

纵向数据分析的回归模型

伍鸥^a, 田欣^b

美国国家心肺血液研究所, 生物统计室

^awuc@nhlbi.nih.gov, ^btianx@nhlbi.nih.gov

余启汾

清华大学, 数学科学系和数学科学中心

kfyu@math.tsinghua.edu.cn

吴密霞

北京工业大学, 应用数理学院

wumixia@bjut.edu.cn

§1 引言

§1.1 纵向研究的目的

在生物医学与流行病学研究中, 通常需要评价治疗、剂量、危险因子或其它生物或环境协变量对感兴趣的结果, 如病情、人群健康状况等随时间改变的效应。由于每个个体的结果和协变量随时间的变化常常可提供重要的科学信息, 因此, 对研究对象在不同时间点进行重复测量, 获得的纵向样本 (longitudinal samples) 通常比经典的截面样本 (cross-sectional samples) 要好。通过结合随机抽样和时间序列观察的特征, 纵向样本的应用大大超出了生物医学与流行病学的范畴, 在经济学、心理学、社会学和自然与社会科学的许多其它领域经常也能见到这类数据。

一般来讲, 在生物医学研究中主要有两种途径获得纵向的观测: (a) 带重复测量的临床试验 (clinical trials); (b) 流行病学研究, 通常被称为观察性队列研究 (ob-

servational cohort studies). 二者的主要区别在于其设计的不同。在临床试验中, 尽管在一些情况下, 出于伦理和逻辑原因, 需要给某些个体非随机化的伴随处理或干预, 但个体的选择, 随机化的处理的类型, 实验期的长度、访问时间和测量方法都可由研究者决定; 而观察性队列研究要比临床试验更为复杂, 因为观察性队列研究的危险因子、处理和测量依赖于研究的参与者, 不受研究者的控制。

在纵向临床试验中, 一个主要研究目的是相对于安慰剂或标准处理而言, 评价试验处理在试验期内对某些健康状态指标等主要结果随时间改变的效应。在许多情形下, 处理期结束后还会增加跟踪期, 除了重复测量健康结果外, 另外一些时间-事件变量, 如住院治疗或死亡时间等, 也可以被作为主要结果被记录。在具体的分析中, 试验期的定义是由分析目标决定的。例如, 如果分析目标是评价治疗期内某治疗对健康指标随时间改变的效应, 则将治疗期当作试验期是合理的。另外, 如果对治疗期外的某些时间-事件变量也感兴趣, 则试验期就包括了治疗期和跟踪期两部分。所研究的处理效应可以通过结果变量的条件均值, 条件分布或条件分位数来评价。虽然基于结果变量条件均值的回归模型是用来分析纵向临床试验的极为流行的方法, 但当结果变量的分布是高度偏的或者其分布不容易由正态分布近似时, 基于结果变量条件分布或条件分位数的回归方法就常常为很重要的统计方法。试验的目的除了评价随机化的处理外, 第二个重要的目的是评价伴随干预或其它协变量对结果变量随时间变化趋势的效应。除了随机化处理外, 还包含协变量的回归分析常常被用来在评价处理与协变量间的交互效应或识别病人总体中对试验治疗方案有效的病人。

纵向观察性队列研究的主要目的不是对随机化的试验处理作检验, 而是评价各个协变量, 如人口和环境因子, 与感兴趣的结果变量之间的潜在的关联关系, 及其随时间变化的趋势。因此, 观测性队列研究常常被用来作数据挖掘, 更多具体科学问题的假设和检验则需做进一步合适设计下的临床试验。然而, 由于变量随时间

被重复测量, 因此纵向观察性队列研究对了解某种疾病在总体范围和某些个体子集内的自然进展情况很有帮助。实际上, 观察性队列研究通常包含了大量个体的充足的不同时间点的重复测量, 因此从该研究中可得到有相当统计精度的新奇结果。类似于纵向临床试验, 在观察性队列研究中分析数据的统计方法的选择依赖于研究目的, 可以包括基于条件均值的回归模型, 也可以包括基于条件分布的回归模型。

§1.2 纵向数据结构

为了勾画典型的纵向数据, 定义 t 为时间的实值变量, $Y(t)$ 为实值结果变量, $\mathbf{X}(t) = (X^{(0)}(t), \dots, X^{(K)}(t))^T$, $K \geq 1$, 为时间 t 处的 R^{K+1} 空间上的协变量向量。时间变量 t 依赖于原点的选择, 不必一定为非负值。作为一般方法的一部分, 采用回归模型进行统计分析的主要目的通常是建立模型和决定 $\{t, \mathbf{X}(t)\}$ 对总体均值、个体与总体均值的偏差、以及 $Y(t)$ 的条件分布或条件分位数的效应。对于 n 个随机选择的个体和每个个体在不同时间点的重复测量, $\{Y(t), t, \mathbf{X}(t)\}$ 的纵向样本可记为 $\{(Y_{ij}, t_{ij}, \mathbf{X}_{ij}); i = 1, \dots, n, j = 1, \dots, n_i\}$, 其中 t_{ij} 是第 i 个体第 j 测量的时间点, Y_{ij} 和 $\mathbf{X}_{ij} = (X_{ij}^{(0)}, \dots, X_{ij}^{(D)})^T$ 分别是第 i 个体在 t_{ij} 时间点的观察结果和协变量向量, n_i 为第 i 个体的重复测量数。于是试验测量总数为 $N = \sum_{i=1}^n n_i$ 。与经典截面研究中的独立同分布 (independent identically distributed, 简称 i.i.d.) 样本不同, 尽管假设个体间的测量值是独立的, 但每一个体内的测量值间可能是相关的。

如果所有个体在设计的一组时间点上都有测量值, 即对某个 $m \geq 1$ 有 $n_i = m$, 且对所有 $j = 1, \dots, m$, 有 $t_{1j} = \dots = t_{nj}$, 则纵向样本为平衡设计。如果不同个体的设计时间 $\mathbf{t}_i = \{t_{ij}; 1 \leq j \leq n_i\}$ 也不同则为非平衡设计。事实上, 当平衡设计出现缺失值或时间设计点有随机变动时均会导致非平衡设计。在长期的临床试验或流行病学的研究中, 所研究的个体常常被指定了一组预先规定的“设计访问时间点”, 但由于各种原因会出现缺失或改变实际的受访时间, 从而导致真正的访问时间点与指定的访问时间点不同。对于一个很好控制的纵向临床试验, 观测到平衡的纵

向数据是有可能的，因为随机化的处理和访问的时间可由研究者决定。然后，观察性队列研究中的许多因素不受研究者的控制，因此大多数的观察性队列研究是非平衡的纵向设计。

§1.3 纵向研究实例

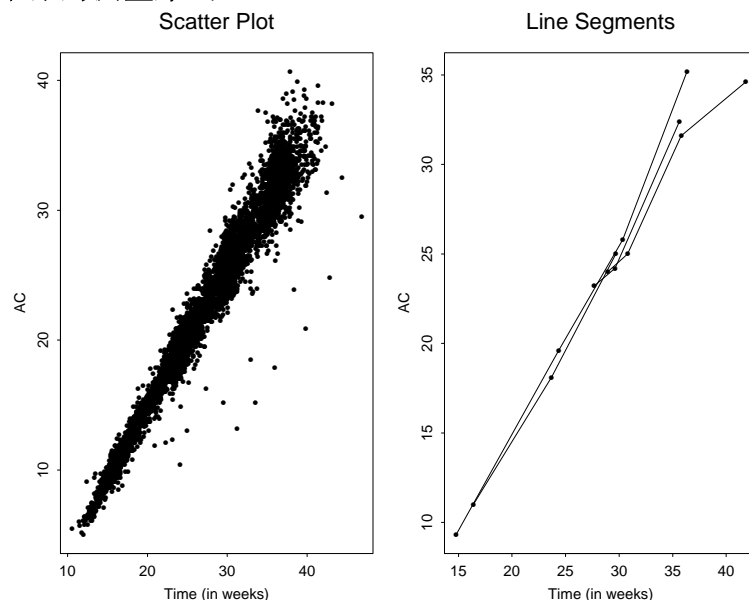
下面采用三个流行病学实例和一个纵向临床试验来阐明纵向研究的一些典型特征和研究目的。尽管这些例子在设计上有些相似，但其数据结构和研究目标却不相同。

例 1. 阿拉巴马州小孕龄 (ASGA) 研究

ASGA (Alabama Small-for-Gestational-Age) 研究是关于危险因子与宫内生长迟缓的一个前瞻性研究，研究采用超声重复测量了 1475 名孕妇的胎儿，规定所有孕妇在大约怀孕 17、25、31、36 周接受测量。但是她们并没有按设计的时间点接受随访，而是分散在孕期的第 12~43 周，怀孕期内孕妇接受访问次数为 1~7 次。由于不是所有的个体在相同的时间设计点上都有测量值，故该数据为非平衡设计。影响胎儿发育的有关协变量包括：吸烟、饮酒、吸毒等母亲的行为因素，和母亲的孕前体重、身高、体块指数 (body mass index，简称为 BMI) 以及不同孕期的胎盘厚度等母亲的人体测量值。在这些协变量中，母亲的人体测量值为时变 (time-dependent) 变量，但诸如母亲的行为因素等其它协变变量，可能为时变变量，也可能为时不变 (time-invariant) 变量，这取决于分析中这些变量的定义。胎儿发育的结果变量自然都属于时变变量，它包括了胎儿腹围、顶骨间径、股骨长等超声测量值。

图 1 显示了不同孕龄 (周) 的胎儿腹围 (厘米)，为了了解个体的重复测量值的时间趋势，采用线段表示 3 个随机选择个体的测量序列。由图可见胎儿的生长呈直线上升趋势。但是，目前尚未发现关于该总体的线性生长模型或协变量对胎儿生长影响的其它统计模型的拟合优度检验的相关研究。然而一个统计分析应该有两个目

图 1 胎儿腹围 (厘米) 与孕龄 (周) 的关系. 左图: 个体测量值; 右图: 随机选择的 3 个个体的测量序列



的: (1) 建立适当的统计模型, 弄清协变量对结果变量的影响; (2) 基于选择的统计模型, 给出合理估计和推断协变量效应的方法。

例 2. 巴尔的摩多中心爱滋病队列研究 (BMACS)

资料来自 1984 至 1991 年间感染人体免疫缺乏病毒 (HIV) 的 400 名同性恋男子的多中心爱滋病队列研究 (multicenter AIDS cohort study, 简称为 MACS), 因为 CD4 细胞 (辅助淋巴细胞) 决定了免疫功能的大小, 所以本研究的一个重要内容是评价危险因子 (如吸烟、药物滥用等) 和健康状况 (如感染前 CD4 细胞水平等) 对感染后 CD4 细胞的损耗百分率 (淋巴细胞的 CD4 百分率) 的影响。尽管规定所有个体每半年接受一次访问与测量, 但实际上个体并没有精确按规定时间接受访问, 而且研究期间感染 HIV 是随机发生的, 故该研究为非平衡设计, 个体重复测量数在 1~14 的范围内, 中位数为 6。与上例的 ASGA 研究相比, 该资料的研究个体数较少, 但

重复测量数的范围更宽。该数据中感兴趣的协变量也包括时间依赖变量和时间固定变量两种情况。关于这类研究的设计和其在医学上的重要性详见 Kaslow 等 (1987)。

例 3. 国民发育和健康研究 (NGHS)

这是一个 3 中心总体的队列研究，共研究了 2379 名女孩，其中 1213 (51%) 非裔美国人，1166 (49%) 白种人，她们是在 9 或 10 岁注册的，然后每年记录一次她们的身高、体重、血压以及其它体征的和实验室的测量值，一直记录到她们 19 岁为止。该研究的主要目标是评价儿童期和青少年期肥胖对心血管风险的影响模式。三个中心分别为加利福尼亚大学伯克利分校，辛辛那提大学辛辛那提儿童医药研究中心，以及马里兰罗克维尔的 Westat 公司，它们之所以被选择是因为人口调查数据中它们包含了各个种族的家庭收入和父母教育程度较广分布。本研究基于标准草案 (NGHSRG, 1992) 收集了从 1986 年到 1997 年在临床中心或家庭随访所观测的数据，该草案包含了研究目标、合理性、研究设计和测量的方法。

目前关于该研究的主要结果有如 Daniels 等 (1998), Kimm 等 (2000, 2001, 2002), Thompson 等 (2007) 以及 Obarzanek 等 (2010)。在这些结果中，Daniels 等 (1998) 研究了种族、身高、体块指数 (body mass index, 简称为 BMI)(kg/m^2) 对舒张压 (systolic blood pressure, 简称为 SBP) 和收缩压 (diastolic blood pressure, 简称为 DBP) 的纵向的平均效应。Kimm 等 (2000, 2001, 2002) 研究了种族对肥胖的效应，以及体育活动改变和 BMI 的关系。Thompson 等 (2007) 研究了儿童期肥胖与高血压、不健康的低密度胆固醇 (low-density lipoprotein cholesterol, 简称为 LDL) 和甘油三酯 (triglyceride, 简称为 TG) 水平等心血管病风险因子间的关联关系。Obarzanek 等 (2010) 研究了种族、肥胖和社会经济与营养变量的范围对高血压和前期高血压 (定义见 NHBPEP 工作组, 2004) 的患病率和发病率的影响。这些作者所采用的统计方法是通过它们的条件均值或者由某些门限值规定的不健康风险水平的概率来概括儿童期肥胖和其

它变量对心血管风险变量的影响。尽管这些结果展现了儿童期肥胖对心血管风险有害影响，但如果采用其它不健康风险水平，儿童期肥胖和心血管风险因子是否仍有类似的关联关系，这个问题还需进一步研究。令人期待的是有一个灵活的基于条件分布的回归方法，能够刻画协变量对风险变量的整个分布的效应。该研究就是促使基于条件分布和条件分位数的纵向估计与推断方法发展的一个典型例子。

例 4. 提高冠心病康复 (ENRICHD) 的研究

该研究是一个随机化的临床试验，来评估相对于通常心脏护理 (usual cardiologic care, 简写为 UC) 而言，认知行为疗法 (cognitive behavior therapy, 简写为 CBT) 对患急性心肌梗塞后患抑郁症和 / 或低知觉的社会支持的 2481 病人的生存和抑郁严重性的功效。关于该研究的试验设计、研究目标以及试验的重要结果详见 ENRICHD (2001, 2003)。该研究的首要目的是针对犯心肌梗塞 (myocardial infarction, 简写为 MI) 后的 28 天内注册病人，判断是否可以通过用治疗抑郁症和低知觉的社会支持症的认知行为疗法，同时根据需要辅助用选择性 5- 羟色胺再摄取抑制剂 (selective serotonin reuptake inhibitor, 简写为 SSRI) 或其它抗抑郁剂来降低其死亡率和心肌梗塞复发率。认知行为疗法是在心肌梗塞后 17 天的中间开始的，在 6 个月内组织的 11 次个人会话中，要求参加 5-6 次。抑郁严重性是由汉密顿抑郁量表 (Hamilton Rating Scale for Depression, 简写为 HRSD) 和贝克抑郁测量表 (Beck Depression Inventory, 简写为 BDI) 测量的，HRSD 和 BDI 越高表明抑郁越严重。当可行时，即针对用 SSRIs 等抗抑郁剂药物疗法五周后 HRSD 得分高于 24 或 BDI 得分的降低比率低于 50% 的病人，组疗法开始实施。由于抗抑郁剂可根据病人或他的主治医生的要求给病人服用，因此抗抑郁剂可看作是试验设计规定的随机化 CBT 心理治疗或通常心脏护理的伴随治疗。主要测量结果包括死亡或 MI 复发的复合主要终点，和包含抑郁者六个月内的 HRSD 和 BDI 的改变量或低知觉的社会支持者的 ENRICHD 的社会支持测定得分 (ENRICHD, 2001 和 2003) 的次要结果。

除了试验的主要目的外，另一个重要问题是是否使用抗抑郁剂对减少该病人总体的死亡率与 MI 复发率以及降低 BDI 得分有好处。为了评价使用抗抑郁药物的药物疗法的效应，Taylor 等 (2005) 比较了该试验中 1834 个抑郁病人的死亡生存率和心血管疾病发病率和死亡率，发现使用抗抑郁药物 SSRI 似乎可以减少随后的心血管疾病率和死亡率。Bang 和 Robins (2005) 也分析了该数据，不过他们仅做了截面分析。通过将病人使用药物疗法的开始时间当作特定个体的“变点时间”，在变点前后分别采用不同的模型来评价了药物疗法随时间对病人的 BDI 得分的效应，Wu, Tian 和 Bang (2008) 与 Wu, Tian 和 Jiang (2011) 表明使用 SSRIs 对降低病人的抑郁程度（即 BDI 得分）有好处。

在第 5 节整节，这个研究将被作为一个推动带结果自适应协变量的回归模型及其估计与推断方法发展的例子。在这种情形，由于协变量是自适应依赖于前面的结果轨迹，因此通常所使用的参数或非参混合效应模型可能被错误指定，从而导出偏的和错误的结论。

§1.4 纵向数据的回归模型

一般来说，合理的纵向分析应至少达到 3 个目的：

- (1) 所考虑的模型必须能合理描述数据的实际关联性，且充分简单、灵活，便于实际应用。在生物医学和流行病学研究中，一个合理的模型应该具有清晰、有意义的生物学解释和简单的数学结构。
- (2) 应具有适当、有效的模型诊断方法来评价统计模型对所给定数据的有效性，两个重要的诊断方法为置信区间和统计学假设检验。
- (3) 所采用的方法必须具有理论和实际意义的，可处理个体内相关的数据。实际上，由于个体内相关结构通常是未知的，且难以被估计，因此人们更乐意采用不依赖于特殊相关结构的建模方法来做估计与推断。

§1.4.1 参数模型

自然, 在纵向分析中最常用的方法是参数回归模型, 如广义线性和非线性混合效应模型。在这些模型中最简单的例子是边缘 / 际线性模型

$$Y_{ij} = \sum_{l=0}^K \beta_l X_{ij}^{(l)} + \epsilon_i(t_{ij}), \quad (1.1)$$

其中 β_0, \dots, β_K 为线性常数系数, 描述相应协变量的效应, $\epsilon_i(t)$ 是均值为 0 的随机过程 $\epsilon(t)$ 在时间点 t 处的实现, 且 \mathbf{X}_{ij} 与 $\epsilon_i(t_{ij})$ 独立。类似于含固定截距项的所有回归模型, 令 $X^{(0)} \equiv 1$ 就产生了一个基线系数 β_0 , 它表示所有协变量 $X^{(l)}(t)$ 为 0 时 $Y(t)$ 的均值。误差过程的一个常见的特殊情形是令 $\epsilon(t)$ 是均值为 0 的高斯平稳过程 (Gaussian stationary process)。尽管对于许多实际问题, 模型 (1.1) 显得过于简单, 但对该模型的扩展可导出许多有用的模型, 构成了纵向分析中所使用的模型的主要部分。

基于参数模型的估计和推断方法包括加权最小二乘法 (weighted least squares)、拟似然法 (quasi-likelihood) 和广义估计方程 (generalized estimating equations), 这些方法在文献中已被广泛地研究, 如 laird 和 Ware(1982)、Pantula 和 Pollock(1985)、Ware(1985)、Liang 和 Zeger(1986)、Diggle(1988)、Zeger, Liang 和 Albert(1988)、Jones and Ackerson(1990)、Jones 和 Boadi-Boteng(1991)、Diggle, Liang 和 Zeger(1994)、Davidian 和 Giltinan(1995)、Vonesh 和 Chinchilli(1997)、Verbeke 和 Molenberghs(1997)、Verbeke 和 Molenberghs (2000) 以及 Diggle, Heagerty, Liang 和 Zeger (2002)。参数模型的主要优点是简单、易解释, 可用 SAS、S-Plus 等流行统计软件包中的程序进行计算。但如果这些参数模型被错误指定, 则可能导致错误的结论。因此, 至少在探索性研究中, 有必要放宽一些参数约束。

§1.4.2 半参数模型

Zeger 和 Diggle(1994) 以及 Moyeed 和 Diggle(1994) 所研究的一个很有用的半参数模型为部分线性模型

$$Y_{ij} = \beta_0(t_{ij}) + \sum_{l=1}^K \beta_l X_{ij}^{(l)} + \epsilon_i(t_{ij}), \quad (1.2)$$

这里 $\beta_0(t)$ 为 t 的未知光滑函数, β_l 为未知常数, $\epsilon_i(t)$ 和 \mathbf{X}_{ij} 如模型 (1.1) 中所定义。因为 $\beta_0(t)$ 不是一个常数, 允许随时间 t 改变, 所以与边缘线性模型 (1.1) 相比该模型更为通用; 此外, 模型 (1.2) 包括了线性项 $X_{ij}^{(l)}$, 与 Hart 和 Wehrly(1986)、Altman(1990)、Hart(1991) 以及 Rice 和 Silverman(1991) 中仅涉及 (t_{ij}, Y_{ij}) 的非参数回归模型相比, 该模型同样更为通用。然而, 由于模型 (1.2) 通过固定的线性系数描述 $X_{ij}^{(l)}$ 对 Y_{ij} 的效应, 所以该模型在一定程度上仍然为了数学上的方便, 而不是数据的实际关联形式。例如, 我们没有任何理由希望整个研究期间, ASGA 研究中的母亲危险因子对胎儿发育的影响, 或 HIV/CD4 损耗数据中的吸烟和感染前 CD4 水平对感染后 CD4 细胞百分率的影响呈固定的线性趋势。因此, 在许多情况下需要进一步扩展模型 (1.2)。

§1.4.3 非参数模型

尽管原则上可以通过完全非参数高维函数为 $\{Y_{ij}, t_{ij}, \mathbf{X}_{ij}\}$ 建模, 但该方法由于众所周知的 “维数祸根” (curse of dimensionality) 问题常常在实际应用中不可行。此外, 由高维非参数拟合所获得的结果通常也很难解释。这些问题驱使我们考虑具有一定意义结构的非参数模型。其中最为重要的一类结构非参数回归模型为变系数模型, 可表示为

$$Y_{ij} = \mathbf{X}_{ij}^T \beta(t_{ij}) + \epsilon_i(t_{ij}), \quad (1.3)$$

这里 $\beta(t) = (\beta_0(t), \dots, \beta_K(t))^T$ 为 $(K+1)$ 的 t 的光滑函数向量, $\epsilon_i(t)$ 和 \mathbf{X}_{ij} 如 (1.1) 中所定义。因为对于每个固定的时间点 t , 模型 (1.3) 给出了 $Y(t)$ 与 $\mathbf{X}(t)$ 间的一

个线性模型, 所以线性系数 $\beta_l(t)$, $l = 0, \dots, K$, 可以与模型 (1.2) 有相同的解释。令 $X_{ij}^{(0)} \equiv 1$, 则 $\beta_0(t)$ 为时间 t 处的截距。另一方面, 因为所有线性系数可随 t 改变, 所以在不同的时间点将得到不同的线性模型, 模型 (1.3) 为 Hastie 和 Tibshirani(1993) 中所讨论的一般变系数模型的一个特例。

关于这类模型的估计与推断问题, 文献中已有大量的研究, 已提出了许多估计 $\beta(t)$ 的不同光滑方法, 如一般最小二乘局部多项式 (ordinary least squares local polynomials)、惩罚最小二乘法 (penalized least squares)、两步与逐元法 (two-step and componentwise methods) 以及基近似法 (basis approximation approaches)。在实践中这些方法各有其优缺点。在第 3.2 与第 3.3 节我们将给出这些估计与推断方法的一个概述, 并将在第 3.4 节阐述了这些方法的应用。

§1.4.4 条件分布的回归模型

以上所介绍的回归模型大部分都是针对重复测量结果变量的条件均值建立的。而当响应变量不服从正态分布时, 在纵向分析中常常采用基于条件分布的回归模型。分析非高斯数据的著名方法, 如广义线性模型 (如, Molenberghs 和 Verbeke, 2005) 常常可被用来分析离散的结果变量或由某些指定的门限值离散化的连续结果变量, 此外, 在一些情况, 当结果变量的条件分布为非高斯分布时, 存在一个已知的变换, 如 Box-Cox 变换, 能够使得变化后的变量的条件分布近似为正态的, 然而在许多情形, 没有合适的门限值或变换可供结果变量考虑, 此时更好的方法就是直接针对条件分布进行建模。

基于两个核光滑方法: 局部 logistic 估计方法和调整的 Nadaraya-Watson 方法, Hall, Wolff 和 Yao (1999) 针对独立同分布的数据和不带任何结构的时间序列数据, 研究了条件分布函数的非参数估计问题。由于当协变量个数较大时, 基于无结构非参数模型的估计方法数值计算很不稳定, 而且相应的结果也很难被解释, Wu, Tian 和 Yu (2010) 通过扩展 Cheng, Wei 和 Ying (1995, 1997) 所研究的线性变换建模结构

到纵向数据, 提出了一种降维的方法, 并采用两步局部多项式方法研究了时变线性变换模型 (time-varying linear transformation model) 下协变量效应和条件分布的估计问题. 结果表明当基于条件均值的回归模型不能充分满足研究目标需要时, Cheng, Wei 和 Ying (1995, 1997) 提出的线性变换时变模型给出了更为灵活、方便的结构非参数体系来刻画结果变量的条件分布. 将这个基于条件分布的结构非参数建模方法应用到例 3 中的 NGHS 数据, 由第 4 节所述的相关统计估计、预测以及推断方法得到了关于儿童和青少年血压的时间趋势的有益见解.

