\hat{\alpha}_1\frac{\sum_{i=1}^{n}x_i}{n},\qquad \hat{\sigma}^2=\frac{\sum_{i=1}^{n}(y_i-\hat{\alpha}_{10}-\hat{\alpha}_{11}x_i)}{n}.$$

则似然比检验统计量为：

$$\lambda=\frac{\sup_{\theta\in\Theta}L(\theta)}{\sup_{\theta\in\Theta_0}L(\theta)}=\frac{\hat{\sigma}_1^k\hat{\sigma}_2^{n-k}}{\hat{\sigma}^n}$$

其中Θ表示整个参数空间，$\Theta_0\in\Theta$表示在H_0下的参数空间。

Quandt (1960)从对检验统计量的经验分布的分位数的分析中，指出$-2\log\lambda$的渐近分布不可能是自由度为4的χ^2分布，这是因为要使得当$n\to\infty$时，$-2\log\lambda\to\chi_4^2$，必须要假设似然函数$L(k)$在最大值处的导函数必须为零，而在上面我们也说明了$L(k)$ 中的变量k取整数，求其最大值时不能采用微分的方法。Quandt 还指出不管是大样本还是有限样本，k越接近$n/2$,得到的功效越大。

Hinkley (1969)也是从经验的角度指出由Quandt (1960)提出的$-2\log\lambda$的渐近分布非常接近自由度为3的χ^2分布。

Feder (1975b)证明了若回归模型在零假设下可辨，则检验统计量服从合适自由度的χ^2分布，具体结论可参考Feder (1975b)中的定理3.2。

Horváth (1995)给出了似然比检验统计量的大样本性质，Csörgő (1997)对此性质作出了进一步的总结。

Quandt (1958)利用参数似然的方法仅分析了简单回归模型中的变点问题，但是把它推广到多元回归模型是没有任何困难的，下面给出具体的介绍并给出检验统计量的渐近分布。

§2.1.1 检验统计量在零假设下的渐近分布

1. 仅仅回归系数发生了变化

设$(y_i, x_{i1}, x_{i2}), i=1,\ldots,n$ 满足下面的回归模型：

$$y_i = \begin{cases} x_{i1}^T\beta + x_{i2}^T\gamma + \epsilon_i & 1 \le i \le k_0; \\ x_{i1}^T\beta^* + x_{i2}^T\gamma + \epsilon_i & k_0 < i \le n. \end{cases} \tag{2.5}$$

其中$x_{i1} \in R^d, x_{i2} \in R^p, 1 \le i \le n$是已知的列向量，回归系数$\beta, \beta^*, \gamma$是未知的列向量，$\epsilon_1, \epsilon_2, \ldots, \epsilon_n \overset{i.i.d}{\sim} N(0, \sigma^2)$。

考虑下面的假设检验问题：

$$H_0: k_0 \ge n, \quad \longleftrightarrow \quad H_1: 1 \le k_0 < n. \tag{2.6}$$

为了计算上的方便，我们引入下面记号：

$$\mathbf{Y_k} = \begin{pmatrix} y_1 \\ \vdots \\ y_k \end{pmatrix}, \quad \mathbf{Y_k^*} = \begin{pmatrix} y_{k+1} \\ \vdots \\ y_n \end{pmatrix}, \quad \epsilon_\mathbf{k} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_k \end{pmatrix}, \quad \epsilon_\mathbf{k}^* = \begin{pmatrix} \epsilon_{k+1} \\ \vdots \\ \epsilon_n \end{pmatrix},$$

$$\mathbf{X_{11k}} = \begin{pmatrix} x_{11}^T \\ \vdots \\ x_{k1}^T \end{pmatrix}, \quad \mathbf{X_{12k}} = \begin{pmatrix} x_{12}^T \\ \vdots \\ x_{k2}^T \end{pmatrix}, \quad \mathbf{X_{21k}} = \begin{pmatrix} x_{(k+1)1}^T \\ \vdots \\ x_{n1}^T \end{pmatrix},$$

$$\mathbf{X_{22k}} = \begin{pmatrix} x_{(k+1)2}^T \\ \vdots \\ x_{n2}^T \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X_{11k}} & \mathbf{X_{12k}} \\ \mathbf{X_{21k}} & \mathbf{X_{22k}} \end{pmatrix}.$$

则在H_0成立时，$\beta = \beta^*, \gamma, \sigma^2$的最大似然估计$\hat{\beta}_n, \hat{\gamma}_n, \hat{\sigma}_n^2$满足

$$\begin{aligned} \mathbf{X_{11n}^T X_{11n}}\hat{\beta}_n + \mathbf{X_{11n}^T X_{12n}}\hat{\gamma}_n &= \mathbf{X_{11n}^T Y_n} \\ \mathbf{X_{12n}^T X_{11n}}\hat{\beta}_n + \mathbf{X_{12n}^T X_{12n}}\hat{\gamma}_n &= \mathbf{X_{12n}^T Y_n} \\ \hat{\sigma}_n^2 &= \frac{1}{n}\|\mathbf{Y_n} - (\mathbf{X_{11n}}\hat{\beta}_n + \mathbf{X_{12n}}\hat{\gamma}_n)\|^2 \end{aligned} \tag{2.7}$$

从而H_0成立时，似然函数

$$\begin{aligned} \Lambda_n^* &= \sup_{\beta,\gamma,\sigma^2} \big(\frac{1}{2\pi\sigma^2}\big)^{\frac{n}{2}} \exp\big(-\frac{1}{2\sigma^2}\|Y_n - (\mathbf{X_{11n}}\beta + \mathbf{X_{12n}}\gamma)\|^2\big) \\ &= (2\pi\hat{\sigma}_n^2)^{-\frac{n}{2}} e^{-\frac{n}{2}} \end{aligned}$$

当H_1成立时，$\beta, \beta^*, \gamma, \sigma^2$的最大似然估计$\hat{\beta}_k, \hat{\beta}_k^*, \hat{\gamma}_k, \hat{\sigma}_k^2$满足

$$\mathbf{X_{11k}^T X_{11k}}\hat{\beta}_k + \mathbf{X_{11k}^T X_{12k}}\hat{\gamma}_k = \mathbf{X_{11k}^T Y_k}$$

$$\mathbf{X_{21k}^T X_{21k}}\hat{\beta}_k^* + \mathbf{X_{21k}^T X_{22k}}\hat{\gamma}_k = \mathbf{X_{21k}^T Y_k^*}$$

$$\mathbf{X_{12k}^T X_{11k}}\hat{\beta}_k + \mathbf{X_{22k}^T X_{21k}}\hat{\beta}_k^* + (\mathbf{X_{12k}^T X_{12k}} + \mathbf{X_{22k}^T X_{22k}})\hat{\gamma}_k = \mathbf{X_{12k}^T Y_k} + \mathbf{X_{22k}^T Y_k^*}$$

$$\hat{\sigma}_{1k}^2 = \frac{1}{n}\|\mathbf{Y_k} - (\mathbf{X_{11k}}\hat{\beta}_k + \mathbf{X_{12k}}\hat{\gamma}_k)\|^2,$$

$$\hat{\sigma}_{2k}^2 = \frac{1}{n}\|\mathbf{Y_k^*} - (\mathbf{X_{21k}}\hat{\beta}_k^* + \mathbf{X_{22k}}\hat{\gamma}_k)\|^2,$$

$$\hat{\sigma}_k^2 = \hat{\sigma}_{1k}^2 + \hat{\sigma}_{2k}^2$$

从而，当H_1成立时，似然函数

$$\begin{aligned}\mathbf{\Lambda}_k^* = &\sup_{\beta,\beta^*,\gamma,\sigma^2}\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}}\exp\Big(-\frac{1}{2\sigma^2}(\|\mathbf{Y_k} - (\mathbf{X_{11n}}\beta + \mathbf{X_{12n}}\gamma)\|^2 \\ &+ \|\mathbf{Y_k^*} - (\mathbf{X_{21k}}\beta^* + \mathbf{X_{22k}}\gamma)\|^2)\Big) \\ =&(2\pi\hat{\sigma}_k^2)^{-\frac{n}{2}}e^{-\frac{n}{2}}\end{aligned}$$

所以对任意固定的k，似然比

$$\mathbf{\Lambda_k} = \frac{\mathbf{\Lambda}_n^*}{\mathbf{\Lambda}_k^*} = (\hat{\sigma}_k^*/\hat{\sigma}_n^*)^{\frac{n}{2}}$$

由于变点k未知，所以我们采用检验统计量

$$Z_n = \max_{d+p\le k\le n-(d+p)}(-2\log\mathbf{\Lambda_k})$$

假设

$C.1$ $rank(\mathbf{X_{11k}}) = d, \forall\ d \le k \le n$.

$C.2$ $rank(\mathbf{X_{12k}}) = P, \forall\ P \le k \le n$.

$C.3$ $rank(\mathbf{X_{21k}}) = d, \forall\ 1 \le k \le n-d$.

$C.4$ $rank(\mathbf{X_{22k}}) = P, \forall\ 1 \le k \le n-p$.

$C.5$ $r_1(t) = (\log t)^{-v}, v > 0, r_2(t) = (\log\log t)^{-\frac{3}{2}}$，存在矩阵$A, B, C$使得

$$\|\frac{1}{k}\mathbf{X_{11k}^T X_{11k}} - A\| = o(r_1(k)),\ k \longrightarrow \infty,$$

$$\|\frac{1}{k}\mathbf{X_{12k}^T X_{11k}} - B\| = o(r_2(k)),\ k \longrightarrow \infty,$$

$$\|\frac{1}{k}\mathbf{X_{12k}^T X_{12k}} - C\| = o(r_2(k)),\ k \longrightarrow \infty,$$

$$\|\frac{1}{n-k}\mathbf{X}_{\mathbf{21k}}^{\mathbf{T}}\mathbf{X}_{\mathbf{21k}}-A\|=o(r_1(n-k)),\ n-k\longrightarrow\infty,$$

$$\|\frac{1}{n-k}\mathbf{X}_{\mathbf{21k}}^{\mathbf{T}}\mathbf{X}_{\mathbf{22k}}-B\|=o(r_2(n-k)),\ n-k\longrightarrow\infty,$$

$$\|\frac{1}{n-k}\mathbf{X}_{\mathbf{22k}}^{\mathbf{T}}\mathbf{X}_{\mathbf{22k}}-C\|=o(r_2(n-k)),\ n-k\longrightarrow\infty.$$

$C.6$ 矩阵

$$\begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$$

的秩为$d+p$.

Horvath (1995)关于检验统计量Z_n的渐近分布给出了下面的定理：

定理2.1.1. *在H_0和$C.1-C.6$，存在某个$v>12$成立时，且$\epsilon_1,\epsilon_2,\ldots,\epsilon_n\overset{i.i.d}{\sim}N(0,\sigma^2)$时，有*

$$\lim_{n\to\infty}\ P\{A(\log n)Z_n^{1/2}\le t+D_d(\log n)\}=\exp(-2e^{-t}),\ \forall\ t. \tag{2.8}$$

其中$A(x)=(2\log x)^{1/2}, D_d(x)=2\log x+\frac{d}{2}\log\log x-\log\Gamma(d/2)$，$\Gamma(t)$为$Gamma$函数。

事实上，当随机误差服从更广的一类分布时，(2.8)式仍然成立。

$$\epsilon_1,\epsilon_2,\ldots,\epsilon_n, i.i.d\ E\ (\epsilon_i)=0,\ 0<\sigma^2=Var\ (\sigma_i)<\infty,\ E\ |\epsilon_i|^{2+\delta}<\infty,\delta>0. \tag{2.9}$$

则有下面的定理：

定理2.1.2. *当H_0和$C.1-C.6$, 存在某个$v>2+27/\min(1,\delta)$成立时，且随机误差$\epsilon_1,\epsilon_2,\ldots,\epsilon_n$满足(2.9)式，则(2.8)式仍然成立。*

注: 当用$\max\limits_{c_1\le k\le n-c_2}(-2\log\mathbf{\Lambda_k}), c_1\ge d+p, c_2\ge d+p$ 代替Z_n时，定理2.1.1 和定理2.1.2仍然成立。

2. 回归系数和随机误差的方差均发生变化

在实际应用中，随机误差的方差也许会随着时间的变化而发生了变化，所以接着我们讨论回归系数和方差均发生变化时的假设检验。为

了记号上的方便，我们假设不存在冗余参数γ，即考虑下面的模型

$$y_i = \begin{cases} x_{i1}^T\beta + \epsilon_i, & 1 \le i \le k_0; \\ x_{i1}^T\beta^* + \epsilon_i^*, & k_0 < i \le n. \end{cases} \tag{2.10}$$

其中$x_i \in R^d, 1 \le i \le n$是已知的列向量，回归系数$\beta, \beta^*$是未知的列向量。

$$\epsilon_1, \epsilon_2, \ldots, \epsilon_n \ \ i.i.d, E(\epsilon_i) = 0,$$
$$Var(\epsilon_1) = \cdots = Var(\epsilon_k) = \sigma_1^2 \ne Var(\epsilon_{k+1}) = \cdots = Var(\epsilon_n) = \sigma_2^2. \tag{2.11}$$

记

$$\mathbf{X_{1k}} = \begin{pmatrix} x_1^T \\ \vdots \\ x_k^T \end{pmatrix}, \qquad \mathbf{X_{2k}} = \begin{pmatrix} x_{k+1}^T \\ \vdots \\ x_k^T \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} \mathbf{X_{1k}} \\ \mathbf{X_{2k}} \end{pmatrix}.$$

则$\beta, \sigma_1^2, \beta^*, \sigma_2^2$的最大似然估计$\hat{\beta}_k, \hat{\sigma}_k^2, \tilde{\beta}_k^*, \tilde{\sigma}_k^2$为

$$\hat{\beta}_k = (\mathbf{X_{1k}^T X_{1k}})^{-1}\mathbf{X_{1k}^T Y_k}, \quad d \le k \le n.$$

$$\hat{\sigma}_k^2 = \frac{1}{k}\|\mathbf{Y_k} - \mathbf{X_{1k}}\hat{\beta}_k\|^2, \quad d \le k \le n.$$

$$\tilde{\beta}_k^* = (\mathbf{X_{2k}^T X_{2k}})^{-1}\mathbf{X_{2k}^T Y_k^*}, \quad 1 \le k \le n-d.$$

$$\tilde{\sigma}_k^2 = \frac{1}{n-k}\|\mathbf{Y_k^*} - \mathbf{X_{2k}}\tilde{\beta}_k^*\|^2, \quad 1 \le k \le n-d.$$

从而

$$\begin{aligned} \mathbf{\Lambda}_n^{**} &= \sup_{\beta,\sigma^2}\Big(\frac{1}{2\pi\sigma^2}\Big)^{\frac{n}{2}}\exp\Big(-\frac{1}{2\sigma^2}\|\mathbf{Y_n} - \mathbf{X_{1n}}\beta\|^2\Big) \\ &= (2\pi\hat{\sigma}_n^2)^{-\frac{n}{2}}e^{-\frac{n}{2}} \end{aligned}$$

$$\begin{aligned} \mathbf{\Lambda}_k^{**} &= \sup_{\beta,\beta^*,\sigma^2,\tau^2}(\frac{1}{2\pi\sigma^2})^{\frac{k}{2}}\exp(-\frac{1}{2\sigma^2}\|\mathbf{Y_k} - \mathbf{X_{1n}}\beta\|^2) \times (\frac{1}{2\pi\sigma^2})^{\frac{n-k}{2}}\exp(-\frac{1}{2\tau^2}\|\mathbf{Y_k^*} - \mathbf{X_{2n}}\beta^*\|^2) \\ &= (2\pi\hat{\sigma}_k^2)^{-\frac{k}{2}}(2\pi\tilde{\sigma}_k^2)^{-\frac{n-k}{2}}e^{-\frac{n}{2}} \end{aligned}$$

所以对任意固定的k，似然比

$$\tilde{\mathbf{\Lambda}}_{\mathbf{k}} = \frac{\mathbf{\Lambda}_n^{**}}{\mathbf{\Lambda}_k^{**}} = \frac{\hat{\sigma_k^k}\ \tilde{\sigma}_k{}^{n-k}}{\hat{\sigma_n}{}^n}, d \le k \le n-d.$$

由于变点k未知，所以我们采用检验统计量

$$\tilde{Z}_n = \max_{d+p \le k \le n-(d+p)}(-2\log\tilde{\mathbf{\Lambda}}_{\mathbf{k}}).$$

当$\tilde{Z}_n$比较大时，拒绝H_0，我们承认回归模型中存在变点，且变点k的估计

$$\hat{k} = \arg\max_k(-2\log\tilde{\mathbf{\Lambda}}_{\mathbf{k}}).$$

由于这里没有冗余参数，所以条件$C.1-C.6$可由下面的条件代替：

$C.7$ $rank(\mathbf{X_{1k}}) = d,\ \forall\ d \le k \le n$.

$C.8$ $rank(\mathbf{X_{2k}}) = p,\ \forall\ 1 \le k \le n-d$.

$C.9$ $r_1(t) = (\log t)^{-v}, v > 0$, 存在矩阵$A$使得

$$\|\frac{1}{k}\mathbf{X_{1k}^T X_{1k}} - A\| = o(r_1(k)),\ k \to \infty.$$

$$\|\frac{1}{n-k}\mathbf{X_{2k}^T X_{1k}} - A\| = o(r_1(n-k)),\ n-k \to \infty.$$

$C.10$ $rank(A) = d.$

定理2.1.3. 当H_0和$C.7-C.10, v > 12$成立，且随机误差$\epsilon_1, \ldots, \epsilon_n$满足(2.11)时，则

$$\lim_{n\longrightarrow\infty} P\{A(\log n)\tilde{Z}_n^{1/2} \le t + D_{d+1}(\log n)\} = \exp(-2e^{-t}),\ \forall\ t.$$

注： 当模型变化的参数的个数是$d+1$，定理2.1.3中$d+1$起的作用类似于χ^2分布自由度$d+1$。

同只有回归系数变化时一样，定理2.1.3中随机误差服从更广的一类分布时，定理2.1.3的结论仍然成立。

$$\begin{aligned} &\epsilon_1, \epsilon_2, \ldots, \epsilon_n, i.i.d\ E(\epsilon_i) = 0,\ 0 < \sigma^2 = Var(\sigma_i) < \infty,\ E(\epsilon_i^3) = 0, \\ &\qquad E(\epsilon^4) = 3\sigma^4,\ E\ (|\epsilon_i|^{4+\delta}) < \infty,\ \exists\ \delta > 0 \end{aligned} \tag{2.12}$$

定理2.1.4. 当H_0和$C.7-C.10, v > 2 + 27/\min(1,\delta)$成立，且随机误差$\epsilon_1, \ldots,$ ϵ_n满足(2.12)时，有

$$\lim_{n\longrightarrow\infty} P\{A(\log n)\tilde{Z}_n^{1/2} \le t + D_{d+1}(\log n)\} = \exp(-2e^{-t}),\ \forall\ t.$$

注： 2.12式中的条件 $E(\epsilon_i) = 0,\ E(\epsilon_i^3) = 0,\ E\ (\epsilon^4) = 3\sigma^4$也就是说$\epsilon_i/\sigma$的前四阶矩与标准正态分布的一样。当$\max\limits_{c_1 \le k \le n-c_2}(-2\log\tilde{\mathbf{\Lambda}}_{\mathbf{k}}), c_1 \ge d, c_2 \ge d$ 代替$\tilde{Z}_n$时，定理2.1.3 和定理2.1.4仍然成立。

§2.1.2 统计量在备择假设下的相合性

下面我们讨论统计量在备择假设下的相合性。为了方便我们均假设在没有冗余参数的模型下讨论相合性，即考虑模型(2.10)。假设在备择假设下真正的变点k_0依赖n，并且

$$\min(k_0, n-k_0) \to \infty. \tag{2.13}$$

$$\epsilon_1, \ldots, \epsilon_n,\ \ i.i.d,\ E\,(\epsilon_i) = 0,\ 0 < \sigma^2 = E\,(\epsilon_i^2) < \infty. \tag{2.14}$$

定理2.1.5. 假设(2.10), (2.13), (2.14)成立，且

$$\frac{1}{n}\mathbf{X_{1n}^T X_{1n}} \to A,\ \frac{1}{k_0}\mathbf{X_{1k_0}^T X_{1k_0}} \to A,\ \frac{1}{n-k_0}\mathbf{X_{2k_0}^T X_{2k_0}} \to A. \tag{2.15}$$

其中$rank(A) = d,\ \delta = \beta^* - \beta.$

$$\|\delta\|^2 \min(k_0, n-k_0) \to \infty. \tag{2.16}$$

则在H_1成立时

$$\frac{-2\log \mathbf{\Lambda_{k_0}}}{n\log(\frac{1}{\sigma^2}\frac{k_0(n-k_0)}{n^2}\delta^T A\delta + 1)} \xrightarrow{P} 1.$$

记

$$\mathbf{X_{kk_0}} = \begin{pmatrix} x_k^T \\ \vdots \\ x_{k_0}^T \end{pmatrix} \qquad \mathbf{Y_{kk_0}} = \begin{pmatrix} y_k \\ \vdots \\ y_{k_0} \end{pmatrix} \qquad \epsilon_{kk_0} = \begin{pmatrix} \epsilon_k \\ \vdots \\ \epsilon_{k_0} \end{pmatrix}, \qquad 1 \le k < k_0.$$

$$\mathbf{X_{k_0k}} = \begin{pmatrix} x_{k_0+1}^T \\ \vdots \\ x_{k_n}^T \end{pmatrix} \qquad \mathbf{Y_{k_0k}} = \begin{pmatrix} y_{k_o+1} \\ \vdots \\ y_n \end{pmatrix} \qquad \epsilon_{k_0k} = \begin{pmatrix} \epsilon_{k_0+1} \\ \vdots \\ \epsilon_n \end{pmatrix}, \qquad k_0 < k < n.$$

定理2.1.6. 假设(2.9), (2.10), (2.13)，$C.7 - C.10$成立，且

$$\frac{(n-k_0)k_0}{n}\delta^T A\delta / \log\log n \to \infty. \tag{2.17}$$

$$\|\frac{1}{k_0-k}\mathbf{X_{kk_0}^T X_{kk_0}} - A\| = o(r_1(k_0-k)),\ \ k_0 - k \to 0. \tag{2.18}$$

$$\|\frac{1}{k-k_0}\mathbf{X}_{\mathbf{k_0k}}^{\mathbf{T}}\mathbf{X}_{\mathbf{k_0k}}-A\| = o(r_1(k-k_0)), \quad k-k_0 \to 0. \tag{2.19}$$