§2 参数、半参数方法概述

§2.1 线性混合效应模型

作为研究协变量对纵向结果变量的效应的一种流行的回归方法, 混合效应模型一般有两个目的: (a) 描述处理与其它因素对平均响应剖面的效应, (b) 描述个体间响应剖面的差异. 服务于第一个目的的回归模型一般被归类为边际模型 (marginal model) 或总体均值模型 (population average model); 服务于第二个目的的回归模型为随机效应模型 (random effects model) 或个体指定模型 (subject specified model)(如 Zeger, Liang 和 Albert, 1988). 而混合效应模型则是将边际与随机效应结合起来, 特别当边际与随机效应可加且呈线性关系时, 便得到线性混合效应模型 (linear mixed effects model)。

为方便, 采用矩阵描述模型. 令 $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$ 为第 i 个体的 $[n_i \times 1]$ 反应向量, $t_i = (t_{i1}, \dots, t_{in_i})^T$ 为设计的个体时间点, \mathbf{X}_i 为相应的 $[n_i \times (k+1)]$ 协变量矩阵, 其中第 j 行是 $(1, X_{ij}^{(1)}, \dots, X_{ij}^{(k)})$, $j = 1, \dots, n_i$. 假设模型 (1.1) 的误差项 $\epsilon_i(t)$ 服从均值向量为 0, 协方差阵为 $\mathbf{V}_i(t_i)$ 的正态分布, 则 \mathbf{Y}_i 为独立随机高斯向量

$$\mathbf{Y}_i \sim \mathbf{N}(\mathbf{X}_i\beta, \mathbf{V}_i(t_i)), \quad (2.1)$$

其中 $\beta = (\beta_0, \dots, \beta_k)^T$, β_j 如 (1.1) 中所定义, $\mathbf{N}(\mathbf{a}, \mathbf{b})$ 表示均数向量为 \mathbf{a} , 协方差阵为 \mathbf{b} 的多元正态分布。注意到 (2.1) 是通过 $\mathbf{X}_i\beta$ 给出了给定 \mathbf{X}_i 下 \mathbf{Y}_i 的条件均值, 所以它是一个边际模型。

模型 (2.1) 的协方差结构通常受随机效应、序列相关和测量误差这三个因素的影响。随机效应刻画了总体中个体间的随机变动。尤其当协变量线性影响响应变量时, 其中部分线性系数可以因人而异; 序列相关是同一个体在不同时间点的测量值间相关的结果, 生物医学研究中, 这种相关一般为正的, 且随着测量的时间间隔增大而变小; 最后, 测量误差通常被假设为正态的且个体内和个体间都相互独立, 它是由测量过程或个体内的随机变动产生的。

对于每一个体 i , 假设在测量时间点 t_{ij} 的解释变量 \mathbf{U}_{ij} 为 $[r \times 1]$ 向量, 它与原协变量向量 \mathbf{X}_{ij} 可以相同, 也可以不相同。假设随机效应、序列相关和测量误差三个成分是线性可加, 则 $\epsilon_i(t_{ij})$ 可表示为

$$\epsilon_i(t_{ij}) = \mathbf{U}_{ij}^T \mathbf{b}_i + W_i(t_{ij}) + Z_{ij}, \quad (2.2)$$

其中 \mathbf{b}_i 为 $[r \times 1]$ 随机向量, 服从多元正态分布 $\mathbf{N}(\mathbf{0}, \mathbf{D})$, \mathbf{D} 为 $[r \times r]$ 协方差阵, 其第 (p, q) 元素为 $d_{pq} = d_{qp}$, $W_i(t_{ij})$, $i = 1, \dots, n$, 是均值为 0 的独立高斯过程, 其在时间点 t_{ij_1} 和 t_{ij_2} 出对应的协方差为 $\rho_W(t_{ij_1}, t_{ij_2})$, Z_{ij} , $i = 1, \dots, n$, $j = 1, \dots, n_i$, 为独立同分布随机变量, 并服从 $N(0, \tau^2)$ 分布。记 $\delta_i(t_{ij}) = W_i(t_{ij}) + Z_{ij}$, $\delta_i = (\delta_i(t_{i1}), \dots, \delta_i(t_{in_i}))^T$, \mathbf{U}_i 为 $[n_i \times r]$ 矩阵, 其第 j 行为 \mathbf{U}_{ij}^T , 则由 (2.1) 与 (2.2) 可得到 Laird 和 Ware(1982) 所给出的线性混合效应模型

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{U}_i\mathbf{b}_i + \delta_i. \quad (2.3)$$

由于边际效应 β 表示 \mathbf{X}_i 对总体平均响应剖面的影响, 而 \mathbf{b}_i 描述了给定解释变量 \mathbf{U}_i 的条件下第 i 个体与总体的差异, 因此, 在给定 \mathbf{X}_i 和 \mathbf{U}_i 的条件下, (2.3) 表明

\mathbf{Y}_i , $i = 1, \dots, n$, 为独立的高斯向量, 且

$$\mathbf{Y}_i \sim \mathbf{N}(\mathbf{X}_i\beta, \mathbf{U}_i\mathbf{D}\mathbf{U}_i^T + \mathbf{P}_i + \tau^2\mathbf{I}_i), \quad (2.4)$$

其中, \mathbf{P}_i 为 $[n_i \times n_i]$ 协方差矩阵, 第 (j_1, j_2) 元素为 $\rho_W(t_{ij_1}, t_{ij_2})$, \mathbf{I}_i 为 $[n_i \times n_i]$ 单位矩阵。

由模型 (2.2) 的协方差结构可派生出许多特例。如针对独立的截面数据 (或独立同分布数据) 的经典线性模型就是 (2.2) 中的 $\epsilon_i(t_{ij})$ 仅受测量误差 Z_{ij} 影响的特例。当模型中既无随机效应, 也无测量误差时, 误差项就是纯序列相关 $\epsilon_i(t_{ij}) = W_i(t_{ij})$, 进一步, 如果 $W_i(t_{ij})$ 来自均值为 0 的平稳高斯过程, 则 $\epsilon_i(t_{ij_1})$ 与 $\epsilon_i(t_{ij_2})$ 的协方差, 即为 Y_{ij_1} 与 Y_{ij_2} 的协方差, 可写为

$$\text{Cov}(\epsilon_i(t_{ij_1}), \epsilon_i(t_{ij_2})) = \sigma^2 \rho(|t_{ij_1} - t_{ij_2}|), \quad (2.5)$$

其中 σ 是正的常数。 $\rho(\cdot)$ 为连续性函数, 可供选择的 $\rho(\cdot)$ 有: 指数相关 $\rho(s) = \exp(-as)$ (其中 a 为大于 0 的常数), 高斯相关 $\rho(s) = \exp(-as^2)$ 等。当 $\epsilon_i(t_{ij})$ 受均值为 0 的平稳高斯过程和均值为 0 的高斯白色噪声 (测量误差) 的影响时, Y_{ij} 的方差为 $\sigma^2 \rho(0) + \tau^2$, 对于, Y_{ij_1} 和 Y_{ij_2} ($j_1 \neq j_2$) 的协方差为 $\sigma^2 \rho(|t_{ij_1} - t_{ij_2}|)$, 这里 $\sigma > 0$, $\tau > 0$, $\rho(\cdot)$ 为连续相关函数。当无序列相关时, 个体内的相关仅来自于随机效应, 此时模型 (2.4) 中不出现 \mathbf{P}_i 。

§2.2 基于似然的估计与推断

§2.2.1 条件极大似然估计

假设 (2.2) 的协方差阵 $\mathbf{V}_i(t_i)$ 是由 R^q 的参数向量 α 决定。记 $\mathbf{V}_i(t_i; \alpha)$ 是由 α 参数化的协方差阵。于是 (2.1) 的对数似然函数为

$$L(\beta, \alpha) = c + \sum_{i=1}^n \left\{ -\frac{1}{2} \log |\mathbf{V}_i(t_i; \alpha)| - \frac{1}{2} (\mathbf{Y}_i - \mathbf{X}_i\beta)^T \mathbf{V}_i^{-1}(t_i; \alpha) (\mathbf{Y}_i - \mathbf{X}_i\beta) \right\}, \quad (2.6)$$

其中 $c = \sum_{i=1}^n [(-n_i/2) \log(2\pi)]$. 对于给定的 α ,

$$\hat{\beta}(\alpha) = \left[\sum_{i=1}^n (\mathbf{X}_i^T \mathbf{V}_i^{-1}(t_i; \alpha) \mathbf{X}_i) \right]^{-1} \left[\sum_{i=1}^n (\mathbf{X}_i^T \mathbf{V}_i^{-1}(t_i; \alpha) \mathbf{Y}_i) \right] \quad (2.7)$$

极大化 (2.6)。

容易验证在模型 (2.1) 下, $\hat{\beta}(\alpha)$ 为 β 的无偏估计。直接计算得 $\hat{\beta}(\alpha)$ 的协方差阵为

$$\begin{aligned} \text{Cov} [\hat{\beta}(\alpha)] &= \left[\sum_{i=1}^n (\mathbf{X}_i^T \mathbf{V}_i^{-1}(t_i; \alpha) \mathbf{X}_i) \right]^{-1} \left[\sum_{i=1}^n (\mathbf{X}_i^T \mathbf{V}_i^{-1}(t_i; \alpha) \text{Cov}(\mathbf{Y}_i) \mathbf{V}_i^{-1}(t_i; \alpha) \mathbf{X}_i) \right] \\ &\quad \times \left[\sum_{i=1}^n (\mathbf{X}_i^T \mathbf{V}_i^{-1}(t_i; \alpha) \mathbf{X}_i) \right]^{-1} \\ &= \left[\sum_{i=1}^n (\mathbf{X}_i^T \mathbf{V}_i^{-1}(t_i; \alpha) \mathbf{X}_i) \right]^{-1}. \end{aligned} \quad (2.8)$$

当指定的方差协方差矩阵的结构不正确时, (2.8) 的第二个等号不成立。进一步结合 (2.1), (2.7) 和 (2.8), 可推得出 $\hat{\beta}(\alpha)$ 服从多元正态分布, 即

$$\hat{\beta}(\alpha) \sim \mathbf{N} \left\{ \beta, \left[\sum_{i=1}^n (\mathbf{X}_i^T \mathbf{V}_i^{-1}(t_i; \alpha) \mathbf{X}_i) \right]^{-1} \right\}. \quad (2.9)$$

当 α 已知时, 利用以上结果可给出 β 的置信域、检验统计量等推断方法。

§2.2.2 极大似然估计

在大多数实际情况下, α 是未知的, 必会用到 α 的相合估计。估计 β 和 α 的一个直观方法是关于 β 和 α 同时极大化 (2.6)。这类极大似然 (maximum likelihood, 简称 ML) 估计可通过将 (2.7) 代入 (2.6) 中, 然后关于 α 极大化 (2.6) 获得。记 $\hat{\beta}_{ML}$ 与 $\hat{\alpha}_{ML}$ 为所得的 ML 估计。采用大样本理论下的标准方法, 就可得到 $(\hat{\beta}_{ML}, \hat{\alpha}_{ML})$ 的渐近分布。

正如绝大多数基于似然的方法那样, 尽管 $(\hat{\beta}_{ML}, \hat{\sigma}_{ML})$ 具有一些合理的统计性质, 但它在实际应用中可能不令人满意。为了显示另一种估计方法为什么在某些情况下可行, 下面我们考虑具有独立误差且 $n_1 = \cdots = n_n = m$ 的一个简单的线性回归模型

$$\mathbf{Y}_i \sim \mathbf{N}(\mathbf{X}_i\beta, \sigma^2\mathbf{I}_m), \quad (2.10)$$

这里 \mathbf{I}_m 为 $[m \times m]$ 单位矩阵, β 和 σ 为模型的参数. 记 $\hat{\beta}_{ML}$ 和 $\hat{\sigma}_{ML}$ 分别为 β 和 σ 的 ML 估计, RSS 为残差平方和, 有

$$\text{RSS} = \sum_{i=1}^n \left(\mathbf{Y}_i - \mathbf{X}_i\hat{\beta}_{ML} \right)^T \left(\mathbf{Y}_i - \mathbf{X}_i\hat{\beta}_{ML} \right).$$

则 σ^2 的 ML 估计为 $\hat{\sigma}_{ML}^2 = \text{RSS}/(nm)$. 对于有限的 n 与 m , 总所周知 $\hat{\sigma}_{ML}^2$ 为 σ^2 的有偏估计, 但稍加修改的 $\hat{\sigma}_{REML}^2 = \text{RSS}/[nm - (k+1)]$ 就是 σ^2 的无偏估计。此处 $\hat{\sigma}_{REML}^2$ 为模型 (2.10) 中 σ^2 的限制极大似然 (restricted maximum likelihood, 简称为 REML) 估计。

§2.2.3 限制极大似然估计

为了估计线性模型中方差分量, Patterson 和 Thompson (1971) 提出了限制极大似然 (REML) 估计方法。其主要思想是对原响应变量作线性变换, 使得变换后的变量的分布与 β 无关。令 $\mathbf{Y} = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_n^T)^T$, $\mathbf{X} = (\mathbf{X}_1^T, \dots, \mathbf{X}_n^T)^T$, \mathbf{V} 为除第 i 主对角块 $\mathbf{V}_i(t_i)$ 外其它元素皆为 0 的块对角矩阵。对于由 α 参数化的 V , 模型 (2.1) 等价于

$$\mathbf{Y} \sim \mathbf{N}(\mathbf{X}\beta, \mathbf{V}(\alpha)), \quad (2.11)$$

其中 α 被称为模型 (2.11) 的方差分量。 α 的 REML 估计就是通过极大化 $\mathbf{Y}^* = \mathbf{A}^T\mathbf{Y}$ 的似然函数得到的, 这里 \mathbf{A} 为 $[N \times (N-k-1)]$ 的列满秩矩阵, 且满足条件 $\mathbf{A}^T\mathbf{X} = \mathbf{0}$, 其中 $N = \sum_{i=1}^n n_i$ 。关于 \mathbf{A} 的构造, 可参见 Diggle, Liang 和 Zeger(1994, 见第 4.5

节)。于是由 (2.11) 可知: \mathbf{Y}^* 服从均值为 0, 协方差阵为 $\mathbf{A}^T \mathbf{V}(\alpha) \mathbf{A}$ 的多元正态分布, Harville(1974) 显示 \mathbf{Y}^* 的似然函数为

$$L^*(\alpha) = \left| \sum_{i=1}^n \mathbf{X}_i^T \mathbf{X}_i \right|^{1/2} \left| \sum_{i=1}^n \mathbf{X}_i^T \mathbf{V}_i^{-1}(t_i; \alpha) \mathbf{X}_i \right|^{-1/2} \left\{ \prod_{i=1}^n |\mathbf{V}_i(t_i; \alpha)|^{-1/2} \right\} \quad (2.12)$$

$$\times \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left(\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}(\alpha) \right)^T \mathbf{V}_i^{-1}(t_i; \alpha) \left(\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}(\alpha) \right) \right\}$$

的常数倍。 α 的 REML 估计 $\hat{\alpha}_{REML}$ 极大化 (2.12)。用 $\hat{\alpha}_{REML}$ 取代 (2.7) 中的 α , 便获得 β 的 REML 估计。事实上, (2.12) 不依赖于 \mathbf{A} 的选择, 故 $\hat{\beta}_{REML}$ 和 $\hat{\alpha}_{REML}$ 也不依赖于具体的线性变换。

\mathbf{Y}^* 的对数似然函数 $\log[L^*(\alpha)]$ 与 \mathbf{Y} 的似然函数 $L(\hat{\beta}, \alpha)$ 仅差一个不依赖 α 的常数和—个不依赖于 β 的函数

$$-\frac{1}{2} \log \left| \sum_{i=1}^n \mathbf{X}_i^T \mathbf{V}_i^{-1}(t_i; \alpha) \mathbf{X}_i \right| A! \#$$

因为 REML 方法与 ML 方法均基于似然原理, 所以它们都具有相合性、渐近正态性和渐近有效性等重要理论性质。然而, 在实际中, 这两者中的任一估计方法都不能一致地优于另一个, 而且两者的算法也不同, ML 法是同时估计固定效应与方差分量, 而 REML 法仅是估计方差分量。

§2.2.4 推断

用上述获得的结果可构造 β 的相关统计推断。关于推断和诊断的详细的叙述可参考文献 Diggle(1988)、Zeger, Liang 和 Albert(1988)、Diggle, Liang 和 Zeger(1994) 以及 Vonesh 和 Chinchilli(1997) 等。这里仅用几个特例来说明。

假设 $\hat{\alpha}$ 为 α 的一个相合估计, 可以是 ML 估计 $\hat{\alpha}_{ML}$, 也可以是 REML 估计 $\hat{\alpha}_{REML}$ 。用 $\hat{\alpha}$ 替代 (2.9) 中的 α 。则至少当 n 很大时, $\hat{\beta}(\hat{\alpha})$ 的分布可由正态分布近似, 即

$$\hat{\beta}(\hat{\alpha}) \sim \mathbf{N}(\beta, \hat{\mathbf{V}}), \quad (2.13)$$

其中 $\hat{\mathbf{V}} = [\sum_{i=1}^n (\mathbf{X}_i^T \mathbf{V}_i^{-1}(t_i; \hat{\alpha}) \mathbf{X}_i)]^{-1}$. 假设 \mathbf{C} 为已知的 $[r \times (k+1)]$ 行满秩矩阵。由 (2.13) 立得：当 n 充分大时， $\mathbf{C}\hat{\beta}(\hat{\alpha})$ 的分布渐近于

$$\mathbf{C}\hat{\beta}(\hat{\alpha}) \sim \mathbf{N}(\mathbf{C}\beta, \mathbf{C}\hat{\mathbf{V}}\mathbf{C}^T). \quad (2.14)$$

于是 $\mathbf{C}\beta$ 的置信系数为 $[100 \times (1-a)]\%$ (其中 $0 < a < 1$) 的近似置信区间为

$$\mathbf{C}\hat{\beta}(\hat{\alpha}) \pm Z_{1-a/2} (\mathbf{C}\hat{\mathbf{V}}\mathbf{C}^T)^{1/2}.$$

令上式中 \mathbf{C} 为只有第 l 元素为 1，其余皆为 0 的 $(k+1)$ 的行向量，便得到 β_l 的置信系数为 $[100 \times (1-a)]\%$ 近似置信区间

$$\hat{\beta}_l(\hat{\alpha}) \pm Z_{1-a/2} \sqrt{\hat{V}_l}, \quad (2.15)$$

这里 \hat{V}_l 为 $\hat{\mathbf{V}}$ 的第 l 个对角元素。

同样利用 (2.14) 的渐近性，我们可以构造线性统计假设的检验统计量，如检验零假设 $\mathbf{C}\beta = \theta_0$ ，备择假设 $\mathbf{C}\beta \neq \theta_0$ (θ_0 为已知常数向量)。一个自然的检验统计量为

$$\hat{T} = [\mathbf{C}\hat{\beta}(\hat{\alpha}) - \theta_0]^T (\mathbf{C}\hat{\mathbf{V}}\mathbf{C}^T)^{-1} [\mathbf{C}\hat{\beta}(\hat{\alpha}) - \theta_0], \quad (2.16)$$

在零假设下，该式近似服从自由度为 r 的 χ^2 分布，记作 χ_r^2 。于是当 $\hat{T} > \chi_r^2(a)$ 时，就以 $(100 \times a)\%$ 检验水平拒绝零假设，其中 $\chi_r^2(a)$ 为 χ_r^2 的上 a 分位点。特例，当零假设为 $\beta_l = 0$ ，备择假设为 $\beta_l \neq 0$ 时，类似于 (2.16) 的推导，可得到零假设的拒绝域，即当

$$|\hat{\beta}_l(\hat{\alpha})| > Z_{1-a/2} \sqrt{\hat{V}_l},$$

拒绝零假设，其中 $Z_{1-a/2}$ 和 \hat{V}_l 的定义见 (2.15)。

§2.3 部分线性模型

正如在 1.4 节所讨论的，作为推广边际线性模型的一种手段，Zeger 和 Diggle(1994) 与 Moyeed 和 Diggle(1994) 已研究了这类模型。对误差过程进一步约束，模

型 (1.2) 等价于

$$Y(t) = \beta_0(t) + \sum_{l=1}^D \beta_l X^{(l)}(t) + \epsilon(t), \quad (2.17)$$

这里 $\epsilon(t)$ 是均值为 0、方差为 σ^2 且相关函数为 $\rho(t)$ 的随机过程, $X^{(l)}(t), l = 1, \dots, D$, 和 $\epsilon(t)$ 独立。模型 (1.2) 中的误差 $\epsilon_i(t_{ij})$ 为 $\epsilon_i(t)$ 的独立重复。为了了解 $\epsilon_i(t_{ij})$, 我们考虑如下分解式

$$\epsilon_i(t_{ij}) = W_i(t_{ij}) + Z_{ij}, \quad (2.18)$$

其中 $W_i(t)$ 是均值为 0、协方差函数为 $\sigma_W^2 \rho(t)$ 的平稳过程 $W(t)$ 的独立重复, Z_{ij} 是均值为 0、方差为 σ_Z^2 的独立同分布的测量误差。测量值 $Y_{ij}, i = 1, \dots, n, j = 1, \dots, n_i$, 的协方差结构为

$$\text{Cov}(Y_{i_1 j_1}, Y_{i_2 j_2}) = \begin{cases} \sigma_Z^2 + \sigma_W^2, & \text{如果 } i_1 = i_2, \quad j_1 = j_2, \\ \sigma_W^2 \rho(t_{i_1 j_1} - t_{i_2 j_2}), & \text{如果 } i_1 = i_2, \quad j_1 \neq j_2, \\ 0, & \text{其它.} \end{cases} \quad (2.19)$$

尽管以上模型可被归类为 (1.3) 的一个特例, 一类结构非参数模型, 该类模型将在下一节讨论, 但它们的估计方法却截然不同。这是因为这两类模型的结构不同。接下来, 本节将重点讨论 $\beta_0(t), \beta_1, \dots, \beta_k$ 的估计的迭代方法。在纵向数据分析中, 主要的科研活动集中在统计推断方法和估计方法上, 目前关于该模型下的这些方法的仍然不好理解, 值得进一步研究。

§2.4 平均响应的无结构光滑方法

假设模型 (2.17) 的平均响应不包含除时间外任何其它协变量, 则模型 (2.17) 就可归结为

$$Y(t) = \beta_0(t) + \epsilon(t), \quad (2.20)$$

其中 $\epsilon(t)$ 如 (2.17) 所定义, $\beta_0(t)$ 为给定时间 t 的条件下 $Y(t)$ 的平均响应, 即 $\beta_0(t) = E[Y(t)|t]$.

一个自然的方法是借用经典独立同分布 (i.i.d) 下的光滑技术得到 $\beta_0(t)$ 的非参估计, 并通过考虑个体内相关的影响来评价所得估计的统计性质。一个简单方法是采用核光滑 (kernel smoothing) 方法, 相当于用一个核函数定义的、 t 邻域内的测量值的加权平均来估计 $\beta_0(t)$ 。令 $K(u)$ 为一个连续的核函数, 通常为定义在实轴上的一个连续的概率密度函数, h 为一个正的带宽序列, 且当 n 趋于无穷大时, 该序列将趋于 0。类似于著名的 i.i.d 情形下的 Nadaraya - Watson 型核估计, $\beta_0(t)$ 的一个核估计为

$$\hat{\beta}_0^K(t) = \frac{\sum_{i=1}^n \sum_{j=1}^{n_i} \{Y_{ij} K[(t - t_{ij})/h]\}}{\sum_{i=1}^n \sum_{j=1}^{n_i} K[(t - t_{ij})/h]}. \quad (2.21)$$

由于 (2.21) 对每一测量值采用了相同权, 因此对具有不同重复测量次数的个体不作区分。然而, 重复测量次数较多的个体通常比重复测量次数较少个体能提供更多的信息量。一个自然的方法就是给每个个体分配特定的权, 用

$$\hat{\beta}_0^K(t; w) = \frac{\sum_{i=1}^n \sum_{j=1}^{n_i} \{Y_{ij} w_i K[(t - t_{ij})/h]\}}{\sum_{i=1}^n \sum_{j=1}^{n_i} \{w_i K[(t - t_{ij})/h]\}} \quad (2.22)$$

来估计 $\beta_0(t)$, 其中, 权 $w = (w_1, \dots, w_n)$ 满足 $w_i \geq 0$, $i = 1, \dots, n$, 且对某些 $1 \leq i \leq n$, 该不等式严格成立。显然, 当 $w_i = 1/N$ 时, 则 (2.22) 就归结为 (2.21)。不同 $w_i = 1/N$ 的另一个直观可选择的权是对每个个体, 而不是对每次测量, 给予相同的权, 相应的核估计就是 (2.22) 中的 $w_i = 1/(nn_i)$ 的情形。

Hart 和 Wehrly(1986)、Müller(1988)、Altman(1990)、Hart(1991) 以及 Rice 和 Silverman(1991) 等也研究了 (2.20) 的其它估计方法, 这里不再讨论, 感兴趣的读者可参见相关文献。包括 (2.22) 在内的以上所介绍的方法, 这些方法本质上都是基于局部光滑的基本思想, 故在实际中通常会得出类似结果。这与下一节将要讨论的光滑方法形成了鲜明的对比, 由于所针对的模型更为复杂, 因此不同的光滑方法常常会导致截然不同的结果。

获得 $\beta_0(t)$ 的一个足够好的核估计关键在于选择适当的窗宽 h , 而核函数的选择相对而言不那么重要。对于核光滑方法以外的其它估计方法而言, 如样条 (spline)

方法, 这就相当于选择一个适当的光滑参数。Rice 和 Silverman(1991) 建议用简单交叉验证 (cross-validation) 方法来选择数据驱动的、不依赖于个体内相关结构的光滑参数。将他们的交叉验证方法应用于核估计 (2.22), 首先, 记 $\hat{\beta}_0^{(-i,K)}(t; w)$ 为用 (2.22) 和删除第 i 个体的所有重复测量值后的数据计算的 $\beta_0(t)$ 的估计, 并用 $\hat{\beta}_0^{(-i,K)}(t; w)$ 来预测第 i 个体在 t 时刻的结果, 则 (2.22) 的交叉验证得分为

$$CV(h) = \sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ w_i \left[Y_{ij} - \hat{\beta}_0^{(-i,K)}(t_{ij}; w) \right]^2 \right\}. \quad (2.23)$$

假设 (2.23) 有唯一的最小值, 则 “留一交叉校验法” 的窗宽 h_{cv} 为其最小值点。因为极小化 (2.23), 相当于近似极小化了平均预测误差 (2.22), 因此应用 h_{cv} 是合理的。更多关于这种交叉校验法的应用和推广的内容将在第 4.7 节讨论。

由于每次重复都需要删除一个新个体, 因此直接计算 (2.23) 通常很费时间。

令 $K_{ij} = K[(t - t_{ij})/h]$,

$$K_{ij}^* = \frac{w_i K[(t - t_{ij})/h]}{\sum_{i=1}^n \sum_{j=1}^{n_i} w_i K[(t - t_{ij})/h]}, \quad K_i^* = \sum_{j=1}^{n_i} K_{ij}^*, \quad i = 1, \dots, n.$$

Rice 和 Silverman(1991) 又提出了一个简单计算 $[Y_{ij} - \hat{\beta}_0^{(-i,K)}(t_{ij}; w)]$ 的方法, 即

$$\begin{aligned} & Y_{ij} - \hat{\beta}_0^{(-i,K)}(t_{ij}; w) \\ &= Y_{ij} - \left[\hat{\beta}_0^K(t_{ij}; w) - \sum_{j=1}^{n_i} (Y_{ij} K_{ij}^*) \right] \left(1 + \frac{K_i^*}{1 - K_i^*} \right) \\ &= \left[Y_{ij} - \hat{\beta}_0^K(t_{ij}; w) \right] + \sum_{j=1}^{n_i} (Y_{ij} K_{ij}^*) - \left[\hat{\beta}_0^K(t_{ij}; w) - \sum_{j=1}^{n_i} (Y_{ij} K_{ij}^*) \right] \left(\frac{K_i^*}{1 - K_i^*} \right) \\ &= \left[Y_{ij} - \hat{\beta}_0^K(t_{ij}; w) \right] + \left(\frac{K_i^*}{1 - K_i^*} \right) \left[\frac{\sum_{j=1}^{n_i} (Y_{ij} K_{ij}^*)}{K_i^*} - \hat{\beta}_0^K(t_{ij}; w) \right]. \end{aligned} \quad (2.24)$$

以上这个表达式是专门针对 (2.22) 所定义的核估计的, 当采用样条等其它光滑方法时, 我们不可能获得像 (2.24) 等号右侧那样明确的表达式, 因此不得不采用一次删除一个个体的方法来直接计算 (2.23)。

基于 $\hat{\beta}_0^K(t; w)$ 的均值、方差和渐近分布, 可获得其大样本推断。注意到 $\hat{\beta}_0^K(t; w)$ 为 Y_{ij} 的线性统计量, 故其均值和方差可直接计算, 从而其渐近分布也容易获得, 事实上, 只要验证考虑了个体内相关之后的三角阵列的中心极限定理成立即可, 参见 Wu, Chiang 和 Hoover(1998) 和 Wu 和 Chiang(2000)。因为 $\hat{\beta}_0^K(t; w)$ 为第 4 节核估计的特例, 对于 $\beta_0(t)$ 的逐点以及同时的推断细节将在第 5.1 节讨论。

§2.5 部分线性模型下协变量效应的估计

当将除时间外的协变量引入模型后, 则可通过联合光滑与参数估计技巧的迭代获得 $(\beta_0(t), \beta_1, \dots, \beta_k)$ 的估计。假设对于所有 $\mathbf{t}_i = (t_{i1}, \dots, t_{in_i})$ 和 $i = 1, \dots, n$, (2.18) 中误差项 $\epsilon_i(t)$ 的方差协方差阵 $\mathbf{V}_i(\mathbf{t}_i)$ 已知。迭代可按如下方法进行:

- (a) 令 $\beta_0(t)$ 为 0, 用 $\mathbf{V}_i(\mathbf{t}_i)$ 取代 $\mathbf{V}_i(\mathbf{t}_i; \alpha)$ 后, 采用广义最小二乘法 (2.7) 计算 $(\beta_1, \dots, \beta_k)^T$ 的初始值。
- (b) 基于当前估计 $(\hat{\beta}_1, \dots, \hat{\beta}_k)$, 计算残差 $r_{ij} = Y_{ij} - \sum_{l=1}^k \hat{\beta}_l X_{ij}^{(l)}$; 然后用 r_{ij} 取代 (2.21) 中的 Y_{ij} 后计算 $\beta_0(t)$ 的核估计 $\hat{\beta}_0^K(t; w)$ 。
- (c) 基于所得的核估计 $\hat{\beta}_0^K(t; w)$, 计算残差 $r_{ij} = Y_{ij} - \hat{\beta}_0^K(t_{ij}; w)$, 并用 $(\mathbf{V}_i(\mathbf{t}_i), r_{ij})$ 取代 $(\mathbf{V}_i(\mathbf{t}_i; \alpha), Y_{ij})$ 后由 (2.7) 更新 $(\beta_1, \dots, \beta_k)$ 的估计。
- (d) 重复步骤 (b) 和 (c), 直到估计收敛。

该算法为 Hastie 和 Tibshirani(1993) 所描述的最一般的后退拟合 (backfitting) 算法的一个特例。

假设相关结构已知是不实际的, 可放宽这一假设。尽管不正确指定相关结构会损失估计效能, 但且并不影响估计的相合性。当方差协方差阵可由 α 参数化, 且误差项来自均值为 0 的高斯平稳过程时, 以上迭代方法可结合上一节的似然和约束似然方法一起使用, 即将第 (a)、(c) 步所采用的广义最小二乘估计替换成基于似

然的估计 $\hat{\beta}_{ML}$ 或 $\hat{\beta}(\hat{\alpha}_{REML})$ 。关于该估计的计算、统计性质以及修改的估计方法的细节, 可参见 Zeger 和 Diggle(1994) 与 Moyeed 和 Diggle(1994)。但基于该估计的推断方法尚未被系统研究, 还有待进一步发展。

§3 变系数非参数模型

本节给出了估计模型 (1.2) 中系数曲线 $\beta(t) = (\beta_0(t), \dots, \beta_k(t))^T$ 的一系列不同光滑方法, 以及基于 $\beta(t)$ 的光滑估计的渐近和 Bootstrap 的推断方法. 通过例 1 的阿拉巴马州小孕龄 (ASGA) 研究和例 2 的巴尔的摩多中心艾滋病列队 (BMACS) 研究来显示这些估计和推断的应用。

§3.1 一些有用的表达式

尽管原则上协变量可以是随机的也可以是固定的, 但在观察研究中, 协变量通常是随机的, 就像个体是被随机选择一样。为了通用性, 假设 $\mathbf{X}(t)$ 是随机的, 且 $E[\mathbf{X}(t)\mathbf{X}^T(t)] \equiv E_{\mathbf{X}\mathbf{X}^T}(t)$ 存在。因为经过适当的符号变化, 以下方法就可修改成适用于协变量非随机的情形。模型 (1.3) 的一个等价表达式为

$$Y(t) = \mathbf{X}^T(t)\beta(t) + \epsilon(t), \quad (3.1)$$

其中 $\epsilon(t)$ 是均值为 0 的随机过程, $\epsilon(t)$ 与 $\mathbf{X}(t)$ 独立。假设 $E_{\mathbf{X}\mathbf{X}^T}(t)$ 可逆, 其逆为 $E_{\mathbf{X}\mathbf{X}^T}^{-1}(t)$ 。则由 (3.1) 可直接推得 $\epsilon(t)$ 的二阶矩

$$E \left\{ [Y(t) - \mathbf{X}^T(t)\beta(t)]^2 \right\} = \inf_{\text{所有 } b(\cdot)} E \left\{ [Y(t) - \mathbf{X}^T(t)b(t)]^2 \right\} \quad (3.2)$$

能够关于 $\beta(t)$ 唯一极小化, 且最小值点为

$$\beta(t) = E_{\mathbf{X}\mathbf{X}^T}^{-1}(t)E[\mathbf{X}(t)Y(t)]. \quad (3.3)$$

当协变量不依赖于时间时, 则有 $\mathbf{X}(t) \equiv \mathbf{X}$ 和 $E_{\mathbf{X}\mathbf{X}^T}(t) \equiv E_{\mathbf{X}\mathbf{X}^T}$, 因此 (3.3) 就归结为

$$\beta_r(t) = E \left[\left(\sum_{l=0}^k e_{rl} X^{(l)} \right) Y(t) \right], \quad (3.4)$$

其中 e_{rl} 为 $E_{\mathbf{X}\mathbf{X}^T}^{-1}$ 的第 r 行第 l 列上的元素。

§3.2 基于最小二乘的光滑

§3.2.1 一般公式

直观上, (3.2) 表明 $\beta(t)$ 可由在 t 邻域内的观察测量值通过局部最小二乘法的方法来估计。对于每个 l 和某个整数 $p \geq 0$, 假定 $\beta_l(t)$ 为 p 次可微, 且其 p 阶导数连续。则对于每个 $l = 0, \dots, k$, $\beta_l(t_{ij})$ 可由一个 p 阶多项式 $\sum_{r=0}^p \{b_{lr}(t)(t_{ij} - t)^r\}$ 近似。于是基于近邻核的 $\beta(t) = (\beta_0(t), \dots, \beta_k(t))^T$ 的局部多项式估计为 $\hat{b}_0(t) = (\hat{b}_{00}(t), \dots, \hat{b}_{k0}(t))^T$, 这里 $\{\hat{b}_{lr}(t); l = 0, \dots, k, r = 0, \dots, p\}$ 极小化

$$L_p(t) = \sum_{i=1}^n \sum_{j=1}^{n_i} w_i \left\{ Y_{ij} - \sum_{l=0}^k \left[X_{ij}^{(l)} \left(\sum_{r=0}^p b_{lr}(t)(t_{ij} - t)^r \right) \right] \right\}^2 K\left(\frac{t_{ij} - t}{h}\right), \quad (3.5)$$

其中 w_i 为 (2.22) 中所定义的非负权重, $K(\cdot)$ 为一个核函数, 通常选择概率密度, h 为非负窗宽。(3.5) 也表明我们可用 $(r!) \hat{b}_{lr}(t)$ 来估计 $\beta_l(t)$ 的第 r 阶导数 $\beta_l^r(t)$, $r = 1, \dots, p$ 。

§3.2.2 最小二乘核估计

(3.5) 的最简单情形为一般最小二乘核估计, 也称其为局部固定拟合, 它是通过极小化 $p = 0$ 情形的 (3.5) 得到的。采用矩阵表示, 记 $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$,

$$\mathbf{X}_i = \begin{pmatrix} 1 & X_{i1}^{(1)} & \cdots & X_{i1}^{(k)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{in_i}^{(1)} & \cdots & X_{in_i}^{(k)} \end{pmatrix}, \quad \mathbf{K}_i(t) = \begin{pmatrix} K_{i1} & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & K_{in_i} \end{pmatrix}$$

其中 $K_{ij} = K[(t_{ij} - t)/h]$ 。如果 $\sum_{i=1}^n \mathbf{X}_i^T \mathbf{K}_i(t) \mathbf{X}_i$ 可逆, 则 $p = 0$ 情形的 (3.5) 能被唯一极小化, 其极小值点, 即 $\beta(t)$ 的核估计, 为

$$\hat{\beta}^{LSK}(t) = \left(\sum_{i=1}^n w_i \mathbf{X}_i^T \mathbf{K}_i(t) \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n w_i \mathbf{X}_i^T \mathbf{K}_i(t) \mathbf{Y}_i \right). \quad (3.6)$$

当模型中不包含除时间外的其它协变量时, 即 $k = 0$, (3.6) 就归为条件期望 $E[Y(t)|t]$ 的一个 Nadaraya-Watson 型核估计, 可参见 Härdle (1990)。

§3.2.3 最小二乘局部线性估计

尽管 (3.6) 有简单的数学表达式, 但当 t 在其支撑的边界时, 通常会导致显著的偏。减少这种边界偏的一般方法是采用高阶局部多项式拟合。但由于高阶局部多项式拟合通常需要大样本, 且计算强度较大, 故在许多应用中不可行。一个能自动调正边界偏的实用方法是采用极小化 $p = 1$ 情形的 (3.5) 的局部线性拟合。记

$$\mathcal{N}_{lr} = \begin{pmatrix} \sum_{i,j} \left[w_i X_{ij}^{(l)} X_{ij}^{(r)} K_{ij} \right] & \sum_{i,j} \left[w_i X_{ij}^{(l)} X_{ij}^{(r)} (t_{ij} - t) K_{ij} \right] \\ \sum_{i,j} \left[w_i X_{ij}^{(l)} X_{ij}^{(r)} (t_{ij} - t) K_{ij} \right] & \sum_{i,j} \left[w_i X_{ij}^{(l)} X_{ij}^{(r)} (t_{ij} - t)^2 K_{ij} \right] \end{pmatrix},$$

$$\mathcal{N}_r = (\mathcal{N}_{0r}, \dots, \mathcal{N}_{kr}), \mathcal{N} = (\mathcal{N}_0^T, \dots, \mathcal{N}_k^T)^T,$$

$$\mathcal{M}_r = \left(\sum_{i,j} \left[w_i X_{ij}^{(r)} Y_{ij} K_{ij} \right], \sum_{i,j} \left[w_i X_{ij}^{(r)} (t_{ij} - t) Y_{ij} K_{ij} \right] \right)^T,$$

$$\mathcal{M} = (\mathcal{M}_0^T, \dots, \mathcal{M}_k^T)^T, b_l(t) = (b_{l0}(t), b_{l1}(t))^T, \quad b(t) = (b_0(t), \dots, b_k(t))^T, \quad r, l = 0, \dots, k.$$

令 $L_1(t)$ 关于 $b_{lr}(t)$ 的偏导数等于 0, 在 $p = 1$ 情形下 (3.5) 的正则方程为

$$\mathcal{N}b(t) = \mathcal{M}. \quad (3.7)$$

假设矩阵 \mathcal{N} 在时间 t 处可逆, 则 (3.7) 有唯一解, 且为 $\hat{b}(t) = \mathcal{N}^{-1}\mathcal{M}$ 。因此 $\beta_l(t)$ 的最小二乘局部线性估计 $\hat{\beta}_l^{LSL}(t)$ 为

$$\hat{\beta}_l^{LSL}(t) = e_{2l+1}^T \hat{b}(t), \quad (3.8)$$

其中 e_q 为第 q 元素为 1 其余皆为 0 的 $[2(k+1) \times 1]$ 列向量。类似地可得到一般的高阶最小二乘局部线性估计的显示表达式, 参见 Hoover 等 (1998)。由于对于文献中出现的几乎所有生物医学相关问题, 局部线性拟合已经足够, 所以这里不详细讨论一般高阶的估计。

§3.2.4 中心化协变量的最小二乘

在一些情况下, 由于模型 (3.1) 中的某些协变量不能取到 0 点, 故基线系数曲线 $\beta_0(t)$ 没有实际意义。如 ASGA 研究 (第 1.2.1 节) 中的母亲胎盘厚度、孕前身高,

HIV/CD4 损耗资料 (第 1.2.2 节) 中的个体感染前 CD4 水平, 均为严格的正值协变量。这种情况下, 一个有用的补救方法是将模型中的协变量中心化, 使得相应的基线系数可解释为中心化后的协变量在零点时 $Y(t)$ 的条件均值。

令 $X^{(*l)}(t) = X^{(l)}(t) - E[X^{(l)}(t)]$ 为 $X^{(l)}(t)$ 的中心化, $\mathbf{X}^{(*)}(t)$, 为其部分或全部元素被中心化的协变量向量。则模型 (3.1) 的一个等价形式为

$$Y(t) = \left(\mathbf{X}^{(*)}(t) \right)^T \beta^*(t) + \epsilon(t), \quad (3.9)$$

其中 $\beta^*(t) = (\beta_0^*(t), \beta_1(t), \dots, \beta_k(t))^T$ 。注意到模型 (3.9) 的基线系数曲线 $\beta_0^*(t)$ 就是令 $X^{(*l)}(t)$, $l = 1, \dots, k$, 等于 0 时 $Y(t)$ 的均值。 (3.9) 中其它系数曲线的解释可类似于 (3.1)。

首先估计 $X_{ij}^{(l)}$ 的中心化协变量 $X_{ij}^{(*l)}$, 然后用 $X_{ij}^{(*l)}$ 替代 (3.5) 中 $X_{ij}^{(l)}$, 并极小化 (3.5) 便可获得 $\beta^*(t)$ 的估计。如果 $X^{(l)}(t)$ 为时间依赖协变量, 那么采用核光滑, 由 $X_{ij}^{(*l)} = X_{ij}^{(l)} - \hat{\mu}_l(t_{ij})$ 来估计 $X_{ij}^{(l)}$ 的中心化形式, 这里

$$\hat{\mu}_l(t) = \frac{\sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ w_i X_{ij}^{(l)} \Gamma_l[(t - t_{ij})/\gamma_l] \right\}}{\sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ w_i \Gamma_l[(t - t_{ij})/\gamma_l] \right\}}, \quad (3.10)$$

其中 $(\Gamma_l(\cdot), \gamma_l)$ 为一组核和窗宽。此外, 如果 $X^{(l)}(t) \equiv X^{(l)}$ 为时间不变的, 则 $X_{ij}^{(l)} \equiv X_i^{(l)}$, $j = 1, \dots, n_i$, $X_i^{(*l)}$ 可取为 $X_i^{(l)} - \bar{X}^{(l)}$, 其中 $\bar{X}^{(l)} = n^{-1} \sum_{i=1}^n X_i^{(l)}$ 为 $X^{(l)}$ 的加权样本平均。令 $\mathbf{X}_i^{(*)}$ 为 $n_i \times (k+1)$ 中心化协变量矩阵, 其第 j 行为 $(1, X_{ij}^{(*1)}, \dots, X_{ij}^{(*k)})$ 。

则 $\beta^*(t)$ 的一个最小二乘核估计为

$$\hat{\beta}^{LSK}(t) = \left[\sum_{i=1}^n w_i \left(\mathbf{X}_i^{(*)} \right)^T \mathbf{K}_i(t) \left(\mathbf{X}_i^{(*)} \right)^T \right]^{-1} \left[\sum_{i=1}^n w_i \left(\mathbf{X}_i^{(*)} \right)^T \mathbf{K}_i(t) \mathbf{Y}_i \right], \quad (3.11)$$

其中 $\mathbf{K}_i(t)$ 和 \mathbf{Y}_i 如 (3.6) 中所定义。

Wu, Yu 和 Chiang (2000) 研究了 $\hat{\beta}^{LSK}(t)$ 的大样本性质。他们的结果表明 $\hat{\beta}^{LSK}(t)$ 和 $\hat{\beta}^{*LSK}(t)$ 中任何一个都不能一致地优于另一个, 特别是当协变量不随时

间变换时, $\hat{\beta}^{LSK}(t)$ 与 $\hat{\beta}^{*LSK}(t)$ 渐近相等。然而对于 $l \geq 1$, 若 $X^{(l)}(t)$ 随时间变化显著, 则可通过中心化 $X^{(l)}(t)$ 得到 $\beta_l(t)$ 的理论上和实际上更优的估计。

当然, 将协变量中心化后, 模型的基线系数曲线将有所改变。协变量是否中心化主要依赖于相应基线系数曲线的生物学解释。当研究者所感兴趣是变量效应的估计性质, 而不是基线系数曲线时, 就可中心化协变量。显然, 除核光滑外的其它方法也可以用于中心化协变量情形的估计。但由于光滑协变量的复杂性, 除 (3.11) 外的估计的统计性质文献中尚未被研究。

§3.2.5 一个简单修正

以上提到的估计, 不管是针对中心化还是未中心化的协变量, 都是依赖于单一窗宽来估计采所有 $(k+1)$ 条系数曲线。当所有曲线基本属于相同的光滑族时, 这种简单方法具有很好的效果, 但这种理想情况在实际中并不常见。因此, 更为实用的方法是能够根据不同的系数曲线对光滑的不同要求自动调整窗宽的灵活方法。

关于单一窗宽的潜在缺陷, 文献中已有许多记载, 如 Hoover 等 (1998)、Fan 和 Zhang(2000)、Wu, Yu 和 Chiang(2000) 等。这些作者还提出了许多其它方法 (见第 4.3-4.6 节) 来克服这些潜在不足, 其中 Wu, Yu 和 Chiang(2000) 提出的简单方法是用线性组合:

$$\hat{\beta}(t; \mathbf{K}, \mathbf{h}) = \sum_{l=0}^k e_{l+1}^T \hat{\beta}(t; K_l, h_l), \quad (3.12)$$

这里 $\mathbf{K}(\cdot) = (K_0(\cdot), \dots, K_k(\cdot))$, $\mathbf{h} = (h_0, \dots, h_k)$, e_p 为 $[(k+1) \times 1]$ 的向量, 且只有第 p 元素为 1, 其余皆为 0, $\hat{\beta}(t; K_l, h_l)$ 为分别由 (3.6) 或 (3.11) 给出的 $\beta(t)$ 或 $\beta^*(t)$ 的核估计, 其中 $K_l(\cdot)$ 为核, h_l 为窗宽。直观上, $\hat{\beta}(t; \mathbf{K}, \mathbf{h})$ 依赖于具体的一对核和窗宽来估计相应的 $\beta(t)$ 或 $\beta^*(t)$ 的分量, 其实可作为一个一般方法, (3.12) 不仅限于核估计, 也可用于其它局部多项式估计。

§3.2.6 w_i 的选择

影响 $\beta(t)$ 的最小二乘局部多项式估计的理论性质和实际效果的一个重要因素是 (3.5) 中 w_i 的选择。对于独立同分布的截面数据研究, 选择一致的权 $w_i \equiv 1/N$ 常常是可取的; 而对于现时抽样, 权 w_i 的选择应依赖于个体内的相关结构和重复测量数 n_i 。然而, 实际上数据的相关结构通常完全未知, 且很难被估计, 因此通常考虑主观地选择权, 如 $w_i = 1/N$ 和 $w_i = 1/(nn_i)$, 其中 $N = \sum_{i=1}^n n_i$ 。直观上, $w_i = 1/N$ 是对每个观察点分配相同权, 而 $w_i = 1/(nn_i)$ 是对每个个体分配相同权。理论上, 当某些 n_i 比其它大许多时, 选择 $w_i = 1/N$ 可能会得到不相合的最小二乘核估计, 但不管 n_i 间的差别多大, 基于 $w_i = 1/(nn_i)$ 的最小二乘核估计且总是相合的 (Hoover 等, 1998、Wu 和 Chiang, 2000)。

§3.3 惩罚最小二乘法

假设 $\beta(t)$ 的所有分量关于 t 都二次连续可微, 且其二阶导数有界、平方可积。则一般惩罚最小二乘 (penalized least squares) 准则是关于 $\beta_l(t)$ 极小化

$$J(\beta, \lambda) = \sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ Y_{ij} - \sum_{l=0}^k X_{ij}^{(l)} \beta_l(t_{ij}) \right\}^2 + \sum_{l=0}^k \lambda_l \int [\beta_l''(t)]^2 dt, \quad (3.13)$$

其中 $\lambda = (\lambda_0, \dots, \lambda_k)^T$, λ_l 为正的光滑参数。(3.13) 的最小值的存在性和唯一性依赖于 t_{ij} 与 $X_{ij}^{(l)}$ 。假定 (3.13) 有唯一最小值点, 则 (3.13) 的最小值点就定义为 $\beta(t)$ 的惩罚最小二乘估计 $\hat{\beta}^{PLS}(t) = (\hat{\beta}_0^{PLS}(t), \dots, \hat{\beta}_k^{PLS}(t))^T$ 。采用类似于单变量的光滑技术, 可以看到 $\hat{\beta}_l^{PLS}(t)$ 为以不同 $\{t_{ij} : i = 1, \dots, n, j = 1, \dots, n_i\}$ 为结点的自然能三次样条, 能够被表示为 $\{Y_{ij} : i = 1, \dots, n, j = 1, \dots, n_i\}$ 的线性函数。

不同于由 (3.5) 所得的估计, $\hat{\beta}^{PLS}(t)$ 的惩罚项采用了多个光滑参数 λ_l 。(3.13) 中惩罚项的所有 $(k+1)$ 个光滑参数 $\lambda_l, l = 0, \dots, k$, 都可以调节。Hoover 等 (1998) 通过数值结果显示: 与仅依靠一个光滑参数的最小二乘局部多项式相比较, 多重光滑参数所带来的更多灵活性的确能使得其获得更好的估计。然而, 因为在所有不同

时间点 $\hat{\beta}^{PLS}(t)$ 都有结点，所以当不同时间点的个数较大时，如不平衡纵向研究中常见的情形，其计算强度会非常大。

§3.4 两步光滑方法

对于大容量的纵向数据，为了提供计算量适中、灵活的光滑估计，Fan 和 Zhang(2000) 提出估计 $\beta(t)$ 的两步光滑法 (two-step smoothing method)，该方法也采用了 $(k+1)$ 个光滑参数，但方法不同于 (3.12) 与 (3.13)。具体步骤如下：

- (i) 计算 $\beta(s)$ 在不同的时间点的初始估计 $\hat{\beta}^{RAW}(s)$ ，时间点为 s_1, \dots, s_m ，其中 m 可依赖于 n 和 $n_i, i = 1, \dots, n$ 。
- (ii) 通过光滑初始估计 $\hat{\beta}_l^{RAW}(s_r), r = 1, \dots, m$ ，来估计每个系数曲线 $\beta_l(t), l = 1, \dots, m$ 。

尽管 Fan 和 Zhang(2000) 是通过局部多项式来阐述这一方法，但原则上，该方法也可用于其它光滑方法，如样条。

对于平衡纵向数据，即所有个体在相同时间点 $s_j, j = 1, \dots, m, m = n_i, i = 1, \dots, n$ ，都有观测值，则初始估计可通过线性模型拟合 $(Y_{ij}, X_{ij}), j = 1, \dots, m$ 得到。但是当设计为不平衡时，如大多数实际情况会出现在某些时间点能观测到个体数很稀少，此时计算初始估计就必需结合邻近时间点的观测。特别是可首先采用具有较小窗宽的局部多项式 (3.5) 计算 $\hat{\beta}_l^{RAW}(s_r), l = 0, \dots, k$ ，然后将 $\hat{\beta}_l^{RAW}(s_r)$ 作为新数据，通过关于 $b_{lr}(t)$ 极小化

$$L_{p,l}^{TS}(t) = \sum_{j=1}^m \left\{ \hat{\beta}_l^{RAW}(s_j) - \sum_{r=0}^p b_{lr}(t)(s_j - t)^r \right\}^2 K_l \left(\frac{s_j - t}{h_l} \right) \quad (3.14)$$

得到 $\beta_l(t)$ 的估计，其中 $(K_l(\cdot), h_l)$ 为一系列核和窗宽。与 (3.5) 式类似，如果 $\hat{b}_{lr}^{TS}(t), r = 0, \dots, k$ 是 (3.14) 的唯一最小值点，则 $\hat{b}_{l0}^{TS}(t)$ 为 $\beta_l(t)$ 的两步第 p 阶局部多项式估计，而 $(r!) \hat{b}_{lr}^{TS}(t)$ 可用来估计 $\beta_l(t)$ 的第 r 阶导数。

(3.5) 是对所有 $\beta_l(t)$ 采用了单一窗宽来获得估计, 而两步光滑法与之不同, 原则上它可根据每条系数曲线对光滑的特殊需要灵活调节其窗宽。然而使用当前的两步光滑的主要困难是在构建非平衡纵向数据的初始估计方面还缺乏具体的、实用的指导。某个数据驱动的窗宽在计算初始估计和最终估计将更为实用。关于不同初始估计对最终两步估计的理论性质和实践的影响问题目前尚不清楚, 还有待进一步研究。

§3.5 时不变协变量的系数曲线的逐元光滑

当协变量为时不变时, 如在整个研究期间临床试验的处理保持固定, 由 (3.3) 启发得到的一个更灵活、计算可行估计 $\beta(t)$ 的方法是分别对 $\beta(t)$ 的每个元素光滑。

令 $Z^{(r)}(t) = [\sum_{l=0}^k e_{rl} X^{(l)}] Y(t)$, $\mathbf{X}_i = (1, X_i^{(1)}, \dots, X_i^{(k)})^T$ 是第 i 个个体的协变量向量, \hat{e}_{rl} 是矩阵 $(\hat{E}_{\mathbf{X}\mathbf{X}^T})^{-1}$ 的第 (r, l) 元素, 其中 $\hat{E}_{\mathbf{X}\mathbf{X}^T} = (1/n) \sum_{i=0}^n \mathbf{X}_i \mathbf{X}_i^T$. 故 $Z^{(r)}(t)$ 的一个很自然的估计为 $Z_{ij}^{(r)} = [\sum_{l=0}^k \hat{e}_{rl} X_i^{(l)}] Y_{ij}$. 由 (3.3) 可知, $\beta_r(t)$ 的逐元光滑估计可通过光滑 $Z_{ij}^{(r)}$ ($i = 1, \dots, n, j = 1, \dots, n_i$) 得到. 特别是 $\beta_r(t)$ 的 $p(\geq 0)$ 阶局部多项式估计为 $\hat{b}_{r0}^{COM}(t)$, 其中 $\hat{b}_{rl}^{COM}(t), l = 0, \dots, p$, 为

$$L_{p,r}^{COM}(t) = \sum_{i=1}^n \sum_{j=1}^{n_i} w_i \left\{ Z_{ij}^{(r)} - \sum_{l=0}^p b_{rl}(t) (t_{ij} - t)^l \right\}^2 K_r \left(\frac{t_{ij} - t}{h_r} \right), \quad (3.15)$$

关于 $b_{rl}(t)$ 的唯一极小值点。对于 $p = 0$ 的局部常数拟合, (3.15) 导出的逐元核估计为

$$\hat{\beta}_r^{COM}(t) = \frac{\sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ w_i Z_{ij}^{(r)} K_r [(t_{ij} - t)/h_r] \right\}}{\sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ w_i K_r [(t_{ij} - t)/h_r] \right\}}. \quad (3.16)$$

Wu 和 Chiang(2000) 给出了 $\hat{\beta}_r^{COM}(t)$ 的大样本均方误差, Wu, Yu 和 Yuan(2000) 提出了构建 $\beta_r(t)$ 的近似渐近逐点和同时置信区间的方法。尽管在 $p \geq 1$ 情形还没有给出具体的渐近风险和渐近分布, 但这些结果给高阶估计 $\hat{b}_{r0}^{COM}(t)$ 的渐近性带来了曙光。Wu 和 Chiang(2000) 与 Wu, Yu 和 Yuan(2000) 的结果表明 $\hat{\beta}_r^{COM}(t)$ 在理

论收敛速度和实际应用的灵活性方面都优于比核估计 (3.6)。对于逐元局部多项式估计, 也期待具有最小二乘法 (3.5) 的类似优点。

很明显, 极小化 (3.15) 并非仅有的逐元光滑方法。假设设计时间点在一个紧集 $[a, b]$ 范围内, 且 $\beta_r(t)$ 在 $[a, b]$ 上关于 t 是二次可微的。则另一个可行的方法是用惩罚最小二乘估计 $\tilde{\beta}_r^{COM}(t)$ 来估计 $\beta_r(t)$, 其中 $\tilde{\beta}_r^{COM}(t)$ 为

$$J_r^{COM}(\beta_r, \lambda_r) = \sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ w_i \left[Z_{ij}^{(r)} - \beta_r(t_{ij}) \right]^2 \right\} + \lambda_r \int_a^b \left[\beta_r''(s) \right]^2 ds \quad (3.17)$$