其中$r_1(t)=(\log t)^{-\beta}, \beta > 2+27/\delta$,则当$H_1$成立时，有

$$\frac{\max_{d\le k\le n-d} -2\log \mathbf{\Lambda_k}}{n\log(\frac{\delta^T A\delta}{\sigma^2}\frac{k_0}{n}\frac{n-k_0}{n}+1)} \xrightarrow{P} 1.$$

以上我们讨论了只有回归系数变化时，统计量的相合性。接下来我们讨论回归系数和方差均发生改变时，统计量的相合性。

假设在模型(2.10)下，随机误差满足下面的条件

$$\begin{aligned} &\{\epsilon_i\},\ i.i.d,\ E\ \epsilon_i=0,\ 0<\sigma_1^2 = E\ \epsilon_i^2 < \infty, 1,\ \ 1\le i\le k_0;\\ &\{\epsilon_i\},\ i.i.d,\ E\ \epsilon_i=0,\ 0<\sigma_2^2 = E\ \epsilon_i^2 < \infty, 1,\ \ k_0 < i\le n;\\ &\{\epsilon_i,\ \ 1\le i\le k_0\}, \{\epsilon_i,\ \ k_0 < i\le n\}\text{独立} \end{aligned} \tag{2.20}$$

定理2.1.7. 当模型(2.10), (2.13), (2.15), (2.16)成立时, 且随机误差满足(2.20), 则在H_1成立时，有

$$\frac{-2\log\tilde{\Lambda}_{k^*}}{n\log(\frac{k_0}{n}\frac{n-k_0}{n}\delta^T A\delta+\frac{k_0}{n}\sigma_1^2+\frac{n-k_0}{n}\sigma_2^2)-k_0\log\sigma_1^2-(n-k_0)\log\sigma_2^2} \xrightarrow{P} 1,$$

设

$$\eta_1,\dots,\eta_n,\ i.i.d,\ E\,(\eta_i) = 0, E\,(\eta_i^2) = 1, E\,\|\eta_i\|^{2+\delta} < \infty,\ 0 < \delta \le 1. \tag{2.21}$$

$$\epsilon_i = \begin{cases} \sigma\eta_i, & 1 \le i \le k_0;\\ (\sigma^2+h)^{1/2}\eta_i, & k_0 < i \le n, h=\sigma_2^2-\sigma_1^2. \end{cases} \tag{2.22}$$

$$\frac{(n-k_0)k_0}{n}h^2/\log\log n \to \infty. \tag{2.23}$$

定理2.1.8. 当(2.10), (2.13), (2.15), (2.16), (2.21), (2.22, (2.23)成立，且$C.7-C.10$对某个$\beta > 2+27/\delta$成立时，则在H_1成立时，当$n \to \infty$有

$$\frac{\tilde{Z}_n}{n\log(\frac{k_0}{n}\frac{n-k_0}{n}\delta^T A\delta+\frac{k_0}{n}\sigma_1^2+\frac{n-k_0}{n}\sigma_2^2)-k_0\log\sigma_1^2-(n-k_0)\log\sigma_2^2} \xrightarrow{P} 1,$$

§2.1.3 变点k的估计的统计性质

当检验统计量Z_n比较大时，拒绝H_0。我们承认回归模型中确实存在变点，这时应该给出变点k_0的一个估计，很自然的一个选择就是：

$$\hat{k} = \hat{k}(n) = \inf\{k : \max_{d \le i \le n-d} -2\log \mathbf{\Lambda_i} = -2\log \mathbf{\Lambda_k}\}$$

以下我们在模型(2.10)下讨论$\hat{k}$的统计性质，假设变点的真值$k_0 = [n\tau]$, $0 < \tau < 1$，下面定理给出了只有回归系数发生变化时$\hat{k}$的弱相合性和渐近分布。

定理2.1.9. 假设(2.10), (2.13), (2.14), (2.17), (2.18), (2.19), $C.7 - C.10$成立，则当$n \to \infty$时，有

$$\hat{k}/n \xrightarrow{P} \tau.$$

更进一步，若还有$\|\delta\| \to 0$, 当$n \to \infty$时，$n\|\delta\|^2/(\log\log n) \to \infty$，则

$$\|\delta\|^2|\hat{k} - k_0| = O_p(1).$$

$$\frac{1}{\sigma^2}\delta^T A\delta(\hat{k} - k_0) \xrightarrow{d} \hat{\eta}.$$

其中$\hat{\eta} = \inf\{t : \hat{W}(t) = \sup\limits_{-\infty < s < \infty} \hat{W}(s)\}$

$$\hat{W}(t) = \begin{cases} W_1(-t) - \frac{1}{2}|t| & if\ t < 0, \\ 0 & if\ t = 0, \\ W_2(-t) - \frac{1}{2}|t| & if\ t > 0. \end{cases}$$

$\{W_1(t),\ t \ge 0\}, \{W_2(t),\ t \ge 0\}$是两个独立的$Wiener$过程。

类似地，当回归系数和随机误差的方差都发生变化时，变点k_0的估计为：

$$\tilde{k} = \tilde{k}(n) = \inf\{k : \max_{d \le i \le n-d} -2\log \tilde{\mathbf{\Lambda}}_\mathbf{i} = -2\log \tilde{\mathbf{\Lambda}}_\mathbf{k}\}.$$

定理2.1.10. 当(2.10), (2.13), (2.17), (2.18), (2.19), (2.21), (2.22), (2.23), $C.7-C.10$, $\beta > 2 + 27/\delta$成立时，则当$n \longrightarrow \infty$时，有

$$\tilde{k}/n \xrightarrow{P} \tau.$$

若还有

$$\|\delta\| \longrightarrow 0,$$

$$\frac{n\|\delta\|^2}{\log\log n} \to \infty, n \to \infty.$$

$$\frac{nh^2}{\log\log n} \to \infty, n \to \infty.$$

则进一步有

(i)

$$(\|\delta\|^2 + h^2)|\hat{k} - k_0| = O_p(1).$$

(ii) 令$m_3 = E(\eta_i^3) = 0, m_4 = E(\eta_i^4)$,

$$\frac{\frac{1}{\sigma^2}\delta^T A\ \delta + \frac{1}{2}(\frac{h}{\sigma^2})^2)^2}{\frac{1}{\sigma^2}\delta^T A\ \delta + (\frac{h}{\sigma^2})^2(m_4 - 1)}(\tilde{k} - k_0) \xrightarrow{d} \hat{\eta}.$$

§2.2 Union-Intersection 检验

Union-Intersection 检验最早是由Hawkins (1989)提出来的，它的基本思想是基于参数估计值的比较，如果确实存在变点，那么发生变化的参数的估计值之间也应该存在较大的差异。下面对模型(2.10)只有回归系数发生变化且随机误差的分布为(2.9)的情况，我们给出具体的介绍。

采用§2.1.1的记号，经过简单的计算可得回归系数β, β^*的最小二乘估计为:

$$\hat{\beta}_k = (\mathbf{X_{1k}^T X_{1k}})^{-1}\mathbf{X_{1k}^T Y_k}, \quad d \le k \le n.$$
$$\hat{\beta}_k^* = (\mathbf{X_{2k}^T X_{2k}})^{-1}\mathbf{X_{2k}^T Y_k^*}, \quad 1 \le k \le n - d.$$

从而当H_0成立时

$$\begin{aligned} Cov(\hat{\beta}_k, \hat{\beta}_k^*) &= E\left((\hat{\beta}_k - \beta)^T(\hat{\beta}_k^* - \beta)\right) \\ &= \sigma^2 H_k. \end{aligned}$$

其中

$$H_k = (\mathbf{X_{1k}^T X_{1k}})^{-1} + (\mathbf{X_{2k}^T X_{2k}})^{-1}.$$

当$\hat{\beta}_k - \hat{\beta}_k^*$比较大时，我们拒绝$H_0$，承认确实存在变点。从而对$\hat{\beta}_k - \hat{\beta}_k^*$实施标准化后，得到如下的Union-Intersection 检验统计量:

$$T_n^* = \max_{d \le k \le n-d} \frac{1}{\sigma^2}(\hat{\beta}_k - \hat{\beta}_k^*)^T H_k^{-1}(\hat{\beta}_k - \hat{\beta}_k^*) \tag{2.24}$$

由于σ^2未知，下面我们给出σ^2的三种估计：

$$\hat{\sigma}_k^2 = \frac{1}{n}\left\{\sum_{i=d}^{k}(y_i - x_i^T\hat{\beta}_k)^2 + \sum_{i=k+1}^{n-d}(y_i - x_i^T\hat{\beta}_k^*)^2\right\}.$$

$$\bar{\sigma}_n^2 = \frac{1}{n} \min_{d \le k \le -n-d} \left\{ \sum_{i=d}^{k} (y_i - x_i^T \hat{\beta}_k)^2 + \sum_{i=k+1}^{n-d} (y_i - x_i^T \hat{\beta}_k^*)^2 \right\}.$$

$$\tilde{\sigma}_n^2 = \min \left\{ \frac{2}{n} \sum_{i=d}^{n/2} (y_i - x_i^T \hat{\beta}_k)^2, \frac{2}{n} \sum_{i=n/2+1}^{n-d} (y_i - x_i^T \hat{\beta}_k^*)^2 \right\}.$$

将(2.24)中未知的σ^2分别用上面的三种估计代替，则得到下面三个统计量：

$$T_{n1} = \max_{d \le k \le n-d} \frac{1}{\hat{\hat{\sigma}}_k^2} (\hat{\beta}_k - \hat{\beta}_k^*)^T H_k^{-1} (\hat{\beta}_k - \hat{\beta}_k^*).$$

$$T_{n2} = \frac{1}{\bar{\sigma}_n^2} \max_{d \le k \le n-d} (\hat{\beta}_k - \hat{\beta}_k^*)^T H_k^{-1} (\hat{\beta}_k - \hat{\beta}_k^*).$$

$$T_{n3} = \frac{1}{\tilde{\sigma}_n^2} \max_{d \le k \le n-d} (\hat{\beta}_k - \hat{\beta}_k^*)^T H_k^{-1} (\hat{\beta}_k - \hat{\beta}_k^*).$$

由于在零假设下，$\hat{\hat{\sigma}}_k^2, d \le k \le n-d, \bar{\sigma}_n^2, \tilde{\sigma}_n^2$都是$\sigma^2$的相合估计，那么$T_{n1}, T_{n2}, T_{n3}$是不是依分布渐近等价呢？为此我们先给出依分布渐近等价的定义。

定义2.2.1. *设观测的样本为*$(x_1, \ldots, x_n)$，*统计量*$A_n = A_n(x_1, \ldots, x_n), B_n = B_n(x_1, \ldots, x_n), P_n(x) = P(A_n \le x), Q_n(x) = P(B_n \le x)$，*若*

$$\lim_{n \longrightarrow \infty} \sup |P_n(x) - Q_n(x)| = 0$$

则称A_n, B_n*依分布渐近等价.*

注: A_n, B_n以分布渐近等价充要条件为A_n, B_n有相同的渐近分布。

经计算可得似然比检验统计量与Union-Intersection 检验统计量有下面的关系：

$$2 \log \Lambda_k = \frac{n}{2} \log \left\{ 1 - \frac{\hat{\beta}_k - \hat{\beta}_k^*)^T H_k^{-1} (\hat{\beta}_k - \hat{\beta}_k^*}{\|Y_n - X_{1n} \hat{\beta}_n\|^2} \right\}. \tag{2.25}$$

根据(2.25)式和定理2.1.2,可以得到$T_n^*, T_{n1}, T_{n2}, T_{n3}$的渐近分布，见下一定理：

定理2.2.1. *若随机误差的分布为*(2.9)，$C.7 - C.10$*对某个*$\beta > 2 + 27/\delta$*成立，则在*H_0*成立时有*:

$$\lim_{n \longrightarrow \infty} P\{A(\log\ n)(T_n^*)^{1/2} \le t + D_d(\log\ n)\} = exp(-2e^{-t}), \forall\ t.$$

$$\lim_{n\longrightarrow\infty} P\{A(\log\ n)(T_{n1})^{1/2} \le t + D_d(\log\ n)\} = exp(-2e^{-t}), \forall\ t.$$
$$\lim_{n\longrightarrow\infty} P\{A(\log\ n)(T_{n2})^{1/2} \le t + D_d(\log\ n)\} = exp(-2e^{-t}), \forall\ t.$$
$$\lim_{n\longrightarrow\infty} P\{A(\log\ n)(T_{n3})^{1/2} \le t + D_d(\log\ n)\} = exp(-2e^{-t}), \forall\ t.$$

下面给出统计量在备择假设下的相合性。记k_0为变点k的真值。

定理2.2.2. 假设(2.13), (2.14), (2.15), (2.16),则在H_1成立时有

$$\frac{(\hat{\beta}_{k_0} - \hat{\beta}^*_{k_0})^T H_{k_0}^{-1} (\hat{\beta}_{k_0} - \hat{\beta}^*_{k_0})}{\frac{k_0(n-k_0)}{n}\delta^T A\delta} \xrightarrow{P} 1.$$

定理2.2.3. 假设(2.9).(2.10), (2.13), (2.17, (2.18), (2.19)和*C.7-C.10*成立，则当$n \to \infty$, H_1成立时，有

$$\frac{T_n^*}{\frac{k_0(n-k_0)}{n}\frac{1}{\sigma^2}\delta^T A\delta} \xrightarrow{P} 1.$$

$$\frac{T_{ni}}{\frac{k_0(n-k_0)}{n}\frac{1}{\sigma^2}\delta^T A\delta} \xrightarrow{P} 1, i = 1, 2, 3.$$

§2.3 基于复发回归残差的检验方法

复发回归残差(recursive regression residuals)是由Brown, Durbin and Evans (1975)提出来的，Brown, Durbin and Evans (1975)基于复发回归残差对回归模型中的变点问题给出了两种检验方法，下面逐一介绍。

模型为

$$y_i = \begin{cases} x_{i1}^T\beta + \epsilon_i, & 1 \le i \le k_0; \\ x_{i1}^T\beta^* + \epsilon_i, & k_0 < i \le n. \end{cases} \tag{2.26}$$

其中$x_i \in d, 1 \le i \le n$是已知的列向量，回归系数β和β^*是未知的列向量。我们对模型中是否存在变点进行检验，即：

$$H_0:\ k_0 \ge n \leftrightarrow 1 < k_0 < n.$$

记

$$\mathbf{X_r} = \begin{pmatrix} x_1^T \\ \vdots \\ x_r^T \end{pmatrix}, \qquad \mathbf{Y_r} = \begin{pmatrix} y_1 \\ \vdots \\ y_r \end{pmatrix}.$$

基于前r个观测值得到的β的最小二乘估计为

$$\hat{\beta}_r = (\mathbf{X_r}^T\mathbf{X_r})^{-1}\mathbf{X_r}^T\mathbf{Y_r}.$$

则复发回归残差定义为：

$$W_r = \frac{y_r - x_r^T\hat{\beta}_{r-1}}{(1 + x_r^T(\mathbf{X_{r-1}}^T\mathbf{X_{r-1}})^{-1}x_r)^{1/2}},\ r = d+1,\ldots,n,\ d \geq 1.$$

为了计算上的方便，我们先给出以下引理：

引理2.3.1. *(i)* $(\mathbf{X_r}^T\mathbf{X_r})^{-1} = (\mathbf{X_{r-1}}^T\mathbf{X_{r-1}})^{-1} - \frac{(\mathbf{X_{r-1}}^T\mathbf{X_{r-1}})^{-1}x_r x_r^T(\mathbf{X_{r-1}}^T\mathbf{X_{r-1}})^{-1}}{1+x_r^T(\mathbf{X_{r-1}}^T\mathbf{X_{r-1}})^{-1}x_r}$.
(ii) $\hat{\beta}_r = \hat{\beta}_{r-1} + (\mathbf{X_r}^T\mathbf{X_r})^{-1}x_r(y_r - x_r^T\hat{\beta}_{r-1})$.
(iii) $\|\mathbf{Y_r} - \mathbf{X_r}\hat{\beta}_r\|^2 = \|\mathbf{Y_{r-1}} - \mathbf{X_{r-1}}\hat{\beta}_{r-1}\|^2 + W_r^2$.

下面的引理说明复发回归残差的性质与随机误差的性质十分相似。

引理2.3.2. 如果随机误差$\epsilon_1,\cdots,\epsilon_n,\overset{i.i.d}{\sim} N(0,\sigma^2)$，则在$H_0$成立时，复发回归误差$W_{d+1},\cdots,W_n \overset{i.i.d}{\sim} N(0,\sigma^2)$.

引理2.3.3. 如果随机误差$\epsilon_1,\cdots,\epsilon_n$不相关且$E(\epsilon_i) = 0, E(\epsilon_i^2) = \sigma^2$, 则在$H_0$成立时，复发回归误差$W_{d+1},\cdots,W_n$也不相关且$E(W_r) = 0, E(W_r^2) = \sigma^2, r = d+1,\cdots,n$.

注:从ϵ_i到W_i的变换实际上是一种广义的Helmert变换(见Kendell and Stuart (1969)).

1. 复发残差和检验(the cusum test)

定义

$$S^*(r) = \sum_{d<j<r} W_j.$$

对$S^*(r)$施以标准化得到检验统计量

$$T_n^1 = \frac{1}{\sigma}\max_{d<r\leq n}\left\{(r-d)^{-1/2}|S^*(r)|\right\}.$$

则当T_n^1比较大时，拒绝H_0。

定理2.3.1. 如果随机误差$\epsilon_1,\ldots,\epsilon_n,\overset{i.i.d}{\sim} N(0,\sigma^2)$, σ已知, 则在H_0成立时, 有

$$\lim_{n\longrightarrow\infty} P\{A(\log n)T_n^1 \leq t + D(\log n)\} = \exp(-2e^{-t}),\ \forall\ t.$$

其中$A(x), D(x)$定义见§2.1.

定理2.3.2. 当引理2.3.2中的条件满足时，若σ未知，则以$\hat{\sigma}_n^2 = \frac{1}{n-d}\lim_{d < i \leq n}(y_i - x_i^T\hat{\beta}_n)^2$代替$T_n^1$中的$\sigma^2$时，定理2.3.1中的结论仍然成立。

2. 复发残差平方和检验（the cusum of squares test）

为了避免对σ^2的估计，我们引入统计量：

$$S_r = \frac{\sum\limits_{d < j \leq r} W_j^2}{\sum\limits_{d < i \leq n} W_i^2}, \quad d < r \leq n.$$

定理2.3.3. 如果随机误差$\epsilon_1, \ldots, \epsilon_n \overset{i.i.d}{\sim} N(0, \sigma^2)$，且$n-d$为偶数，则在$H_0$成立时，有

$$\{S_{d+2}, S_{d+4}, \ldots, S_{n-2}\} \overset{d}{=} \{U_{1,(n-d)/2-1}, U_{2,(n-d)/2-1}, \ldots, U_{(n-d)/2-1,(n-d)/2-1}\}.$$

其中$U_{1,(n-d)/2-1} \leq U_{2,(n-d)/2-1} \leq \ldots, \leq U_{(n-d)/2-1,(n-d)/2-1}$是来自$U(0,1)$的$(n-d)/2-1$个独立顺序样本。

§2.4 基于回归残差的检验

本小节的模型为§2.3中的模型(2.26)，记

$$\mathbf{X_n} = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix}, \qquad \mathbf{Y_n} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

则在H_0成立时，基于$\{(x_i, y_i), i = 1, \ldots, n\}$得到回归系数的最小二乘估计为：

$$\hat{\beta}_n = (\mathbf{X_n}^T\mathbf{X_n})^{-1}\mathbf{X_n}^T\mathbf{Y_n}.$$

记

$$\bar{y}_n = \frac{\sum\limits_{i=1}^n y_i}{n}, \bar{x}_n = \frac{\sum\limits_{i=1}^n x_i}{n}, \bar{y}_k = \frac{\sum\limits_{i=1}^k y_i}{k}, \bar{x}_k = \frac{\sum\limits_{i=1}^k x_i}{k}.$$

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x}_n)^T(x_i - \bar{x}_n).$$

$$
\begin{aligned}
G_k &= \sum_{i=1}^{k}(y_i - \bar{y}_n - (x_i - \bar{x}_n)^T\hat{\beta}_n) \\
&= \{\sum_{i=1}^{k}(x_i - \bar{x}_n)\}^T(\beta - \hat{\beta}_n) + S^{**}(k) - \tfrac{k}{n}S^{**}(n).
\end{aligned}
$$

其中$S^{**}(k) = \sum_{i=1}^{k}\epsilon_i$.

$$R_n(k) = \Big(\frac{n}{k(n-k)}\Big)^{1/2}G_k,\ 1 \le\ k\ < n.$$

$$U_n(k) = (\frac{k}{1-k/n})^{1/2}\frac{\bar{y}_k - \bar{y}_n - (\bar{x}_k - \bar{x}_n)^T\hat{\beta}_n}{(1 - k(\bar{x}_k - \bar{x}_n)^T(\bar{x}_k - \bar{x}_n)/(S_{xx}(1-k/n)))^{1/2}}.$$

Brown, Durbin and Evans (1975)提出用$\max\limits_{1\le k\le n}|R_n(k)|$作为检验统计量，则当$\max\limits_{1\le k\le n}|R_n(k)|$显著大时，拒绝$H_0$。James, James and Siegmund (1987)建议采用$\max\limits_{1\le k\le n}|U_n(k)|$作为检验统计量，于是当$\max\limits_{1\le k\le n}|U_n(k)|$大时，拒绝$H_0$。Gombay and Horváth (1994)给出了更广的一类检验统计量$Z_n(i,j) = \max\limits_{i\le k\le j}|U_n(k)|$和$T_n(i,j) = \max\limits_{i\le k\le j}|R_n(k)|$，$i,j$任意。

下面接着讨论这些检验统计量的关系及统计性质。

记

$$w_n(k) = \{1 - k(\bar{x}_k - \bar{x}_n)^T(\bar{x}_k - \bar{x}_n)/(S_{xx}(1-\frac{k}{n}))\}^{-1/2}.$$

则

$$U_n(k) = w_n(k)R_n(k),\ 1 \le\ k\ < n.$$

从而$\max\limits_{1\le k\le n}|R_n(k)|$与$\max\limits_{1\le k\le n}|U_n(k)|$依分布渐近等价。注意到$\max\limits_{nt_1\le k\le nt_2}|R_n(k)|$与$\max\limits_{nt_1\le k\le nt_2}|U_n(k)|$, $0 < t_1 < t_2 < 1$当n较大时,统计性质也许会不同。

假设

(i)

$$\epsilon_1, \epsilon_2, \ldots, \epsilon_n, i.i.d, E(\epsilon_i) = 0, 0 < \sigma^2 = Var(\epsilon_i), E\,|\epsilon_i|^p < \infty, p > 2. \tag{2.27}$$

(ii) 存在函数$f^T(t) = (f_1(t), f_2(t), \ldots, f_d(t))$，满足

$$\max_{1\le j\le d}\sup_{0\le t\le 1}|f_j'(t)| < \infty. \tag{2.28}$$

其中f_j'是f_j的导函数，满足

$$x_{j,i} = f_j(i/n), 1 \le j \le d, 1 \le i \le n. \tag{2.29}$$

(2.29)式说明设计矩阵是非随机的。

(iii) 令

$$d_{i,j} = \int_0^1 f_i(t)f_j(t)dt, \quad A = \{d_{i,j}, 1 \le i,j \le d\}.$$

满足

$$rank(A) = d \tag{2.30}$$

(iv) 令$h_(j)(t) = \int_0^t f_j(u)du - t\int_0^1 f_j(u)du$,

$$h^T(t) = (h_1(t), \ldots, h_d(t)), \quad u^T(t) = (\int_0^t f_1(u)du, \ldots, \int_0^t f_d(u)du).$$

$$Q = \sum_{j=2}^d \left\{ \int_0^1 f_j^2(s)ds - \left(\int_0^1 f_j(s)ds \right)^2 \right\}, \; g(t) = \frac{1}{Q(1-t)} \sum_{j=2}^d \left(th_(j)(t) \right)^2$$

$$\sup_{0 \le t \le 1} g(t) < 1. \tag{2.31}$$

引理2.4.1. 假设(2.27), (2.28), (2.29), (2.30)成立，则在H_0成立时，有

$$\hat{\beta}_n - \beta = \frac{1}{n} A^{-1} X_n{}^T \epsilon_n + 0_p(\frac{1}{n^{3/2}})$$

其中$\epsilon_n{}^T = (\epsilon_1, \ldots, \epsilon_n)$.