的极小值点, 这里 λ_r 为一个非负的光滑参数。采用与 2.3 节相同的原理, 容易证明: $\tilde{\beta}_r^{COM}(t)$ 为一个自然三次样条, 结点为不同时间点 t_{ij} ; $i = 1, \dots, n, j = 1, \dots, n_i$ 。进一步, 采用等价核方法, Chiang, Rice 和 Wu(2001) 导出了 $\tilde{\beta}_r^{COM}(t)$ 的渐近均方误差和渐近分布。与 (3.13) 的多个惩罚最小二乘法相比, (3.17) 大大简化了计算, 因为前者是对一个包含所有 $(k+1)$ 元素的大线性系统求解, 后者则是分别对 $(k+1)$ 个线性系统求解。计算上的优点使得 (3.17) 在大多数情况可以实施, 而 (3.13) 的计算量很大, 使得优化该式通常不切实际。

§3.6 通过基近似的光滑

以上描述的所有光滑方法均为局部光滑, 某种意义上讲, 即仅在 t 的某个邻域内的测量值被有效用来估计 $\beta(t)$ 。尽管当所有系数曲线 $\beta_r(t)$ 为非参数时局部光滑很有效, 但当一些系数曲线有已知参数形式时, 如部分线性模型 (1.1), 局部光滑将不适用。

与局部光滑相比, 采用基近似 (basis approximations) 估计有三个重要的优点:
(I) 无论 $\beta(t)$ 的元素为参数还是非参数, 均可用该方法估计 $\beta(t)$, 因此对于参数与非参数的变系数模型均可采用该方法。(II) 当随机效应出现时, 可自然均值将随机效应合并到非参数或半参数模型中。(III) 与光滑样条相比, 截断多项式或 B-样条等流行的基估计通常依赖更少的结点或近似项, 因此计算上大大优于光滑样条或局部

多项式。尽管混合效应的估计在各种情况下都很受重视，但这里仅讨论边际模型。

至于其在混合效应模型的扩展可参见 Rice 和 Wu(2000)。

主要思想是：首先通过扩展带有 K_r 项的基函数近似 $\beta_r(t)$ ，其中 K_r 可随 n 趋于无穷大而趋于无穷大，也可以不趋于无穷大；然后通过估计该扩展的系数来估计 $\beta_r(t)$ 。对于每个 $r = 0, \dots, k$ ，令 $B_{rs}(t)$, $s = 1, \dots, K_r$ ，为一组基函数。如果 $\beta_r(t)$ 可通过 $B_{rs}(t)$, $s = 1, \dots, K_r$ 的一个扩展来近似，则存在一组常数 γ_{rs} 使得

$$\beta_r(t) \approx \sum_{s=1}^{K_r} \gamma_{rs} B_{rs}(t). \quad (3.18)$$

将 (3.18) 代入 (1.2)，得到变系数模型的一个近似，即

$$Y_{ij} \approx \sum_{r=0}^k \sum_{s=1}^{K_r} X_{ij}^{(r)} \gamma_{rs} B_{rs}(t) + \epsilon_i(t_{ij}). \quad (3.19)$$

如果对于所有 $r = 0, \dots, k$ ， $\beta_r(t)$ 都属于由 $\{B_{rs}(t); s = 1, \dots, K_r\}$ 生成的线性空间，则 (3.19) 中的近似符号可用等号取代。

利用 (3.19)，通过极小化

$$\ell(\gamma) = \sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ w_i \left[Y_{ij} - \sum_{r=0}^k \sum_{s=1}^{K_r} \left(X_{ij}^{(r)} \gamma_{rs} B_{rs}(t_{ij}) \right) \right]^2 \right\}, \quad (3.20)$$

便可得到 γ_{rs} 最小二乘估计 $\hat{\gamma}_{rs}$ ，其中 $\gamma = (\gamma_0^T, \dots, \gamma_k^T)^T$ ， $\gamma_r = (\gamma_{r1}, \dots, \gamma_{rK_r})^T$ 。如果 (3.20) 存在唯一的极小值点，则 $\beta_r(t)$ 的基函数估计为

$$\hat{\beta}_r^{BAS}(t) = \sum_{s=1}^{K_r} [\hat{\gamma}_{rs} B_{rs}(t)], \quad (3.21)$$

这里 K_r 可以依赖于 n 和 n_i , $i = 1, \dots, n$ 。显然，如果 K_r 为已知且有限的，且 $\beta_r(t)$ 属于由 $B_{rs}(t)$, $s = 1, \dots, K_r$ 张成的线性空间，那么由 (3.21) 将得到 $\beta_r(t)$ 的一个参数估计。另一方面，如果 (3.18) 中的 K_r 未知，则由 (3.21) 仍可得到相合的非参数估计须要求 K_r 为 n 和 n_i ($i = 1, \dots, n$) 的函数，且当 n 趋于无穷大时， K_r 可趋于无穷。

依赖于数据的基本实际意义，许多不同的基可用来近似 $\beta(t)$ 的分量。在经典线性模型中，最受欢迎的基是多项式基 $\{1, t, \dots, t^{K_r-1}\}$ 。比以上多项式基在计算上有优越的一类基是分段多项式基。分段多项式基的例子包括线性、二次、三次样条等 B-样条基，或一类截断幂级数；分段多项式的显式表达式及其数值性质的具体内容参见 de Boor (1978)。如果相信 $\beta_r(t)$ 具有周期性，则通常选择傅立叶序列作为基。

Huang, Wu 和 Zhou(2000) 证明了 (3.21) 的相合性，并通过大量模拟研究了 B-样条下 (3.21) 的实际性能。Huang, Wu 和 Zhou (2004) 给出了 (3.21) 的 B-样条估计的进一步的渐近性质。总之，B-样条估计要求光滑参数由三个方面组成，即多项式阶数、结的个数和位置。尽管希望如此，但由数据同时决定这三个方面通常有很大难度。Rice 和 Wu(2000) 表明：采用等空间结的简单方法在实际应用中通常有效可行，Huang, Wu 和 Zhou(2000) 的模拟研究证实了这一结论。

§3.7 交叉验证

影响以上所有光滑方法的最重要因素是如何选择恰当的光滑参数，如窗宽、正的惩罚权 λ 和结的个数与位置。因此直接由数据选取这些值具有理论与实际的意义。

为独立同分布数据的非参数回归，选取数据驱动的光滑参数已是文献中研究非常多的问题。目前广泛运用的方法是 Rice 和 Silverman(1991) 提出的交叉验证法，该方法每次删除的是一个体的全部重复测量值，而不是一个的测量值。对除时间外无其它协变量的简单非参数回归，Hart 和 Wehrly(1993) 下交叉验证的相合性。记 ξ 为光滑参数向量， $\hat{\beta}(t; \xi)$ 为基于 ξ 和 $\hat{\beta}^{(-i)}(t; \xi)$ 的光滑估计， $\hat{\beta}^{(-i)}(t; \xi)$ 为删除第 i 个个体的测量值后用与 $\hat{\beta}(t; \xi)$ 同样的方法计算的估计。 $\hat{\beta}(t; \xi)$ 的交叉验证得分为

$$CV(\xi) = \sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ w_i \left[Y_{ij} - X_{ij}^T \hat{\beta}^{(-i)}(t; \xi) \right]^2 \right\}, \quad (3.22)$$

它测量了 $\hat{\beta}(t; \xi)$ 的预测误差。假定 $CV(\xi)$ 存在唯一的最小值，则交叉验证光滑参数 ξ_{cv} 就是 $CV(\xi)$ 的最小值点。以上交叉验证准则可直接用于除 2.4 节的两步光滑外的以上所有光滑方法。对于第 2.2, 2.3 和 2.5 节的估计和等空间结的 B-样条，极小化相应的交叉验证得分可获得单变量窗宽或 R^{k+1} -值向量。自动寻找全局最小值点通常需要很好的优化软件。实际上，尤其当光滑参数为多变量时，采用使其交叉验证得分接近全局最小的光滑参数常常是合理的。

采用交叉验证准则 (3.22) 有三个直观的原因：首先，一次删除一个个体，可保持数据的相关结构；其次，与 AIC、BIC 和广义交叉验证 (如 Akaike, 1970、Schwarz, 1978、Shibata, 1981 和 Wahba, 1990) 等相比，(3.22) 不依赖于个体内的相关结构，因此几乎可在所有实际情况下都可被采用；最后，当个体数足够大时，极小化 (3.22) 获得光滑参数可近似极小化均方误差

$$ASE(\hat{\beta}(\cdot; \xi)) = \sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ w_i \left[X_{ij}^T (\beta(t_{ij}) - \hat{\beta}(t_{ij}; \xi)) \right]^2 \right\}. \quad (3.23)$$

最后的这个结论可有如下分解式通过分解

$$\begin{aligned} CV(\xi) &= \sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ w_i [Y_{ij} - X_{ij}^T \beta(t_{ij})]^2 \right\} \\ &\quad + 2 \sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ w_i [Y_{ij} - X_{ij}^T \beta(t_{ij})] \left[X_{ij}^T (\beta(t_{ij}) - \hat{\beta}^{(-i)}(t_{ij}; \xi)) \right] \right\} \\ &\quad + \sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ w_i \left[X_{ij}^T (\beta(t_{ij}) - \hat{\beta}^{(-i)}(t_{ij}; \xi)) \right]^2 \right\} \end{aligned} \quad (3.24)$$

导出。事实上，结合 (3.23) 和 $\hat{\beta}^{(-i)}(t; \xi)$ 的定义可看出 (3.24) 中等号右边的第三项近似等于 $ASE(\hat{\beta}(\cdot; \xi))$ 。这是因为 (3.24) 中等号右边的第一项不依赖于光滑参数，且第二项近似为 0，故 ξ_{cv} 可近似极小化了 $ASE(\hat{\beta}(\cdot; \xi))$ 。

§3.8 核估计的近似推断

基于估计的近似分布或 Bootstrap 法都可获得置信区间。目前只得到了核估计 (3.6) 和 (3.16) 的渐近分布的显式表达式。

§3.8.1 逐点置信区间

(3.6) 和 (3.16) 的渐近分布的推导都是基于以下两个重要假设: (a) 重复测量数 n_i 为非随机的, 且当 n 趋于无穷时, 它可能趋于无穷, 也可能不; (b) 时间设计点 t_{ij} 为随机、独立同分布, 但其密度函数 $f(\cdot)$ 是未知的。这些假设不但是为了实际应用方便, 同时也是为了数学上的可处理。

首先考虑基于 (3.6) 的置信区间。在以上假设和另外一些的温和的正则条件下, Wu, Chiang 和 Hoover(1998) 表明: 如果 $w_i = 1/N$, $h = N^{-1/5}h_0$, 且对于某些常数 $h_0 > 0$ 和 $0 \leq \theta < \infty$, 有

$$\lim_{n \rightarrow \infty} N^{-6/5} \sum_{i=1}^n n_i^2 = \theta$$

则 $\hat{\beta}^{LSK}(t)$ 服从渐近多元正态分布, 且有

$$(Nh)^{1/2} \left[\hat{\beta}^{LSK}(t) - \beta(t) \right] \longrightarrow \mathbf{N}(B(t), D^*(t)), \quad (3.25)$$

则偏差 $B(t)$ 和协方差阵 $D^*(t)$ 分别为

$$B(t) = [f(t)]^{-1} E_{XX^T}^{-1}(t) (b_0(t), \dots, b_k(t))^T \quad (3.26)$$

和

$$D^*(t) = [f(t)]^{-2} E_{XX^T}^{-1}(t) D(t) E_{XX^T}^{-1}(t), \quad (3.27)$$

这里 $D(t)$ 为 $(k+1) \times (k+1)$ 的矩阵, 其第 (l, r) 元素为

$$\begin{aligned} D_{lr}(t) = & \sigma^2(t) E \left[X^{(l)}(t) X^{(r)}(t) \right] f(t) \left\{ \int [K(u)]^2 du \right\} \\ & + \theta h_0 \rho_\epsilon(t) E \left[X^{(l)}(t) X^{(r)}(t) \right] [f(t)]^2, \end{aligned}$$

其中 $\sigma^2(t) = E[\epsilon^2(t)]$, $\rho_\epsilon(t) = \lim_{a \rightarrow 0} E[\epsilon(t+a)\epsilon(t)]$,

$$\begin{aligned} b_l(t) = & h_0^{3/2} \sum_{c=0}^k \left\{ \left[\int u^2 K(u) du \right] \left\{ \beta'_c(t) \left[E \left[X^{(l)}(t) X^{(c)}(t) \right] \right]' f(t) \right. \right. \\ & \left. \left. + \beta'_c(t) E \left[X^{(l)}(t) X^{(c)}(t) \right] f'(t) + (1/2) \beta''_c(t) E \left[X^{(l)}(t) X^{(c)}(t) \right] f(t) \right\} \right\}. \end{aligned}$$

故对于任何已知向量 $A = (a_0, a_1, \dots, a_k)'$ ，上、下端点 $L_\alpha(t)$ 和 $U_\alpha(t)$ 为

$$\left\{ A^T \hat{\beta}^{LSK}(t) - (Nh)^{-1/2} A^T B(t) \right\} \pm Z_{\alpha/2} (Nh)^{-1/2} [A^T D^*(t) A]^{1/2}, \quad (3.28)$$

这里 $Z_{\alpha/2}$ 为标准正态分布的上 $\alpha/2$ 分位点，使得

$$\lim_{n \rightarrow \infty} P \{ L_\alpha(t) \leq A^T \beta(t) \leq U_\alpha(t) \} = 1 - \alpha. \quad (3.29)$$

因为 $B(t)$ 与 $D^*(t)$ 依赖于未知分位数，(3.28) 实际上是不可行的。如果 $B(t)$ 与 $D^*(t)$ 有相合估计 $\hat{B}(t)$ 与 $\hat{D}^*(t)$ ，则 $A^T \beta(t)$ 的逐点置信系数为 $(1 - \alpha)$ 的置信区间可由 $(\hat{L}_\alpha(t), \hat{U}_\alpha(t))$ 近似，其中上、下限 $\hat{L}_\alpha(t)$ 和 $\hat{U}_\alpha(t)$ 为

$$\left\{ A^T \hat{\beta}^{LSK}(t) - (Nh)^{-1/2} A^T \hat{B}(t) \right\} \pm Z_{\alpha/2} (Nh)^{-1/2} [A^T \hat{D}^*(t) A]^{1/2}. \quad (3.30)$$

Wu, Chiang 和 Hoover(1998) 建议通过用相应的核估计替代 (3.26) 与 (3.27) 中的 $f(t)$, $\sigma^2(t)$, $\rho_\epsilon(t)$, $E[X^{(l)}(t)X^{(r)}(t)]$ 和所需的导数的来计算 $\hat{B}(t)$ 与 $\hat{D}^*(t)$ 。假设核函数 $K(\cdot)$ 在其支撑的内部至少两次连续可微分。他们建议用

$$\begin{aligned} \hat{f}(t) &= (Nh)^{-1} \sum_{i=1}^n \sum_{j=1}^{n_i} K\left(\frac{t_{ij} - t}{h}\right), \\ \hat{\sigma}^2(t) &= \frac{1}{Nh \hat{f}(t)} \sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ \hat{\epsilon}_i^2(t_{ij}) K\left(\frac{t_{ij} - t}{h}\right) \right\}, \\ \hat{\rho}_\epsilon(t) &= \frac{\sum_{i=1}^n \sum_{j_1 \neq j_2} \left\{ \hat{\epsilon}_i(t_{ij_1}) \hat{\epsilon}_i(t_{ij_2}) K\left(\frac{t_{ij_1} - t}{h}\right) K\left(\frac{t_{ij_2} - t}{h}\right) \right\}}{\sum_{i=1}^n \sum_{j_1 \neq j_2} \left\{ K\left(\frac{t_{ij_1} - t}{h}\right) K\left(\frac{t_{ij_2} - t}{h}\right) \right\}} \end{aligned}$$

和

$$\hat{E} [X^{(l)}(t)X^{(r)}(t)] = \frac{1}{Nh \hat{f}(t)} \sum_{i=1}^n \sum_{j=1}^{n_i} \left\{ X_i^{(l)}(t_{ij}) X_i^{(r)}(t_{ij}) K\left(\frac{t_{ij} - t}{h}\right) \right\}$$

来估计 $f(t)$, $\sigma^2(t)$, $\rho_\epsilon(t)$ 和 $E[X^{(l)}(t)X^{(r)}(t)]$ ，其中 $\hat{\epsilon}_i(t_{ij}) = Y_{ij} - X_i^T(t_{ij})\hat{\beta}(t_{ij})$ 为残差，然后通过 $\hat{f}(t)$ 、 $\hat{\beta}_t^{LSK}(t)$ 和 $\hat{E}[X^{(l)}(t)X^{(r)}(t)]$ 的相应导数来估计 $f(t)$ 、 $\beta_l(t)$ 和 $E[X^{(l)}(t)X^{(r)}(t)]$ 的第一与第二阶导数。通过大量模拟，这些作者表明总的来说，尽

管以上估计使用不同窗宽，但由 (3.22) 得到的交叉验证窗宽可以用来计算以上所有估计。

当协变量为时间不变时，也可将以上插入 (plug-in) 方法扩展到 (3.16) 中的 $\hat{\beta}_r^{COM}(t)$ 。Wu, Yu 和 Yuan (2000) 导出了 $\hat{\beta}_r^{COM}(t)$ 的偏差 $B(\hat{\beta}_r^{COM}; t)$ 和标准差 $SD(\hat{\beta}_r^{COM}; t)$ 的显式表达式，而且建议采用 $\beta_r(t)$ 的近似 $(1 - \alpha)$ 置信区间

$$\left\{ \hat{\beta}_r^{COM}(t) - \hat{B}(\hat{\beta}_r^{COM}; t) \right\} \pm Z_{1-\alpha/2} \widehat{SD}(\hat{\beta}_r^{COM}; t),$$

其中 $\hat{B}(\hat{\beta}_r^{COM}; t)$ 和 $\widehat{SD}(\hat{\beta}_r^{COM}; t)$ 分别为 $B(\hat{\beta}_r^{COM}; t)$ 和 $SD(\hat{\beta}_r^{COM}; t)$ 的插入估计。由于 $\hat{\beta}^{LSK}(t)$ 与 $(\hat{\beta}_r^{COM}; t)$ 类似，这里省略其细节。

以上的渐近区间不同于独立同分布数据下的区间，主要是因为它们的方差项中包括个体内相关。当 n_i 相对于 n 不可被忽略时，则 (3.27) 中的 θ 不可被忽略，从而由相关带来的影响也不可被忽略。对于 HIV/CD4 数据 (第 1.2.2 节)，重复测量数范围为 $1 \sim 14$ ，而个体数为 400，不考虑个体内相关的近似结果不能获得令人满意的近似。这种情况下，应直接由数据来估计相关。当重复测量数相对于个体数可以忽略时，如 ASGA 数据 (第 1.2.1 节)，在核估计的方差中可忽略个体内相关结构，此时可采用类似独立同分布样本的方法得到置信区间。

§3.8.2 同时置信区间

在大多数应用中，推断的主要目的是在适当的 t 范围内获得 $\beta_l(t)$ 的全部置信区间，而不是某一特定时间点的置信区间。当数据来自独立同分布样本时，回归曲线的同时置信区带可采用高斯过程极值理论 (如 Eubank 和 Speckman, 1993) 或通过网格点的逐点区间连接的可变动带 (如 Knafl, Sacks 和 Ylvisaker, 1985; Hall 和 Titterton, 1988; Härdle 和 Marron, 1991) 来构造。对于纵向样本，文献中还没有类似的极值渐近理论，后一种方法是目前纵向分析中仅有的、实用的同时推断的工具。为了构建基于最小二乘核估计 $\hat{\beta}^{(LSK)}(t)$ 的 $A^T \beta(t)$ 在 $t \in [a, b]$ 上的同时置信带，我们

可选择一个正整数 M , 将 $[a, b]$ 分割成 M 个相等区间, 即网点 $a = \xi_1 < \dots < \xi_{M+1} = b$ 满足 $\xi_{j+1} - \xi_j = (b - a)/M$, $j = 1, \dots, M$. $A^T \beta(\xi_j)$, $j = 1, \dots, M + 1$, 的一组近似 $(1 - \alpha)$ 同时置信区间为区间集合 $(\hat{l}_\alpha(\xi_j), \hat{u}_\alpha(\xi_j))$, $j = 1, \dots, M + 1$, 满足

$$\lim_{n \rightarrow \infty} P \left\{ \hat{l}_\alpha(\xi_j) \leq A^T \beta(\xi_j) \leq \hat{u}_\alpha(\xi_j) \text{ 对所有 } j = 1, \dots, M + 1 \right\} \geq 1 - \alpha. \quad (3.31)$$

Bonferroni 调整建议

$$(\hat{l}_\alpha(\xi_j), \hat{u}_\alpha(\xi_j)) = (\hat{L}_{\alpha/(M+1)}(\xi_j), \hat{U}_{\alpha/(M+1)}(\xi_j)), \quad (3.32)$$

其中 $(\hat{L}_\alpha(\xi_j), \hat{U}_\alpha(\xi_j))$ 在 (3.30) 中已定义.

为了建立一个覆盖网点 ξ_j , $j = 1, \dots, M + 1$ 间所有点的区间, 首先考虑 $A^T \beta(\xi_j)$ 的内插法, 即对于 $t \in [\xi_j, \xi_{j+1}]$, 有

$$(A^T \beta)^{(I)}(t) = \left\{ \frac{M(\xi_{j+1} - t)}{b - a} \right\} [A^T \beta(\xi_j)] + \left\{ \frac{M(t - \xi_j)}{b - a} \right\} [A^T \beta(\xi_{j+1})]. \quad (3.33)$$

则 $(A^T \beta)^{(I)}(t)$ 在 $t \in [a, b]$ 上的同时置信带为 $(\hat{l}_\alpha^{(I)}(t), \hat{u}_\alpha^{(I)}(t))$, 其中 $\hat{l}_\alpha^{(I)}(t)$ 与 $\hat{u}_\alpha^{(I)}(t)$ 为 $\hat{l}_\alpha(\xi_j)$ 与 $\hat{u}_\alpha(\xi_j)$ 的线性插值, 其定义类似于 (3.33). 然后网点间的间隙通过 $A^T \beta(t)$ 的光滑条件连接. 如果 $A^T \beta(t)$ 满足

$$\sup_{t \in [a, b]} \left| (A^T \beta)'(t) \right| \leq c_1, \quad c_1 > 0 \text{ 为已知常数}, \quad (3.34)$$

则对于所有 $t \in [\xi_j, \xi_{j+1}]$, 有

$$\left| A^T \beta(t) - (A^T \beta)^{(I)}(t) \right| \leq 2c_1 \left[\frac{M(\xi_{j+1} - t)(t - \xi_j)}{b - a} \right],$$

因此, $A^T \beta(t)$ 的近似 $(1 - \alpha)$ 的置信带为

$$\left(\hat{l}_\alpha^{(I)}(t) - 2c_1 \left[\frac{M(\xi_{j+1} - t)(t - \xi_j)}{b - a} \right], \hat{u}_\alpha^{(I)}(t) + 2c_1 \left[\frac{M(\xi_{j+1} - t)(t - \xi_j)}{b - a} \right] \right). \quad (3.35)$$

若 $A^T \beta(t)$ 满足

$$\sup_{t \in [a, b]} \left| (A^T \beta)''(t) \right| \leq c_2, \quad c_2 > 0 \text{ 为一个已知常数}, \quad (3.36)$$

则对于所有 $t \in [\xi_j, \xi_{j+1}]$, 有

$$\left| A^T \beta(t) - (A^T \beta)^{(I)}(t) \right| \leq \frac{c_2}{2} \left[\frac{M(\xi_{j+1} - t)(t - \xi_j)}{b - a} \right],$$

$A^T \beta(t)$ 的近似 $(1 - \alpha)$ 的置信带为

$$\left(\hat{l}_\alpha^{(I)}(t) - \frac{c_2}{2} \left[\frac{M(\xi_{j+1} - t)(t - \xi_j)}{b - a} \right], \hat{u}_\alpha^{(I)}(t) + \frac{c_2}{2} \left[\frac{M(\xi_{j+1} - t)(t - \xi_j)}{b - a} \right] \right). \quad (3.37)$$

其它不同于 (3.34) 与 (3.36) 的光滑条件, 也可类似地建立相应的置信区间。当协变量为时间不变时, 可基于 $\hat{\beta}^{COM}(t)$ 采用同样方法建立同时置信区间, 详见 Wu, Yu 和 Yuan(2000)。

§3.9 Bootstrap 可变动带

以上渐近推断受到两个约束: 第一, 由于目前仅得到两种核估计 $\hat{\beta}^{LSK}(t)$ 和 $\hat{\beta}^{COM}(t)$ 的渐近分布, 因此还得不到基于其它估计的置信区间。注意到样条和局部多项式等光滑方法比核方法具有更多理论与实际的优势, 尤其在 t 支撑的边界处, 故实际中也很需要基于这些光滑方法的推断方法。第二, 因为插入估计需要对出现在估计偏差与方差项中的设计密度、协方差函数以及其它量进行估计, 故该方法计算量通常很大, 也会在包含概率方面引入额外误差。以上这两个约束在某种程度上限制了渐近推断在纵向分析中的应用。

文献中一个更有吸引力的推断方法是 ”再抽样” Bootstrap 法。由于该方法是从原始数据的个体中再抽样, 保持了个体内的相关结构, 故该方法在纵向分析中具有广泛的魅力。尽管 ”再抽样” Bootstrap 法的理论性质还不清楚, 但其实际性能已经有大量的模拟研究。令 $\hat{\beta}(t) = (\hat{\beta}_0(t), \dots, \hat{\beta}_k(t))^T$ 是基于前面的任一个光滑方法构建的 $\beta(t)$ 的估计值。为了构建 $A^T E[\hat{\beta}(t)]$ 的近似的置信系数为 $(1 - \alpha)$ 的逐点的置信区间, 可采用以下步骤:

1. 从原数据集中有放回地随机抽取 n 个个体, 记所得的 Bootstrap 样本为 $\{(Y_{ij}^*, t_{ij}^*, X_{ij}^*), i = 1, \dots, n, j = 1, \dots, n_i\}$.

2. 采用以上 Bootstrap 样本和 $\hat{\beta}(t)$ 所采用的光滑方法, 计算 Bootstrap 估计 $\hat{\beta}^{boot}(t)$, 然后计算 $A^T \hat{\beta}^{boot}(t)$.
3. 重复以上两步骤 B 次, 便获得 B 个 Bootstrap 估计值 $A^T \hat{\beta}^{boot}(t)$ 。
4. 分别计算 B 个 Bootstrap 估计值 $A^T \hat{\beta}^{boot}(t)$ 的上、下第 $[100 \times (\alpha/2)]$ 百分位数 $L_{\alpha}^{boot}(t)$ 和 $U_{\alpha}^{boot}(t)$, 则近似的 $(1 - \alpha)$ 的 Bootstrap 区间为 $(L_{\alpha}^{boot}(t), U_{\alpha}^{boot}(t))$.