引理2.4.2. 若(2.28), (2.29)成立，则

$$\sup_{0 \le t \le 1} \left\| \frac{\sum_{i=1}^{[nt]} (x_i - \bar{x_n})}{n} - h(t) \right\| = O(\frac{1}{n}).$$

令$G_n(t) = \frac{1}{\sigma n^{1/2}} G((n+1)t), 0 \le t < 1, G_n(1) = 0$，则有

定理2.4.1. 假设(2.27), (2.28), (2.29), (2.30)成立,则在H_0成立时，有

$$G_n(t) \xrightarrow{d[0,1]} \upsilon(t)$$

其中$\{\upsilon(t), 0 \le t \le 1\}$是一$Gauss$过程，满足$E(\upsilon(t)) = 0$，$E(\upsilon(t)\upsilon(s)) = h^T(t)A^{-1}h(s) + t(1-s) - h^T(s)A^{-1}(u(t) - tu(1)) - h^T(t)A^{-1}(u(s) - su(1)), 0 \le t \le s \le 1$

令

$$I_{0,1}(q,c) = \int_0^1 \frac{1}{t(1-t)} exp(-\frac{cq^2(t)}{t(1-t)})dt$$

定理2.4.2. 假设(2.27), (2.28), (2.29), (2.30)成立,$q(t)$是定义在$(0,1)$上的非负函数，且在0的邻域内非降，在1的邻域内非增，则在H_0成立时，有
(i) 存在*Gaussian*过程$\{\upsilon_n(t), 0 \le t \le 1\}$，使得

$$\{\upsilon_n(t), 0 \le t \le 1\} \stackrel{d}{=} \{\upsilon(t), 0 \le t \le 1\}, \forall\, n.$$

且

$$\sup_{0\le t\le 1} |G_n(t) - \upsilon_n(t)|/q(t) = o_P(1).$$

当且仅当$I_{0,1}(q,c) < \infty, \forall\, c > 0$.
(ii)

$$\sup_{0\le t\le 1} |G_n(t)|/q(t) \xrightarrow{d} \sup_{0\le t\le 1} |\upsilon(t)|/q(t).$$

当且仅当$I_{0,1}(q,c) < \infty, \forall\, c > 0$.

注: 取$q(t) = (t(1-t))^{1/2}$，则由$I_{0,1}(q,c)$的定义可知$I_{0,1}(q,c) = \infty$，所以根据(2.4.2)中的(ii)知，$\max\limits_{1\le k<n} |R_n(k)|$不可能以分布收敛到Gauss 过程的最大值，事实上，根据Brown,Durbin and Evans(1975)可知

$$\max_{1\le k<n} |R_n(k)| \xrightarrow{P} \infty.$$

下面给出$\max\limits_{1\le k<n} |R_n(k)|$标准化后的渐近分布。

定理2.4.3. 若(2.27), (2.28), (2.29), (2.30)成立，则在H_0成立时，有

$$\lim_{n\to\infty} P\left\{A(\log n)\frac{1}{\sigma}\max_{1\le k<n} |R_n(k)| \le t + D(\log n)\right\} = \exp(-2e^{-t}),\ \forall\, t.$$

$$\lim_{n\to\infty} P\left\{A(\log n)\frac{1}{\sigma}\max_{1\le k<n} |U_n(k)| \le t + D(\log n)\right\} = \exp(-2e^{-t}),\ \forall\, t.$$

更进一步若$\lambda_1(n) \to 0, \lambda_2(n) \to 0, m_1 = n\lambda_1(n) \to \infty, m_2 = n\lambda_2(n) \to \infty$,且(2.31)成立，则有

$$\lim_{n\to\infty} P\left\{A(\log m_1)\frac{1}{\sigma}Z_n(1,m_1) \le t + D(\frac{1}{2}\log m_1)\right\} = \exp(-2e^{-t}),\ \forall\, t.$$

$$\lim_{n\to\infty} P\left\{A(\log m_1)\frac{1}{\sigma}T_n(1,m_1) \le t + D(\frac{1}{2}\log m_1)\right\} = \exp(-2e^{-t}),\ \forall\, t.$$

$$\lim_{n\to\infty} P\left\{A(\log m_2)\frac{1}{\sigma}Z_n(n-m_2,n) \le t + D(\frac{1}{2}\log m_2)\right\} = \exp(-2e^{-t}),\ \forall\, t.$$

$$\lim_{n\to\infty} P\left\{A(\log m_2)\frac{1}{\sigma}T_n(n-m_2,n) \le t + D(\frac{1}{2}\log m_2)\right\} = \exp(-2e^{-t}),\ \forall\, t.$$

定理2.4.4. 若(2.27), (2.28), (2.29), (2.30), (2.31)成立，$0 < t_1 < t_2 < 1$，则在H_0成立时，有

$$
\begin{aligned}
& \lim_{n\to\infty} P\left\{ A(\log n)\frac{1}{\sigma}Z_n(1,n) \le t + D(\log n),\ \frac{1}{\sigma}Z_n(nt_1, nt_2) \le y \right\} \\
= & \exp(-2e^{-t}) \lim_{n\to\infty} P\left\{ \frac{1}{\sigma}Z_n(nt_1, nt_2) \le y \right\}. \\
& \lim_{n\to\infty} P\left\{ A(\log m_1)\frac{1}{\sigma}T_n(1,n) \le t + D(\log n),\ \frac{1}{\sigma}T_n(nt_1, nt_2) \le y \right\} \\
= & \exp(-2e^{-t}) \lim_{n\to\infty} P\left\{ \frac{1}{\sigma}T_n(nt_1, nt_2) \le y \right\}.
\end{aligned}
$$

由定理2.4.4知$Z_n(nt_1, nt_2)$与$Z_n(1,n)$是渐近独立的，$T_n(nt_1, nt_2)$与$T_n(1,n)$也有一样的结论。

§2.5 信息标准的方法

我们判断模型是否存在变点的过程，也可以看成是对模型进行选择的过程，Chen and Gupta (1997)和Chen, Gupta and Pan (2006)利用模型选择的方法对分布函数中的变点问题进行了统计推断。对于回归中变点问题，我们同样可以用信息标准的方法进行统计推断。

1. AIC和SIC

K-L信息是模型选择的基础，下面先给出它的定义:

定义2.5.1.【*K-L信息(距离)*】
假设模型F和G,是两个定义在(Ω, Λ)概率测度,且F, G关于σ-有限测度μ有密度函数。记$f(x) = \frac{dF(x)}{d\mu}, g(x) = \frac{dG(x)}{d\mu}$，则称

$$
\begin{aligned}
I(f,g) &= E_f log\ \tfrac{f(x)}{g(x)} \\
&= \int f(x) log\ \tfrac{f(x)}{g(x)} d\mu(x) \\
&= \int f(x) log(f(x)) d\mu(x) - \int f(x) log(g(x)) d\mu(x)
\end{aligned}
$$

为*K-L*信息(距离)。

假设f是真模型，模型选择的目标就是从一系列候选模型中选出与真模型最接近的模型。根据信息理论，K-L信息$I(f,g)$表示用模型g逼近模型f时的损失的信息，从模型选择的角度，损失的信息越小越好，也

就是说$I(f,g)$越小，模型g越接近真模型f。所以，在寻找真模型f的一个逼近时，应该寻找使得$I(f,g)$达到最小的g。

注:

(1)从直观上看，$I(f,g)$是g到f的距离，所以有时也称$I(f,g)$为K-L距离，但由于$I(f,g)\neq I(g,f)$，所以$I(f,g)$并非是严格意义上的距离。

(2)$I(f,g)\geq 0$,$I(f,g)=0$当且仅当$f=g$, $a.e.$

下面假设g 属于正则参数族分布$G=\{g_\theta,\theta\in\Theta\}$，所以

$$\begin{aligned} I(f,g) &= \int f(x)\log\tfrac{f(x)}{g(x|\theta)}d\mu(x) \\ &= \int f(x)\log(f(x))d\mu(x)-\int f(x)\log(g(x|(\theta)))d\mu(x). \end{aligned}$$

由于θ和f均未知，我们无法直接计算$I(f,g)$，但是我们可以从给定的样本数据出发给出它的估计值。上式右边第一项是常数，对所有的候选模型都一样，所以只需给出第二项的估计即可。

Akaike (1973)将极大似然方法和K-L距离相结合，并利用似然估计的性质，推导出了最佳模型选择的AIC准则。

设y来自真模型$f(x)$的样本，$\hat{\theta}_y$为θ的最大似然估计。Akaike指出应用K-L距离进行模型选择，也就是在G中寻找使得$E_yI(f,g(x|\hat{\theta}(y)))$,达到最小的g。又由于

$$E_yI\Big(f,g(x|\hat{\theta}(y))\Big)=\int f(x)\log(f(x))d\mu(x)-E_yE_x[\log(g(x|\hat{\theta_y}))]$$

所以等价于在G中寻找使得$E_yE_x[log(g(x|\hat{\theta_y}))]$达到最大的g。

由于f未知，所以寻找使得$E_yE_x[log(g(x|\hat{\theta_y}))]$的渐近无偏估计$\log(g(x|\hat{\theta}))-K$,$K=dim(\theta)$达到最大的g即可，由此得到了AIC(Akaike's Information Criterion)。

定义2.5.2.【*AIC(Akaike's Information Criterion)*】

$$AIC=-2\log(g(x|\hat{\theta}))+2K \tag{2.32}$$

其中K表示模型中参数的个数。

AIC 使得K-L距离$I(f,g)$和最大对数似然函数之间有了联系，为模型选择提供了一种更简单、更有效的一种方法。当从一组候选模型中选

取模型时，使得AIC达到最小的模型是可取的；如果两个模型有很大的差异，这个差异应该体现在(2.32)式右边第一项，而当第一项不出项显著差异时，第二项就起作用，应该选参数个数较少的模型，从而第二项可以解释为对增加模型参数的一种惩罚。

Schwarz (1978)指出AIC准则选出的模型并非是渐近相合的，当$f(x)$属于指数族分布时，Schwarz (1978)从贝叶斯的观点提出了SIC(Schwarz Information Criterion)或称为(BIC，Bayesian Information Criterion)模型选择标准，$SIC = 2\log(g(x|\hat{\theta})) + K\log n$，其中n为样本的个数。SIC相对于AIC而言，加大了对模型参数的个数的惩罚。

2. 利用SIC研究分布函数中的变点问题

设随机变量$X_1, X_2, \ldots, X_n$互相独立，其中$X_1, \ldots, X_k, \sim\ f(x, \theta_1)$，$X_{(k+1)}$，$\ldots, X_n, \sim\ f(x, \theta_2)$，且$f(x, \theta_1), f(x, \theta_2)$ 都属于参数族分布$\{f(x, \theta) : \theta \in R^d\}$。有关变点问题的提法是：检验是否存在变点，如果存在，给出变点k的位置。从假设检验的角度问题化为检验

$$H_0 : \theta_1 = \theta_2 = \theta\ \leftrightarrow\ H_1 : \theta_1 \neq \theta_2, 1 \leq k\ < n.$$

从模型选择的角度就是从H_0和H_1中选出一个模型。
考虑似然函数

$$l_n(\theta_1, \theta_2, k) = \sum_{i=1}^{k} \log f(X_i, \theta_1) + \sum_{i=k+1}^{n} \log f(X_i, \theta_2).$$

对固定的k，记$\hat{\theta}_{1k}, \hat{\theta}_{2k}$为$\theta_1, \theta_2$的最大似然估计，若直接将SIC和AIC用于变点模型，则得到

$$SIC(k) = -2l_n(\hat{\theta}_{1k}, \hat{\theta}_{2k}, k) + (2dim(\hat{\theta_{1k}}) + 1)\log n$$

$$AIC(k) = -2l_n(\hat{\theta}_{1k}, \hat{\theta}_{2k}, k) + 2(2dim(\hat{\theta_{1k}}) + 1)$$

当H_0成立时

$$SIC(n) = -2l_n(\hat{\theta}, n) + dim(\hat{\theta})\log n$$

其中$\hat{\theta}$为在H_0成立时θ的最大似然估计。

当$SIC(n) < \min\limits_{2\leq k\leq n-2} SIC(k)$时接受零假设；当存在$2 \leq k \leq n-2$，使得$SIC(n) > SIC(k)$时，接受备择假设，此时变点k的估计$\hat{k}$为

$$\hat{k} = \arg\min_{2\leq k\leq n-2} SIC(k) \tag{2.33}$$

Chen and Gupta (1997)给出了有关统计量的渐近性质。

定理2.5.1. *若k_0为变点的真值，$\hat{k}$由(2.33)式给出，则在备择假设H_1成立时，有*

$$\hat{k} \longrightarrow k_0 \qquad a.e$$

令

$$\Delta_n = \min_{2\leq k\leq n-2}\Big(SIC(k) - SIC(n)\Big)$$

则当Δ_n较小时，拒绝零假设。

定理2.5.2. *在零假设H_0成立时*

$$\lim_{n\longrightarrow\infty} P[a(\log n)(2\log n - \Delta_n)^{1/2}) - b(\log n) \leq x] = \exp(-2e^{-x}).$$

其中$a(\log n) = (2\log\log n)^{\frac{1}{2}}, b(\log n) = 2\log\log n + \log\log\log n$

3. 利用MIC研究分布函数中的变点问题

实际上，在正则参数模型(满足Wald (1949)中的条件)中，模型中参数的个数可以看成是对模型复杂程度的一种度量，模型中含的参数越多，模型当然越复杂，从而AIC和SIC中对模型中参数个数的惩罚在非正则参数模型中可以推广为对模型复杂程度的惩罚。

因为变点模型并非正则模型，并且当k接近1或n时，θ_1或θ_2变成多余的参数了，所以Chen, Gupta and Pan (2006) 给出了一种修改后的信息准则。由于当变点为k时，

$$Var\hat{\theta}_{2k} \propto \frac{1}{k}, Var\hat{\theta}_{2k} \propto \frac{1}{n-k}.$$

这时总方差与

$$\frac{1}{k} + \frac{1}{n-k} = \frac{1}{n}[\frac{1}{4} - (\frac{k}{n} - \frac{1}{2})^2]^{-1}.$$

称正比。所以Chen, Gupta and Pan (2006)选择了采用

$$2dim(\hat{\theta}_{1k}) + (\frac{2k}{n} - 1)^2 + C.$$

来度量模型的复杂度。修改后的信息准则为
当H_1成立时

$$MIC(k) = -2l_n(\hat{\theta}_{1k}, \hat{\theta}_{2k}, k) + [2dim(\hat{\theta}_{1k}) + (\frac{2k}{n} - 1)^2]\log n,$$

当H_0成立时

$$MIC(n) = -2l_n(\hat{\theta}, n) + dim(\hat{\theta}) \log n$$

其中$\hat{\theta}$为在H_0成立时θ的最大似然估计。

当$MIC(n) > \min_{1 \le k < n} MIC(k)$时，拒绝$H_0$，认为模型中确实存在变点。此时给出变点的一个估计为：

$$\hat{k} = \{k_0 : MIC(k_0) = \min_{1 \le k < n} MIC(k)\}$$

令

$$S_n = MIC(n) - \min_{1 \le k < n} MIC(k) + dim(\hat{\theta}) \log n$$

上式加上$dim(\hat{\theta}) \log n$是为了消除$MIC(n)$和$\min_{1 \le k < n} MIC(k)$中相差的常数项。

定理2.5.3. 当参数族分布$\{f(x, \theta) : \theta \in R^d\}$满足*Chen, Gupta and Pan (2006)*中的条件$W_1 - W_7$和正则条件$R_1 - R_3$时，则
*(i)*在零假设成立时有

$$S_n \xrightarrow{d} \chi_d^2, n \to \infty.$$

其中$d = dim(\theta)$.
*(ii)*如果存在k_0满足$k_0/n \longrightarrow c \in (0, 1)$, 则当$n \to \infty$时，有

$$S_n \xrightarrow{P} \infty.$$

关于变点k_0的估计$\hat{k}$的收敛速度和极限分布有下面的定理。

定理2.5.4. 当参数族分布$\{f(x, \theta) : \theta \in R^d\}$满足*Chen, Gupta and Pan (2006)*中的条件$W_1 - W_7$和正则条件$R_1 - R_3$时，当$n \longrightarrow \infty$时，$k_0 = [n\lambda], 0 < \lambda < 1$,则有

$$\hat{k} - k_0 = O_p(1).$$

设$\{Y_i, i = -1, -2, -3, \dots\} \overset{i.i.d}{\sim} f(x, \theta_{10})$，$\{Y_i, i = 1, 2, 3, \dots\} \overset{i.i.d}{\sim} f(x, \theta_{20})$且$\{Y_i, i = -1, -2, -3, \dots\}$与$\{Y_i, i = 1, 2, 3, \dots\}$相互独立，$Y_0$满足$f(Y_0, \theta_{10}) = f(Y_0, \theta_{20})$。令

$$W_k = \sum_{j=0}^{k} sgn(k)[\log f(Y_j, \theta_{20}) - \log f(Y_j, \theta_{10})], k = 0, \pm 1. \pm 2, \cdots$$

定理2.5.5. 若参数族分布$\{f(x,\theta):\theta\in R^d\}$满足*Chen, Gupta and Pan (2006)*中的条件W_1-W_7和正则条件R_1-R_3，存在k_0满足$k_0/n\to c\in(0,1)$ 则当$n\to\infty$时，有

$$\hat{k}-k_0\xrightarrow{d}\zeta.$$

其中θ_{10},θ_{20}是θ_1,θ_2在备择假设下的真值，$\zeta=\underset{-\infty<k<\infty}{\arg}\min\{W_k\}$。

4. 模拟比较SIC和MIC

下面对有限样本通过模拟比较MIC和SIC的功效，模拟过程如下：

- 对正态分布均值发生变化；正态分布方差发生变化；指数分布均值发生变化三种情况进行模拟。
- 样本量n=100和n=200。
- 变点$k_0=0.25n,0.5n,0.75n$。
- SIC和MIC对应的行分别表示拒绝零假设的频率(the null rejection rates)。
- 由于在零假设成立时，SIC和MIC拒绝零假设的频率不同，这样直接比较功效是不准确的，所以SIC^*对应的行表示将SIC和MIC拒绝零假设的频率调整一样后，SIC的功效。

从表2.1可以看出：当样本量增加时，犯第一类错误的概率减少而功效增加了；在第一类错误的概率相同的条件下，MIC的功效比SIC的功效大；当变点位于序列中点时，功效最大。

对$\hat{k}-k$的极限分布的模拟安排如下：

- 对正态分布均值发生变化，正态分布方差发生变化，指数分布均值发生变化四种情况进行模拟，正态分布均值和方差均发生变化。
- 样本量n=100和n=1000。
- 变点$k_0=0.5n$。

从表2.2看出当样本量为100时，$P(|\hat{k}_{MIC}-k_0|\le\delta)\ge P(|\hat{k}_{SIC}-k_0|\le\delta)$，随着n的增大，它们之间的差距缩小，当$n=1000$时，二者几乎相同，并且此时$\zeta$是$\hat{k}-k_0$很好的一个逼近。

表 2.1 SIC和MIC功效的比较

	n=100				n=200			
	normal model: change in the mean (c=0.5), variance=1							
k	0	25	50	75	0	50	100	150
MIC	19.98	65.5	80.32	65.8	15.9	84.1	95.88	84.14
SIC	5.44	36.56	49.5	36.76	3.62	60.02	76.56	58.74
SIC^*		62.32	72.16	63.02		81	90.3	79.8
	normal model: change in the variance(c=2), mean =0							
k	0	25	50	75	0	50	100	150
MIC	20.48	64.02	79.72	62.68	16.46	83.24	94.06	83.04
SIC	6.52	31.92	47.44	39.08	4.16	52.76	72.5	61.54
SIC^*		61.16	71.28	59.6		82.44	91.58	81.26
	exponential model: changes in the mean ($c=\sqrt{2}$)							
k	0	25	50	75	0	50	100	150
MIC	20.7	45.02	56.26	43.38	16.48	59.28	74.36	60.08
SIC	6.42	19.12	24.28	20.42	4.26	26.86	56.9	

表 2.2 $\hat{k}-k$与ζ 分布的比较

$P_\zeta = P(\|\zeta\| \le \delta), P_M = P(\|\hat{k_{MIC}} - k_0\| \le \delta), P_S = P(\|\hat{k_{SIC}} - k_0\| \le \delta)$						
Models	Prob	$\delta = 5$	$\delta = 10$	$\delta = 20$	$\delta = 30$	$\delta = 40$
	n=100					
Model 1	P_ζ	0.4622	0.6614	0.8394	0.9242	0.9592
	P_M	0.5136	0.6886	0.8668	0.9392	0.9774
	P_S	0.3970	0.5468	0.7078	0.7904	0.8818
Model 2	P_ζ	0.4590	0.6280	0.8218	0.9034	0.9620
	P_M	0.4824	0.6596	0.8484	0.9178	0.9608
	P_S	0.3730	0.4994	0.6652	0.7512	0.8476
Model 3	P_ζ	0.3324	0.4876	0.6858	0.8142	0.8982
	P_M	0.3840	0.5710	0.7554	0.8798	0.9372
	P_S	0.2302	0.3386	0.6858	0.6130	0.7308
Model 4	P_ζ	0.5884	0.7686	0.9124	0.9622	0.9874
	P_M	0.5242	0.6942	0.8424	0.8978	0.9312
	P_S	0.4368	0.5696	0.7058	0.7754	0.8338
	n=1000					
Model 1	P_ζ	0.4664	0.6456	0.8150	0.8942	0.9406
	P_M	0.4638	0.6418	0.8012	0.8898	0.9374
	P_S	0.4746	0.6390	0.8072	0.8850	0.9230
Model 2	P_ζ	0.4462	0.6246	0.8004	0.8822	0.9236
	P_M	0.4378	0.6134	0.7912	0.8738	0.9218
	P_S	0.4328	0.6150	0.7870	0.8704	0.9242
Model 3	P_ζ	0.2888	0.4578	0.6138	0.7300	0.8068
	P_M	0.2998	0.4456	0.6214	0.7260	0.7912
	P_S	0.2906	0.4298	0.6082	0.6964	0.7762
Model 4	P_ζ	0.5880	0.7580	0.9084	0.9560	0.9790
	P_M	0.5724	0.7654	0.9050	0.9530	0.9770
	P_S	0.5896	0.7502	0.9044	0.9532	0.9752

第三章　经验似然

Thomas and Grunkemerier (1975)就利用经验似然的思想对随机删失数据下的生存概率进行了区间估计，但是直到Owen (1988, 1990)才系统地研究了经验似然，利用经验似然方法给出了独立同分布且完全样本情形下的未知参数的非参数统计推断。

经验似然方法属于非参数统计推断。它既具有非参数方法的稳键性又融合了似然方法的有效性和灵活性，所以相比其它经典的统计方法，有许多突出的优点。比如当样本来自非正态分布或方差估计不稳定时，经验似然方法较之经典的正态逼近方法要精确的多。它具有bootstrap的抽样特性，但是所构造的置信域由样本自行决定，具有Bartlett纠偏性，域保持性和变换不变性等等。正是由于具有这些突出的优点，自Owen (1988, 1990)以来，经验似然方法备受统计学家的青睐，并且在统计的许多领域都得到了广泛应用。经验似然推断不仅可以用于完全样本，而且对有偏抽样和不完全数据，比如随机缺失，删失和有测量误差的数据都可以应用。本章对经验似然方法做简单介绍。

§3.1　经验似然比

定义3.1.1. 设$X_1,\ldots,X_n\in R^d$, $d\le 1$,则$X_1,\ldots,X_n$的经验分布函数(EDF)为

$$F_n=\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}.$$

其中$\delta_x(A)=I_{\{x\in A\}}$

特别地，若$X_1,\ldots,X_n\in R$, 则$X_1,\ldots,X_n$的累积经验分布函数(ECDF)$F_n(x)=\frac{1}{n}\sum\limits_{i=1}^{n}I_{\{X_i\le x\}}$, $-\infty<x<\infty$.

若$X\in R^d\sim F$，则$F(A)$表示$P(X\in A), A\subseteq R^d$。

定义3.1.2. 设$X_1,\ldots,X_n\in R^d$, $d\le 1$，相互独立且具有共同的分布函数F，那么分布函数F的非参数似然定义为

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

特别地，若$X_1,\ldots,X_n \in R$，则$L(F)=\prod_{i=1}^{n}(F(X_i)-F(X_i-))$.

注:非参数似然$L(F)$表示从分布函数F中得到观测样本$X_1,\ldots,X_n$的概率，如果F是连续型分布，则$L(F)=0$。

定理3.1.1. $X_1,\ldots,X_n \in R^d$, $d \leq 1$，相互独立且具有共同的分布函数F_0，F_n是其经验分布函数，F是任一分布函数，若$F \neq F_n$，则

$$L(F) < L(F_n).$$

定理3.1.1表明经验分布函数F_n是F的非参数极大似然估计，从而类似于参数似然比，我们也可以定义非参数似然比

$$R(F)=\frac{L(F)}{L(F_n)}.$$

在参数模型中，如果我们对参数θ感兴趣，并且$\theta=\theta(\eta)$，若η的最大似然估计为$\hat{\eta}$，则θ的最大似然估计$\hat{\theta}=\theta(\hat{\eta})$。在参数推断中，我们往往以似然比为基础进行假设检验和区间估计，如果$L(\theta_0)<L(\hat{\theta})$，则拒绝假设$\theta=\theta_0$，取$\theta$的置信域为$\{\theta|L(\theta)\geq cL(\hat{\theta})\}$。根据Wilk定理知在一定的正则条件下

$$-2\log\left(\frac{L(\theta_0)}{L(\hat{\theta})}\right)\xrightarrow{d}\chi^2.$$

所以可以利用Wilk定理选择阀值c。

类似参数推断，我们在进行非参数推断时，也利用非参数似然比进行假设检验和构造置信域。一般来说我们感兴趣的参数如均值、方差、分位数等等都可以表示成分布函数的统计泛函，即$\theta=T(F), F\in \mathcal{F}$，其中 θ 是我们感兴趣的参数，$T(\cdot)$是某个已知的泛函。由于F_n是F的非参数极大似然估计，所以取$\hat{\theta}=T(F_n)$为θ的非参数极大似然估计(NPMLE)。

在参数似然比推断中，我们会把冗余参数清除掉，所以在非参数推断中，我们类似地定义轮廓似然比函数(profile likelihood ratio function)为

$$\mathcal{R}(\theta)=\sup_F\{R(F)\mid T(F)=\theta, F\in\mathcal{F}\}$$

若$\mathcal{R}(\theta_0)<r_0$，则拒绝假设$T(F_0)=\theta_0$，$\theta$的经验似然置信域为$\{\theta|\mathcal{R}(\theta)\geq r_0\}$

若$T(F)=\int XdF$, $F=(1-\epsilon)F_n+\epsilon\delta_x$，其中$\delta_x$为单点分布，$x$异于$X_1,\ldots,X_n$的点，这时经验似然比

$$R(F)=\frac{\prod_{i=1}^{n}(1-\epsilon)/n}{\prod_{i=1}^{n}1/n}=(1-\epsilon)^n,\quad T(F)=(1-\epsilon)\bar{X}+\epsilon x.$$

对任意阀值$r < 1$，只要取ϵ足够小，就可使得$R(F) > r$，但是若令$x \to \pm\infty$,将会导致$T(F)$的置信区间的长度会无限长，从而我们需要对F加以限制，很自然的一个选择是$F \ll F_n$。

$F \ll F_n$等价于F的支撑为$[X_{(1)}, X_{(n)}]$,也就是说$\sum\limits_{i=1}^{n} p_i = 1$，并且Owen (2001)证明了若$\sum\limits_{i=1}^{n} p_i < 1$，则可以找到$\widetilde{F}$满足$\int FdF = \int \widetilde{F}d\widetilde{F}), L(\widetilde{F}) > L(F)$，从而一般假设$\sum\limits_{i=1}^{n} p_i = 1$。

当样本中无结点时，即$\forall\ i \neq j,\ X_i \neq X_j$设$F$在$X_i$处的概率为$p_i \geq 0$，即$p_i = F(\{X_i\})$，则$\sum\limits_{i=1}^{n} p_i \leq 1$，所以

$$R(F) = \frac{L(F)}{L(F_n)} = \prod_{i=1}^{n} np_i, \quad p_i \geq 0, \quad \sum_{i=1}^{n} p_i \leq 1. \tag{3.1}$$

当样本中有结点时，即存在$i \neq j$但$X_i = X_j$。设不同的值z_j在样本中出现n_j次，并且在z_j处的概率为p_j，即$n_j = \sharp\{X_i = z_j\}$，$p_j = F(\{z_j\})$，$k$为样本中出现不同值的个数，所以此时

$$R(F) = \prod_{j=1}^{k} (\frac{p_j}{\hat{p}_j})^{n_j} = \prod_{j=1}^{k} (\frac{np_j}{n_j})^{n_j}, \quad \sum_{j=1}^{k} n_j = n, \quad \sum_{j=1}^{k} p_j \leq 1. \tag{3.2}$$

显然采用(3.1)式要比用(3.2)式简单得多，下面的引理表明有结点和无结点时的轮廓似然比函数$\mathcal{R}(\theta)$是一样的。

引理3.1.1. 对$\forall\ r \in [0,1]$,

$$\{F|R(F) \geq r\} \;=\; \{F|\prod_{i=1}^{n} np_i \geq r, \quad p_i \geq 0, \quad \sum_{i=1}^{n} p_i \leq 1\}.$$

所以不论样本中是否有结点，似然比的形式我们都采用(3.1)式，此时轮廓似然函数为

$$\mathcal{R}(\theta) = \sup\left\{\prod_{i=1}^{n} np_i \mid p_i \geq 0, \ \sum_{i=1}^{n} p_i \leq 1, \ T(F) = \theta\right\}.$$

§3.2 统计泛函的经验似然推断

1. 有关均值的经验似然推断

假设我们感兴趣的参数为$\mu = T(F) = \int XdF$，则关于μ的轮廓似然比函数为

$$\mathcal{R}(\mu) = \max\{\prod_{i=1}^{n} np_i \mid \sum_{i=1}^{n} p_i X_i = \mu,\ p_i \geq 0, \sum_{i=1}^{n} p_i = 1\}. \tag{3.3}$$

μ的经验似然置信域为

$$\{\mu | \mathcal{R}(\mu) \geq r\}.$$

令

$$C_{r,n} = \left\{ \sum_{i=1}^{n} p_i X_i | \prod_{i=1}^{n} np_i \geq r,\ p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\}$$

定理3.2.1. $\mu \in C_{r,n}$当且仅当$\mathcal{R}(\mu) \geq r$.

证明:

必要性; 若$\mu \in C_{r,n}$，则存在$\widetilde{p_i}$，满足$\prod_{i=1}^{n} n\widetilde{p_i} \geq r$，$\widetilde{p_i} \geq 0, \sum_{i=1}^{n} \widetilde{p_i} = 1$，使得$\mu = \sum_{i=1}^{n} \widetilde{p_i} X_i$ 所以$\mathcal{R}(\mu) \geq \prod_{i=1}^{n} n\widetilde{p_i} \geq r$

充分性: 若$\mathcal{R}(\mu) \geq r$，则存在$p_i^*, \sum_{i=1}^{n} p_i^* X_i = \mu$，$p_i^* \geq 0, \sum_{i=1}^{n} p_i^* = 1$ 使得$\prod_{i=1}^{n} np_i^* \geq r$，从而$\mu = \sum_{i=1}^{n} p_i^* X_i \in C_{r,n}$.

为了求得(3.3)的最大值，Owen(1988, 1990)采用拉格朗日乘子法，求得p_i的最优值为

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda^T (X_i - \mu)},\ i = 1, \cdots, n$$

其中拉格朗日乘子$\lambda \in R^d$满足

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

从而轮廓似然比函数

$$\mathcal{R}(\mu) = \prod_{i=1}^{n} np_i = \prod_{i=1}^{n} (1 + \lambda^T (X_i - \mu)).$$

进一步可以得到类似于参数似然比的Wilks定理:

定理3.2.2. 设$X_1, \ldots, X_n \in R^d$ $i.i.d$, $EX = \mu_0, Cov(X) = \Sigma_0$，且$rank(\Sigma_0) = q > 0$，则$C_{r,n}$为凸集且

$$-2 \log \mathcal{R}(\mu_0) \xrightarrow{d} \chi^2_{(q)}, \quad n \longrightarrow \infty.$$

注: 定理3.2.2中的χ^2分布，与Wilks (1938)给出的参数似然比的极限分布是一样的，并且取$r = \exp(-\frac{1}{2}\chi_q^2((1-\alpha)))$，就可得到均值$\mu$的可信度为$100(1-\alpha)\%$的置信域。

2. 均值光滑函数的经验似然推断

随机变量X的方差$\sigma^2 = E(X^2) - (EX)^2$也就是说$\sigma^2$可以写成向量$(X, X^2)^T$均值的光滑函数，类似地，随机变量$X, Y$的相关系数

$$\rho = \frac{E(XY) - E(X)E(Y)}{\sqrt{Var(X)Var(Y)}}$$

也可以写成随机向量$(X, Y, X^2, Y^2, XY)^T$均值的光滑函数，下面我们就将经验似然推断用于均值的光滑函数的统计推断中。

方差的非参数最大似然估计为$(1/n)\sum\limits_{i=1}^{n}(X_i - \bar{X})^2$，轮廓似然比为

$$\mathcal{R}(\sigma^2) = \max\{\prod_{i=1}^{n} np_i \mid \sum_{i=1}^{n} p_i(X_i - \mu(F))^2 = \sigma^2,\ p_i \geq 0,\ \sum_{i=1}^{n} p_i = 1\}$$

其中$\mu(F) = \sum\limits_{i=1}^{n} p_i X_i$.

一般地，设$X_i, \ldots, X_n \in R^d$具有共同的分布函数F_0，均值为μ_0，$h: R^d \to R^q (1 \leq q \leq d)$是光滑函数。我们感兴趣的参数$\theta = h(\mu)$，则$\theta$的非参数最大似然估计为$\hat{\theta} = h(\bar{X})$，轮廓似然比函数为

$$\mathcal{R}(\theta) = max\{\prod_{i=1}^{n} np_i | h(\sum_{i=1}^{n} p_i X_i) = \theta,\ p_i \geq 0,\ \sum_{i=1}^{n} p_i = 1\}.$$

定理3.2.3. 设$X_1, \ldots, X_n \in R^d$, $i.i.d$, $E(X) = \mu_0, Cov(X) = \Sigma_0$，且$rank(\Sigma_0) = q > 0$，$h: R^d \to R^q$, $1 \leq q \leq d$是光滑函数，并且h 在μ_0处的*Frechet*导数$G_{d\times q}$在μ_0处不为*0*，$rank(G_{d\times q}) = p$，定义$\theta_0 = h(\mu_0)$,

$$\begin{aligned} C_{r,n}^{(1)} &= \{\sum_{i=1}^{n} p_i X_i | \prod_{i=1}^{n} np_i \geq r,\ p_i \geq 0, \sum_{i=1}^{n} p_i = 1\}, \\ C_{r,n}^{(2)} &= \{h(\mu) | \mu \in C_{r,n}^{(1)}\}, \\ C_{(r,n)}^{(3)} &= \{\theta_0 + G^T(\mu - \mu_0) | \mu \in C_{r,n}^{(1)}\}, \end{aligned} \tag{3.4}$$

则当$n \longrightarrow \infty$时，有

$$P(\theta \in C_{(r,n)}^{(3)}) \longrightarrow P(\chi_{(p)}^2 \leq -2\log(r)).$$

$$\sup_{\mu\in C_{r,n}^{(1)}} \|h(\mu)-\theta_0-G^T(\mu-\mu_0)\| = o_p(n^{-1/2}).$$

注： $D(I)=\{f;f$为I上左连右极的实函数$I\subseteq R\}$，$\|f\|=\sup_{x\in I}|f(x)|$。称$T:D(I)\to R$在$F_0\in D(I)$处Frechet 可微，若存在有界线性函数$T_0':D(I)\to R$，使得当$\|F-F_0\|\to 0$时，

$$\frac{|T(F)-T(F_0)-T_0'(F_0-F)|}{\|F-F_0\|}\longrightarrow 0.$$

称T_0'为T在F_0处的Frechet 导数。

3. 估计方程的经验似然推断

设随机向量$X_1,\ldots,X_n\in R^d, i.i.d$，具有共同的分布函数$F_{\theta_0}$，与$F_{\theta_0}$有关的参数 $\theta\in R^p$，定义函数$m(X,\theta):R^d\times R^p\longrightarrow R^s$，且满足

$$E_{\theta_0}(m(X,\theta_0))=0 \tag{3.5}$$

当且仅当$\theta=\theta_0$.
则根据大数定理，可以得到

$$\frac{1}{n}\sum_{i=1}^{n} m(X_i,\theta_0)\approx 0. \tag{3.6}$$

根据方程(3.6)，若$p=s$可以得到真值θ_0的估计$\hat{\theta}$，称$m(X,\theta)$为估计函数，(3.6)为估计方程。估计方程的应用非常广泛，矩估计，极大似然估计，最小二乘估计，M-估计都可以从估计方程出发得到，包含前面提到的均值及均值的函数都是可以表示成估计方程的形式。对满足(3.5)式的θ，定义其经验似然比函数为

$$\mathcal{R}(\theta)=max\Big\{\prod_{i=1}^{n} np_i|\sum_{i=1}^{n} p_i m(X_i,\theta)=0, p_i\geq 0, \sum_{i=1}^{n} p_i=1\Big\}.$$

定理3.2.4. 设随机向量$X_1,\ldots,X_n\in R^d, i.i.d$，具有共同的分布函数$F_{\theta_0}$，$\forall\,\theta\in\Theta\subseteq R^p$，定义函数$m(X,\theta):R^d\times R^p\longrightarrow R^s$，若$\theta_0\in\Theta$，使得$Var(m(X_i,\theta_0))$有限，且秩为$q>0$，$E(m(X,\theta_0))=0$，则

$$-2\log\mathcal{R}(\theta_0)\longrightarrow\chi_q^2.$$

定理3.2.4并没有说明θ_0有一个很好的估计，更没有保证(3.5)式有唯一的解，但是这并不影响定理3.2.4的运用。当我们可以证明(3.5)式有唯

一的解，并且由估计方程可以得到θ_0一个相合估计，则通过定理3.2.4可以得到θ_0的置信域并且可以进行相应的检验。

在参数估计中，如果有$p+q$个估计方程，但是仅有q个未知参数时，我们可以从$p+q$个估计方程中选择q个，忽略其它p个进行估计，但是更常用的是利用$p+q$个估计方程的q个线性组合进行估计，也即选取$(p+q)\times q$ 阶矩阵$A(\theta)$，且$rank(A(\theta))=p,\ \forall\ \theta$，使得

$$E(m(X,\theta)A(\theta))=0. \tag{3.7}$$

Qin and Lawless (1994)证明了如果用经验似然的方法处理这种超定(overdetermined)问题，并且得到的经验似然估计的渐近方差比由(3.7)式得到的估计的渐近方差要小。

定理3.2.5. 设随机向量$X_1,\ldots,X_n\in R^d, i.i.d$，$\theta_0\in R^p$由$E(m(X,\theta))$唯一确定，其中$m(X,\theta):R^d\times R^p\to R^{(p+q)}$，$q\geq 0$,

$$\mathcal{R}(\theta)=max\left\{\prod_{i=1}^n np_i|\sum_{i=1}^n p_i m(X_i,\theta)=0,\ p_i\geq 0,\ \sum_{i=1}^n p_i=1\right\}$$

令$\widetilde{\theta}=\arg\max_\theta \mathcal{R}(\theta)$，假设存在$\theta_0$的邻域$\Theta$及函数$M(x)$, $E(M(X))<\infty$，满足条件

1. $E(\partial m(X,\theta_0)/\partial\theta)$的秩为$p$.
2. $E(m(X,\theta_0)m(X,\theta_0)^T)$为正定阵.
3. $\partial m(x,\theta)/\partial\theta$关于$\theta\in\Theta$连续.
4. $\partial^2 m(x,\theta)/\partial\theta\partial\theta^T$对$\forall\ \theta\in\Theta$连续.
5. $\|m(x,\theta)\|^3\leq M(x),\forall\ \theta\in\Theta$.
6. $\|\partial m(x,\theta)/\partial\theta\|\leq M(x),\forall\ \theta\in\Theta$.
7. $\|\partial^2 m(x,\theta)/\partial\theta\partial\theta^T\|\leq M(x),\forall\ \theta\in\Theta$.