当 $A^T E[\hat{\beta}(t)]$ 满足光滑条件 (3.34) 或 (3.36) 时, 将 (3.35) 和 (3.37) 中的 (3.32) 用 $(L_{\alpha/(M+1)}^{boot}(\xi_j), U_{\alpha/(M+1)}^{boot}(\xi_j))$ 取代后便得到 $A^T E[\hat{\beta}(t)]$ 的近似的 Bootstrap 同时置信区间。

Bootstrap 法的主要优点是通用和简单。它不受核估计类型的限制, 不依赖于数据的相关与设计。尽管它具有很大的应用潜能, 但有关它的几个理论与实际问题尚有待解决。因为估计的偏差没有得到修正, 从而使得得到 $A^T \beta(t)$ 的置信区间或置信带并非总是具有期待的覆盖概率。如果也可得到偏差的相合估计, 则可通过纠正 $(L_{\alpha/(M+1)}^{boot}(\xi_j), U_{\alpha/(M+1)}^{boot}(\xi_j))$ 中所出现的偏差, 改进 $A^T \beta(t)$ 的置信域。目前仅具体情况逐一获得了偏差的相合估计, 尚无一般方法。第 4 步中百分位数的端点的另一自然方法是考虑采用正态近似区间界值 $A^T \hat{\beta}(t) \pm z_{(1-\alpha/2)} \hat{se}^{boot}(t)$, 其中 $\hat{se}^{boot}(t)$ 为 B 个 Bootstrap 估计值 $A^T \hat{\beta}^{boot}(t)$ 的样本标准差。关于百分位数和正态近似的这两种 Bootstrap 法的渐近性质, 尚有待进一步研究。

§3.10 阿拉巴马州胎儿生长研究

胎儿是否正常生长自然被认为会影响婴儿的生存和儿童的正常发育。本研究的目的是为了了解母亲危险因素和身体测量值对胎儿生长方式的影响。尽管胎儿腹围、顶骨间径、股骨长度的测量结果均依赖于时间, 但感兴趣的协变量既可以是时变的也可以是时不变的, 每次访问由超声波所测量的母体胎盘厚度就是一个典型的时变协变量, 怀孕初期母亲的身高、体重和体重指数是时不变协变量。还有一些其

它变量，如母亲抽烟、饮酒习惯等，可能为时变也可能为时不变的，这就看如何定义这些变量。将母亲吸烟和饮酒状况定义为时不变的一个简单方法是将母亲分为吸烟（曾在孕期吸烟）与不吸烟（在孕期从不吸烟），不饮或少饮酒（在孕期平均每天饮一瓶啤酒或一杯白酒，或更少）与大量饮酒（在孕期平均每天饮一瓶以上啤酒或一杯以上白酒）。正如大多数的调查问卷，数据包括了每人每天的吸烟和饮酒平均数，当被调查对象在研究期间改变其行为时，吸烟和饮酒的实际消耗就明显变成了时变变量。分析中可根据特定的研究问题，来考虑吸烟与饮酒为分类（时不变）或实际消耗（时变）协变量。

为叙述方便，以下分析将母亲吸烟与饮酒视为分类协变量，来考虑母亲吸烟、饮酒和胎盘厚度对胎儿腹围生长的影响。只要模型具有清晰、有意义的生物学解释，则可采用类似地研究其它协变量对测量结果的影响。尽管图 1 的总趋势为上升趋势，但几乎不能为胎儿生长和感兴趣协变量之间的关系提供任何线索。一般可先由 (1.2) 进行非参数分析。

令 $Y(t)$ 和 $X^{(1)}(t)$ 分别为怀孕 t 周的胎儿腹围和胎盘厚度， $X^{(2)}$ 和 $X^{(3)}$ 分别为母亲饮酒和吸烟分类，定义为

$$X^{(2)} = \begin{cases} 1 & \text{如果她不吸烟 / 少吸烟,} \\ 0 & \text{如果她大量吸烟,} \end{cases} \quad X^{(3)} = \begin{cases} 1 & \text{如果她吸烟,} \\ 0 & \text{其它,} \end{cases}$$

$X^{(4)}$ 为孕初母亲的身高（厘米）。

鉴于胎盘发育也可能受到饮酒与吸烟的影响，首先考虑时不变协变量向量 $X = (1, X^{(2)}, X^{(3)}, X^{(4)})^T$ 的效应。尽管可用 $(Y(t), t, X)$ 直接拟合 (1.2)，用 $\beta(t) = (\beta_0(t), \beta_2(t), \beta_3(t), \beta_4(t))^T$ 描述协变量效应，但如果用中心化后的变量 $X^{(*4)} = X^{(4)} - E[X^{(4)}]$ 取代 $X^{(4)}$ ，则可获得更好的生物学解释，其中 $\beta^*(t) = (\beta_0^*(t), \beta_2(t), \beta_3(t), \beta_4(t))^T$ 为相应的协变量效应。对于后者，基线系数曲线 $\beta_0^*(t)$ 代表了不吸烟、不饮或少饮酒、平均身高的母亲在怀孕 t 周时胎儿的平均腹围；而对于前者， $\beta_0(t)$ 本身无实际生物学意义。

为了用 $(Y(t), t, X^{(*)})$ 拟合模型 (1.2), 其中 $X^{(*)} = (1, X^{(2)}, X^{(3)}, X^{(*)4})^T$, $X_i^{(*)4}$, $i = 1, \dots, 1475$, 是由 $X_i^{(4)}$ 减去 $\{X_j^{(4)}; j = 1, \dots, 1475\}$ 的样本均值计算的。图 2 给出了估计的系数曲线, 包括基线曲线、饮酒、吸烟和母亲身高的效应曲线, 以及它们相应的 95% 同时置信带。这些系数曲线是由 (3.16) 采用 Epanechnikov 核、交叉验证窗宽 (3.22) 和 $w_i = 1/(nn_i)$ 的逐条估计计算的。值得注意的是数据集中重复测量的次数大多数为 4 左右, 这相对于个体数 $n = 1475$ 而言非常小。因此, 通过假定 n 趋于无穷大, 而 n_i 为有限, 所得到的渐近结果可望获得很好的近似。对于核光滑估计, 这就意味着 $w_i = 1/(nn_i)$ 和 $w_i = 1/N$ 都可产生十分类似的结果, 且在估计的渐近方差中可以忽略个体内相关。因此, 在构建渐近的近似置信区间时就不需要估计协方差。在系数曲线估计中基于同样的核与窗宽, 采用渐近近似 (3.35) 和 Bonferroni 调整 ($M = 40$, $c_1 = 5$) 可计算同时置信带。这些图形提示基线曲线 $\beta_0^*(t)$ 是线性向上的, 而饮酒、吸烟和母亲身高的效应探测不到。这可能是因为这里采用的置信区间趋于保守, 故它们没有足够的灵敏性来探测到小的协变量效应。于是我们也采用 (3.6) 的最小二乘核估计方法计算了曲线的估计及其相应的置信区带, 然而这些结果与图 2 显示的结果类似, 所有在此省略。

以上图 2 的非参数结果表明: 胎儿腹围 $Y(t)$ 与孕龄 t 、饮酒 $X^{(2)}$ 、吸烟 $X^{(3)}$ 和中心化母亲身高 $X^{(*)4}$ 之间的关系可用线性模型来描述, 即

$$Y(t) = \beta_{00} + \beta_{01}t + \beta_2X^{(2)} + \beta_3X^{(3)} + \beta_4X^{(*)4} + \epsilon(t)$$

, 其中 $(\beta_{00}, \beta_{01}, \beta_2, \beta_3, \beta_4)$ 为未知参数, $\epsilon(t)$ 为均值为 0 的误差过程。可利用 S-plus 的混合效应程序 (mixed-effects procedure) 拟合模型 (Bates 和 Pinheiro, 1999)。表 1 给出了由以上线性模型和 S-plus 程序计算的参数估计和相应的标准差。从该线性模型分析结果看: 饮酒和母亲身高的效应不显著, 吸烟具有微弱正效应。而线性分析中显示的微弱吸烟效应可能是由数据的随机变动导致的, 而不是显示胎儿大小与吸烟间有任何实质的关联性。这些结果与上述非参数分析获得的结果大致相同。

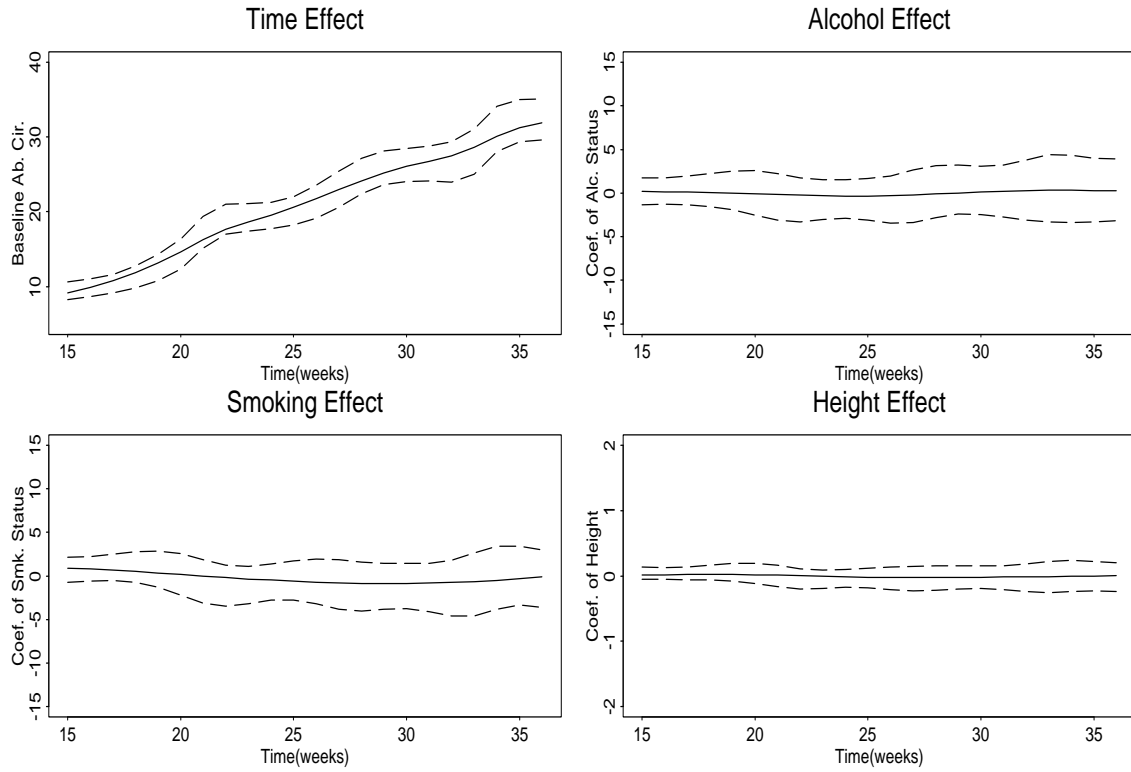


图 2 实线：用 Epanechnikov 核，交叉验证窗宽和 $w_i = 1/(nn_i)$ 计算的协变量系数曲线的逐元核估计。虚线：置信系数为 95% Bonferroni 型置信带。

当将胎盘厚度 $X^{(1)}(t)$ 加入模型时，对时变协变量必须进行光滑。为了使基线系数曲线具有生物学意义，我们采用了中心化协变量 $X^{(*1)}(t) = X^{(1)}(t) - E[X^{(1)}(t)]$ ，即时间 t 处个体胎盘厚度与条件均值的差。为了避免模型一开始就有太多的协变量，首先将 $(t, X^{(*1)}(t), X^{(*4)})$ 作为协变量向量来拟合 (1.2)。图 3(a) 是采用了具有标准高斯核、交叉验证窗宽和 $w_i = 1/N$ 的两步核方法 (3.11) 估计了 $X^{(*1)}(t)$ 的系数曲线。正如它没有清楚的生物学解释一样，这一估计表现出不光滑。而图 3(b) 是采用了窗宽向量 $(\gamma_1, h_0, h_1, h_4) = (1.5, 1.0, 2.0, 1.0)$ 外，其它均与图 3(a) 相同的计算方法估计的 $X^{(*1)}(t)$ 的系数曲线，它就具有较清晰的生物学解释。选择这一窗宽向量主要是因为其交叉验证得分与交叉验证窗宽的交叉验证得分十分接近。这里采用 Bootstrap 百

表 1 用 S-plus 的混合效应程序计算的参数估计及其标准差.

	参数估计	标准差	Z-ratio
β_{00}	-6.5496	0.0614	-106.5880
β_{01}	1.0645	0.0021	496.1262
β_2	0.0026	0.0551	0.0478
β_3	0.1009	0.0516	1.9555
β_4	0.0007	0.0035	0.1996

分位数区间来显示估计的变动情况, 因为目前尚无基于渐近近似的该类型估计的统计推断。

图 3 至少得出了胎盘厚度与胎儿腹围间具有一定正相关的定性结论。估计的中心化母亲身高 $X^{(*4)}(t)$ 的系数曲线接近于 0, 提示母亲身高效应无统计意义。由于估计的基线系数曲线与图 2 接近, 因此这里省略。将母亲饮酒 $X^{(2)}$ 与吸烟 $X^{(3)}$ 状况加入模型后获得的效应与图 2 非常相似, 故这里也省略了该分析。

§3.11 MACS 的 CD4/HIV 研究

令 t_{ij} 代表第 i 个体, 第 j 次测量的已感染 HIV 的时间长度 (年). 研究的目的是为了评价感染前 CD4 百分率 $X^{(1)}$ 与吸烟状况 $X^{(2)}(t)$ 两因素对 HIV 感染后 CD4 损耗百分率 $Y(t)$ 随时间的效应。第一个协变量 $X^{(1)}$ 不依赖于已 HIV 感染的时间。第二个协变量 $X^{(2)}(t)$ 等于 1 如果个体在时间点 t 为吸烟者, 否则等于 0。注意到某些个体在研究期间会改变它们的吸烟习惯, 故 $X^{(2)}(t)$ 为时变变量。由于缺乏现存的描述这些变量间关系的参数与半参数模型, 所以初步分析采用非参数模型 (1.2) 是合理的。

与 ASGA 研究分析采用的原理相同, 为了便于生物学解释, 模型 (1.2) 采用中心化变量 $X^{(*1)} = X^{(1)} - E[X^{(1)}]$ 要比未中心化变量 $X^{(1)}$ 更好。但由于 $X^{(2)}(t)$ 为时变的二分变量, 所以没有必要被中心化。因此, 首先用 $X_i^{(1)}$ 减去其样本均值后获得

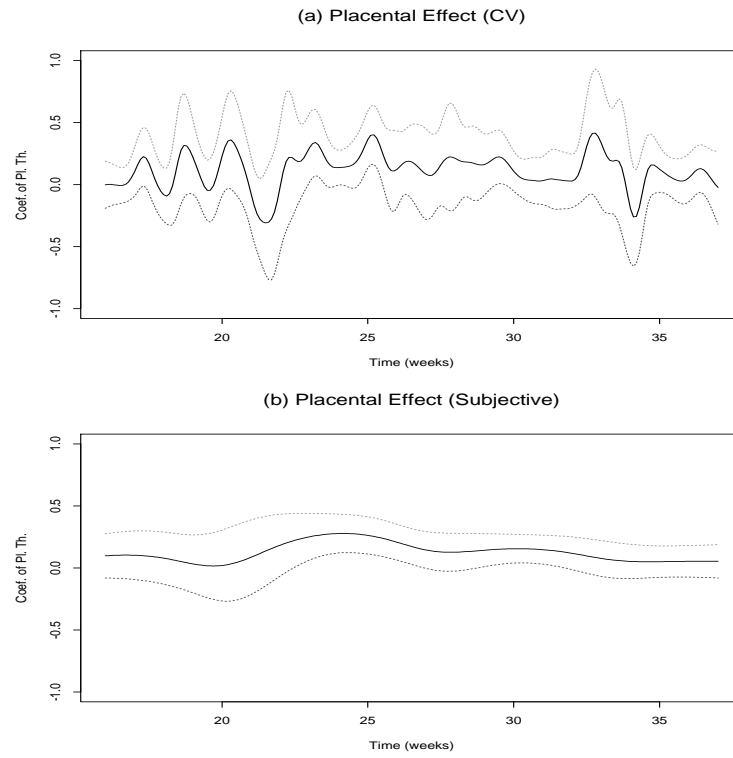


图 3 实线：采用标准高斯核、 $w_i = 1/N$ 、交叉验证窗宽 (图 (a)) 和窗宽向量 $(\gamma_1, h_0, h_1, h_4) = (1.5, 1.0, 2.0, 1.0)$ (图 (b))，利用 (3.11) 计算的胎盘厚度的系数曲线 (即协变量效应)。虚线：采用”再抽样”Bootstrap 百分位数点计算的 95% 逐点区间

$X_i^{(*1)}$ ，再采用数据 $\{(Y_{ij}, t_{ij}, X_{ij}^*); i = 1, \dots, 400, j = 1, \dots, n_i\}$ 来拟合模型 (1.2)。基线系数曲线 $\beta_0^*(t)$ 表示在时间 t 处不吸烟，且具有平均感染前 CD4 百分率水平的个体，在感染后 t 年时的平均 CD4 百分率。 $\beta_1^*(t)$ 和 $\beta_2(t)$ 分别为 $X^{(*1)}$ 和 $X^{(2)}(t)$ 的效应，其解释与一般情形相同。

该例与上例的估计与推断相比，除了协变量中心化有所不同外，还存在着另一个重要的不同，即该数据集中个体重复测量次数相对于个体数而言不能被简单忽略。因此，尽管 $w_i = 1/(nn_i)$ 和 $w_i = 1/N$ 仍看起来是两个合理的权的选择，但至少对于已知的核估计情形，在假定 n 趋于无穷且 n_i 有界的条件下的渐近近似可能得不到方差的适当估计。由于数据的相关结构完全未知，且很难被准确估计，所以 Wu, Chiang 和 Hoover(1998) 建议在这种情况下用 $|\rho_\epsilon(t)|$ 的上限，即方差 $\sigma^2(t)$ 取代 (3.28) 中的协方差 $\rho_\epsilon(t)$ 获得保守的 Bonferroni 型置信带是合理的。图 4 给出了个体随时间的 CD4 损耗百分率、估计的系数曲线和相应的保守 Bonferroni 型 95% 渐近置信带，其中估计系数曲线是采用了 Epanechnikov 核、交叉验证窗宽和权 $w_i = 1/N$ 的 (3.6) 计算的，置信带是采用 $M = 138$, $c_1 = 3$ 并将 $\rho_\epsilon(t)$ 用 $\sigma^2(t)$ 取代后的 (3.30) 和 (3.35) 计算的，在计算所有插值核估计 (3.30) 所要求的核和窗宽可采用与 (3.6) 中的相同。

由图 4(b) 可见，随 HIV 感染时间的增长，基线 CD4 百分率曲线逐渐下降，这与图 4(a) 的基本趋势一致。感染前 CD4 百分率的系数曲线的同时置信带至少在 HIV 感染后的前 4 年为正值，明确表明了高的感染前的 CD4 水平在感染初期是很有好处的。但在感染后期，感染前 CD4 百分率对感染后 CD4 百分率的正效应就逐渐消失。尽管图 4(c) 表明在这个数据集所考虑的整个 7 年时间里，估计的曲线都是正的，但该曲线的置信带没有显示吸烟与感染后 CD4 水平间存在显著的正相关。这可能是由于两变量间存在弱相关或我们给的置信区间较为保守所造成。显然，这里仅获得了对数据的探索性分析结果，而能提供其它有意义的生物学机制的生物学

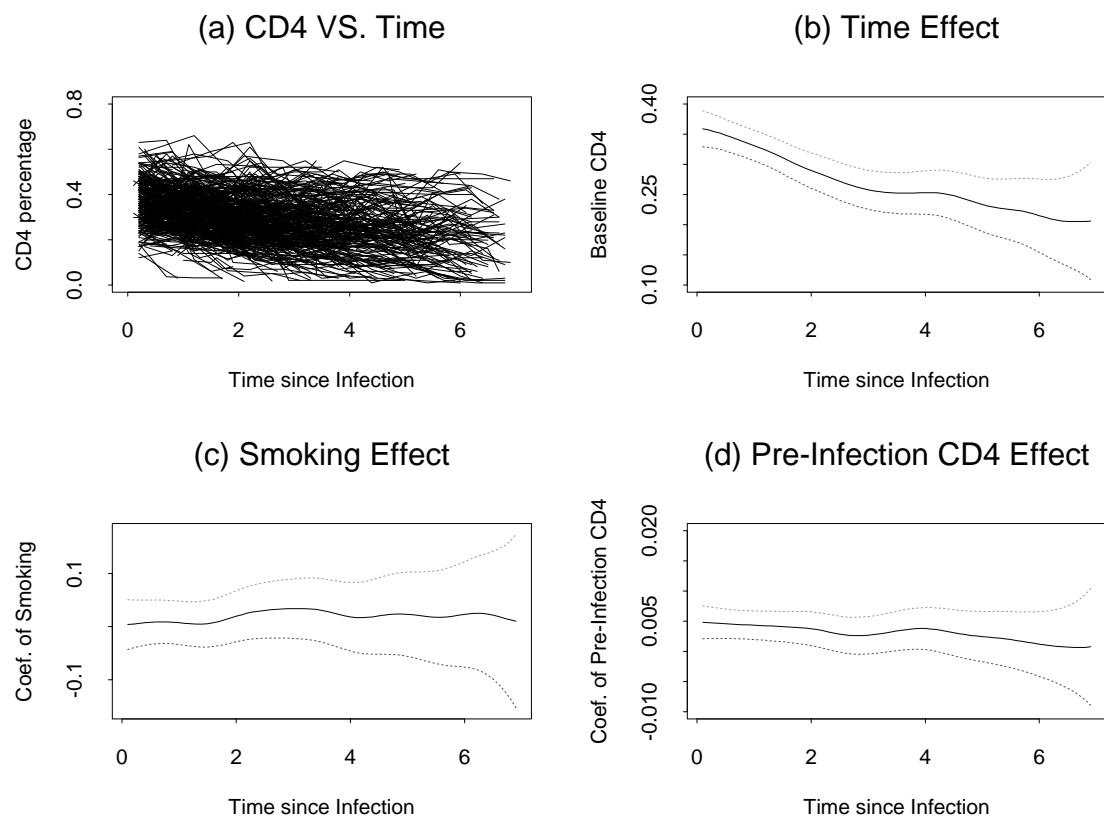


图 4 (a) 个体 CD4 百分率与 HIV 感染后的时间 (年), (b)-(d) 估计的 CD4 百分率基线、吸烟的系数曲线与 HIV 感染前 CD4 百分率的系数曲线 (实线) 和其相应的 95% 同时置信带 (点线).

关联关系和参数模型还有待于其它试验进一步研究与证实。不过如前所述,采用非参数回归,尤其是变系数模型,所得到的纵向数据的初步探索是很清晰的。

§3.12 小结与讨论

本章给出了一系列纵向资数据分析的参数、半参数、非参数模型,以及相应的估计和推断方法,这些方法在生物医学研究中具有广泛的应用。文献中,参数模型的理论与方法,尤其是线性模型,已得到了广泛的研究。基于参数模型的估计与推断也很容易通过 SAS、S-plus 等现有的统计软件来实现。然而基于半参数和非参数模型的方法且代表了这个活跃研究领域的最新进展。

这里介绍的非参数估计方法与推断方法均基于变系数模型的总体框架。对于大的纵向研究,这些方法具有灵活性的优势。在实践中,采用局部多项式与样条对模型进行光滑的方法各有其优缺点。总的来说,当协变量为时不变时,逐元光滑方法较灵活,且容易计算,但基于一般、惩罚最小二乘方法和基近似则可用于协变量为时变或时不变的两种情况。系数曲线的逐点和同时置信区间可采用渐近近似或“Bootstrap”再抽样”来构造,其中渐近置信区间仅有由核方法导出的相关结果。原则上,“Bootstrap”再抽样”法可用于任意光滑估计。然而尽管有大量模拟研究表明了 Bootstrap 的可用性,但其理论性质还未被研究。两步光滑方法看起来是克服一般最小二乘估计的某些缺点,但为了将该方法用于非平衡纵向研究,我们还需进一步研究计算初始估计值的具体方法以及最终估计的渐近性质。最后,采用统一的权 $w_i = 1/N$ 还是统一的个体权 $w_i = 1/(nn_i)$,要由实际应用决定。尽管没有哪一个权在所有纵向设计中一致的优于另一个,理想的权可能依赖于未知的相关结构以及相对于 n , $n_i, i = 1, \dots, n$, 有多快能趋于无穷,但目前文献中的大量模拟研究表明:当所有个体的重复测量数差不多时,选择两种权都是适合的,当个体重复测量数彼此相差很大时,通常采用权 $w_i = 1/(nn_i)$ 更好。

此外,还有大量课题值得进一步研究。首先,尽管估计方法与置信区间方法

在纵向分析中重要，但在生物医学中更为重要的是评价不同假设的统计显著性检验方法。这些方法将可以解释科学假设的参数子模型从一般非参数模型中区分出来。而作判断的关键是给出合适的检验统计量的分布。另一个重要的实际问题就是改进置信程序。本文所给的方法很保守，这常常妨碍其实际应用。进一步的工作是在保持覆盖概率满意的同时能缩小置信带的宽度。最后，鉴于变系数模型对许多纵向数据仍不适用，故有必要进一步扩展这些模型。一个有用的扩展就是考虑回归模型结果变量依赖于协变量的历史的和当前取值的情形。对于这种扩展所有估计与推断方法都必须重新建立。

§4 基于分布函数的非参数模型

§4.1 动机和合理性

在第 2 节和第 3 节所讨论的回归模型普遍是基于给定一组协变量下对响应变量的条件均值和协方差结构进行建模，其中协变量可以是时变的，也可以是时不变的。由于该均值和协方差结构能被参数化或非参数化，基于回归模型的条件均值无疑成为了纵向分析的首要工具，并构成了文献中已有方法的主要部分。尽管该方法很流行，但实际上当条件均值和协方差不适合回答所研究的科学问题时，这类方法就不再适合。当结果变量的条件分布不能用正态分布近似，且研究的目标是通过其条件分布来定义时，这种现象就会出现。

1.3 节的 NGHS 数据 (例 3) 就是一个典型的例子，其中许多相关的科学问题要通过评估条件分布才能较好地回答。如 1.3 节中所描述的，NGHS 研究的一个重要目标是评估青少年期年龄、种族和肥胖对由相应的血压和高血压等风险因子决定的几个心血管风险因子和暂时的心血管健康趋势的影响。由于 NGHS 研究中所观测到的心血管风险因子的条件分布未知的，通常为非高斯分布，故推断年龄、种族和肥胖对这些风险因子的条件均值的效应可能无有意义的临床试验解释，这是因为对条

件均值的统计效应与关于感兴趣总体的心血管健康状态的分布并没有直接联系。一个更有意义的方法是研究年龄、种族和肥胖对风险因子的分布的影响，这样就可以用各种临床干预来减少向不预期的健康状态发展的可能。

§4.2 无结构的条件分布模型

令 $F_t[y|\mathbf{X}(t)] = P[Y(t) \leq y|\mathbf{X}(t), t]$ 为给定 $\{t, \mathbf{X}(t)\}$ 下的 $Y(t)$ 的累积分布函数 (CDF). 如果没有对 $\{t, \mathbf{X}(t)\}$ 和 $F_t[y|\mathbf{X}(t)]$ 之间的关系作任何结构假设, 则基于纵向样本 $\{(Y_{ij}, t_{ij}, \mathbf{X}_{ij}) : i = 1, \dots, n, j = 1, \dots, n_i\}$ 的 $F_t[y|\mathbf{X}(t)]$ 的估计可以通过扩展 Hall, Wolff 和 Yao (1999) 提出的核方法到当前的抽样框架来给出。当 $\mathbf{X}(t)$ 中涉及的协变量的个数很小时, 如 $K = 0$ 或 1 , 此时该扩展就很简单, 尤其是当 $K = 0$ 时, 就可直接应用在 Hall, Wolff 和 Yao (1999) 的第 3 节所描述的一元的局部 logistic 方法和调整的 Nadaraya-Watson 估计。而当 $K \geq 1$, 就必须考虑在 Hall, Wolff 和 Yao (1999) 的第 3.2 节所介绍的多元估计, 当协变量的个数 K 很大时, 由于著名的“维数祸根”(curse of dimensionality) 问题, 多元和光滑方法在计算上就很不稳定 (Fan 和 Gijbels, 2006, p.264)。此外, 在实际情形下, 也很难解释基于 $F_t[y|\mathbf{X}(t)]$ 的完全无结构非参数估计得到的统计结果和推断。这些潜在的缺点, 即计算的不稳定性和难于解释, 使得 $F_t[y|\mathbf{X}(t)]$ 的无结构的光滑估计不实用。本节剩下的部分将介绍基于 $F_t[y|\mathbf{X}(t)]$ 的时变变换模型, 这类结构非参数模型的估计和推断问题。