则

$$\lim_{n\longrightarrow\infty} nVar(\widetilde{\theta})=[E(\frac{\partial m}{\partial\theta})^T(E(mm^T))^{-1}E(\frac{\partial m}{\partial\theta})]^{-1}.$$

且上式得到的渐近方差至少比任何由(3.7)式得到的估计的渐近方差要小. 更进一步，当$n\to\infty$时，有

$$-2\log(\mathcal{R}(\theta_0)/\mathcal{R}(\widetilde{\theta}))\xrightarrow{d}\chi_p^2,$$

$$-2\log\mathcal{R}(\widetilde{\theta})\xrightarrow{d}\chi_q^2.$$

设$X\in R^p, Y\in R^q$，$E(X)=\mu_{x0}$已知，我们感兴趣的参数是$\mu_{y0}=E(Y)$，由于已知X的信息，所以在对μ_{y0}进行参数推断时，很自然的应该

限定p_i满足$\sum_{i=1}^n p_i X_i = \mu_{x0}$，下面的定理表明如果利用这些辅助信息(auxiliary information)可以给出其它统计泛函更为准确的推断。

定义

$$\mathcal{R}_{XY}(\mu_x, \mu_y) = \max\{\prod_{i=1}^n np_i | \sum_{i=1}^n p_i X_i = \mu_x,\ \sum_{i=1}^n p_i Y_i = \mu_y,\ p_i \geq 0, \sum_{i=1}^n p_i = 1\},$$

$$\mathcal{R}_X(\mu_x) = \max\{\prod_{i=1}^n np_i | \sum_{i=1}^n p_i X_i = \mu_x,\ p_i \geq 0, \sum_{i=1}^n p_i = 1\},$$

$$\mathcal{R}_{Y|X}(\mu_y|\mu_x) = \frac{\mathcal{R}_{XY}(\mu_x, \mu_y)}{\mathcal{R}_X(\mu_x)}.$$

定理3.2.6. 设$Z_1, \ldots, Z_n \in R^{p+q}, i.i.d$ 服从分布F，$Z_i = (X_i^T, Y_i^T)^T, X_i \in R^p, Y_i \in R^q, E(Z_1) = (\mu_{x0}^T, \mu_{y0}^T)^T$,

$$Var(Z_1) = \Sigma_{zz} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}.$$

满秩，且Z_1四阶矩有限。如果$\mu_x = \mu_{x0} + O_p(n^{-1/2}), \mu_y = \mu_{y0} + O_p(n^{-1/2})$，则

(i)

$$-2\log \mathcal{R}_{Y|X}(\mu_y|\mu_x) = n((\bar{Y}-\mu_y)-\beta_{y.x}(\bar{x}-\mu_x))^T \Sigma_{y|x}^{-1}(\bar{Y}-\mu_y)-\beta_{y.x}(\bar{x}-\mu_x))+o_p(n^{-1/2})$$

其中$\Sigma_{y|x} = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}\Sigma_{xy}$，$\beta_{y.x} = \Sigma_{yx}\Sigma_{xx}^{-1}$

(ii) $-2\log \mathcal{R}_{Y|X}(\mu_y|\mu_x) \xrightarrow{d} \chi^2_{(q)}, \quad n \longrightarrow \infty$

从上面的定理看出，如果知道μ_{x0},我们用$\mathcal{R}_{Y|X}(\mu_y|\mu_{x0})$来代替$\mathcal{R}_Y(\mu_y)$可以得到对$\mu_{y0}$更准确的推断，这时$\mu_{y0}$ 的最大经验似然估计与$\bar{Y}-\beta_{y.x}(\bar{x}-\mu_x)$很接近，置信域与条件协方差阵有关。如果$X,Y$有很强的相依关系时，带有辅助信息的置信域比没有限制条件的信息小得多，当X,Y不相关时，二者是一样的，这与参数似然比推断中是否存在冗余参数的情况很相似。

§3.3 线性回归模型的经验似然推断

线性回归是统计学中应用最为广泛的方法之一，本节用经验似然方法研究线性回归模型中有关参数的统计推断。线性回归有两种用的

比较多的模型：一种是将预报变量看成固定的向量，一种是把预报变量看成随机向量。下面针对这两种情形，我们分别讨论。

1. 预报变量随机

当预报变量为随机向量时，设$X^T \in R^p, Y \in R$，(x_i, y_i)为向量(X, Y)n个独立的观测值，假设$\mu(X) = E(Y|X) = X\beta_0$，其中

$$\beta_0 = \arg\min_{\beta} E(Y - X\beta)^2 = E(X^T X)^{-1} E(X^T Y),\ \beta_0 \in R^p.$$

则β_0的NPMLE

$$\hat{\beta}_0 = (\frac{1}{n}\sum_{i=1}^{n} x_i^T x_i)^{-1}(\frac{1}{n}\sum_{i=1}^{n} x_i^T y_i).$$

一方面β_0可以看成是均值的光滑函数。从而可以运用定理3.2.3对其进行统计推断，只需$E(\|X\|^4) < \infty,\ E(\|x\|^2 Y^2) < \infty, E(X^T X)$满秩即可。

另一方面β_0可以看成是由估计方程得到的。β_0满足方程

$$E(X^T(Y - X\beta_0)) = 0$$

从而$\hat{\beta}_0$满足

$$\frac{1}{n}\sum_{i=1}^{n} x_i^T (y_i - x_i\hat{\beta}_0) = 0$$

此时若定义变量$Z = Z(\beta) = X^T(Y - X\beta)$，则$\beta_0 = \beta$当且仅当$E(Z) = 0$，所以根据定理3.2.4，只需假设$Z(\beta_0)$具有有限方差即$E\|X^T(Y - X\beta)\|^2 < \infty$即可，$\beta$的轮廓经验似然比函数为

$$\mathcal{R}(\beta) = \max\{\prod_{i=1}^{n} np_i | \sum_{i=1}^{n} p_i x_i^T (y_i - x_i\beta) = 0,\ p_i \geq 0,\ \sum_{i=1}^{n} p_i = 1\}.$$

当$-2\log\mathcal{R}(\beta)$较小时，拒绝$\beta_0 = \beta$，根据定理3.2.4，在$\beta_0 = \beta$成立的条件下，$-2\log\mathcal{R}(\beta) \to \chi_p^2$。

如果我们仅对β_0中的r个分量$\beta_1, \ldots, \beta_r$感兴趣，则可以采用Owen (1990)中的嵌套算法(nested algorithm) 将冗余的$p - r$个参数$\beta_{r+1}, \ldots, \beta_p$清除掉，也即求$\max\limits_{\beta_{r+1}, \ldots, \beta_p} \mathcal{R}(\beta)$,则仍有

$$-2\log \max_{\beta_{r+1}, \ldots, \beta_p} \mathcal{R}(\beta) \longrightarrow \chi_p^2$$

2. 预报变量非随机

设样本为(x_i, y_i), $1 \le i \le n$，其中$x_i^T \in R^p, y_i \in R$，y_i是预报变量x_i对应的响应变量，y_i也可以看成是

$$Y_i = x_i\beta_0 + \epsilon_i,\ Y_i\ i.i.d$$

的观测值，其中$\beta_0 \in R^P$,随机误差ϵ_i相互独立，ϵ_i的分布与$x_j, i \ne j$无关，且$E(\epsilon_i) = 0,\ Var(\epsilon_i) = \sigma^2(x_i) < \infty$.

记$F_x(y) = P(Y_i < y|x_i = x)$，$\mu(x) = E(Y_i|x_i = x) = \int YdF_x = x\beta_0$.

类似预报变量为随机时的情况，引入辅助变量$Z_i = Z_i(\beta) = x_i(Y_i - x_i\beta)$，则$\beta_0 = \beta$当且仅当$E(Z_i) = 0$,由于$x_i$非随机，所以$Z_i$独立但不同分布，$Var(Z_i) = x_i^T x_i \sigma^2(x_i)$, 此时不可以运用以前给出的i.i,d情形的ELT。

下面针对独立但不同分布的变量给出三角形阵列(triangular array) ELT。$maxeig(A), mineig(A)$分别表示对称矩阵A的最大和最小特征根。

定理3.3.1. 设$Z_{in} \in R^p, 1 \le i \le n, p \le n < \infty$是一列随机向量，$\forall\ n, Z_{1n}, \ldots, Z_{nn}$独立，且$E(Z_{in}) = \mu_n, Var(Z_{in}) = V_{in}$，记$V_n = (1/n)\sum_{i=1}^{n} V_{in}$，$\sigma_{1n} = maxeig(V_n)$, $\sigma_{pn} = mineig(V_n)$, $\mathrm{H_n} = \{\sum_{i=1}^{n} \mathrm{w_i Z_{in}} | \sum_{i=1}^{n} \mathrm{w_i} = 1, \mathrm{w_i} \ge 0\}$.
假设当$n \to \infty$时

$$P(\mu_n \in \mathrm{H_n}) \to 1, \tag{3.8}$$

且

$$\frac{1}{n^2}\sum_{i=1}^{n} E(\|Z_{in} - \mu_n\|^4 \sigma_{1n}^{-2}) \longrightarrow 0. \tag{3.9}$$

$\exists\ c > 0$,对$\forall\ n \ge p$,有

$$\sigma_{pn}/\sigma_{1n} \ge c. \tag{3.10}$$

则

$$-2\log \mathcal{R}(\mu_n) \longrightarrow \chi^2_{(p)}.$$

其中

$$\mathcal{R}(\mu_n) = \max\{\prod_{i=1}^{n} np_i | \sum_{i=1}^{n} p_i(Z_{in} - \mu_n) = 0,\ p_i \ge 0,\ \sum_{i=1}^{n} p_i = 1\}.$$

对于预报变量非随机的回归模型，运用定理3.3.1时，$Z_{in} = Z_i(\beta) = x_i(Y_i - x_i\beta)$，当$\beta_0 = \beta$时，$\mu_n = 0$,对所有的$n$，且$V_{in} = x_i^T x_i \sigma^2(x_i)$,$V_n = (1/n)\sum_{i=1}^{n} x_i^T x_i \sigma^2(x_i)$, Owen (1990)给出了满足定理3.3.1中条件的充分条件。

引理3.3.1. $P \triangleq \{x_i | Y_i - x_i\beta - 0 > 0\}$, $N \triangleq \{x_i | Y_i - x_i\beta - 0 < 0\}$,$ch(A)$表示集合$A$的凸组合，如果

$$ch(N) \cap ch(P) \neq \emptyset \tag{3.11}$$

则$0 \in \mathrm{H_n}$

引理3.3.2. 记$\mu_4(x) = \int (Y - x\beta_0)^4 dF_x$，如果对于充分大的$n$,$mineig(V_n) > a > 0$

$$n^{-2} \sum_{i=1}^{n} \|x_i\|^4 \mu_4(x_i) \longrightarrow 0 \quad n \longrightarrow \infty \tag{3.12}$$

则

$$\frac{1}{n^2} \sum_{i=1}^{n} E(\|Z_{in} - \mu_n\|^4 \sigma_{1n}^{-2}) \longrightarrow 0$$

推论3.3.1. 设$n_0 \geq p, \alpha \geq 0, a > 0, b > 0$，若当$n \longrightarrow \infty$时，(3.12)式成立，(3.11)式以概率1成立，$\forall\ i, a < \sigma^2(x_i) < b\|x_i\|^\alpha$, $\forall\ n \geq n_0, a < mineig((1/n)\sum_{i=1}^{n} x_i^T x_i)$, $(1/n)\sum_{i=1}^{n} \|x_i\|^{2+\alpha} < b$, 则

$$-2\log \mathcal{R}(\beta_0) \xrightarrow{d} \chi^2_{(p)}, \quad n \longrightarrow \infty.$$

经验似然方法应用到方差分析、非线性回归、广义线性模型等诸多领域，本章就不一一列举，有兴趣的可以参考Owen (2001).

第四章　利用经验似然对逐段线性回归中的变点进行统计推断

经验似然是一种具有非常好的统计性质的非参数方法，第二章中的参数方法都是在假设随机误差独立同分布的条件下进行的。本章在随机误差独立的条件下，我们利用经验似然方法对逐段线性回归中的变点问题进行研究，探讨了Liu and Qian (2010)提出的特设检验统计量的大样本性质，得到了其在零假设下的渐近分布，并且通过改进特设经验似然方法中残差的相依性以及充分使用预报变量的信息有效地进行回归分析提出了一种新的改进方法。蒙特卡洛模拟结果表明新方法极大地改进了Liu and Qian (2010)的特设经验似然方法在有限样本下的有效性。从一个实际例子的应用中也说明了改进后的方法更有效。更进一步地，我们提出了采用自助法(bootstrap)来逼近新方法的P-值。

§4.1　特设(ad hoc)经验似然方法

首先考虑逐段简单线性回归模型(segmented simple linear regression model)，设$\{(x_i, y_i)\}_{i=1}^n$是一列来自(X,Y)的独立观测值，服从下面的模型

$$y_i = (\alpha_0 + \alpha_1 x_i) I(x_i \le \tau) + (\beta_0 + \beta_1 x_i) I(x_i > \tau) + \epsilon_i, i = 1, \cdots, n, \tag{4.1}$$

其中随机误差ϵ_i独立，且$E(\epsilon_i) = 0, Var(\epsilon_i) = \sigma^2, I(\cdot)$为示性函数，$\alpha_0 + \alpha_1\tau = \beta_0 + \beta_1\tau$。

不失一般性，假设$x_1 \le, \cdots, \le x_n$，则模型(4.1)变为

$$y_i = (\alpha_0 + \alpha_1 x_i) I(i \le k) + (\beta_0 + \beta_1 x_i) I(i > k) + \epsilon_i, i = 1, \cdots, n, \tag{4.2}$$

在模型(4.2)中，假设k未知，我们的主要任务是检测是否存在变点k，也即对$H_0 : k \ge n$作出检验，如果存在，则需要给出k的一个估计。

为了方便，引入记号

$y = (y_1, \cdots, y_n)^T$, $\alpha = (\alpha_0, \alpha_1)^T$, $\beta = (\beta_0, \beta_1)^T$, $\epsilon = (\epsilon_1, \cdots, \epsilon_n)^T$，令

$$\theta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \quad X_k = \begin{pmatrix} X_{1k} & 0 \\ 0 & X_{2k} \end{pmatrix},$$

其中$X_{1k}=(J_k,x_{1k})$, $J_k=(1,\cdots,1)^T$ 长度为k, $x_{1k}=(x_1,\cdots,x_k)^T$, $X_{2k}=(J_{n-k},x_{2k})$, $x_{2k}=(x_{k+1},\cdots,x_n)^T$.

则模型(4.2)用矩阵的形式可以表示为

$$y=X_k\theta+\epsilon, \tag{4.3}$$

Dong (2004)在假设变点k已知和随机误差服从正态分布的条件下，给出了ELw(Empirical Likelihood type wald)检验统计量,

$$ELw=W=(\hat{\alpha}-\hat{\beta})^T\{\hat{\sigma}_1^2(X_{1k}^TX_{1k})^{-1}+\hat{\sigma}_2^2(X_{2k}^TX_{2k})^{-1}\}^{-1}(\hat{\alpha}-\hat{\beta}),$$

其中$\hat{\alpha}$, $\hat{\beta}$ 是α , β的最小二乘估计，$\hat{\sigma}_i$ 是第i阶段随机误差方差的经验验似然估计$i=1,2$。Dong (2004)指出当H_0： $\alpha=\beta$成立时，$ELw\xrightarrow{d}\chi_2^2$。

Liu and Qian (2010)在变点k未知和未假设随机误差服从正态分布的条件下，给出了另外一种用经验似然的方法对逐段简单线性回归模型中的变点问题进行了统计推断。

设$\hat{\alpha}(k)=(\hat{\alpha}_0(k)),\hat{\alpha}_1(k))^T$是从样本$\{(x_i,y_i)\}_{i=1}^k$得到的$\alpha$的最小二乘估计，$\hat{\beta}(k)=(\hat{\beta}_0(k),\hat{\beta}_1(k))$是从样本$\{(x_i,y_i)\}_{i=k+1}^n$计算得到的$\beta$的最小二乘估计,则残差为

$$\hat{e}_i(k)=\begin{cases} y_i-(\hat{\alpha}_0(k)+\hat{\alpha}_1(k)x_i), & i=1,\cdots,k; \\ y_i-(\hat{\beta}_0(k)+\hat{\beta}_1(k)x_i), & i=k+1,\cdots,n. \end{cases}$$

由于在H_0： $\alpha=\beta$成立时，$\hat{\alpha}(k)$与$\hat{\beta}(k)$应该很接近，所以Liu and Qian (2010)建议在估计随机误差时，将$\hat{\alpha}(k)$与$\hat{\beta}(k)$交换。定义

$$\tilde{e}_i(k)=\begin{cases} y_i-(\hat{\beta}_0(k)+\hat{\beta}_1(k)x_i), & i=1,\cdots,k; \\ y_i-(\hat{\alpha}_0(k)+\hat{\alpha}_1(k)x_i), & i=k+1,\cdots,n \end{cases}$$

由于$E(\tilde{e}_i(k))=0$当且仅当$\alpha=\beta$, $i.e\ H_0$成立，所以根据§3.1，当ELR(empirical likelihood ratio)

$$\mathcal{R}(k)=\sup\{\prod_{i=1}^n np_i|\sum_{i=1}^n p_i\tilde{e}_i(k)=0,p_i\geq 0,\sum_{i=1}^n p_i\}$$

比较小时，拒绝H_0，即认为回归模型中确实存在变点。ELR取对数后为

$$-2\log\mathcal{R}(k)=-2\sup\{\sum_{i=1}^n\log(np_i)|\sum_{i=1}^n p_i\tilde{e}_i(k)=0,p_i\geq 0,\sum_{i=1}^n p_i\}$$

事实上变点k的真值k_0未知，Liu and Qian (2010)推荐使用下面的检验统计量进行检验：

$$M_n = \max_{[ln\ n]^2 \leq k \leq n-[ln\ n]^2} \{-2\log\mathcal{R}(k)\}$$

当对$\forall\, k, -2\log\mathcal{R}(k)$都很小时，$M_n$也很小；当变点$k_0$确实存在时，$-2\log\mathcal{R}(k_0)$和$M_n$都很大。所以当$M_n$比较大时，拒绝$H_0$, 承认确实存在变点，此时变点的真值$k_0$的估计为

$$\hat{k}_0 = \min\{k : M_n = -2\log\mathcal{R}(k)\}$$

由于上述用于经验似然检验中的回归系数是通过最小二乘方法得到的，所以暂时称这种方法为特设(ad hoc)经验似然。

对任意固定的k，根据定理3.2.2知，$-2\log\mathcal{R}(k) \xrightarrow{d} \chi_1^2$. $-2\log\mathcal{R}([ln\,n])$, $\cdots, -2\log\mathcal{R}(n-[ln\ n])$并不独立，Liu and Qian(2010)通过模拟表明$Z_n = \sqrt{M_n}$在零假设下的渐近分布为Gumbel 极值分布，这一结论与Csörgó and Horváth(1997) 给出的参数似然比的结论是一致的，下一节我们将给出这个结果的证明。

注: Gumbel极值分布属于广义极值分布(Generalized Extreme Value distribution)。广义极值分布的累积分布函数为

$$F(x;\mu,\sigma,\varsigma) = \exp\left\{ -\left[1+\varsigma(\frac{x-\mu}{\sigma})\right]^{-1/\varsigma}\right\},\ \ x\in R,\ 1+\varsigma(x-\mu)/\sigma > 0.$$

其中$\mu\in R$是位置参数，$\sigma > 0$是尺度参数，$\varsigma\in R$是形状参数。形状参数$\varsigma\in R$ 决定了分布函数尾部的性质。当$\varsigma\in R\to 0$时，广义极值分布的极限分布是Gumbel极值分布，其分布函数为

$$F_G(x,\mu,\sigma) = \exp\left\{ -\exp\left[\frac{-(x-\mu)}{\sigma}\right]\right\}.$$

Gumbel极值分布的均值和方差分别为$\mu+\sigma a, \sigma^2\pi^2/6$。在计算中我们可以用**R**中软件包**evd**的**fgev**函数来计算有关F_G的参数。

§4.2 大样本性质

这一节，我们主要给出逐段多元线性回归中似然比检验统计量在零假设下的渐近分布，很显然包含了§4.1逐段简单线性回归模型的情形。

设$\{(x_i,y_i)\}_{i=1}^n$是一列来自$(X,Y), X\in R^d, Y\in R$的独立观测值，服从下面的模型：

$$y_i = x_i^T\alpha I(i\le k) + x_i^T\beta I(i>k) + \epsilon_i \tag{4.4}$$

其中随机误差ϵ_i i.i.d，且$E(\epsilon_i)=0, Var(\epsilon_i)=\sigma^2$。类似§4.1，将模型(4.4)表示成矩阵的形式。
令
$y=(y_1,\cdots,y_n)^T$，$\alpha=(\alpha_0,\alpha_1)^T$，$\beta=(\beta_0,\beta_1)^T$，$\theta=(\alpha^T,\beta^T)^T\in R^{2d}$，$\epsilon=(\epsilon_1,\cdots,\epsilon_n)^T$，

$$\theta=\begin{pmatrix}\alpha\\ \beta\end{pmatrix},\quad X_k=\begin{pmatrix}X_{1k} & 0\\ 0 & X_{2k}\end{pmatrix},$$

其中$X_{1k}^T=(x_1,\cdots,x_k)$，$X_{2k}^T=(x_{(k+1)},\cdots,x_n)$，则模型(4.4) 可表示为

$$y = X_k\theta+\epsilon, \tag{4.5}$$

则α,β的最小二乘估计$\hat\alpha(k),\hat\beta(k)$可表示为

$$\hat\alpha(k) = (X_{1k}^TX_{1k})^{-1}X_{1k}^T(I_k,0)y,$$
$$\hat\beta(k) = (X_{2k}^TX_{2k})^{-1}X_{2k}^T(0,I_{n-k})y.$$

定义

$$\tilde y_i(k)=\begin{cases} x_i^T\hat\beta(k) & i\le k,\\ x_i^T\hat\alpha(k) & k+1\le i.\end{cases}$$

则

$$\tilde e_i(k) = y_i-\tilde y_i(k),\quad i=1,\cdots,n.$$

$$\mathcal{R}(k)=\sup\{\prod_{i=1}^n np_i|\sum_{i=1}^n p_i\tilde e_i(k)=0, p_i\ge 0,\sum_{i=1}^n p_i=1\} \tag{4.6}$$

根据Lagrange乘子法，得

$$-2\log\mathcal{R}(k) = -2\{\sum_{i=1}^n\log[1+\hat\lambda(k)\tilde e_i(k)]\}, \tag{4.7}$$

其中$\hat\lambda(k)$是方程(4.8)的根。

$$\sum_{i=1}^k\frac{\tilde e_i(k)}{1+\lambda\tilde e_i(k)}=0. \tag{4.8}$$

根据§4.1,

$$M_n=\max_{d\le k\le n-d}\{-2\log\mathcal{R}(k)\}.$$

所需要的正则条件如下:

$C.1$ $rank(X_{1k_n}) = rank(X_{2k_n}) = d.$

$C.2$ 存在$\nu > 0, \tau_1^2 > 0, \tau_2^2 > 0$及正定阵$\Sigma_1, \Sigma_2$，使得当$k \to \infty$, $n \to \infty$时

$$|\frac{1}{k} X_{1k}^T X_{1k} - \Sigma_1| = o(r(k)), \qquad |\frac{1}{n-k} X_{2k}^T X_{2k} - \Sigma_2| = o(r(n-k)). \tag{4.9}$$

$$|(\sum_{i=1}^{k} x_i)^T (X_{2k}^T X_{2k})^{-1} (\sum_{i=1}^{k} x_i) - \tau_1^2| = O(r(n-k)). \tag{4.10}$$

$$|(\sum_{j=k+1}^{n} x_j)^T (X_{1k}^T X_{1k})^{-1} (\sum_{j=k+1}^{n} x_j) - \tau_2^2| = o(r(k)). \tag{4.11}$$

其中$r(t) = 1/(\log t)^\nu$, $|\cdot|$为一般的范数：$|(a_{ij})| = (\sum_i \sum_j a_{ij}^2)^{1/2}$.

$C.3$ 存在$\delta > 0$,使得当$n \to \infty$时，

$$\max_{1 \le i \le n} |x_i| = o(n^{1/(2+\delta)}), \ E|\epsilon_i|^{2+\delta} < \infty.$$

由条件$C.1$知$rank(X_{1k}) = d$，对所有满足$k_n \le k \le n$的k都成立，$rank(X_{2k}) = d, \forall\ 1 \le k \le k_n$; 条件$C.2$比Csörgó and Horváth (1997)中的假设$\Sigma_1 = \Sigma_2$弱一些。由于在逐段回归模型中，我们比较关心的是模型在不同的阶段有不同的形式，所以希望$(1/k)X_{1k}^T X_{1k}$与$(1/(n-k))X_{2k}^T X_{2k}$ 有不同的极限。比如在逐段简单回归模型中(d=2)，若k^*为变点的真值，则希望对所有$i \le k^*, j \ge k^*+1$，$x_i \le x_j$。在一般的回归模型中，$(y_i, x_i^T), i = 1, \cdots, n$是独立同分布的样本，且存在某个$\delta > 0$使得$E|(y_i, x_i^T)|^{2+\delta} < \infty$,此时条件$C.2$和$C.3$以概率1成立，所以条件$C.1, C.2$和$C.3$很自然。

定理4.2.1. *若存在$\nu > 2 + 27/\min(1, \delta)$,使得条件$C.1 - C.3$成立，则在$H_0$成立时，对任意$t$，有*

$$\lim_{n \longrightarrow \infty} P\{A(\log n)\sqrt{M_n} \le t + D_d(\log n)\} = \exp(-2e^{-t}). \tag{4.12}$$

其中$A(x) = (2\log x)^{1/2}, \quad D_d(x) = 2\log x + (d/2)\log\log x - \log\Gamma(d/2), \Gamma(t)$为$Gamma$ 函数。

定理4.2.1的主要证明思路就是利用Owen (1991)得到$\mathcal{R}(k)$的二次逼近，然后类似第二章参数似然比的方法就可以得到此极限。而Owen(1991)中关键就是在得到$\mathcal{R}(k)$的Taylor展开式中的主要项的过程中，得到$\hat{\lambda}(k)$的阶为$o_p(n^{-1/2})$的逼近。为了证明定理4.2.1，我们先给出下面几个引理。

引理4.2.1给出$\max|\tilde{e}_i(k)|$阶的估计，记$\bar{e}(k) = (1/n)\sum_{i=1}^{n}\tilde{e}_i(k)$, $s^2(k) = (1/n)\sum_{i=1}^{n}\tilde{e}_i^2(k)$

引理4.2.1. 假设H_0和$C.1-C.3$成立，则当$n\to\infty$时，

$$\max_{1\le i\le n}\{\max_{d\le k\le n-d}|\tilde{e}_i(k)|\} = o_P(n^{1/(2+\delta)})$$

证明:

当$H_0:\alpha=\beta$成立时，$E\hat{\alpha}(k)=E\hat{\beta}(k)$.令$\gamma_1=(I_k,0)\epsilon$, $\gamma_2=(0,I_{n-k})\epsilon$, 则在$H_0:\alpha=\beta$成立时，当$i\le k$时

$$\begin{aligned}\tilde{e}_i(k) &= y_i - \tilde{y}_i(k)\\ &= y_i - x_i^T\hat{\beta}(k)\\ &= \epsilon_i + x_i^T\alpha - x_i^T(X_{2k}^TX_{2k})^{-1}X_{2k}^T(0,I_{n-k})y\\ &= \epsilon_i - x_i^T(X_{2k}^TX_{2k})^{-1}X_{2k}^T\Big((0,I_{n-k})y - X_{2k}\alpha\Big)\\ &= \epsilon_i - x_i^T(X_{2k}^TX_{2k})^{-1}X_{2k}^T(0,I_{n-k})\epsilon\end{aligned} \tag{4.13}$$

同理当$i>k$时，有$\tilde{e}_i(k)=\epsilon_i - x_i^T(X_{2k}^TX_{2k})^{-1}X_{2k}^T(I_k,0)\epsilon$。所以$\tilde{e}_i$可表示为

$$\tilde{e}_i = \epsilon_i - x_i^T(X_{2k}^TX_{2k})^{-1}X_{2k}^T\gamma_2 I(i\le k) - x_i^T(X_{1k}^TX_{1k})^{-1}X_{1k}^T\gamma_1 I(i>k) \tag{4.14}$$

根据条件$C.3$，$E\{|\epsilon_i|^{(2+\delta)/2}\}^2<\infty$,从而,由*Owen (2001)*中的引理*11.2*, 知$\max|\epsilon_i|^{(2+\delta)/2}=o_P(n^{1/2})$，从而根据连续映射定理可得

$$\max\{|\epsilon_i|\} = o_P(n^{1/(2+\delta)}) \tag{4.15}$$

另外，根据条件$C.2$和重对数律，有

$$\max_{1\le i\le n}\left\{\max_{d\le k\le n-d}(n-k)^{1/2}\frac{|(X_{2k}^TX_{2k})^{-1}X_{2k}^T\gamma_2|}{[\log\log(n-k)]^{1/2}}\right\} = O_p(1) \tag{4.16}$$

所以由条件$C.3$ 和(4.16)式，得

$$\begin{aligned}&\max_{1\le i\le n}\left\{\max_{d\le k\le n-d}|x_i^T(X_{2k}^TX_{2k})^{-1}X_{2k}^T\gamma_2 I(i\le k)|\right\}\\ =&o_P(n^{1/(2+\delta)})\max_{1\le i\le n}\left\{\max_{d\le k\le n-d}(n-k)^{-1/2}[\log\log(n-k)]^{1/2}\right\}\\ =&o_P(n^{1/(2+\delta)})\end{aligned} \tag{4.17}$$

类似地，可得到

$$\max_{1\le i\le n}\left\{\max_{d\le k\le n-d}|x_i^T(X_{1k}^TX_{1k})^{-1}X_{1k}^T\gamma_1 I(i>k)|\right\} = o_P(n^{1/(2+\delta)}) \tag{4.18}$$

所以根据(4.14), (4.15), (4.17), (4.18),引理的结论得证。

引理4.2.2. 假设H_0和$C.1-C.3$成立，则
(a) 当$n\to\infty$时，$\max\limits_{d\le k\le n-d}|\bar{e}(k)|=O_P(n^{-1/2}(\log\log n)^{1/2})$
(b) 当$n\to\infty$时，$\max\limits_{d\le k\le n-d}s^2(k)=O_P(1)$且$\liminf\limits_{n\longrightarrow\infty}\max\limits_{d\le k\le n-d}s^2(k)\ge\sigma^2>0$以概率1成立。
(c) 若当$n\to\infty$时，$k_n\to\infty$，则有

$$\max_{k_n\le k\le n-k_n}|s^2(k)-\sigma^2|=o_P(1)$$

证明:

令$\bar{\epsilon}=(1/n)\sum\limits_{i=1}^n\epsilon_i,\gamma_1=(I_k,0)\epsilon,\ \gamma_2=(0,I_{n-k})\epsilon$, 则在$H_0$成立时，有

$$\bar{e}=\bar{\epsilon}-\frac{1}{n}\left\{(\sum_{i=1}^k x_i)^T(X_{2k}^TX_{2k})^{-1}X_{2k}^T\gamma_2+(\sum_{j=k+1}^n x_j)^T(X_{1k}^TX_{1k})^{-1}X_{1k}^T\gamma_1\right\}\quad(4.19)$$

根据条件$C.2$和重对数律，有

$$\max_{d\le k\le n-d}\left\{\frac{|(\sum\limits_{i=1}^k x_i)^T(X_{2k}^TX_{2k})^{-1}X_{2k}^T\gamma_2|}{[(n-k)\log\log(n-k)]^{1/2}}\right\}=O_P(1)$$

从而

$$\frac{1}{n}\max_{d\le k\le n-d}|(\sum_{i=1}^k x_i)^T(X_{2k}^TX_{2k})^{-1}X_{2k}^T\gamma_2|=O_P(n^{-1/2}(\log\log n)^{1/2})\qquad(4.20)$$

类似地，可得到

$$\frac{1}{n}\max_{d\le k\le n-d}|(\sum_{i=k+1}^n x_i)^T(X_{1k}^TX_{1k})^{-1}X_{1k}^T\gamma_1|=O_P(n^{-1/2}(\log\log n)^{1/2})\qquad(4.21)$$

根据(4.19), (4.20), (4.21)，可得

$$\max_{d\le k\le n-d}|\bar{e}(k)|=O_P(n^{-1/2}(\log\log n)^{1/2})$$

(a)得证。

下面考虑$s^2(k)$,

$$\begin{aligned}s^2(k)=&\frac{1}{n}\sum_{i=1}^n\epsilon_i^2+\frac{1}{n}\gamma_1^TX_{1k}(X_{1k}^TX_{1k})^{-1}X_{2k}^TX_{2k}(X_{1k}^TX_{1k})^{-1}X_{1k}^T\gamma_1\\&+\frac{1}{n}\gamma_2^TX_{2k}(X_{1k}^TX_{1k})^{-1}X_{1k}^TX_{1k}(X_{2k}^TX_{2k})^{-1}X_{2k}^T\gamma_2\\&-\frac{2}{n}\gamma_2^TX_{2k}(X_{1k}^TX_{1k})^{-1}X_{1k}^T\gamma_1-\frac{2}{n}\gamma_1^TX_{1k}(X_{2k}^TX_{2k})^{-1}X_{2k}^T\gamma_2\end{aligned}\qquad(4.22)$$

由条件$C.2$和重对数律，得

$$\begin{aligned}
&\max_{d\le k\le n-d} n^{-1}|\gamma_1^T X_{1k}(X_{2k}^T X_{2k})^{-1} X_{2k}^T \gamma_2| \\
=&\max_{d\le k\le n-d} O_P\left\{\frac{[k(\log\log k)\log\log(n-k)]^{1/2}}{n(n-k)^{1/2}}\right\} \\
=&O_P(n^{-1/2}(\log\log n)^{1/2})
\end{aligned} \tag{4.23}$$

类似得到，

$$\max_{d\le k\le n-d} n^{-1}|\gamma_2^T X_{2k}(X_{1k}^T X_{1k})^{-1} X_{1k}^T \gamma_1| = O_P(n^{-1/2}(\log\log n)^{1/2}) \tag{4.24}$$

再次利用条件$C.2$和重对数律得

$$\begin{aligned}
&\max_{d\le k\le n-d} n^{-1}|X_{2k}(X_{1k}^T X_{1k})^{-1} X_{1k}^T \gamma_1|^2 \\
=&\max_{d\le k\le n-d} n^{-1}\frac{[\log\log k]^{1/2}}{k^{1/2}} O_P^2(1) = O_P^2(1)
\end{aligned} \tag{4.25}$$

和

$$\begin{aligned}
&\max_{d\le k\le n-d} n^{-1}|X_{1k}(X_{2k}^T X_{2k})^{-1} X_{2k}^T \gamma_2|^2 \\
=&\max_{d\le k\le n-d} n^{-1}\frac{[\log\log(n-k)]^{1/2}}{(n-k)^{1/2}} O_P^2(1) = O_P^2(1)
\end{aligned} \tag{4.26}$$

当$k_n\to\infty$时，有

$$\max_{k_n\le k\le n-k_n} \frac{[\log\log k]^{1/2}}{k^{1/2}} = o_P(1), \tag{4.27}$$

$$\max_{k_n\le k\le n-k_n} \frac{[\log\log(n-k)]^{1/2}}{(n-k)^{1/2}} = o_P(1) \tag{4.28}$$

所以(4.25)和(4.26) 变为

$$\max_{d\le k\le n-d} \frac{1}{n}|X_{2k}(X_{1k}^T X_{1k})^{-1} X_{1k}^T \gamma_1|^2 = o_P(1), \tag{4.29}$$

$$\max_{d\le k\le n-d} \frac{1}{n}|X_{1k}(X_{2k}^T X_{2k})^{-1} X_{2k}^T \gamma_2|^2 = o_P(1). \tag{4.30}$$

根据(4.22) − (4.32)，(b)和(c)得证。

引理4.2.3. 假设H_0和$C.1-C.3$成立，则存在$\tau>0$, 当$n\to\infty$时，有

$$\max_{d\le k\le n-d} |\hat\lambda(k) - \bar e(k)/s^2(k)| = o_P(n^{-1/2-\tau})$$

证明:

证明过程类似*Owen (1990)*中的讨论，由于$\hat{\lambda}(k)$为方程(4.8)的解，所以

$$
\begin{aligned}
0 &= \tfrac{1}{n}|\sum_{i=1}^{n}\tfrac{\tilde{e}_i(k)}{1+\hat{\lambda}(k)\tilde{e}_i(k)}| \\
&= \tfrac{1}{n}|\sum_{i=1}^{n}\tilde{e}_i(k)(1-\tfrac{\hat{\lambda}(k)\tilde{e}_i(k)}{1+\hat{\lambda}(k)\tilde{e}_i(k)})| \\
&= |\bar{e}(k)-\tfrac{1}{n}\hat{\lambda}(k)\sum_{i=1}^{n}\tfrac{\tilde{e}_i(k)^2}{1+\hat{\lambda}(k)\tilde{e}_i(k)}| \\
&\geq |\hat{\lambda}(k)|\tfrac{s^2(k)}{1+|\hat{\lambda}_k|\max|\tilde{e}_i(k)|}-|\bar{e}(k)|
\end{aligned}
$$

根据引理4.2.1知，存在$\delta>0$,使得

$$\frac{|\hat{\lambda}(k)|}{1+|\hat{\lambda}(k)|o(n^{1/(2+\delta)})}\leq|\bar{e}(k)|/s^2(k). \tag{4.31}$$

由引理4.2.2得

$$\max_{k_n\leq k\leq n-k_n}|\bar{e}(k)|/s^2(k)=O_P(n^{-1/2}(\log\log n)^{1/2}) \tag{4.32}$$

联合(4.31), (4.32)得

$$\max_{k_n\leq k\leq n-k_n}|\hat{\lambda}(k)|=O_P(n^{-1/2}(\log\log n)^{1/2}) \tag{4.33}$$

令$\eta_i(k)=\hat{\lambda}(k)\tilde{e}_i(k)$,根据(4.33)和引理4.2.1得

$$\max_{1\leq i\leq n}\{\max_{k_n\leq k\leq n-k_n}|\eta_i(k)|\}=o_P(1)$$

利用*Taylor*展开得

$$0=\frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{e}_i(k)}{1+\eta_i(k)}=\bar{e}(k)-s^2(k)\hat{\lambda}(k)+\frac{\hat{\lambda}^2(k)}{n}\sum_{i=1}^{n}\frac{\tilde{e}_i{}^3(k)}{(1+\xi_i(k))^3} \tag{4.34}$$

其中$|\xi_i(k)|\leq|\eta_i(k)|=|\hat{\lambda}(k)\tilde{e}_i(k)|$,根据(4.33)式和引理4.2.1得

$$\max_{1\leq i\leq n}\max_{d\leq k\leq n-d}|\xi_i(k)|=o_P(n^{1/(2+\delta)})O_P(n^{-1/2}(\log\log n)^{1/2})=o_P(1) \tag{4.35}$$

因此根据(4.33)式和引理4.2.1, 4.2.2得,下式对k一致成立,

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{\lambda}^2(k)\tilde{e}_i{}^3(k)}{(1+\xi_i(k))^3}\right| &\leq \hat{\lambda}^2(k)s^2(k)O_P(1)\max_{1\leq i\leq n}|\tilde{e}_i(k)| \\
&= o_P(n^{-(1+\delta)/(2+\delta)}\log\log n)=o_P(n^{-1/2})
\end{aligned} \tag{4.36}
$$

根据(4.34), (4.36)和引理4.2.2，得对任意$0<\tau<\delta/[2(2+\delta)]$，引理4.2.3的结论成立。

定理 4.2.1的证明.

首先我们利用引理4.2.1, 4.2.2 和4.2.3得到$-2\log R(k)$ 对k一致的二次逼近，根据Owen (2001)中的讨论,记$z_i = \hat{\lambda}(k)\tilde{e}_i(k)$. 利用Taylor展开，得

$$-2\log R(k) = 2\sum_{i=1}^{n}\log(1+z_i) = 2\sum_{i=1}^{n}\left\{z_i - \frac{1}{2}z_i^2 + \frac{1}{3}\frac{z_i^3}{(1+\xi_i)^3}\right\}. \qquad (4.37)$$

与(4.35)式一样，$|\xi_i| \le |z_i| = |\hat{\lambda}(k)\tilde{e}_i(k)| = o_P(1)$, 对$k$一致成立. 根据引理4.2.1 和引理4.2.3知,存在$\delta > 0$满足,

$$\begin{aligned}\max_{d\le k\le n-d}\sum_{i=1}^{n}\left|\frac{z_i^3}{(1+\xi_i)^3}\right| &\le n\left\{\max_{d\le k\le n-d}[|\hat{\lambda}^3(k)|s^2(k)]\right\}\max_{1\le i\le n}|\tilde{e}_i(k)| \\ &= o_P\left\{n^{-\frac{3}{2}+1+\frac{1}{2+\delta}}\log\log^{3/2}n\right\} \\ &= o_P(n^{-\delta/(4+2\delta)}\log\log^{3/2}n). \qquad (4.38)\end{aligned}$$

根据引理4.2.3，存在$\tau > 0$满足,

$$2\sum_{i=1}^{n}z_i = 2n\bar{e}^2(k)/s^2(k) + n\bar{e}(k)o(n^{-1/2-\tau}) = 2n\bar{e}^2(k)/s^2(k) + o_P(n^{-\tau}\log\log^{1/2}n). \qquad (4.39)$$

和

$$\sum_{i=1}^{n}z_i^2 = n\bar{e}^2(k)/s^2(k) + no_P(n^{-1/2-\tau})\bar{e}(k) = n\bar{e}^2(k)/s^2(k) + o_P(n^{-\tau}\log\log^{1/2}n). \qquad (4.40)$$

联合(4.37), (4.38), (4.39)和(4.40),得到对$\forall\ 0 < \tau_1 < \min\{\delta/(4+2\delta), \tau\}$,

$$\max_{d\le k\le n-d}\left|-2\log R(k) - n\frac{\bar{e}^2(k)}{s^2(k)}\right| = o_p(n^{-\tau_1}). \qquad (4.41)$$

利用Taylor展开

$$(a+x)^{1/2} = a^{1/2} + x/(2a^{1/2}) + o(x/a^{1/2}),$$

对$\forall\ 0 < \tau_2 < \tau_1$有,

$$M_n^{1/2} = \{\max_{d\le k\le n-d}[-\log R(k)]\}^{1/2} = \max_{d\le k\le n-d}n^{1/2}\{|\bar{e}(k)|/s(k)\} + o_p(n^{-\tau_2}). \qquad (4.42)$$

记$M_n^{1/2} = T_n^{1/2} + o_P(n^{-\tau_2})$, i.e.,

$$T_n^{1/2} = \max_{d\le k\le n-d}n^{1/2}\{|\bar{e}(k)|/s(k)\}.$$

类似Csörgő and Horváth (1997)中定理3.1.2的讨论, 有

$$\lim_{n\to\infty}P\{A(\log n)T_n^{1/2} \le x + D_d(\log n)\} = \exp(-2e^{-x}) \qquad \forall\ x.$$

由于$A(\log n)o_P(n^{-\tau_2})=o(1)$, 根据(4.42)式，有

$$A(\log n)M_n^{1/2}-D_d(\log n)=A(\log n)T_n^{1/2}-D_d(\log n)+o_P(1).$$

证毕. □

Liu and Qian (2010)中通过模拟说明，当k过小或接近样本量n时，$-2\log R(k)$对例外点(outliers)非常敏感，为了克服这一缺点，采用了Csörgő and Horváth (1997)中的截断经验似然。取$d\le k_{n1}$, $k_{n2}\le n-d$.

定义

$$M_n'=\max_{k_{n1}\le k\le k_{n2}}\{-2\log R(k)\}.$$

假设当$n\to\infty$时,有

$$u_n=\frac{n^2-k_{n1}k_{n2}}{k_{n1}(n-k_{n2})}\to\infty. \tag{4.43}$$

利用Csörgő and Horváth (1997)中定理A.3.4，类似定理4.2.1的证明，可以得到

$$\lim_{n\to\infty}P\{A(\log u_n)[M_n']^{1/2}\le x+D_d(\log u_n)\}=\exp(-e^{-x}),\qquad \forall\quad x.$$

当k_{n1} 和$n-k_{n2}$ 为常数时, (4.43)式显然成立; Liu and Qian (2010) 中选择$k_{n1}=\log^2 n$ 和$k_{n2}=n-\log^2 n$，则同样满足(4.43)式; 另外使用较多的为$k_{n1}=2\log n=n-k_{n2}$见Perron and Vogelsang (1992)。特别地，若$0<\lambda<1$，$k_{n1}=[\lambda n]$ 和$k_{n2}=n-[\lambda n]$ ，其中$[x]$是小于或等于x的最大整数。利用Csörgő and Horváth (1997) 中推论A.3.1，有

$$\lim_{n\to\infty}P\{A(\log n)[M_n']^{1/2}\le x+D_d(\log n)\}=\exp(-2e^{-x}),\qquad \forall\quad x$$

注: 在蒙特卡洛模拟的过程中，由于$M_n^{1/2}$精确分布未知，如果不用其渐近分布，那么根据(4.42)式，用$\max_{d\le k\le n-d}n^{1/2}\{|\bar{e}(k)|/s(k)\}$来近似计算P-值也很方便。

§4.3　一种改进

我们知道经验似然方法要求变量间是独立的，更进一步，我们希望最大限度的使用预报变量的信息来进行回归分析。但是§4.1中的$\tilde{e}_i, i=1,\cdots,n$是相关的，而且在经验似然方法中的约束方程没有有效的使用预报变量的信息。这一节我们给出一种改进的新方法，并且通过模拟

说明新方法比Liu and Qian (2010)提出的方法更有效。

定义$G_{ik}=(X_{ik}^TX_{ik})^{-1}X_{ik}^T,\ i=1,2$,其中$X_{ik},i=1,2$如§4.2定义，$y_{1k}=(y_1,\cdots,y_k)^T,y_{2k}=(y_{k+1},\cdots,y_n)^T$,则$\hat{\alpha}=G_{1k}y_{1k},\hat{\beta}=G_{2k}y_{2k}$,

$$\begin{aligned}\tilde{e} &= (\tilde{e}_1,\cdots,\tilde{e}_n)^T=\begin{pmatrix} y_{1k}-X_{1k}\hat{\beta} \\ y_{2k}-X_{2k}\hat{\alpha}\end{pmatrix} \\ &= \begin{pmatrix} I_k & -X_{1k}G_{2k} \\ -X_{2k}G_{1k} & I_{n-k}\end{pmatrix}y,\end{aligned}$$

其中I_r为$r\times r$单位阵，令

$$C=\begin{pmatrix} I_k & -X_{1k}G_{2k} \\ -X_{2k}G_{1k} & I_{n-k}\end{pmatrix}. \tag{4.44}$$

则$\tilde{e}=Cy, Var(\tilde{e})=\sigma^2CC^T$, 令$Q$为正交阵，对角阵$\Delta=diag(\delta_1^2,\cdots,\delta_n^2)$，且满足

$$CC^T=Q^T\Delta Q. \tag{4.45}$$

定义

$$z=(z_1,\cdots,z_n)^T=Q\tilde{e}$$

引理4.3.1. y服从模型(4.5),假设X_{1k},X_{2k}的秩为p, C,Q,Δ定义如(4.44), (4.45), 则$z=QCy$服从下面的模型

$$z=X_k^*\theta+\epsilon^*, \tag{4.46}$$

其中$X_k^*=QCX_k$,随机误差$\epsilon^*=(\epsilon_1^*,\cdots,\epsilon_n^*)^T$, $E(\epsilon^*)=0$, $Var(\epsilon_i^*)=\sigma^2\delta_i^2$, 对$i\neq j$, $Cov(\epsilon_i^*,\epsilon_j^*)=0$ $i.e.$, $Var(\epsilon^*)=\sigma^2\Delta$. 更进一步有，$E(z)=0$ 当且仅当$\alpha=\beta$.