§4.3 时变变换模型

针对时间事件数据的变换模型在生存分析中已被广泛地研究, 关于各种右删失时间事件数据的估计和推断的方法可参见 Cheng, Wei 和 Ying (1995, 1997), Lu 和 Ying (2004), Lu 和 Tsiatis (2006), Zeng 和 Lin (2006) 等。与基于条件均值的回归模型相比, 变换模型能够为条件分布提供一类函数结构, 它已被显示是一个近似条件分布函数的有效降维的方法。

通过扩展变系数方法到变换模型, Wu, Tian 和 Yu (2010) 表示 $F_t[y|\mathbf{X}(t)]$ 可通过时变线性变换模型

$$g\{S_t[y|\mathbf{X}(t)]\} = h(y, t) + \mathbf{X}^T(t) \beta(t) \quad (4.1)$$

来建模, 其中 $g(\cdot)$ 是一个已知的、单调减的连接函数, $S_t[y|\mathbf{X}(t)] = 1 - F_t[y|\mathbf{X}(t)]$, $h(\cdot, \cdot)$ 是未知的关于 y 严格增的基线函数, $\beta(t) = (\beta_0(t), \dots, \beta_K(t))^T$, 且 $\beta_k(t)$ 是在时间域 \mathcal{T} 上的光滑函数。对于时间域内的每一个点 t , 模型 (??) 是一个非参数基线函数为 $h(y, t)$ 和线性系数为 $(\beta_0(t), \dots, \beta_K(t))^T$ 的半参数线性变换模型。在任意给定时间 t 处, 协变量的效应 $\beta_k(t)$ 表示 $X^{(k)}(t)$ 每增加一个单位时 $g\{S_t[y|\mathbf{X}(t)]\}$ 对所有 $y \in R$ 的改变量。由于 y 对 $F_t[y|\mathbf{X}(t)]$ 的效应可概括为 $h(y, t)$, 故当 t 固定时, $X^{(k)}(t)$, $k = 0, \dots, K$, 对 $F_t[y|\mathbf{X}(t)]$ 的协变量效应对 y 的所有可能取值都是常数。因此, 与无结构的非参数条件分布模型相比, 通过假定函数的线性结构, (4.1) 的模型复杂性被大大降低。

本节都假设连接函数 $g(\cdot)$ 的形式已知, 是研究者根据具体的分析目标选择的。知名的特殊情形包括比例分线模型 $g\{S_t[y|\mathbf{X}(t)]\} = \log\{-\log[S_t(y|\mathbf{X}(t))]\}$ 和比例优势模型 $g\{S_t[y|\mathbf{X}(t)]\} = -\log\{S_t[y|\mathbf{X}(t)]/F_t[y|\mathbf{X}(t)]\}$ 。事实上, (4.1) 的拟合可以通过图评价, 首先将时间范围分成若干时间小段, 然后对所有的时间小段内的 t , 查看 $g\{S_t[y|\mathbf{X}(t)]\}$ 对 $X^{(k)}(t)$ $k = 0, \dots, K$, 的图的线性性。虽然带未知连接函数的模型 (4.1) 的估计和预测是很值得大量研究的问题, 但是现有的文献中仍无这方面的结果。为符号简单, 记 n 个个体中的每个个体的观测时间点是 $J > 1$ 个不同设计时间点 $\mathbf{t} = (t_{(1)}, \dots, t_{(J)})^T$ 的一个随机选择的子集。由于不是所有的个体在每个时间 $t_{(j)}$ 都能被观测到, 记 \mathcal{S}_j 为在时间 $t_{(j)}$ 处可观测到的个体的集合, 则 $\mathcal{Z} = \{Y_i(t_{(j)}), \mathbf{X}_i(t_{(j)}), t_{(j)}; i \in \mathcal{S}_j, j = 1, \dots, J\}$ 为 $\{Y(t), \mathbf{X}(t), t \in \mathcal{T}\}$ 的纵向样本, $\mathcal{D} = \{\mathbf{X}_i(t_{(j)}), t_{(j)}; i \in \mathcal{S}_j, j = 1, \dots, J\}$ 为观测到的协变量集合, 这里 $Y_i(t_{(j)})$ 和 $\mathbf{X}_i(t_{(j)}) = (X_{i0}(t_{(j)}), \dots, X_{iK}(t_{(j)}))^T$ 分别为当 $i \in \mathcal{S}_j$ 时第 i 个体在时间 $t_{(j)}$ 处的结果

向量和协变量向量。令 $m_j = \#\{i \in \mathcal{S}_j\}$ 为 \mathcal{S}_j 中个体的个数, $m_{j_1 j_2} = \#\{i \in \mathcal{S}_{j_1} \cap \mathcal{S}_{j_2}\}$ 为 \mathcal{S}_{j_1} 和 \mathcal{S}_{j_2} 中都出现的个体的个数, 这里 $j_1 \neq j_2$. 显然, $m_{j_1 j_2} \leq \min(m_{j_1}, m_{j_2})$.

本节的主要结果是关于 (4.1) 中时变协变量效应 $\beta(t)$ 的估计和推断, 以及统计结果的实际解释。除了 $\beta(t)$ 的估计外, 在生物医学研究中, $h(y, t)$, $F_t[y|\mathbf{X}(t)]$ 和它们的函数 (如条件分位数) 的非参数光滑估计与预测也具有重要的应用。例如, 对任何给定的 $\mathbf{X}(t) = \mathbf{x}$, $F_t[y|\mathbf{x}]$ 的合适的非参数预测可以被用来识别不同时期总体中可能处于高风险的子群。但模型 (4.1) 下 $h(y, t)$, $F_t[y|\mathbf{X}(t)]$ 以及它们的函数的非参估计与预测的方法和理论尚在发展中, 因此这里不将讨论这部分内容。在使用线性变换模型的生存分析中, 大部分估计方法是关于随机删失的时间 - 事件数据的, 如 Cheng, Wei 和 Ying (1995, 1997)。然而在纵向研究中, 尽管理论可能, 但重复测量的结果变量和协变量很少会出现删失的情形。因此在模型 (4.1) 在估计 $\beta(t)$ 时, 该数据集里没有变量删失。

§4.4 两步光滑方法

如上讨论, 当 t 固定时, 模型 (4.1) 就归为 Cheng, Wei 和 Ying (1995) 中的线性变换模型, 故可应用他们文章中所给出的估计方程组来估计 $\beta(t)$ 。而当 t 在 \mathcal{T} 内变化时, 估计 (4.1) 中系数曲线的有用的光滑方法是先在时间设计点 \mathbf{t} 处计算出一组系数曲线的初始估计值, 然后基于时间点 \mathbf{t} 处的初始估计值计算系数曲线在 $t \in \mathcal{T}$ 上的光滑估计。

§4.4.1 系数的初始估计

将 $t_{(j)}$ 时间点的观测值代入 Cheng, Wei 和 Ying (1995) 估计方程组 (4.1), 从中便可得到系数 $\beta(t_{(j)})$ 的估计。记

$$\epsilon_{i(j)} = g\left\{S_{t_{(j)}}\left[Y_i(t_{(j)})\middle|\mathbf{X}_i(t_{(j)})\right]\right\}.$$

由

$$P[\epsilon_{i(j)} \leq u | \mathbf{X}_i(t_{(j)}), t_{(j)}] = P\left\{S_{t_{(j)}}[Y_i(t_{(j)}) | \mathbf{X}_i(t_{(j)})] \geq g^{-1}(u) | \mathbf{X}_i(t_{(j)}), t_{(j)}\right\}$$

和类似的 Cheng, Wei 和 Ying (1995) 方程组 (1.4) 的推导, 可证实 (4.1) 等价于

$$h[Y_i(t_{(j)}), t_{(j)}] = -\mathbf{X}_i^T(t_{(j)})\beta(t_{(j)}) + \epsilon_{i(j)}, \quad (4.2)$$

其中 $\epsilon_{i(j)}$ 为随机误差, 其分布函数为 $G(\cdot) = 1 - g^{-1}(\cdot)$. 令

$$Z_{i_1, i_2}(t_{(j)}) = 1_{[Y_{i_1}(t_{(j)}) \geq Y_{i_2}(t_{(j)})]}, \quad \mathbf{X}_{i_1, i_2}(t_{(j)}) = \mathbf{X}_{i_1}(t_{(j)}) - \mathbf{X}_{i_2}(t_{(j)})$$

和 $\xi(s) = \int_{-\infty}^{\infty} \{1 - G(t+s)\} dG(t)$. 由 (4.2) 可得

$$\begin{aligned} E[Z_{i_1, i_2}(t_{(j)}) | \mathbf{X}_{i_1}, \mathbf{X}_{i_2}, t_{(j)}] &= P\left\{h[Y_{i_1}(t_{(j)}), t_{(j)}] \geq h[Y_{i_2}(t_{(j)}), t_{(j)}] | \mathbf{X}_{i_1}, \mathbf{X}_{i_2}, t_{(j)}\right\} \\ &= P[\epsilon_{i_1, (j)} - \epsilon_{i_2, (j)} \geq \mathbf{X}_{i_1, i_2}^T(t_{(j)})\beta(t_{(j)})] \\ &= \xi[\mathbf{X}_{i_1, i_2}^T(t_{(j)})\beta(t_{(j)})]. \end{aligned} \quad (4.3)$$

$\beta(t_{(j)})$ 的一个初始估计就是估计方程

$$\sum_{i_1 \neq i_2 \in \mathcal{S}_j} U_{i_1 i_2} [\tilde{\beta}(t_{(j)})] = 0, \quad (4.4)$$

的解 $\tilde{\beta}(t_{(j)})$, 其中 $w(\cdot)$ 是一个已知的权函数,

$$U_{i_1 i_2} [\beta(t_{(j)})] = w[\mathbf{X}_{i_1, i_2}^T(t_{(j)})\beta(t_{(j)})] \mathbf{X}_{i_1, i_2}(t_{(j)}) \left\{ Z_{i_1, i_2}(t_{(j)}) - \xi[\mathbf{X}_{i_1, i_2}^T(t_{(j)})\beta(t_{(j)})] \right\}.$$

在 Cheng, Wei 和 Ying (1995) 的第 2 节和附录 1 中, 已表明如果权 $w(\cdot)$ 是正的, 他们的估计方程有近似唯一的解, 而当 $w(\cdot) = 1$ 且方程 (2.1) 的矩阵 $\sum \sum Z_{ij} Z_{ij}^T$ 为正定时, 他们的估计方程有唯一解。对每个固定的 $t_{(j)}$, 估计方程 (4.4) 等于无删失的 Cheng, Wei 和 Ying (1995) 的估计方程 (2.2), 因此估计方程 (4.4) 解的唯一性和近似唯一性的结论也成立。尽管仍没有得到 $\tilde{\beta}(t_{(j)})$ 的有限样本的均值和方差的

显示表达式，但对固定点 $t_{(j)}$ 的 $\tilde{\beta}(t_{(j)})$ 渐近性质与 Cheng, Wei 和 Ying (1995) 中得到性质相同。当不同的时间点 \mathbf{t} 都被考虑时，个体观测数据内的潜在相关性暗示不同时间点处的初始估计也是潜在相关的。本节随后给出了 $\tilde{\beta}(t_{(j)})$ 在不同时间点的均值，协方差和方差的大样本渐近性质。

值得注意的是假设有 J 个不同时间点 $\mathbf{t} = (t_{(1)}, \dots, t_{(J)})^T$ 是数学简化，目的是为了简化理论讨论。在实际上生物学样本中，设计时间点通常不是提前规定的， \mathbf{t} 可基于生物学或临床试验的合理性从实际化整后的观测时间点中选择的。当不同时间点的个数很大且在 t_j 处的观测到的个体个数非常少时，初始估计 $\tilde{\beta}(t_j)$ 在某个时间点 $t_{(j)}$ 可能不存在。在这种情形下，一个实用的方法是将观测时间点分组到小的时间面元 (bins)，其第 j 时间面元是以 $t_{(j)}$ 为中点，使得初始估计在每个时间段都能被计算。Fan 和 Zhang (2000) 考虑了基于变系数模型的条件均值的两步估计的面元方法。当使用面元方法时，要求时间面元要小，能使得初始估计，相对于在光滑步计算的光滑参数，是不光滑的。各种面元选择实用的效应是值得但还没有被系统研究。本节全节假设 \mathbf{t} 已包含了选择合适的时间面元的中点，这样就没有必要再重新分化。这个假设的实际意义就是 \mathbf{t} 在所有的时间点处，都有足够多的个体有观测值。

§4.4.2 系数曲线的光滑估计

基于在时间点 $\mathbf{t} = \{t_{(1)}, \dots, t_{(J)}\}$ 的初始估计，对所有 $t \in \mathcal{T}$ ，都能构造 $\beta(t)$ 的一个光滑估计。除了初始估计外，还必须有光滑步的重要的原因是初始估计仅是针对是时间点 $\{t_{(1)}, \dots, t_{(J)}\}$ 的系数，而光滑步能够计算 \mathcal{T} 内所有的 t 的曲线估计。此外，初始估计在不同的实际时间点上有很大的变动，如此“尖”的估计总体上是不会有有意义的生物学解释。通过光滑步，共享临近时间点的信息，从而降低了所得估计的变动性。

应用第 3.2 节的最小二乘核方法于初始估计 $\tilde{\beta}(t_{(j)})$ ， $j = 1, \dots, J$ ，对于 $0 \leq$

$d \leq K$, $\beta_d(t)$ 的一个核估计就为

$$\hat{\beta}_d(t) = \frac{\sum_{j=1}^J \tilde{\beta}_d(t_{(j)}) K_h(t - t_{(j)})}{\sum_{j=1}^J K_h(t - t_{(j)})}, \quad (4.5)$$

这里 $K_h(s) = (1/h)K(s/h)$, $K(\cdot)$ 为非负核函数。鉴于核估计有边界偏差较大的缺点 (Fan 和 Gijbels, 1996), 一个更好对 $\beta_d(t)$ 的估计进行光滑的方法是局部多项式方法。一般来讲, 如果 $\beta_d(t)$ 关于 t 是 $(Q+1)$ 次连续可导的, 则 $\beta_d(t)$ 的第 q 阶导数的局部多项式估计就为

$$\hat{\beta}_d^{(q)}(t) = \sum_{j=1}^J w_{q,Q+1}(t_{(j)}, t) \tilde{\beta}_d(t_{(j)}), \quad (4.6)$$

其中 $w_{q,Q+1}(t_{(j)}, t)$ 由光滑方法决定的。

当 $q = 0$ 时, 则 $\hat{\beta}_d(t) = \hat{\beta}_d^{(0)}(t)$ 为 $\beta_d(t)$ 的局部多项式估计, (4.6) 中 $w_{q,Q+1}(t_{(j)}, t)$ 的具体选择决定了光滑估计的光滑性和统计性质。令 $C_j = (1, t_{(j)} - t, \dots, (t_{(j)} - t)^Q)^T$, $C = (C_1, \dots, C_J)^T$, $W_j = K\{(t_{(j)} - t)/h\}$, $W = \text{diag}(W_1, \dots, W_J)$, 其中 $K(\cdot)$ 为非负核函数。(4.6) 的第 q 阶局部多项式估计的权为

$$w_{q,Q+1}(t_{(j)}, t; h) = q! e_{q+1,Q+1}^T (C^T W C)^{-1} C_j W_j, \quad j = 1, \dots, J, \quad (4.7)$$

其中 $e_{q+1,Q+1} = (0, \dots, 0, 1, 0, \dots, 0)^T$, 即仅第 $(q+1)$ 个元素为 1, 其余元素皆为 0。

(4.7) 表达形式与 Fan 和 Zhang (2000) 的方程 (3.5) 相同。

在条件均值模型中的系数曲线可以用一步光滑方法估计, 如 Hoover 等 (1998), Lin 和 Carroll (2000), 以及 Wu 和 Chiang (2000)。由于 (4.1) 的估计一般是基于秩的方法得到的, 故类似的一步光滑方法仍不能用于现在的情形, 这是因为当前还没有一个基于秩的光滑方法不依赖于初始估计。Wu, Tian 和 Yu (2010) 考虑了两步光滑的方法, 该方法计算简单, 且也能根据不同系数曲线的不同的光滑要求自动作调整。然而, 两步光滑方法有个潜在的缺点, 它常常要求在每个时间设计点 $t_{(j)}$ 都有充分多的观测值。

在实际应用中, 在给定协变量 $\{\mathbf{X}(t), t\}$ 条件下估计和预测相应变量 $Y(t)$ 的条件分位数和累积分布常常是很重要的. 这就意味着需要构造 $h(y, t)$ 的一个对所有 $t \in \mathcal{T}$ 关于 y 是 ”单调增” 或 ”保序” 的非参估计. 原则上, 可以通过光滑在不同时间点 $\{t_{(j)}; j = 1, \dots, J\}$ 处得到的某组初始估计来估计 $h(y, t)$, 而这组初始估计可由 Cheng, Wei 和 Ying (1997) 的方法来构造. 然而, 正如 Hall 和 Müller (2003) 所显示的, 除非是非负核估计这种简单情形, 否则局部多项式方法不能自动导出保序的光滑估计. 因此, 在构造基于 (4.1) 的条件分位数和累积分布函数的保序非参数估计时需要用到不同于本节前面所用到的方法和理论.

§4.4.3 窗宽的选择

在 (4.5) 和 (4.7) 中, 窗宽 h 的选择对得到合适的光滑估计极其重要. 在实际应用中, 往往可通过检查拟合曲线和评估特定的临床背景来选择主观窗宽. 尽管主观窗宽可导出有意义的估计, 然而在实际研究中常常采用数据驱动的窗宽.

对于基于条件均值的回归模型, 纵向数据选择窗宽的一个很普遍的方法是 ”删除个体” 交叉验证 (CV) 方法, 该方法每次删除一个个体的所有观测值, 如 Hoover 等 (1998) 和 Wu 和 Chiang (2000). 推广该交叉验证方法到我们考虑的这种情形, 这里基于数据, 我们考虑了选择 (4.5) 和 (4.6) 的窗宽的两种交叉验证法. 第一个方法是 ”删除个体” 交叉验证 (CV) 窗宽的一个直接推广, 即极小化交叉验证得分

$$CV_Z(\mathbf{h}) = \sum_{j=1}^J \sum_{i_1 \neq i_2 \in \mathcal{S}_j} \left[Z_{i_1, i_2}(t_{(j)}) - \xi \left\{ \mathbf{X}_{i_1, i_2}^T(t_{(j)}) \hat{\beta}_{-(i_1, i_2)}(t_{(j)}; \mathbf{h}) \right\} \right]^2, \quad (4.8)$$

其中 $\hat{\beta}_{-(i_1, i_2)}(t_{(j)}; \mathbf{h})$ 是从删除个体 i_1 和 i_2 后的所有观测值计算的两步局部多项式估计或核估计, $\mathbf{h} = (h_0, \dots, h_D)^T$ 是 $(\hat{\beta}_0(t), \dots, \hat{\beta}_D(t))^T$ 的窗宽向量. 实际上, 当样本量 n 很大时, (4.8) 的计算量就会特别大. 可加快计算速度的另一个方法是用 “ M -fold” 交叉验证得分替代 (4.8), 即通过每次删除一组个体来计算. 这就需要先

将个体随机分成 M 组, $\{b(m); m = 1, \dots, M\}$, 然后通过极小化

$$CV_Z^{(M)}(\mathbf{h}) = \sum_{j=1}^J \sum_{m=1}^M \sum_{i_1 \neq i_2 \in S_j; (i_1, i_2) \in b(m)} \left[Z_{i_1, i_2}(t_{(j)}) - \xi \left\{ \mathbf{X}_{i_1, i_2}^T(t_{(j)}) \hat{\beta}_{-b(m)}(t_j; \mathbf{h}) \right\} \right]^2 \quad (4.9)$$

来计算 \mathbf{h}_{CV_Z} , 其中 $\hat{\beta}_{-b(m)}(t_{(j)}; \mathbf{h})$ 为删除整个组 $b(m)$ 中所有个体的观测值后计算的两步核或局部多项式估计。

在 (4.8) 和 (4.9) 中, 由于 $CV_Z(\mathbf{h})$ 与 $CV_Z^{(M)}(\mathbf{h})$ 是关于 $\mathbf{h} = (h_0, \dots, h_K)^T$ 求极小化, 所以当 K 较大时, 计算量仍然会很大. 另一寻找 CV 窗宽 $h_{d,CV}$ 的方法是依赖于逐元方法, 即对每一个 d ($0 \leq d \leq K$), 极小化

$$CV_{\beta_d}(h_d) = \sum_{j=1}^J \left\{ \tilde{\beta}_d(t_{(j)}) - \hat{\beta}_{d,-(j)}(t_{(j)}; h_d) \right\}^2, \quad (4.10)$$

其中 $\hat{\beta}_{d,-(j)}(t_{(j)}, h_d)$ 由删除时间点 $t_{(j)}$ 处的初始估计 $\tilde{\beta}_d(t_{(j)})$ 后计算的光滑估计。由 (4.10) 可直接看出 (4.10) 是对每个 d 极小化交叉验证得分, 与同时关于 k 个元素极小化交叉验证得分的 (4.8) 和 (4.9) 相比, 后者能大大减少计算量。 (4.10) 清晰地表明该方法每次删除一个时间设计点 $t_{(j)}$ 的初始估计, 故其忽视了数据个体内潜在的相关性, 因此 (4.10) 的理论性质不同于由 (4.8) 和 (4.9) 所得的 CV 窗宽。目前这些 CV 窗宽的理论性质尚未被系统研究。

§4.5 推断

关于该情形下基于条件分布的回归模型, 其光滑估计的渐近分布还没有被系统研究。由于缺乏可靠的渐进结果, 故目前仍没能给出本节中光滑估计的渐近近似的统计推断。建立合适的统计推断主要两个困难问题是 (a) 光滑估计的纠偏问题; (b) 光滑估计的标准差的计算问题。实际上, 光滑估计的偏差常常是很难被估计。不过, 文献 (如 Fan 和 Zhang, 2000; Huang, Wu 和 Zhou, 2002) 显示忽视偏差的 “ $\pm Z_{1-\alpha/2}$ 的标准差区间常常可以用来近似一个光滑估计的 $[100 \times (1 - \alpha)]\%$ 逐点置信区间。如果在给定 \mathcal{D} 下 $\hat{\beta}_d^{(q)}(t)$ 的条件方差存在一个相合估计 $\widehat{\text{var}}\{\hat{\beta}_d^{(q)}(t)|\mathcal{D}\}$, 则 $\beta_d^{(q)}(t)$ 的一

个“ $\pm Z_{1-\alpha/2}$ 标准差区间”可表达为

$$\hat{\beta}_d^{(q)}(t) \pm Z_{1-\alpha/2} \left[\widehat{\text{var}}\{\hat{\beta}_d^{(q)}(t)|\mathcal{D}\} \right]^{1/2}. \quad (4.11)$$

为了得到方差的合适估计 $\widehat{\text{var}}\{\hat{\beta}_d^{(q)}(t)|\mathcal{D}\}$, 一个实用的方法是“个体再抽样”的 bootstrap 方法, 其中一个 bootstrap 样本就是从原样本中重复抽取的 n 个“bootstrap 个体”。令 $\hat{\beta}_{d,b}^{(q)}(t)$ 为由第 b 个 bootstrap 样本计算的 $\beta_d^{(q)}(t)$ 的光滑估计。则条件方差的估计 $\widehat{\text{var}}\{\hat{\beta}_d^{(q)}(t)|\mathcal{D}\}$ 就可通过 $\{\hat{\beta}_{d,b}^{(q)}(t); b = 1, \dots, B\}$ 的样本方差来给出。

不同于在 (4.11) 中采用的正态渐近方法, 另一个构造 $\beta_d^{(q)}(t)$ 的 $[100 \times (1 - \alpha)]\%$ 置信区间的方法是采用 bootstrap 分位数区间

$$\left(L_{\alpha/2}, U_{\alpha/2} \right), \quad (4.12)$$

其中 $L_{\alpha/2}$ 和 $U_{\alpha/2}$ 分别为 $\{\hat{\beta}_{d,b}^{(q)}(t); b = 1, \dots, B\}$ 的下和第 $[100 \times (\alpha/2)]$ 百分数。实际上, (4.11) 和 (4.12) 都可用来作为近似推断的方法。当样本量较大时, 如第 1.3 节的例子, 由 (4.11) 和 (4.12) 构造的近似置信区间很接近, 它们可以交替使用。

§4.6 NGHS 数据的应用

将基于条件分布的回归法方法应用于 NGHS 数据, 用时变变化模型 4.1) 来评价种族, 身高和 BMI 的指定年龄的协变量对收缩压 (systolic blood pressure, SBP) 和舒张压 (diastolic blood pressure, DBP) 的分布函数的效应。在用条件均值回归模型或条件分布回归模型拟合 SBP 和 DBP 数据之前, Wu, Tian 和 Yu (2010) 首先对数据作了一系列预备的正态性检验, 包括 Shapiro-Wilk 检验, Kolmogorov-Smirnov 检验和可视化 Q-Q 图检验, 结果表明对给定年龄, 种族, 身高 (米) 和体块指数 (公斤 / 米²) 的女孩总体, 其 SBP 和 DBP 的条件分布明显不符从正态分布。特别是采用拟合优度检验时, 这些条件分布中许多是偏的, 以 5% 的最显著水平拒绝其正态性。

Wu, Tian 和 Yu (2010) 预研结果表明 Daniels 等 (1998) 或 Thompson 等 (2007) 所给出的关于协变量对 SBP 和 DBP 的整个条件分布的效应的描述可能不合适。尤

其是 Daniels 等 (1998) 采用的是条件均值回归模型, 该模型不适于描述这些协变量对心血管的不健康水平的概率, 而 Thompson 等 (2007) 的结论依赖于结果变量所选的具体门限, 如果采用其它门限值, 该结论可能不成立. 因此, 使用 (4.1) 比使用条件均值回归模型有两大优势: 首先, 它有灵活的稀疏结构可概括协变量对结果变量的整个分布的具体年龄效应; 其次, 年龄依赖的系数曲线的统计推断和 Cheng, Wei 和 Ying(1995) 中的有相同的生物学解释, 且不依赖于结果变量的正态假设或定义不健康风险水平的门限值.

因为重复测量是在女孩 9 岁到 19 岁期间进行的, 所以这里给出分析年龄区间限制在 $\mathcal{T} = [9, 19)$ 上. 对一个年龄为 $t \in \mathcal{T}$ 的女孩, 结果变量 $Y(t)$ 为该女孩在 t 时的 SBP 或 DBP 值, 协变量向量为 $\mathbf{X}(t) = (Race, Height(t), BMI(t))^T$, 其中 $Race = 0$ 或 1 分别表示该女孩为白人或非裔美国人, $Height(t)$ 和 $BMI(t)$ 分别为该女孩 t 岁时的身高和体块指数. 由于这个研究中随访的时间没有一定规则, 年龄向上取整到一年中的十分之一区间内都是有临床意义的, 因此年龄区间由设计时间点 $\mathbf{t} = \{9.0, 9.1, 9.2, \dots, 18.9\}$ 等分为面元 $[9.0, 9.1), \dots, [18.9, 19.0)$. 如果在随访中第 i 个女孩的真实年龄落在面元 $[t_j, t_{j+1})$ 上, 则记该女孩的年龄、协变量和结果变量为 $\{t_j, \mathbf{X}_i(t_j), Y_i(t_j)\}$. 该面元的分割是基于年龄的临床定义, 且假设 $Height(t)$, $BMI(t)$ 和 $Y(t)$ 是年龄 t 的光滑函数.