证明:

由于

$$Var(\epsilon^*)=Var(z)=Var(QCy)=\sigma^2QCC^TQ^T=\sigma^2\Delta$$

所以$E(\epsilon^*)=0$, $Var(\epsilon_i^*)=\sigma^2\delta_i^2$, 对$i\neq j$, $Cov(\epsilon_i^*,\epsilon_j^*)=0$ $i.e.$, $Var(\epsilon^*)=\sigma^2\Delta$.

根据C和X_k的定义得

$$\begin{aligned}CX_k\theta &= \begin{pmatrix} X_{1k}\alpha-X_{1k}G_{2k}X_{2k}\beta \\ X_{2k}\beta-X_{2k}G_{1k}X_{1k}\alpha\end{pmatrix} \\ &= \begin{pmatrix} X_{1k}\alpha-X_{1k}\beta \\ X_{2k}\beta-X_{2k}\alpha\end{pmatrix}\end{aligned} \tag{4.47}$$

所以$E(z)=X_k^*\theta=0$当且仅当$X_{1k}\alpha=X_{1k}\beta, X_{2k}\alpha=X_{2k}\beta$, 从而得$E(z)=0$当且仅当$\alpha=\beta$.

令$z=QCX_k=(g_1,\cdots,g_n)^T$,由于$-2\log\hat{\mathcal{R}}(k)$对于过大或过小的$k$非常不稳定，所以类似参数似然比的情况，我们在检验$H_0:\ \alpha=\beta$时，采取截断(trimmed)检验统计量;

$$\hat{M}=\max_{t_1\le k\le t_2}\{-2\log\hat{\mathcal{R}}(k)\},$$

其中

$$\hat{\mathcal{R}}(k)=\sup\Big\{\prod_{i=1}^n nw_i|w_i\ge 0,\sum_{i=1}^n w_i=1,\sum_{i=1}^n w_iz_ig_i=0\Big\}. \tag{4.48}$$

正如Perron and Vogelsang (1992)中指出的t_1,t_2的选择是任意的，但是在模拟的过程中发现如果t_1,t_2过小，$\hat{M}$仍然不稳定，所以这里我们采取Zou etal(2007)中的$t_1=t_2=2[\ln n]$. 当$\hat{M}$大时，拒绝H_0，承认存在变点，并且变点k估计的估计$\hat{k}=\{\hat{k}:\hat{\mathcal{R}}(\hat{k})=\hat{M}\}$.
注： 约束方程$\sum_{i=1}^n w_iz_ig_i=0$就是通常回归分析中的正规方程，它有效地运用了预报变量的信息。见Owen (1990)。因此我们希望改进后的方法比特设经验似然方法要有效的多。特别指出，新方法与特设经验似然方法在计算上一样的容易，除了基本代数的运算外，只需要用**R**中软件包**emplik**中的**el.test**函数求$\hat{M}$。

下面通过模拟来比较M和$\hat{M}$的有效性。模拟安排如下:

- 样本量: $n=200$.
- 截断: $t_1=2\log n$，$t_2=n-2\log n$.
- 样本: 从正态分布$N(1,1)$产生n个样本量x,并且x在零假设和备择假设保持不变。
- 零假设下的模型: 从$0.5+x_i+\epsilon_i, i=1,\cdots,n$中产生$y_i$，考虑$\epsilon_i$服从正态分布$N(0,1)$和自由度为4的t-分布两种情况。
- 阀值: 通过蒙特卡洛方法得到M和$\hat{M}$在零假设模型下的5% and 10%的阀值。运用R中软件包emplik来实现计算。

- 备择假设下的模型: 从$0.5+x_i+\epsilon_i$中产生y_i, $i=1,\cdots,K$, and 从$0.8+1.5x_j+\epsilon_j$中产生y_j $j=K+1,\cdots,n$, $K=0.25n,0.5n,0.75n$. 考虑ϵ_i服从正态分布$N(0,1)$和自由度为4的t-分布两种情况。

- 蒙特卡洛的次数: 在零假设和备择假设下，模拟重复进行$m=1000$次

- Hit Rate: $m^*=\#\{$ 当H_0被拒绝且$\hat{k}\in[K-5,K+5]\}$，$m=\#\{$ 当H_0被拒绝$\}$，$r=\frac{m^*}{m}$

表 4.1 M **和**$\hat{M}$ **功效的比较，**ϵ_i**服从正态分布**$N(0,1)$ **.**

	$\alpha=5\%$		$\alpha=10\%$	
	M	$\hat{M}$	M	$\hat{M}$
$k=0.25n$	0.423	0.684	0.612	0.857
$k=0.5n$	0.312	0.629	0.478	0.856
$k=0.75n$	0.427	0.652	0.612	0.868

表 4.2 M 和$\hat{M}$ **hit rate** 的比较，ϵ_i服从正态分布$N(0,1)$.

	$\alpha=5\%$		$\alpha=10\%$	
	M	$\hat{M}$	M	$\hat{M}$
$k=0.25n$	0	0.443	0	0.442
$k=0.5n$	0	0.541	0	0.531
$k=0.75n$	0	0.379	0	0.396

表 4.3 M 和$\hat{M}$ 功效的比较，ϵ_i服从自由度为4的t-分布.

	$\alpha=5\%$		$\alpha=10\%$	
	M	$\hat{M}$	M	$\hat{M}$
$k=0.25n$	0.333	0.670	0.545	0.832
$k=0.5n$	0.226	0.681	0.406	0.877
$k=0.75n$	0.354	0.726	0.578	0.870

表 4.4 M 和$\hat{M}$ **hit rate** 的比较，ϵ_i服从自由度为4的t-分布.

	$\alpha=5\%$		$\alpha=10\%$	
	M	$\hat{M}$	M	$\hat{M}$
$k=0.25n$	0	0.403	0	0.421
$k=0.5n$	0	0.529	0	0.531
$k=0.75n$	0	0.362	0	0.371

从表4.1 – 4.4中可以看出我们利用新方法得到的功效比Liu and Qian (2010)中方法要大的多，并且用旧方法得到的hit rate 全部为0. 有关新方法的大样本性质还在讨论中。基于引理4.3.1以及下节描述的经验似然方法，我们认为改进后的新方法所采用的检验统计量在零假设下的渐近分布为Gumbel极值分布。此结论作为本论文的一个猜想。在实际运用中，正如其它非参数方法的应用一样，我们也可以用自助法来逼近P-值。

由Efron (1982)年提出的自助法是现代统计推断中应用非常广泛并且也非常有效的一种重抽样方法。若样本$w_1,\cdots,w_n$来自分布函数F_θ, $\hat{\theta}(x_1,\cdots,x_n)$为$\theta$的一个估计量。我们的目的是在分布$F_\theta$下计算有关统计量$\hat{\theta}(w_1,\cdots,w_n)$的一些值。由于$F_\theta$未知或者即使知道$F_\theta$而计算上非常复杂等原因而导致这些感兴趣的量无法计算时，我们采用自助法来解决这些问题。比如计算偏差$E_F(\hat{\theta})-\theta$时，自助法采用经验函数F_n代替F,从而用$E_{F_n}(\hat{\theta})-\hat{\theta}(w_1,\cdots,w_n)$ 来估计偏差$E_F(\hat{\theta})-\theta$。其中

$$E_{F_n}(\hat{\theta})=\int\hat{\theta}(t_1,\cdots,t_n)dF_n(t_1,\cdots,t_n).$$

在实际应用中往往用蒙特卡洛方法来估计$E_{F_n}(\hat{\theta})$。

在已知$\hat{M}=\hat{m}$时，我们想得到P-值，即在零假设下$\hat{M}>\hat{m}$的概率$P(\hat{M}>\hat{m})$。由于并不知道响应变量$y_1,\cdots,y_n$是否来自零假设(这也正是我们采用统计量$\hat{M}$的原因)，所以此时不能利用这些原始样本。但是根据$\hat{M}$我们可以得到变点k的一个估计$\hat{k}$，即$\hat{k}=\underset{k}{argmax}\,\hat{M}$。此时设$\hat{\alpha}=(\hat{\alpha}_0,\hat{\alpha}_1)^T$是从样本$\{x_i,y_i\}_{i=1}^{\hat{k}}$得到的$\alpha$的最小二乘估计，$\hat{\beta}=(\hat{\beta}_0,\hat{\beta}_1)$是从样本$\{x_i,y_i\}_{i=\hat{k}+1}^{n}$计算得到的$\beta$的最小二乘估计。记

$$\hat{y}_i=\begin{cases}\hat{\beta}_0+\hat{\beta}_1x_i, & i=1,\cdots,\hat{k};\\ \hat{\alpha}_0+\hat{\alpha}_1x_i, & i=\hat{k}+1,\cdots,n\end{cases}$$

则我们可以认为$y_i-\hat{y}_i,i=1,\cdots,n$是来自零假设的样本。如果备择假设成立，则$\hat{k}$是变点$k$的相合估计；如果零假设成立，虽然此时$\hat{k}$没有意义，但是$y_i-\hat{y}_i,i=1,\cdots,n$有相同的均值0，这也正是我们想要的结果。

令$u_1,\cdots,u_{\hat{k}}$是从$\{y_1-\hat{y}_1,\cdots,y_{\hat{k}}-\hat{y}_{\hat{k}}\}$重抽样得到的样本量为$\hat{k}$的样本，$u_{\hat{k}+1},\cdots,u_n$是从$\{y_{\hat{k}+1}-\hat{y}_{\hat{k}+1},\cdots,y_n-\hat{y}_n\}$重抽样得到的样本量为$n-\hat{k}$个样本。基于样本$u_1,\cdots,u_n$计算得到的$\hat{M}$记为$\hat{m}^1$。类似地，重复抽样B次，基于这些重复抽样样本得到的$\hat{M}$依次记为$\hat{m}^1,\cdots,\hat{m}^B$,则可以利用

$$\hat{p}=\frac{1}{B}\sum_{b=1}^{B}I(\hat{m}^b>\hat{m}).$$

逼近P-值。

用自助法逼近P-值的具体步骤如下：

- 计算出$\hat{M}$的观察值$\hat{m}$和变点k的估计值$\hat{k}=\arg\max\limits_{k}\hat{M}$.
- 从样本$\{x_i,y_i\}_{i=1}^{\hat{k}}$和$\{x_i,y_i\}_{i=\hat{k}+1}^{n}$中分别得到$\alpha$和$\beta$的最小二乘估计记为$\hat{\alpha},\hat{\beta}$.
- 计算$\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1x_i, i=1,\cdots,\hat{k}$，$\hat{y}_i=\hat{\alpha}_0+\hat{\alpha}_1x_i,\ i=\hat{k}+1,\cdots,n$.
- 从$\{y_1-\hat{y}_1,\cdots,y_{\hat{k}}-\hat{y}_{\hat{k}}\}$重复抽样得到样本量为$\hat{k}$的自助样本$u_1,\cdots,u_{\hat{k}}$；从$\{y_{\hat{k}+1}-\hat{y}_{\hat{k}+1},\cdots,y_n-\hat{y}_n\}$重复抽样得到样本量为$n-\hat{k}$的自助样本$u_{\hat{k}+1},\cdots,u_n$。利用样本$u_1,\cdots,u_n$计算得到的$\hat{M}$记为$\hat{m}^1$。
- 重复上一步的步骤B次，并将得到的$\hat{M}$依次记为$\hat{m}^1,\cdots,\hat{m}^B$。
- 用$\hat{p}=\frac{1}{B}\sum\limits_{b=1}^{B}I(\hat{m}^b>\hat{m})$逼近P-值。

§4.4 经验似然方法

在§4.3中提到过将经验似然方法直接应用到变点推断，这一节我们给出具体的介绍。§4.1, §4.3中回归系数是通过最小二乘方法得到的，Liu ,Zou and Zhang (2008)利用经验似然的方法得到回归系数和变点的估计，然后将其用于逐段回归模型的经验似然推断。以下考虑预报变量非随机的情形。

设$E(Y_i|X=x_i)=x_i^T\alpha, i=1,\cdots,k, E(Y_i|X=x_i)=x_i^T\beta, i=k+1,\cdots,n$,其中$x_i\in R^d$,$n\geq 2d+1$ $y_1,\cdots,y_n$为独立变量$Y_1,\cdots,Y_n$的观测值，我们想知道是否确实存在k,使得$\alpha\neq\beta$.

记

$p_i=P(Y_i=y_i|X=x_i), i=1,\cdots,k, q_j=P(Y_j=y_j|X=x_j), j=k+1,\cdots,n$,则经验似然为

$$l(p,q|k)=\sum_{i=1}^{k}\log p_i+\sum_{j=k+1}^{n}\log q_j, \tag{4.49}$$

其中p_i 和q_j 满足

$$p_i\geq 0,\sum_{i=1}^{k}p_i=1;\ q_j\geq 0,\sum_{j=k+1}^{n}q_j=1. \tag{4.50}$$

对于固定的k,当$p_i = 1/k$ 和$q_j = 1/(n-k)$时，$l(p,q|k)$达到最大，从而经验似然比定义为

$$r(p,q|k) = \sum_{i=1}^{k} \log\{kp_i\} + \sum_{j=k+1}^{n} \log\{(n-k)q_j\}. \tag{4.51}$$

对于固定的k，通过最小二乘得到α和β的估计，也即解估计方程$Ex^T(Y - x\alpha) = 0, Ex^T(Y - x\beta) = 0$即可。所以在经验似然方法中,当$H_0 : \alpha = \beta$成立时，考虑下面的估计方程：

$$\sum_{i=1}^{k} p_i(y_i - {x_i}^T\beta)x_i = 0 \text{ 和 } \sum_{j=k+1}^{n} q_j(y_j - {x_j}^T\beta)x_j = 0. \tag{4.52}$$

对于固定的k,定义似然比检验统计量

$$R(k) = \max_{\beta}\{\max_{p,q}\{r(p,q|k)|p_i, q_j \text{ 满足 (4.50) 和 (4.52) }\}\}.$$

截断经验似然比统计量定义为：

$$T_n = \max_{k_{n1} \le k \le k_{n2}}\{-2R(k)\},$$

其中$p+1 \le k_{n1} < k_{n2} \le n-p-1$。根据Perron and Vogelsang (1992), k_{n1}和k_{n2}的选择理论上可以任意的，所以这里我们采取Zou *et al.* (2007)中的$k_{n1} = 2\log n$,$k_{n2} = n - 2\log n$.

在随机变量i.i.d情形，Zou *et al.* (2007)证明了用于分布函数变点问题的截断经验似然比统计量在零假设下的渐近分布与Csörgő and Horváth (1997)中参数似然比一样有Gumbel极值分布。Liu, Zou and Zhang (2008)证明了在逐段线性回归中，T_n在零假设下的渐近分布为Gumbel极值分布。令

$A(x) = \{2\log x\}^{1/2}$, $u_n = (n^2 - k_{n1}k_{n2}/(k_{n1}(n-k_{n2})$,

$$D_d(x) = 2\log x + (d/2)\log\log x - \log\Gamma(d/2).$$

其中$\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx$ 为Gamma 函数。

定理4.4.1. *设$y_1, \cdots, y_n$ 为d维预报变量$x_1, \cdots, x_n$相对应的n个独立响应变量. 假设零假设$H_0 : \alpha = \beta$成立, i.e., $E(Y_i|x_i) = x_i^T\beta$, $i = 1, \cdots, n$. $\sigma^2(x_i) = Var(Y_i|x_i)$. 假设下列条件成立:*

(i) 当$n \to \infty$时,

$$n^{-2}\sum_{i=1}^{n} \| x_i \|^4 E\{(y_i - x_i^T\beta)^4|x_i|\} \to 0,$$

(ii) $c \geq 0$, $a, b > 0$, $a < \sigma^2(x_i)) < b \parallel x_i \parallel^c \forall\ i$, 且$a < \lambda_{\min,n}/n$,$(1/n)\sum \parallel x_i \parallel^{2+c} < b$, 其中$\lambda_{\min,n}$ 是矩阵$\sum x_i x_i^T$最小的特征根,

(iii) 当$n \to \infty$时,$k_{n1}, k_{n2} \to \infty$且$k_{n1}/n \to 0$ 和$k_{n2}/n \to 0$。

则

$$\lim_{n\to\infty} P\left\{A(logu_n)\sqrt{T_n} \leq x + D_d(\log(u_n))\right\} = \exp\{-e^{-x}\}, \forall\ \ x \tag{4.53}$$

证明: 见*Liu, Zou and Zhang (2008)*.

第五章　一个实际例子的应用

§5.1　美国黄石国家公园喷泉的资料

我们比较感兴趣的一个例子就是美国黄石国家公园的喷泉。如果能找出每次喷泉持续的时间与两次喷泉的间隔时间之间的关系，这样就可以用来预测下次喷泉发生的时间。该例子曾被Gombay and Horváth (1994)研究过。下面先给出1980年十月份美国国家黄石公园发生的270次喷泉的数据见Sanford,W (2005)。表5.1 具体地描述了这些数据。表格5.1给出了喷泉持续的时间(单位：秒)与间隔时间(单位：分)的散点图。表格5.1中“持续”表示一次喷泉所持续的时间,“间隔”表示两次喷泉所间隔的时间。

表 5.1: 美国国家黄石公园1980年十月270次喷泉.

持续	间隔	持续	间隔	持续	间隔	持续	间隔
216	79	124	65	122	53	262	77
108	54	282	73	224	79	231	84
200	74	242	82	254	81	116	49
137	62	118	56	134	69	270	83
272	85	270	79	272	82	143	71
173	55	240	71	289	77	282	80
282	88	119	62	260	76	112	49
216	85	304	76	119	59	230	75
117	51	121	60	278	80	205	64
261	85	274	78	121	49	254	76
110	54	233	76	306	96	144	53
235	84	216	83	108	53	288	94
252	78	248	75	302	77	120	55
105	47	260	82	240	77	282	83
249	76	246	70	144	65	112	50
105	62	244	73	214	71	105	54
130	52	158	65	276	81	256	82
288	84	296	88	244	70	269	75

续表 5.1

持续	间隔	持续	间隔	持续	间隔	持续	间隔
96	52	237	76	270	81	240	78
255	79	271	80	245	93	247	79
108	51	130	48	108	53	245	78
100	47	240	86	238	89	256	78
207	78	132	60	132	45	235	70
184	69	260	90	249	86	273	79
272	74	112	50	120	58	245	70
216	83	289	78	230	78	145	54
118	55	110	63	210	66	251	86
245	76	258	72	275	76	133	50
231	78	280	84	142	63	267	90
266	79	225	75	300	88	113	54
258	73	112	51	116	52	111	54
268	77	294	82	277	93	257	77
202	66	149	62	115	49	237	79
242	80	262	88	125	57	140	64
230	74	126	49	275	77	249	75
121	52	270	83	200	68	141	47
112	48	243	81	250	81	296	86
290	80	112	47	260	81	174	63
110	59	282	84	270	73	275	85
287	90	107	52	145	50	230	82
261	80	291	86	240	85	125	57
113	58	221	81	250	74	262	82
274	84	284	75	113	55	128	67
105	58	294	89	275	77	261	74
272	73	265	79	255	83	132	54
199	83	102	69	226	83	267	83
230	64	278	81	122	51	214	73
126	53	139	50	266	78	270	73
278	82	276	85	246	84	249	88
120	59	109	51	110	46	229	80
288	75	265	87	265	83	235	71

续表 5.1

持续	间隔	持续	间隔	持续	间隔	持续	间隔
283	90	157	63	131	55	267	83
110	54	244	67	288	81	120	56
290	80	255	77	110	57	257	79
104	54	118	56	288	76	286	78
293	83	276	88	245	84	272	84
223	71	226	81	238	77	111	58
100	64	270	82	254	81	255	83
274	77	136	55	210	87	119	43
259	81	279	90	262	77	135	60
134	59	112	45	135	51	285	75
270	84	250	83	280	78	247	81
105	48	168	56	126	60	129	46
288	82	260	89	261	82	265	90
109	60	110	46	248	91	109	46
264	92	263	82	112	53	268	74
250	78	113	51	276	78		
282	78	296	86	107	46		

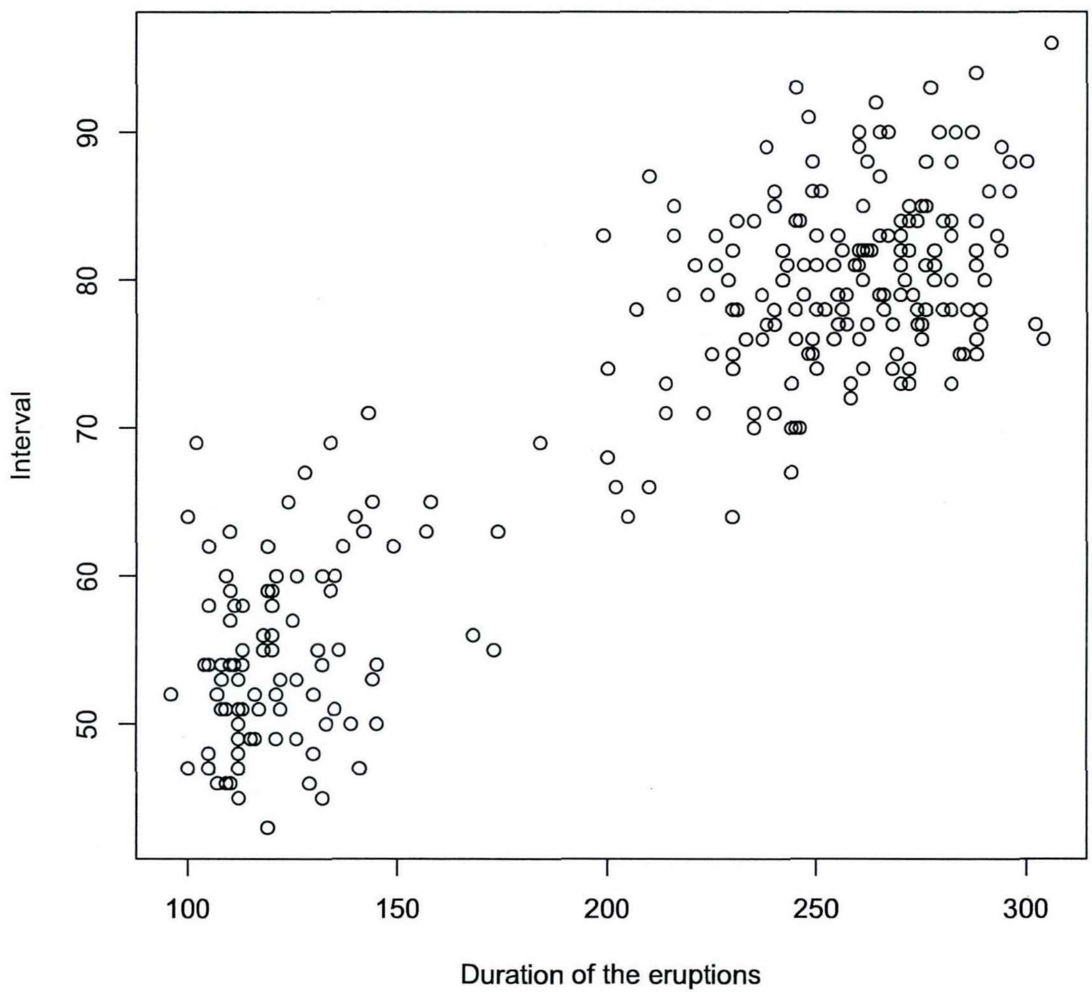

图 5.1 美国黄石国家公园1980年10月发生的270次喷泉的持续时间和间隔时间的散点图

从图5.1喷泉的持续时间与间隔时间的散点图可以看出，它们之间的回归关系可以分为两段，也即持续时间与相隔时间的回归关系至少存在一个变点。我们在§5.2给出具体的分析。