在 Wu, Tian 和 Yu (2010) 探索性分析中, 由年龄、种族、身高和 BMI 各层门限值范围内的 SBP 比值比图显示: 比例比值模型给出了 SBP 的条件分布函数和协变量 $\mathbf{X}(t)$ 之间关系的合理的近似. 因此该分析考虑了如下时变比例比值模型

$$\begin{aligned} & \log \left\{ \frac{P[Y(t) > y | t, Race, Height, BMI]}{P[Y(t) \leq y | t, Race, Height, BMI]} \right\} \\ &= -h(y, t) + \beta_0(t) \times Race + \beta_1(t) \times Height(t) + \beta_2(t) \times BMI(t), \quad (4.13) \end{aligned}$$

其中 $\beta_0(t)$ 的正 (负) 值表示在 t 岁时非裔美国女孩的 SBP 或 DBP 值趋于比白人女

孩的高(低), $\beta_1(t)$ 和 $\beta_2(t)$ 分别表示女孩 t 年龄时的 $Height(t)$ 和 $BMI(t)$ 每增加一个单位时相应的 $Y(t) > y$ 的对数优势的改变量。

应用第 4.4 节的估计方法, 首先用权 $w(\cdot) = 1$ 计算出时间设计点 $\mathbf{t} = \{9.0, 9.1, 9.2, \dots, 18.9\}$ 处的初始估计, 然后用以 Epanechnikov 核和由 (4.9) 和 (4.10) 选择的交叉验证 (CV) 窗宽的局部线性方法计算光滑估计 $\hat{\beta}(t) = (\hat{\beta}_0(t), \hat{\beta}_1(t), \hat{\beta}_2(t))^T$ 。对于 M -区组的 CV 窗宽, 10 个个体组, $M = 10$, 用来极小化 CV 得分 (4.9)。图 5 显示了基于 (4.10) 的 CV 窗宽计算的 $\beta(t) = (\beta_0(t), \beta_1(t), \beta_2(t))^T$ 的两部局部线性估计, 以及由 500 次 bootstrap 重复计算的“ ± 1.96 标准差带”。为了减轻计算负担, 从原样本得到的 CV 窗宽可被用来计算 bootstrap 标准差带。图 5 显示协变量对 SBP 和 DBP 的效应类似; 种族效应 $\beta_0(t)$ 的贡献接近于零, 但随着年龄的增长会轻微增长, 这表明在实验后期非裔美国女孩比白人女孩容易有较高的 SBP 和 DBP 的比值; $\beta_1(t)$ 和 $\beta_2(t)$ 的正估计表明身高和 BMI 都会增大 SBP 和 DBP 的比值; $Height(t)$ 和 $BMI(t)$ 对 SBP 的条件分布的效应呈现出随年龄增大而变小的趋势; $BMI(t)$ 对 SBP 的条件分布和 DBP 的条件分布显示出不同的效应, $BMI(t)$ 对 SBP 的条件分布的效应, $\beta_2(t)$, 显示是时间 t 的线性递减的, 而 $BMI(t)$ 对 DBP 的条件分布的效应, $\beta_2(t)$, 显示是女孩年龄 t 的一个非线性函数。

§4.7 讨论和推广

本节所讨论的时变线性变换模型是用基于条件分布的回归方法为整个条件分布函数建模, 并评价协变量对它们分布的效应。当研究目标依赖于条件分布函数时, 该方法在概念上优于条件均值模型。如应用到 NGHS 血压数据所显示的, 条件分布函数建模的实用优势可以推广到其它典型的生物医学研究。类似于基于条件均值的变系数模型, 时变变换模型能有效地降低“维数祸根”同时又可保持模型的高度灵活性。从计算角度看, 两步估计方法概念简单, 可以通过结合生存分析中现有的秩估计方法和非参数曲线估计中的光滑方法来实施。

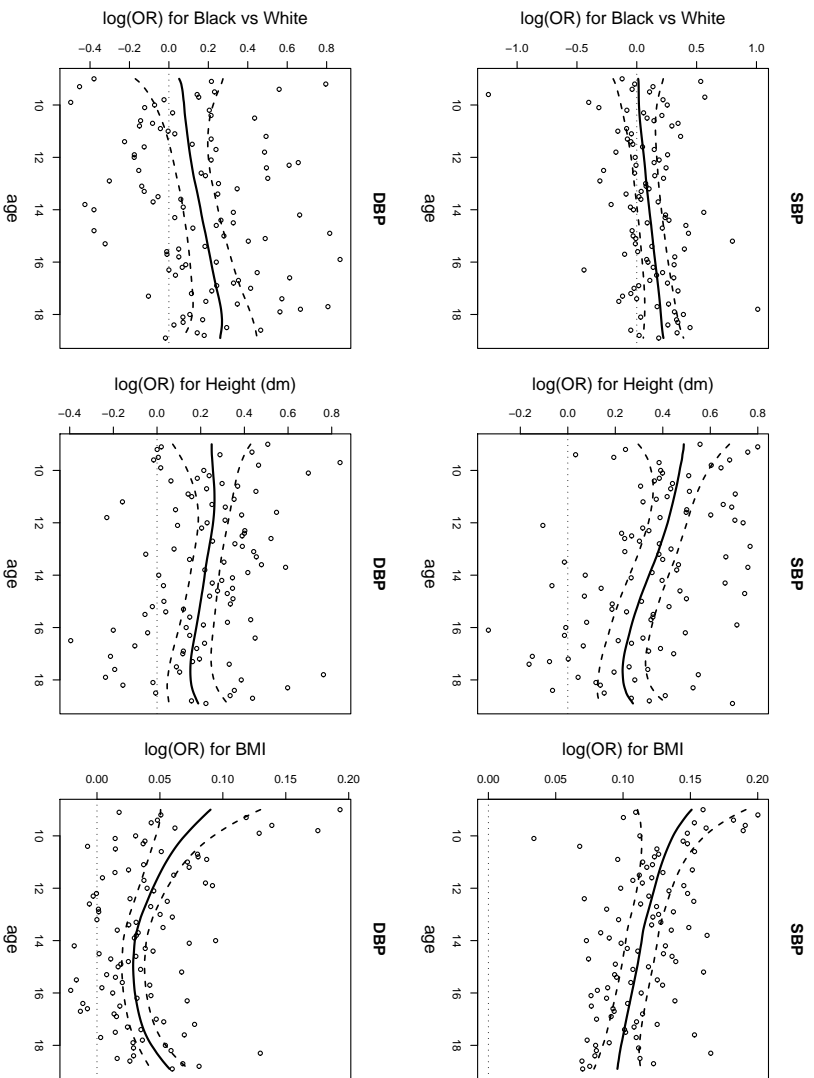


图 5 上排: 收缩压 (SBP). 下排: 舒张压 (DBP). 每排给出 (4.1) 中 $\beta_0(t)$, $\beta_1(t)$ 与 $\beta_2(t)$ 的初始估计、由 CV 窗宽 (4.10) 计算的两部局部线性估计 (实线) 以及它们的 ± 1.96 标准差带 (虚线)。

鉴于条件分布回归模型的现实可行性, 故值得对该方法作进一步的研究。首先, 注意到在 NGHS 血压数据应用中, 模型 (4.1) 的拟合通过一系列的探索性图的 *ad hoc* 方法来检验的, 故有待构造更严格的且恰当的拟合优度检验来检验时变模型的统计假设。其次, 本节所考虑的模型 (4.1) 仅包含一个一元结果变量, 在许多情形下, 研究的目标需要用多元结果变量的联合分布来描述, 因此该方法在多元结果变量情形下的推广是值得进一步研究的; 最后, 本节讨论的两步方法仅限于去估计协变量随时间变化的效应, 但估计和预测条件分布与条件分位数的光滑方法在纵向研究中也很有重要的应用, 因此也应该被很好的去研究。

§5 结果自适应协变量的回归方法

文献 (如 Pepe 和 Anderson, 1994) 中众所周知: 当协变量的值依赖于结果变量在前一时间点的值时, 第 2 节和第 3 节的条件均值回归方法可能会导出令人不满意的结果。这种协变量类型, 因此被称为结果自适应协变量。作为带结果自适应协变量的纵向分析的一个特例, 本节将给出有伴随干预的纵向数据的建模和估计方法。对伴随干预做一个简单介绍之后, 本节下面的部分将给出 Wu, Tian 和 Bang (2008) 与 Wu, Tian 和 Jiang (2011) 建模方法之间的相似点和不同点, 以及他们的相应估计和推断方法。

§5.1 结果自适应协变量

对于随机分配的研究处理的纵向临床试验, 研究处理的纵向效应是通过时不变分类协变量向量建模的, 而感兴趣的其它因子, 如年龄、性别、种族和疾病风险因子, 能够通过时不变协变量和时变协变量来建模的。在许多纵向研究中, 一些时变协变量是结果自适应的, 它们在一个时间点潜在的取值可以依赖于结果变量前一时间点的值或时间趋势。

一个包含结果自适应协变量的典型的情形是纵向临床始中伴随干预的出现。

临床研究处理是随机安排给个体的，与之不同，伴随干预通常是出于道义给那些健康结果呈现令人不满意的发展趋势的研究个体，因此不是随机安排的。

伴随干预的现象与带信息性缺失数据的纵向研究有一些相似。在带信息性缺失数据的情形，结果出现不满意的时间趋势的研究个体比结果出现满意的时间趋势的研究个体倾向于从研究中更早退出试验。与带有伴随干预的研究唯一区别在于开始伴随干预后研究个体的结果仍可以继续被观测到。注意到在一个随机安排研究处理的随机化的纵向临床试验中，除了被安排的研究处理外，另加使用了伴随干预的研究个体一般与不需要伴随干预个体有不同的病理学。因此除了评价研究处理的效应这个首要目标外，第二个重要的目标是评价附加的伴随干预对研究总体的结果变量的效应。

例 4 中介绍的增强冠心病康复研究 (ENRICHD) 是同时包含伴随干预和随机安排处理机制的一个典型例子。这是一个评价六个月认知行为疗法 (CBT) 相对于通常心血管护理 (UC) 有效性的随机临床试验，其中，在 CBT 组的病人的贝克抑郁量表 (BDI) 得分是在治疗期的每周复诊和每年四次的跟踪回访时重复测得，而在 UC 组的病人的 BDI 得分仅在研究起点、第六个月的复诊和每年的跟踪回访中测得。若病人有高的基线 BDI 得分，或注册后五周 BDI 没有降低趋势，或抗抑郁剂被病人或主治医生要求使用时，研究设计 (ENRICHD, 2001) 允许 CBT 组和 UC 组都可使用抗抑郁剂的药物治疗。虽然 Taylor 等 (2005) 表示药物治疗可以提高 1834 个抑郁 ENRICHD 病人中的存活率，然而他们的结果不能回答药物治疗是否有利于降低病人的抑郁严重程度。

采用 CBT 组在治疗期接受药物治疗的 91 个 ENRICHD 病人子样本的重复测量 BDI 得分，Wu, Tian 和 Bang (2008) 显示简单的混合效应模型会给出药物治疗对 BDI 随时间趋势的效应的误导性结果，并提出了一个变系数混合效应模型来降低相关药物治疗效应的估计的潜在的偏。Wu, Tian 和 Bang (2008) 的主要缺点是潜在的

信息损失，因为他们的回归模型不能被应用于试验一开始就已接受药物疗法的病人或在整个研究中都没有使用药物疗法的病人。

作为变系数混合效应模型的一个推广，Wu, Tian 和 Jiang (2011) 提出了一个评价伴随干预效应的综合的回归方法，该方法能够合并纵向研究中所有研究个体的信息。采用 Follmann 和 Wu (1995) 的共享参数模型的框架，Wu, Tian 和 Jiang (2011) 通过变点混合效应模型描述了协变量对相应变量的效应，并通过一系列联合分布将随机效应和干预开始点 (变点时间) 结合起来。将试验一开始就已接受药物疗法的病人或在整个研究期都没有使用药物疗法的病人当作删失。故可基于似然的方法建立其统计估计和推断，其计算可通过两阶段迭代程序来实施。应用他们的方法到 ENRICH 药物疗法数据，Wu, Tian 和 Jiang (2011) 的结果表明当伴随干预出现时，他们提出的方法可以导出合适的估计，而在此情形下简单的混合效应模型可能被错误指定。

§5.2 带一个伴随干预的数据结构

遵循 1.2 节的记号，令 T_0 和 T_1 为研究开始和结束的时间， n 为随机选择的个体的总数。假设第 i 个个体有 n_i 次回访，记他的第 j 次回访的观测值为 $(T_{ij}, Y_{ij}, \mathbf{X}_i)$ ，其中 T_{ij} 为研究时间，是从研究开始到第 j 次回访时的时间， \mathbf{X}_i 为一个时间不变的协变量向量， Y_{ij} 为实值的结果变量。为简单，全节都假设研究仅包含一个伴随干预，且第 i 个个体在研究期由“无伴随干预”到“伴随干预”仅能改变一次，其中伴随干预开始的时间或变点时间记为 S_i 。令

$$\lambda_{ij} = \begin{cases} 0, & \text{当 } T_{ij} \leq S_i, \\ 1, & \text{当 } T_{ij} > S_i, \end{cases}$$

则 λ_{ij} 为 i 个个体的伴随干预的标记。由于研究期内不是每个个体都有变点时间，如果 $T_{i1} \leq S_i \leq T_{in_i}$ ，则 i 个个体的变点可被观测到。如果 $S_i < T_{i1}$ 或 $S_i > T_{in_i}$ ，则个体的变点时间分别被左删失或右删失。删失的指示变量 $\delta_i^{(c)}$ 定义为 $\delta_i^{(c)} = 0$ 如

果 $T_{i1} \leq S_i \leq T_{in_i}$; $\delta_i^{(c)} = 1$ 如果 $S_i > T_{in_i}$; $\delta_i^{(c)} = 2$ 如果 $S_i < T_{i1}$. 观测到的变点时间为 $\{S_i^{(c)} = (S_i^{(c)}, \delta_i^{(c)}); i = 1, \dots, n\}$, 其中 $S_i^{(c)} = S_i$ 如果 $\delta_i^{(c)} = 0$; $S_i^{(c)} = T_{in_i}$ 如果 $\delta_i^{(c)} = 1$; $S_i^{(c)} = T_{i1}$ 如果 $\delta_i^{(c)} = 2$.

§5.3 模型结构和解释

§5.3.1 简单的混合效应变点模型

由于伴随干预不是随机指定的, 所以伴随干预的效应不能通过直接比较受到干预的个体和没受到干预的个体来评价, 实际上这是可以理解的. ENRICHHD (2003) 和 Taylor 等 (2005) 在 ENRICHHD 研究中已注意到这个事实. Wu, Tian 和 Bang (2008) 与 Wu, Tian 和 Jiang(2011) 建议采用变点模型, 假设结果变量 Y_{ij} 在伴随干预前后遵循不同的轨迹, 因此伴随干预的效应能够通过均值轨迹的差来评价.

令 $\mu_0(T_{ij}, \mathbf{X}_i; \mathbf{a}_i)$ 为第 i 个个体伴随干预前的轨迹, 其可通过个体指定参数 $\mathbf{a}_i^T = (a_{i1}, \dots, a_{id_0})^T$, $d_0 \geq 1$ 参数化, $\mu_1(T_{ij}, \mathbf{X}_i, R_{ij}; \mathbf{b}_i)$ 为伴随干预后轨迹的改变量, 其可通过个体指定参数 $\mathbf{b}_i = (b_{i1}, \dots, b_{id_1})^T$, $d_1 \geq 1$ 参数化, 且可依赖于 T_{ij} , \mathbf{X}_i , 以及“干预持续的时间” $R_{ij} = T_{ij} - S_i$. 一般的混合效应模型体系 (Davadian 和 Giltinan, 1995; Verbeke 和 Molenberghs, 2000; Diggle 等 2002) 建议用简单的混合效应模型来评价干预前、后的轨迹, 即

$$\begin{cases} Y_{ij} = \mu_0(T_{ij}, \mathbf{X}_i; \mathbf{a}_i) + \lambda_{ij} \mu_1(T_{ij}, \mathbf{X}_i, R_{ij}; \mathbf{b}_i) + \epsilon_{ij}, \\ (\mathbf{a}_i^T, \mathbf{b}_i^T)^T \sim \text{联合分布} G(\cdot), \end{cases} \quad (5.1)$$

其中, $(\mathbf{a}_i^T, \mathbf{b}_i^T)^T$ 有未知的均值向量 $(\alpha^T, \beta^T)^T$ 和协方差阵 Σ , ϵ_{ij} 为零均值的随机误差向量, $\text{cov}(\epsilon_{ij_1}, \epsilon_{ij_2}) = \sigma_{ij_1 j_2}$, 且 $\epsilon_{i_1 j_1}$ 和 $\epsilon_{i_2 j_2}$ ($i_1 \neq i_2$) 相互独立. 为数学上方便, 假设 of \mathbf{a}_i 和 \mathbf{b}_i 的联合分布为多元高斯分布 $\mathcal{N}\{(\alpha^T, \beta^T)^T, \Sigma\}$. 在 (5.1) 中, $\mu_1(T_{ij}, \mathbf{X}_i, R_{ij}; \mathbf{b}_i)$ 的正 (或负) 值表示干预倾向于增大 (或减小) 给定 $(T_{ij}, \mathbf{X}_i, R_{ij})$ 下 Y_{ij} 的条件均值.

§5.3.2 简单的混合效应模型偏的例子

由于忽视了干预的“可自选性”，Wu, Tian 和 Bang (2008) 显示：即使 $\mu_0(\cdot; \mathbf{a}_i)$, $\mu_1(\cdot; \mathbf{b}_i)$, ϵ_{ij} 和 $G(\cdot)$ 的假设都正确，(5.1) 也是一个错误指定模型。甚至在 (5.1) 看起来有合理的解释的简单情形下，有时也会出现错误的结论。下面是一个用来显示简单的混合效应模型的标准估计程序可导出潜在错误结论的简单的人工例子。

假设纵向研究中有 $n = 24$ 独立的个体，且对于 $1 \leq i \leq 24$, $j = 1, \dots, n_i$, 有 $n_i = 10$ 和 $T_{ij} = j$. 令 S_i 为第 i 个个体由“无伴随干预”转到“伴随干预”时的变点时间， λ_{ij} 为相应的“伴随干预的标记”，即 $\lambda_{ij} = 0$ 如果 $S_i < T_{ij}$ ； $\lambda_{ij} = 1$ 如果 $S_i \geq T_{ij}$. 对于前 12 个个体，如 $1 \leq i \leq 12$, 他们的变点时间为 $S_i = 2$, 他们的结果变量是由 $Y_{ij} = 20 + e_{ij}$ 当 $\lambda_{ij} = 0$ 且 $1 \leq j \leq 2$ 和 $Y_{ij} = 19 + e_{ij}$ 当 $\lambda_{ij} = 1$ 且 $3 \leq j \leq 10$ 产生，其中 e_{ij} 独立同分布于 $N(0, 3)$. 对于剩下的 12 个个体， $13 \leq i \leq 24$, 他们的变点时间为 $S_i = 8$, 他们的结果变量是由 $Y_{ij} = 19 + e_{ij}$ 对于 $\lambda_{ij} = 0$ 且 $1 \leq j \leq 8$ 和 $Y_{ij} = 17 + e_{ij}$ 对于 $\lambda_{ij} = 1$ 且 $9 \leq j \leq 10$ 产生的，其中 e_{ij} 独立同分布于 $N(0, 3)$. 假设对于所有的 $1 \leq i \leq 24$ 和 $1 \leq j \leq 10$ ，已知 Y_{ij} 都不依赖于 T_{ij} . 则可被考虑用于当前情形的 (5.1) 的简单混合效应变点模型的特殊模型为

$$\begin{cases} Y_{ij} = a_i + b_i \lambda_{ij} + \epsilon_{ij}, \\ (a_i, b_i)^T \sim \text{联合分布} G(\cdot), \end{cases} \quad (5.2)$$

其中，对所有的 $1 \leq i \leq 24$ 和 $1 \leq j \leq 10$, ϵ_{ij} 独立同分布于 $N(0, 3)$, $E(a_i) = \alpha$ 和 $E(b_i) = \beta$ 为固定效应， $G(\cdot)$ 为未知的分布。由于对于 $1 \leq i \leq 12$, 伴随干预效应的真值为 $b_i = -1$, 而对于 $13 \leq i \leq 24$, 伴随干预效应的真值为 $b_i = -2$, 总体的固定的均值效应的真值为 $\beta = -1.5$, 但在模型 (5.2) 下它们都是未知的，需要由数据给出估计。

如果样本 $\{(Y_{ij}, T_{ij}, S_i); i = 1, \dots, 24, j = 1, \dots, 10\}$ 是从上面的模式产生的，则可由许多统计分析软件如 SAS 和 R 来估计 (5.1) 的均值和协方差参数。虽然确切的分

布函数 $G(\cdot)$ 一般未知, 但可通过假设 $G(\cdot)$ 为带恰当相关系数结构的多元正态分布, 得到均值参数 $\{\alpha, \beta\}$ 的合适的估计。由于数据的相关结构一般是未知的, 在 R 统计软件中有 3 个 LME 程序能用来估计 $\{\alpha, \beta\}$: 带独立相关结构的 LME(LMEWI), 带随机截距的 LME (LMERI), 和带随机截距和斜率的 LME (LMERIS)。类似地, 用广义估计方程 (GEE) 程序给出参数估计也能三种相关结构: 带独立相关结构的 GEE(GEEWI), 带可交换的相关结构的 GEE(GEEC), 和带无结构相关结构的 GEE(GEEUC)。关于线性混合效应模型的参数估计方法的更详细的介绍可参见 Verbeke 和 Molenberghs (2000)。

为了检查简单的混合效应变点模型 (5.2) 是否能给出平均的伴随干预效应 β 的合适估计, 模拟产生 10000 个 $\{(Y_{ij}, T_{ij}, S_i); i = 1, \dots, 24, j = 1, \dots, 10\}$ 的独立样本。对每个随机样本应用 SAS 的每个估计程序 LMEWI, LMERI, LMERIS, GEEWI, GEEEC 和 GEEUC, 计算出估计 $\hat{\beta}$ 和它们的标准差。表 2 给出了由这 10,000 模拟样本和简单混合效应模型 (5.2) 计算的估计和它们的标准差 (SE) 的平均值, 以及相应的 95% 置信区间包含真值参数 $\beta = -1.5$ 的实际覆盖率。结果表明这些带不同相关结构假设的 LME 和 GEE 给出的 β 的估计类似, 都在 -0.6 和 -0.7 附近, 相对于其标准差而言远离真值 $\beta = -1.5$ 。因此, 表 2 中的 95% 置信区间的经验覆盖率有低, 这表明由模型 (5.2) 得到的 β 的估计具有很大的偏差, 不适合于应用。

是否有能导出平均伴随干预效应 β 的较好估计的简单方法? 显然有。这里为结构简单考虑, 可以基于 (5.2) 考虑 β 的一个“个体拟合”估计, 即

$$\hat{\beta}_{ind} = (1/n) \sum_{i=1}^n \left\{ \frac{\sum_{j=1}^{n_i} (Y_{ij} 1_{[\delta_{ij}=1]})}{\sum_{j=1}^{n_i} 1_{[\delta_{ij}=1]}} - \frac{\sum_{j=1}^{n_i} (Y_{ij} 1_{[\delta_{ij}=0]})}{\sum_{j=1}^{n_i} 1_{[\delta_{ij}=0]}} \right\}. \quad (5.3)$$

将 (5.3) 直接应用于模拟样本, 所得的 β 的估计与其真值 -1.5 很接近。然而, 文献中大家已知道相比, 分别基于每个个体重复测量的“个体拟合”估计方法一般比著名的程序, 如 ML 估计、REML 估计和 GEE(如 Verbeke 和 Molenberghs, 2000; Diggle

表 2 由 10,000 模拟样本和简单混合效应模型 (5.2) 计算的参数估计和它们标准差的平均值, 以及相应的 95% 置信区间 (CI) 包含真值参数 $\beta = -1.5$ 的经验覆盖率

R 程序 相关结构	估计	标准差	95% CI 的 经验覆盖率
LME 独立结构	-0.598	0.395	37.8%
LME 随机截距	-0.731	0.403	51.0%
LME 随机截距 & 斜率	-0.788	0.431	60.4%
GEE 独立结构	-0.598	0.385	36.2%
GEE 可交换	-0.709	0.385	46.7%
GEE 无结构	-0.752	0.572	43.9%

等 2002) 效果差。此外, 也很难将“个体拟合”估计方法推广到比 (5.2) 更复杂的回归模型。

与用模型 (5.2) 产生 $\{(Y_{ij}, T_{ij}, S_i); i = 1, \dots, 24, j = 1, \dots, 10\}$ 的潜在机制相比, 采用 (5.2) 相应的 LME 和 GEE 估计程序的一个潜在缺点是模型没有考虑变点时间 S_i 和变点时间前的 Y_{ij} 的取值之间的关系, 即当 $\delta_{ij} = 0$ 时, Y_{ij} 的取值。事实上, 在该数据产生机制下, 在 $j = 1$ 时 Y_{ij} 的个体指定均值为 20 的个体在 $S_i = 2$ 处有变点时间, 而在 $j = 1$ 时 Y_{ij} 的个体指定均值为 19 的个体在 $S_i = 8$ 处有变点时间。尽管估计阶段这个事实是未知的, 但当 $\delta_{ij} = 0$ 时, 基于简单混合效应模型的 ML 估计、REML 估计和 GEE 不允许 S_i 和 Y_{ij} 之间可能存在潜在关系。为了看将变点时间 S_i 引入模型是否可以降低 β 估计的偏, 我们考虑 (5.2) 的如下简单推广:

$$\begin{cases} Y_{ij} = a_{0i} + a_{1i}S_i + b_i\lambda_{ij} + \epsilon_{ij}, \\ (a_{0i}, a_{1i}, b_i)^T \sim N((\alpha_0, \alpha_1, \beta)^T, \Gamma), \end{cases} \quad (5.4)$$

其中 $(\alpha_0, \alpha_1, \beta)^T$ 为均值参数向量, Γ 为未知的协方差矩阵。 Γ 的结构一般对 $(\alpha_0, \alpha_1, \beta)^T$ 的估计影响不大, 在应用估计程序 (Diggle 等 2002) 时, 许多常用的参数结构都可以被采用。

注意到 β 在 (5.2) 和 (5.4) 中具有相同的解释，这一点很重要。严格上讲，模型 (5.4) 对产生数据的潜在机制而言是不正确的，因为简单线性模型中 Y_{ij} 不依赖于 S_i 。尽管 (5.4) 是产生真实数据的潜在机制的最好的粗略的近似，但这里的主要目的是评价将 S_i 引入模型后是否可以降低伴随干预效应估计的偏。关于 (5.4) 的合理性以及其参数解释的细节，我们将作为 Wu, Tian 和 Bang (2008) 中变系数混合效应模型的一个特例，在第 5.4 节中给出。

表 3 由 10,000 模拟样本和混合效应模型 (5.4) 计算的参数估计和它们的标准差的平均值以及相应的 95% 置信区间 (CI) 包含真值参数 $\beta = -1.5$ 的经验覆盖率