§5.2 变点分析

这一节用§2.4中提到的Gombay and Horváth (1994)基于回归残差的方

法，§4.1中Liu and Qian (2010)中特设经验似然方法和§4.3中改进后的经验似然方法对§5.1中的美国国家黄石公园喷泉的数据进行分析。

Gombay and Horváth (1994)在随机误差服从正态分布的假设下，利用回归残差构造检验统计量，由定理2.4.3知，该统计量的渐近分布与随机误差的方差σ^2有关，而σ未知。Gombay and Horvath (1994)指出，若(2.27)式中的$p \geq 4$，用σ^2的无偏估计$\hat{\sigma}_n^2$代替σ^2，则在零假设下定理2.4.3仍然成立。所以对于美国黄石国家公园的喷泉，Gombay and Horvath (1994)采用统计量$Z_{270}(1,270)/(\hat{\sigma}_{270})$，得到P-值为0.17。

Liu and Qian (2010)先利用最小二乘估计得到回归系数的估计值，然后将其交叉带入逐段回归中，得到交换后的残差。由于在零假设残差的均值为0，接着利用经验似然的方法得到检验统计量。因为参数的估计值并非由经验似然的方法得到，所以§4.1称其为特设经验似然方法。我们用Liu and Qian (2010)中提出的特设经验似然方法分析美国黄石国家公园喷泉，得到的P-值为0.07。

利用改进后的经验似然方法分析美国黄石国家公园喷泉，得到的P-值为0.05。

正如前面，从散点图5.1可以看出，美国黄石国家公园的喷泉持续时间和间隔时间的回归关系可以分为两段，因此我们倾向于认为实际数据中有变点，基于这样的认识和前面得到的结论，我们认为Liu and Qian (2010)的方法优于Gombay and Horváth (1994) 中的方法，而我们提出的改进后的方法又比Liu and Qian (2010)的方法显得更加有效。

第六章　总结与展望

变点问题是统计推断的热点问题之一，它把统计估计和假设检验理论、贝叶斯方法、非贝叶斯方法结合起来，并且它在社会的各个邻域都有广泛的应用，比如在金融学、医学、遗传学、气象学、流行病学和信号过程等方面都有大量的应用背景。本文针对回归模型中的变点问题进行了讨论，总结了参数似然比方法，union-intersection检验，基于回归残差的检验和复发回归的检验以及信息标准等参数方法。我们知道经验似然方法既有非参数方法的稳健性又融合了似然方法的有效性和灵活性，从而相对其它的统计方法有许多突出的优点。

鉴于经验似然方法的优越性，本文着重于用经验似然方法研究回归中的变点问题，探讨了Liu and Qian (2010)提出的特设检验统计量的大样本性质，得到了其在零假设下的渐近分布，并且这一分布与Csörgő and Horváth (1997)中参数似然比检验统计量的渐近分布是一致的。在本章中提出了一种新的改进特设经验似然的方法。该新方法的主要想法是是改进特设经验似然方法中残差的相依性以及充分使用预报变量的信息有效地进行回归分析。通过线性变换使相依的残差成为不相关的量，然后利用回归分析中的正规方程作为经验似然方程的约束方程。因此，提出的新方法被认为是对特设经验似然方法的有效改进。特别指出改进后的方法在计算上与特设经验似然方法是一样的容易计算，并未从本质上加大计算量。蒙特卡洛模拟结果表明新方法极大地改进了Liu and Qian (2010)的特设经验似然方法在有限样本下的有效性。从对一个实际例子的应用中也说明了改进后的方法更有效。此外，本论文给出了检验统计量在零假设下渐近分布的猜想。更进一步地，我们提出了采用自助法(bootstrap)来逼近改进的新方法的P-值。

本文讨论的变点问题均是在假设观测值相互独立的条件下进行的，实际上也可以推广到非独立的情形，比如在自回归-滑动平均的平稳时间序列模型中，在某个时刻，自回归或（和）滑动平均表达式的系数发生了变化。这里我们主要用非贝叶斯的方法讨论了回归中的变点问题，如今也有很多用贝叶斯的方法研究变点问题，比如Ferreira (1975)，Kezim and Abdelli (2004) 和Smith, A. M. F. (1975)等文献中都有讨论。关于§4.3中改进的方法中变点的估计$\hat{k}$的大样本性质还在研究中。目前用变点理论来研究人类基因中copynumber的变化已经成为统计界的热点问题之一。

参考文献

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory* . 267-281.

[2] Anderson, D. (1993). Testing for parameter instability and structural change with unknow chang point. *Econometrica* **61**, 821-856.

[3] Asher, A. and Isreal, Z. (1981a). A new maximum likelihood algorithm for piecewise regression. *J. Am. Stat. Assoc* **76**, 989-987.

[4] Asher, A. and Isreal, Z. (1981b). A maximum likelihood method for piecewise regression models with a continuous dependent variable. *J. Roy. Stat. Soc. ser. C* **30**, 116-124.

[5] Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structure changes. *Econometrica* **66**, 47-78.

[6] Beckman, R. J. and Cook, R. D. (1979). Testing for two-phase regressions. *Technometrics* **21** 65-69.

[7] Benzekri, S, and Brodesu, F. (1991). Asrmptotic results for parametric estimaion in inadequate two phases regression models. *Statistics* **22,** 331-348.

[8] Berman, N., Wong, W., Bhasin, S. and Ipp, E. (1996). Application of segemented regression models for biomedial studies. *Am,J.Physiol* **270**, 723-732.

[9] Bhattacharya, G. K. and Johnson, R. A. (1968). Non-parametrics tests for shift at an unknown time points. *Ann.Math.Sta* **39**, 1731-1743.

[10] Bhattacharya, G. K.(1994). Some aspects of change-point analysis. *IMS Lecture Notes-Monograph Ser* **23**, 28-56

[11] Brodsky, B. E. and Darkhovsky, B. S. (1993). *Nonparametric Methods in Change-point Problems.* Dordrecht,Kluwer Academic Plblisherss.

[12] Brown, R. L., Durbin, J. and Evans, J. M. (1975). Techniques for testing the consistency of regression relationships over time. *Journal of Royal Statistical Society B* **37**, 149-192.

[13] Charlstein, E., Müller, H. G. and Siegmund, D (1994). Change-point Problems. *IMS Lecture Notes* **23**.

[14] Chen, J. and Gupta, A. K. (1997). Testing and locating variance change points hypothesis with application to stock prices. *Journal of the Americal Statistical Association* **92**, 739-747.

[15] Chen, J. and Gupta, A. K. (2000). *Parametric Statical Change Point Analysis.* Springer.

[16] Chen, J. and Gupta, A. K. (2006). Information criterion and change point problem for regular models *Indian Statistical Institute* **68**, 252-282.

[17] Chen, J. and Gupta, A. K. (2001). On change point detection and estimation. *Comm.Stat. Simulat.Comput.* **30**, 665-697.

[18] Chow, G. (1960). Tests of equality between two sets of coefficients in two linear regressions. *Econometrica* **28**, 591-605.

[19] Chow, G. (1960). Testing of equality between two sets of coefficients in two linear regression model. *Stat.Decisions* **28**, 591-605.

[20] Chu, C. S. J. and White, H. (1992). A direct test for changing trending. *J.Business Economic Statist* **10**, 289-299.

[21] Cox, C. (1987). Threshold dose-reponse models in toxicology. *Biometrics* **43**, 523-544.

[22] Csörgő, M. and Horváth, L. (1998). Invariance principles for change-point problems. *J.Multivariate Anal* **27**, 151-168.

[23] Csörgő, M. and Horváth, L. (1997). Limits Theorems in change-point Analysis. Wiley, New York.

[24] Diniz, C. and Brochi, L. (2005). Robust of two-phase regression tests. *REVSTAT Statistical Journal* **3**, 1-18.

[25] Dong, L. (2004). Testing change structure in regression. *Econlmetrics Working Paper EWP0405.*

[26] Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans.* Conf. Series in Appl. Math., No. 38. Philadelphia: SIAM.

[27] Elisabeth, V. (1989). Fitting piecewise linear regression functions to bilogical responses. *J. App. Phy.* **67**, 390-396.

[28] Feder, P. I. (1975a). On asymptotic distribution theory in segmented regression probelem-identified case *Ann. Sta.* **3**, 49-83.

[29] Feder, P. I. (1975b). The log likelihood ratio in segmented regression *Ann,Stat* **3**, 84-97.

[30] Ferreira, P. E. (1975). A Bayesian analysis of switching regression models;knowing number of regimes. *J. Am. Sta. Assoc* **70**, 370-374.

[31] Gbur, E. E. and Dahm, P. F. (1985). Estimation of the linear-linear segmented regression model in the presence of measurement error *Commun.Stat. Theory Methods* **14**, 809-826.

[32] Gibels, I. and Goderniaux, A. C. (2004). Bootstrap test for change-points in noparametric regression. *J. Non. Sta.* **16**, 691-611.

[33] Gombay, E. and Horváth, L. (1994). Limit Theorems for Change in Linear Regression. *J. Multi. Ana* **48**, 43-69.

[34] Hawkins, D. M. (1976). Point estimation of the parameters of piece-wise regression models. *J. Roy. Stat. Soc. Ser. C Appl. Statist* **25**, 51-57.

[35] Hawkins, D. M. (1980). A note on continuous and discontinuous segmented regressions. *Technometrics* **22**, 443-444.

[36] Hawkins, D. L., Gallant. A. R. and Fuller. W. (1986). A simple least squares method for estimating a change in mean. *Comm.Statist.Simulations* **15**, 655-679.

[37] Hawkins, D. L. (1987). A test for a change-point in a parametric model based on a maximal Wald-type statistic. *Sankhya Ser.A* **49**, 368-376.

[38] Hawkins, D. L. (1989). A U-I apporoach to retrospective testing for shift parameters in a linear model. *Comm. Statist-Theory Method* **18**, 3117-3134.

[39] Hinkley, D. V. (1969). Inference zhout the intersection in two-phase regression. *Biometrika* **56** 495-504.

[40] Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika* **56**, 495-504.

[41] Hinkley, D. V. M. (1971). Inference in two-phase regression *J. Am, Stat. Assoc.* **66**, 736-743.

[42] Hudson, D. J. (1966). Fitting segmented curves whose join points have to be estimated. *J. Am, Stat. Assoc.* **61**, 1097-1129.

[43] Huh. J. and Park, B. U. (2004). Detection of a change point with local polynomial fits for the random design case. *Aust. N. Z. J. Stat* **46**, 425-441.

[44] Huskova, M. (1996). Estimation of a change in linear models. *Statistics and Probability Letters* **26**, 13-24.

[45] Jandhyala, V. K. and Minogue, C. D. (1993). Distribution of Bayes-type change-point statistics underpolynomial regeression. *Journal of Statistics Planning and Inference* **37**, 271-290.

[46] James, B. J., James, K. L. and Siegmund, D. (1987). Tests for a change-point. *Biometrika* **74**, 71-84.

[47] James, B. J., James, K. L. and Siegmund, D. (1992). Asymptotic approximations for likelihood ratio tests and confidence regions for a change-point in the mean of a multivariate normal distribution . *Statistica Sinnica* **2**, 69-90.

[48] Kendall, D. G. and Kendall,W. S. (1980). Alignments in two-dimensional random sets of points. *Adv. Appl. Prob.* **12**, 380-424.

[49] Kezim, B. and Abdelli, Z. (2004). A Bayseian analysis of a structural change in the parameters of a time series. *Comm. Stat.Theory and Methods* **33**,1893-1876

[50] Kim, H. J. and Siegmund, D. (1989). The likelihood ratio test for a change point in simple linear regression. *Biometrika* **76**, 409-423.

[51] Kim, H. J. (1993). Two-phase regression with non-homogeneous errors. *Comm. Stat.* **22**, 647-657.

[52] Kim, H. J. and Cai, L. (1993). Robustness of the likelihood ratio for a change in simple linear regression. *J. Am. Stat. Assoc* **88**, 864-871.

[53] Koul. H. L. (2000). Fitting a two phase linear regression model. *Journal of the Indian Statistical Association* **38**, 331-353.

[54] Koul. H. L. and Qian, L. (2002). Asymptotics of maxmimum likelihood estimator in two-phase linear regression model. *Jourmal of Statistical Planning and Inference* **108**, 99-119.

[55] Koul, L. H., Qian, L. and Surgailis, D. (2003). Asymptotics of M-estimators in two-phase linear regression models. *Stachastic Processes and their Applications* **103**, 123-154.

[56] Liu, Y., Zou, C. and Zhang, R. (2008). Empirical likelihood ratio test for a change-point in Linear Regression Models. *Communication in Statistics-Theory and Methods* **37**, 2551-2563.

[57] Liu, Z. and Qian, L. (2010). Changepoint Estimation in a Segmented Linear Regression via Empirical Likelihood. *Communications in Statistics–Simulation and Computation* **39**, 85-100.

[58] Lund, R. and Reeves, J. (2002). Detection of undocumented changepoints: A revision of the two-phase regression model. *Journal of Climate* **1**, 2547-2554.

[59] Loader, C. R. (1996). Change point estimation using nonparametric regression. *The Annals of Statistics* **24**, 1667-1678.

[60] Mahmoud, M. A., Park, P. A., Woodall, W. H. and Hawkins, D. M. (2007). A change point method in linear profile data. *Qua;.Reliab.Eng.Int.* **23**, 247-268.

[61] Muggeo, V. M. R (2003). Estimating regeression modelswith unknown breakpoints. *Statistics in Medicine* **22**, 3055-3071.

[62] Muller, H. G. (1992). Change-point in nonparametric regression analysis. *The Annals of Statistics* **20**, 737-761.

[63] Owen, A. B. (1990). Empirical likelihood confidence regions. *Annals of Statistics* **18**, 90-120.

[64] Owen, A. B. (1991). Empirical likelihood for linear models. *Annals of Statistics* **19**, 1725-1747.

[65] Owen, A. B. (2001). *Empirical Likelihood.* New York: Chapman & Hall/CRC.

[66] Page, E. S. (1954). Continuous Inspection Schemes. *Biometrika* **41**, 100-115.

[67] Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* **42**, 523-527.

[68] Page, E. S. (1957). On the problem in which a change in a parameter occurs ata an unknown point. *Biometrika* **44**, 248-252.

[69] Pastor, R. and Guallar, E. (1998). Use of two-segmented logistic regression to estimate change-pointa in epidemiologic studies. *Am, J. Epidemiol* **148**, 631-642.

[70] Pettit, A. N. (1979). A non-parametric approach to the change-point problem. *Appl. Stat.* **28**, 126-135.

[71] Perron, P. and Vogelsang, T. J. (1992). Nonstationarity and level shifts with an application to purchasing power parity. *J. Bus. Econ. Sta.* **10**, 301-320.

[72] Perron, P. and Vogelsang, T. J. (1992). Testing for a unit root in a time series with a changing mean:corrections and extensions. *J. Bus. Econ. Sta.* **10**, 467-470.

[73] Piegorsch, W. W. and Bailer, A. J. (1997). *Statistics for environmental biology and toxicology.* Chapman and Hall.

[74] Piegorsch, W. W. and Bailer, A. J. (2005). *Analyzing environementao data.* John Wiley & Sons, Ltd.

[75] Piepho, H. P. and Ogutu, J. O. (2003). Inference for the break point in segmented regression with application to longitudinal data *Biometric* **45**, 591-601.

[76] Qian, L. and Yao, Q. (2002). Software project effort estimation using two-phase linear regression models. *Proceedings of The 15th Annual Motorola Software Engineering Symposium.*

[77] Qian, L. and Shao, Q. (2007) *Asymptotic Theory in Probability and Statistics with Applications* .高等教育出版社

[78] Quandt, R. E. (1958). Maximum-likelihood method for estimating the parameters in two separate regression lines that swith ata an unknown point. *J. Am. Stat. Assoc.* **53**, 873-880.

[79] Quandt, R. E. (1960). A likelihood ratio test to test two separate regrssion lines as opposed to the null hypothesis that the data follow only one. *J. Am. Stat. Assoc.* **55**, 324-330.

[80] Sanford, W. (2005).*Applied Linear Regeression 3ed* John Wiley & Sons, Inc., Hoboken, New Jersey.

[81] Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* **6** , 461-464.

[82] Shaban, S. A. (1980). Change point problem and two-phase regression: an annotated bibliography. *Internat. Stat. Rev.* **48**, 83-93.

[83] Siegmund, D. (1985). *Suquential Analysis:Testing and Confidence Intervals.* Springer-Verlag, New York.

[84] Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of Royal Statistical Society B* **47**, 1-52.

[85] Smith, A. M. F. (1975). A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika* **62**, 407-416.

[86] Smith, A. M. F. and Cook, D, G. (1980). Straight lines with a change point: A Bayesian analysian analysis of some renal transplant data. *Appl. Stat.* **29**, 180-189.

[87] Sprent, P. (1961). Some hypotheses concerning two-phase regression lines. *Biometrics* **17**, 634-645.

[88] Thomas, David. R. and Grunkemerier, Gary. L. (1975) Confidence interval estimation of survival probability for censored data. *J. Amer. Statist. Assoc.***70** , 865-871.

[89] Toma, J. D. and Lesperance, M. L. (2003). Piecewise regression: A tool for identifying ecological thresholds. *Ecology* **84**, 2034-2041.

[90] Ulm, K. W. (1991). A statistical method for assessing a threshold in epidemiological studies. *Statistics in mediciline* **10**, 341-349.

[91] Wang, X. L. (2003). Comments on"Detection of Undocuments Changepoints:A Revision of the Two-Phase Regression Model. *J. Clim* **16**, 3383-3386.

[92] Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20**, 365-367.

[93] Wilk, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* **9**, 60-62.

[94] Worsley, K. J. (1983). Test for a two-phase multiple regression. *Technometrics* **25**, 35-42.

[95] Worsley, K. J. (1986). Confidence regions and tests for a chang-point in a sequence of exponential family random variables. *Biometrika* **73**, 91-104.

[96] Wu, Y. (2005). Inference for change point and post change means after a CUSUM test. *Springer.*

[97] Zeileis, A. (2006). Implementing a class of structural change tests:an econometric computing approach. *Comput. Stat. Data Anal* **50**, 2987-3008.

[98] Zhao, H., Wu, X. and Chen, H. (2010). A new algorithm in detecting changepoint in linear regeression models. *2010 3rd International Conference on Biomedical Engineering and Informatics.* 2261-2264.

[99] Zou, C., Liu, Y., Qin, P., and Wang, Z. (2007). Empirical likelihood ratio test for the change-point problem. *Statistics & Probability Letters* **77**, 374-382.

[100] 陈希孺(1991)。变点统计分析简介。数理统计与管理 **12**，1-4。

[101] 谭常春(2007)。变点问题的统计推断及其在金融中的应用。博士学位论文。中国科技大学。

[102] 王松桂，史建红，尹素菊，吴密霞(2003) 线性模型引论。科学出版社.

攻读博士学位期间撰写的论文目录

[1] Zhao, H., Wu, X.and Chen, H. (2010). A new algorithm in detecting change-point in linear regeression models. *2010 3rd International Conference on Biomedical Engineering and Informatics* 2261-2264.

[2] Zhao, H., Wu, X., Zhang, H. and Chen,H. (2010). Estimating the proportion of true null hypotheses in nonparametric exponential mixture model with application to the leukemia gene expression data. *Submitted to Communication in Statistics-Simulaiton and Computation.Current under revising per the editor's invitation.*

[3] 吴小霞，赵华玲，柳福祥 (2010). 多重假设检验问题中一类估计问题的分析。中国科技论文在线精品论文 **3(1)**: 74-78。

[4] Wu, X.and Zhao, H. (2010). Estimating the Proportion of True Null Hypotheses in Multiple Testing from a Parametric Marginal Model. *2010 Intermational Institute of Statistics and Management Engineering Symposium.*

致谢

我的博士求学生涯即将结束，回首这一段难忘的岁月，既有收获成功的喜悦也有经历挫折的失落。在这里我要感谢我的老师们和所有亲朋好友。

首先，衷心的感谢导师陈汉峰教授的淳淳教诲与悉心关怀。陈老师从论文选题到写作等各方面都给予我细心的指导并提出了宝贵的建议。这几年，导师在我的身上花费了大量的心血，在此，我向导师表示我最衷心的感谢！

我要感谢刘禄勤教授。刘老师作为我的硕士导师，在学习中为我打下了坚实的概率统计的基础，同时在生活中都给予了我很大的帮助。

同时我要感谢刘妍岩教授、章逸平教授。感谢他们对我学习给予的关怀和帮助。我从他们指导的讨论班上学到了很多知识。另外，还要感谢高付清教授、胡亦均教授、刘莉老师，邓爱娇老师，丁洁丽老师，冯艳钦老师，王忠海老师，感谢他们给予的帮助。

我还要感谢吴小霞、吴远山、袁中尚、杨青龙、邓世容、陈玉蓉、余吉昌、石小平、马悦、陈娟和所有的同窗好友，在读博期间与他们进行了有益的讨论并得到了他们的帮助和鼓励。

借此机会我还要感谢我的父母和家人，没有他们的关爱和支持，我不可能顺利的完成我的学业。感谢我的丈夫赵波，感谢他无私的支持和无微不至的关怀。

最后，也是非常重要的，我要感谢武汉大学数学与统计学院，感谢她为我们提供的良好的学习环境和科研经费。