R 程序 相关结构	估计	标准差	95% CI 的 经验覆盖率
LME 独立结构	-1.498	0.485	94.7%
LME 随机截距	-1.498	0.483	94.6%
LME 随机截距 & 斜率	-1.498	0.496	95.2%

表 3 概括了由 10,000 模拟样本、混合效应模型 (5.4) 以及与表 2 相同的 LME 程序所计算的参数估计和它们的标准差的平均值、及其相应的 95% 置信区间 (CI) 包含真值参数 $\beta = -1.5$ 的经验覆盖率。表 3 显示 β 估计的平均值于相对表 2，3 中的标准差很接近于真值 $\beta = -1.5$ ；表 3 中 95% 置信区间 (CI) 包含真值参数 $\beta = -1.5$ 的经验覆盖率很接近名义置信水平 95%，且比表 2 中相应的覆盖率大许多。显然，由 (5.4) 导出 β 的估计的偏要比由 (5.2) 导出的小许多。

§5.3.3 一般的共享参数模型

第 5.2 节的数据结构是结果 - 自适应协变量的一个特殊情形，那里只包含了一个伴随干预且每个研究的个体最多只有一个由“无伴随干预”到“伴随干预”的变点。结果 - 自适应性的这个简单结构表明 (5.1) 混合效应方法的一个自然推广是将

伴随干预 initiation 引入回归模型。可通过允许干预起始点与干预前随机系数 \mathbf{a}_i 或更一般 $\{\mathbf{a}_i, \mathbf{b}_i\}$ 相关来实现 (5.1) 的推广。令 $\mu_0(\cdot; \mathbf{a}_i)$ 和 $[\mu_0(\cdot; \mathbf{a}_i) + \mu_1(\cdot; \mathbf{b}_i)]$ 分别为伴随干预开始前、后个体 - 特定的响应曲线, 其中 $\mu_1(\cdot; \mathbf{b}_i)$ 可被解释为伴随干预效应。一个给定数据集 $\{Y_{ij}, T_{ij}, \mathbf{X}_i, S_i\}$ 下的共享参数变点模型为

$$\begin{cases} Y_{ij} = \mu_0(T_{ij}, \mathbf{X}_i; \mathbf{a}_i) + \delta_{ij} \mu_1(T_{ij}, \mathbf{X}_i, R_{ij}; \mathbf{b}_i) + \epsilon_{ij}, \\ (\mathbf{a}_i^T, \mathbf{b}_i^T, S_i)^T \sim \text{联合分布}, \end{cases} \quad (5.5)$$

这里 $R_{ij} = T_{ij} - S_i$, ϵ_{ij} 为均值为零的误差, 满足 $\text{cov}(\epsilon_{ij_1}, \epsilon_{ij_2}) = \sigma_{ij_1 j_2}$, 且对任意 $i_1 \neq i_2$, $\epsilon_{i_1 j_1}$ 和 $\epsilon_{i_2 j_2}$ 独立, S_i 和 $\{T_{ij}, \mathbf{X}_i\}$ 在给定条件 $\{\mathbf{a}_i, \mathbf{b}_i\}$ 下独立。另外假设 $\{\mathbf{a}_i, \mathbf{b}_i\}$ 和 $\{T_{ij}, \mathbf{X}_i\}$ 也独立。用矩阵表示 $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$ 和 $\mathbf{T}_i = (T_{i1}, \dots, T_{in_i})^T$, 基于模型 (5.5), 给定 $\{\mathbf{T}_i, \mathbf{X}_i\}$ 下 $(\mathbf{Y}_i^T, S_i)^T$ 的联合似然函数为

$$f(\mathbf{Y}_i, S_i | \mathbf{T}_i, \mathbf{X}_i) = \int f(\mathbf{Y}_i | \mathbf{T}_i, \mathbf{X}_i, S_i, \mathbf{a}_i, \mathbf{b}_i) f(S_i | \mathbf{a}_i, \mathbf{b}_i) dH(\mathbf{a}_i, \mathbf{b}_i), \quad (5.6)$$

其中 $f(\cdot)$ 表示条件密度, $H(\cdot, \cdot)$ 为 $\{\mathbf{a}_i, \mathbf{b}_i\}$ 的联合分布。正由于在被积函数中多了一项 $f(S_i | \mathbf{a}_i, \mathbf{b}_i)$, 使得 (5.6) 不同于通常的混合效应模型 (Verbeke 和 Molenberghs, 2000, 第 24 页) 的似然函数。

在 (5.5) 中, 参数 $\{\mathbf{a}_i, \mathbf{b}_i\}$ 与 Y_{ij} 的效应曲线和 S_i 的分布有相关。共享参数方法是由 Follmann 和 Wu (1995) 为带信息缺失数据建模时提出的。然而在 (5.5) 中, 个体在变点之后仍可以被观测到。 S_i 和 \mathbf{a}_i 之间的相关表明第 i 个个体的变点时间会受到干预前响应曲线 $\mu_0(\cdot)$ 的影响, 而 S_i 和 \mathbf{b}_i 相关表明 S_i 也可以影响响应曲线 $\mu_1(\cdot)$, 刻画了干预效应。

§5.4 变系数混合效应模型 - 一个特例

§5.4.1 变系数混合效应模型结构

联合似然函数 (5.6) 要求 $\{\mathbf{a}_i, \mathbf{b}_i, S_i\}$ 的分布函数已知, 实际上极大化这个似然函数的计算量通常也很大。当全部个体都能在研究期内观测到变点时间时, 即

对所有的 $1 \leq i \leq n$, 都有 $\mathcal{T}_0 < S_i < \mathcal{T}_1$, 则使用不依赖于 S_i 分布的简单回归方法。

由于对所有 $1 \leq i \leq n$, S_i 都能被观测到, 故考虑条件分布

$$f(\mathbf{Y}_i | S_i, T_i, \mathbf{X}_i) = \int f(\mathbf{Y}_i | T_i, \mathbf{X}_i, S_i, \mathbf{a}_i, \mathbf{b}_i) dG(\mathbf{a}_i, \mathbf{b}_i | S_i), \quad (5.7)$$

并用 $\{\mathbf{a}_i, \mathbf{b}_i\}$ 在 S_i 给定的条件下的条件分布改写 (5.5) 为一个变系数模型。尽管允许 $\mu_0(\cdot)$ 和 $\mu_1(\cdot)$ 采用参数或非参数形式, 但为简单起见, 这里假设 $\mu_0(\cdot)$ 和 $\mu_1(\cdot)$ 为线性函数: $\mu_0(T_{ij}, \mathbf{X}_i; \mathbf{a}_i) = \mathbf{Z}_{ij}^T \mathbf{a}_i$ 和 $\mu_1(T_{ij}, \mathbf{X}_i, S_i; \mathbf{b}_i) = \mathbf{W}_{ij}^T \mathbf{b}_i$, 其中, $\mathbf{Z}_{ij} = (Z_{ij0}, \dots, Z_{ijD_1})^T$ 是由 $\{(T_{ij}, \mathbf{X}_i); 1 \leq j \leq n_i, \delta_{ij} = 0\}$ 产生的, $\mathbf{W}_{ij} = (W_{ij0}, \dots, W_{ijD_2})^T$ 是由 $\{(T_{ij}, \mathbf{X}_i, S_i); 1 \leq j \leq n_i, \delta_{ij} = 1\}$ 产生的。

令 $\alpha(S_i) = E(\mathbf{a}_i | S_i)$, $\beta(S_i) = E(\mathbf{b}_i | S_i)$, $\mathbf{a}_i^* = \mathbf{a}_i - \alpha(S_i)$, $\mathbf{b}_i^* = \mathbf{b}_i - \beta(S_i)$. 针对数据 $\{Y_{ij}, T_{ij}, S_i; 1 \leq i \leq n, 1 \leq j \leq n_i\}$ 的一个变系数混合效应模型为

$$\begin{cases} Y_{ij} = \mathbf{Z}_{ij}^T [\alpha(S_i) + \mathbf{a}_i^*] + \delta_{ij} \mathbf{W}_{ij}^T [\beta(S_i) + \mathbf{b}_i^*] + \epsilon_{ij}, \\ (\mathbf{a}_i^{*T}, \mathbf{b}_i^{*T})^T | S_i \sim G(\cdot | S_i), \end{cases} \quad (5.8)$$

其中, 对于 $S_i = s$, $G(\cdot | s)$ 为均值为 0 协方差阵为 $cov[(\mathbf{a}_i^{*T}, \mathbf{b}_i^{*T})^T | s] = \mathbf{C}(s)$ 的分布函数。感兴趣的总体均值参数为 $\alpha(s)$ 和 $\beta(s)$, 且两者皆为 s 的光滑函数。当 $S_i = s$ 时, $\beta(s)$ 为伴随干预的平均效应。特殊情形: 对所有的 $s \in (\mathcal{T}_0, \mathcal{T}_1)$, 都有 $\beta(s) = 0$, 此时伴随干预对 Y_{ij} 的时间趋势曲线没有总体平均效应。

在实际应用中, 可以通过指定 $\alpha(\cdot)$, $\beta(\cdot)$ 和 $G(\cdot | \cdot)$ 的特定形式, 得到 (5.8) 的许多有意义的特例。显然可选择 $G(\cdot | S_i)$ 为均值为 0 协方差阵为 $\mathbf{C} = cov[(\mathbf{a}_i^{*T}, \mathbf{b}_i^{*T})^T | s]$ 的多元正态分布, 为简单起见, 还可假设其为时不变的。只要通过对 $\mathbf{C}(s)$ 建模就可将其推广到时间变协变量的情形。由于主要目标是评价伴随干预的总体平均效应, 而 $G(\cdot | S_i)$ 的显式形式常常是未知的, 因此给 $\alpha(s)$ 和 $\beta(s)$ 建立合适的模型往往比给 $\mathbf{C}(s)$ 建立恰当的模型更为重要。 $\alpha(s)$ 和 $\beta(s)$ 的线性模型能被表示为 $\alpha(s; \gamma) =$

$(\alpha_0(s; \gamma_0), \dots, \alpha_{D_1}(s; \gamma_{D_1}))^T$ 和 $\beta(s; \tau) = (\beta_0(s; \tau_0), \dots, \beta_{D_2}(s; \tau_{D_2}))^T$, 其中

$$\alpha_d(s; \gamma) = \sum_{l=0}^{L_d} \gamma_{dl} \mathcal{T}_{dl}(s), \quad \beta_d(s; \tau) = \sum_{m=0}^{M_d} \tau_{dm} \mathcal{T}_{dm}^*(s), \quad (5.9)$$

这里 $\{L_d, M_d\}$ 为固定的, $\{\mathcal{T}_{dl}(s), \mathcal{T}_{dm}^*(s)\}$ 为 s 的已知变换. 令 $\mathcal{T}_{dl}(s) = s^l$ 和 $\mathcal{T}_{dm}^*(s) = s^m$, 就导出局部多项式 $\alpha_d(s; \gamma)$ 和 $\beta_d(s; \tau)$.

当 $\alpha(s)$ 与 $\beta(s)$ 的参数形式未知时, 可通过基展开式来逼近 $\alpha(s)$ 与 $\beta(s)$, 进行非参数分析. 如果 $\{\mathcal{B}_{d_1}(s) = (\mathcal{B}_{d_1 0}(s), \dots, \mathcal{B}_{d_1 \mathcal{L}_{d_1}}(s))^T; 0 \leq d_1 \leq D_1\}$ 和 $\{\mathcal{B}_{d_2}^*(s) = (\mathcal{B}_{d_2 0}^*(s), \dots, \mathcal{B}_{d_2 \mathcal{M}_{d_2}}^*(s))^T; 0 \leq d_2 \leq D_2\}$ 是两个预先指定的基函数, 则 $\alpha(s)$ 和 $\beta(s)$ 的基近似分别为

$$\alpha_d(s; \gamma) \approx \sum_{l=0}^{\mathcal{L}_d} \gamma_{dl} \mathcal{B}_{dl}(s), \quad \beta_d(s; \tau) \approx \sum_{m=0}^{\mathcal{M}_d} \tau_{dm} \mathcal{B}_{dm}^*(s), \quad (5.10)$$

其中 \mathcal{L}_d 和 \mathcal{M}_d 可随 $n \rightarrow \infty$ 时趋于无穷. 通常可选的基函数包括截断多项式基, 傅立叶基和 B-样条. 由于 B-样条近似具有数值计算稳定性的优势, 故目前文献 (Wu, Tian 和 Bang, 2008) 中仅对模型 (5.8) 下带固定节点序列的 B-样条进行了研究. 另一个光滑方法是通过光滑样条近似 $\alpha(s)$ 和 $\beta(s)$ (Lin 和 Zhang, 1999; Chiang, Rice 和 Wu, 2001). 但由于光滑样条估计的显式表达式和统计性质与 B-样条截然不同, 因此在 (5.8) 中采用光滑样条的非参估计和推断尚未被研究.

§5.4.2 最小二乘估计

如果 $G(\cdot|S_i)$ 的显式形式和 ϵ_{ij} 分布未知, 则 (5.8) 中 $\alpha(s)$ 和 $\beta(s)$ 基于似然的估计就无法计算. 当 $\alpha(s) = \alpha(s; \gamma)$ 和 $\beta(s) = \beta(s; \tau)$ 可分别被 Euclidean 值参数 γ 和 τ 参数化时, 一个的实用的方法是通过极小化

$$\begin{aligned} \ell(\gamma, \tau) = & \sum_{i=1}^n \left\{ [\mathbf{Y}_i - (\mathbf{Z}_i^T \alpha(S_i; \gamma) + (\delta \mathbf{W})_i^T \beta(S_i; \tau))]^T \right. \\ & \left. \times \Lambda_i [\mathbf{Y}_i - (\mathbf{Z}_i^T \alpha(S_i; \gamma) + (\delta \mathbf{W})_i^T \beta(S_i; \tau))] \right\} \end{aligned} \quad (5.11)$$

来计算加权最小二乘估计 $\hat{\gamma}_{LS}$ 和 $\hat{\tau}_{LS}$ ，该方法不要求分布函数具有显式表达式，其中 $\mathbf{Z}_i = (\mathbf{Z}_{i1}, \dots, \mathbf{Z}_{in_i})^T$, $(\delta \mathbf{W})_i = (\delta_{i1} \mathbf{W}_{i1}, \dots, \delta_{in_i} \mathbf{W}_{in_i})^T$, Λ_i 为预先指定的 $n_i \times n_i$ 的非奇异的对称权矩阵。加权最小二乘估计的显式表达式为

$$\begin{pmatrix} \hat{\gamma}_{LS}(\mathcal{T}) \\ \hat{\tau}_{LS}(\mathcal{T}) \end{pmatrix} = \left\{ \sum_{i=1}^n [\mathcal{W}_i \mathcal{T}_i]^T \Lambda_i [\mathcal{W}_i \mathcal{T}_i] \right\}^{-1} \left\{ \sum_{i=1}^n [\mathcal{W}_i \mathcal{T}_i]^T \Lambda_i \mathbf{Y}_i \right\}, \quad (5.12)$$

其中 $\sum_{i=1}^n [\mathcal{W}_i \mathcal{T}_i]^T \Lambda_i [\mathcal{W}_i \mathcal{T}_i]$ 是非奇异的，且 \mathcal{W}_i 的第 j 行是 $(\mathbf{Z}_{ij}^T, \delta_{ij} \mathbf{W}_{ij}^T)$ 。

对于 $\alpha(s)$ 和 $\beta(s)$ 的基近似 (5.10) 非参数函数， $\alpha(s)$ 和 $\beta(s)$ 的加权最小二乘非参数估计可通过将 (5.11) 中的 $\alpha(s)$ 和 $\beta(s)$ 用 (5.10) 的右边项替换后来计算，即得

$$(\tilde{\alpha}_{LS}^T(s; \mathcal{B}), \tilde{\beta}_{LS}^T(s; \mathcal{B}))^T = \mathcal{B}(s) (\tilde{\gamma}_{LS}^T(\mathcal{B}), \tilde{\tau}_{LS}^T(\mathcal{B}))^T, \quad (5.13)$$

其中 $\{\tilde{\gamma}_{LS}(\mathcal{B}), \tilde{\tau}_{LS}(\mathcal{B})\}$ 是将 (5.12) 中的 $\mathcal{T}(s)$ 用 $\mathcal{B}(s)$ 替代后得到的。

当 $\Lambda_i = \mathbf{V}_i^{-1}$ 且假设分布函数为正态时，(5.12) 和 (5.13) 与极大似然估计或它们基于 B-样条近似结果相同。实际中 \mathbf{V}_i 常常是未知的，此时可以主观选择 Λ_i 。一个可能的“插入”方法是由数据来估计 \mathbf{V}_i ，然后通过用 \mathbf{V}_i^{-1} 的估计替代 Λ_i 来计算估计。但实际中，常常很难估计 \mathbf{V}_i ，而且也不清楚是否“插入”估计比用主观选择 Λ_i 得到的估计具有更好的统计性质。

§5.5 共享参数模型的估计方法

§5.5.1 线性可加共享参数模型

在许多有伴随干预的情形，变点时间 S_i 依赖于干预前 Y_{ij} 的趋势，由于一些个体的变点可能出现在研究期前或后，所以并非所有研究个体都能被观测到其变点时间。在此情形下，第 5.2 节将 S_i 视为双边删失，一个自然的方法就是将 (5.2) 进一步简化为更一般的实用模型。假设伴随干预效应仅依赖于结果变量干预前的趋势 \mathbf{a}_i ，而不依赖于变点时间 S_i 。则 (5.2) 的一个有用的特殊形式为

$$\begin{cases} Y_{ij} = \mu_0(T_{ij}, \mathbf{X}_i; \mathbf{a}_i) + \lambda_{ij} \mu_1(T_{ij}, \mathbf{X}_i, R_{ij}; \mathbf{b}_i) + \epsilon_{ij}, \\ \mathbf{a}_i \sim F_a(\cdot), S_i | \mathbf{a}_i \sim F_s(\cdot | \mathbf{a}_i), \mathbf{b}_i | \mathbf{a}_i \sim F_b(\cdot | \mathbf{a}_i), \end{cases} \quad (5.14)$$

其中 $F_a(\cdot)$ 为 \mathbf{a}_i 的累积分布函数 (CDF), $F_s(\cdot|\mathbf{a}_i)$ 和 $F_b(\cdot|\mathbf{a}_i)$ 分别为给定 \mathbf{a}_i 条件下 S_i 和 \mathbf{b}_i 的条件累积分布函数, 在给定 \mathbf{a}_i 下 \mathbf{b}_i 和 S_i 条件独立. 对照变系数模型 (5.8), 模型 (5.8) 用到了给定 S_i 下 \mathbf{a}_i 和 \mathbf{b}_i 的条件均值, 而 (5.14) 是借助于 $F_s(\cdot|\mathbf{a}_i)$ 将 S_i 加入模型的. 通过对给定 \mathbf{a}_i 下的 S_i 的条件分布建模, 因此, 模型 (5.14) 允许 S_i 被左删失或有删失. 尽管进一步的推广允许 (5.14) 中的 \mathbf{b}_i 的分布依赖于 (S_i, \mathbf{a}_i) , 但为简单起见, 这里假设 \mathbf{b}_i 不依赖于 S_i .

为了平衡计算的可行性和模型的灵活性, 在实际中可考虑 $F_a(\cdot)$, $F_s(\cdot|\mathbf{a}_i)$ 和 $F_b(\cdot|\mathbf{a}_i)$ 的更多规定. (5.14) 的一个有用的易计算的规定是假设 S_i 和 \mathbf{b}_i 仅通过它们的条件均值依赖于 \mathbf{a}_i , 是 \mathbf{a}_i 的线性函数. 则 (5.14) 的一个线性共享参数模型为

$$\begin{cases} Y_{ij} = \mu_0(T_{ij}, \mathbf{X}_i; \mathbf{a}_i) + \lambda_{ij} \mu_1(T_{ij}, \mathbf{X}_i, R_{ij}; \mathbf{b}_i) + \epsilon_{ij}, \\ \mathbf{a}_i = \alpha + e_i^{(a)}, S_i = \gamma^T(1, \mathbf{a}_i^T)^T + e_i^{(s)}, \mathbf{b}_i = \beta^T(1, \mathbf{a}_i^T)^T + e_i^{(b)}, \end{cases} \quad (5.15)$$

其中 $\alpha = (\alpha_1, \dots, \alpha_{d_0})^T$, $\alpha_d \in R$, $\beta = (\beta_1^T, \dots, \beta_{d_1}^T)^T$, $\beta_l = (\beta_{l0}, \dots, \beta_{ld_0})^T$, $\beta_{ld} \in R$, $\gamma = (\gamma_0, \dots, \gamma_{d_0})^T$, 和 $\{\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{in_i})^T, e_i^{(a)}, e_i^{(b)}, e_i^{(s)}\}$ 是独立的均值为零的随机误差, 其协方差阵分别为 $\{\mathbf{V}_y, \mathbf{V}_a, \mathbf{V}_b, \sigma_s^2\}$. (5.15) 中未知参数为均值分量 $\theta = (\alpha^T, \beta_1^T, \dots, \beta_{d_1}^T, \gamma^T)^T$ 和协方差结构 $\mathbf{V} = \{\mathbf{V}_y, \mathbf{V}_a, \mathbf{V}_b, \sigma_s^2\}$.

当 (5.14) 中的 S_i 和 \mathbf{a}_i 的关系未知时, $\{S_i, \mathbf{a}_i\}$ 的一个非参模型为 $S_i = \mu^{(s)}(\mathbf{a}_i) + \epsilon_i^{(s)}$, 其中 $\mu^{(s)}(\mathbf{a}_i) = E(S_i|\mathbf{a}_i)$ 是 \mathbf{a}_i 的一个光滑函数. 由于当 \mathbf{a}_i 是高维向量时, 将很难给出 $\mu^{(s)}(\mathbf{a}_i)$ 的无结构估计, 故应用一个简单辅助方法, 将 (5.15) 中 S_i 和 \mathbf{a}_i 的关系用

$$S_i = \sum_{d=0}^{d_0} \mu_d^{(s)}(a_{id}) + \epsilon_i^{(s)} \quad (5.16)$$

替换, 其中 $\mu_d^{(s)}(a_{id})$ 为 a_{id} 的一个光滑函数, 于是 (5.14) 的一个可加共享参数模型为

$$\begin{cases} Y_{ij} = \mu_0(T_{ij}, \mathbf{X}_i; \mathbf{a}_i) + \lambda_{ij} \mu_1(T_{ij}, \mathbf{X}_i, R_{ij}; \mathbf{b}_i) + \epsilon_{ij}, \\ \mathbf{a}_i = \alpha + e_i^{(a)}, S_i = \sum_{d=0}^{d_0} \mu_d^{(s)}(a_{id}) + \epsilon_i^{(s)}, \mathbf{b}_i = \beta^T(1, \mathbf{a}_i^T)^T + e_i^{(b)}. \end{cases} \quad (5.17)$$

虽然理论上可以进一步推广 (5.17), 但其相应的计算将非常复杂.

§5.5.2 极大似然估计方法

如果分布函数是一个已知的参数化形式, 则 (5.14) 中的参数就可由极大似然方法来估计. 记 $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$, $\mathbf{T}_i = (T_{i1}, \dots, T_{in_i})^T$, $\mathcal{D}_i = (\mathbf{T}_i, \mathbf{X}_i)$, 且 $f_y(\cdot)$, $f_b(\cdot)$, $f_s(\cdot)$ 和 $f_a(\cdot)$ 分别为 Y_{ij} , \mathbf{b}_i , S_i 和 \mathbf{a}_i 的密度函数. 则 $(\mathbf{b}_i, S_i, \mathbf{a}_i)$ 的联合密度函数可表示为

$$f(\mathbf{b}_i, S_i, \mathbf{a}_i) = f_b(\mathbf{b}_i|\mathbf{a}_i)f_s(S_i|\mathbf{a}_i)f_a(\mathbf{a}_i).$$

于是给定 $\mathcal{D}_i = (\mathbf{T}_i, \mathbf{X}_i)$ 下, (\mathbf{Y}_i, S_i) 的条件密度函数可通过关于 \mathbf{a}_i 和 \mathbf{b}_i 求积分得到, 即为

$$f_{(y,s)}(\mathbf{Y}_i, S_i|\mathcal{D}_i) = \iint f_y(\mathbf{Y}_i|\mathcal{D}_i, S_i, \mathbf{a}_i, \mathbf{b}_i)f_b(\mathbf{b}_i|\mathbf{a}_i)f_s(S_i|\mathbf{a}_i)f_a(\mathbf{a}_i) d\mathbf{a}_i d\mathbf{b}_i. \quad (5.18)$$

由于观测到的变点时间是 $\{S_i^{(c)}, \delta_i^{(c)}\}$ 双侧删失的, 故 (5.18) 的条件密度函数不可以直接用来做估计, 必须考虑用给定 \mathbf{a}_i 下 $\mathcal{S}_i^{(c)} = (S_i^{(c)}, \delta_i^{(c)})$ 的如下条件密度

$$f_s(S_i^{(c)}, \delta_i^{(c)}|\mathbf{a}_i) = \begin{cases} f_s(S_i|\mathbf{a}_i), & \text{如果 } \delta_i^{(c)} = 0, \\ 1 - F_s(T_{in_i}|\mathbf{a}_i), & \text{如果 } \delta_i^{(c)} = 1, \\ F_s(T_{i1}|\mathbf{a}_i), & \text{如果 } \delta_i^{(c)} = 2. \end{cases} \quad (5.19)$$

由 (5.18) 和 (5.19), $(\mathbf{Y}_i, \mathcal{S}_i^{(c)})$ 在给定条件 \mathcal{D}_i , $i = 1, \dots, n$ 下的条件似然函数为

$$L_c = \frac{1}{n} \sum_{i: \delta_i^{(c)}=0} \log f_{(y,s)}(\mathbf{Y}_i, S_i|\mathcal{D}_i) + \frac{1}{n} \sum_{l=1,2} \sum_{i: \delta_i^{(c)}=l} \log f_{(y,l)}(\mathbf{Y}_i|\mathcal{D}_i) \quad (5.20)$$

其中 $f_{(y,s)}(\cdot|\cdot)$ 的定义见 (??), $f_{(y,1)}(\cdot|\mathcal{D}_i, \mathbf{a}_i)$ 和 $f_{(y,2)}(\cdot|\mathcal{D}_i, \mathbf{a}_i, \mathbf{b}_i)$ 分别为给定 $\{\mathcal{D}_i, \mathbf{a}_i, \delta_i^{(c)} = 1\}$ 和 $\{\mathcal{D}_i, \mathbf{a}_i, \mathbf{b}_i, \delta_i^{(c)} = 2\}$ 下 \mathbf{Y}_i 的条件密度

$$f_{(y,1)}(\mathbf{Y}_i|\mathcal{D}_i) = \int f_{(y,1)}(\mathbf{Y}_i|\mathcal{D}_i, \mathbf{a}_i)\{1 - F_s(T_{in_i}|\mathbf{a}_i)\}f_a(\mathbf{a}_i) d\mathbf{a}_i$$

和

$$f_{(y,2)}(\mathbf{Y}_i|\mathcal{D}_i) = \iint f_{(y,2)}(\mathbf{Y}_i|\mathcal{D}_i, \mathbf{a}_i, \mathbf{b}_i)F_s(T_{i1}|\mathbf{a}_i)f_b(\mathbf{b}_i|\mathbf{a}_i)f_a(\mathbf{a}_i) d\mathbf{a}_i d\mathbf{b}_i.$$

若记 $f_{(y,s)}(\cdot|\cdot)$ 的参数族为 $\{f_{(y,s)}(\cdot|\phi|\mathcal{D}_i); \phi = (\theta, \mathbf{V})\}$, 则极大化对数似然函数 (5.20) 便可得到 $\phi = (\theta, \mathbf{V})$ 的 ML 估计.

§5.5.3 近似极大似然估计方法

可加共享参数模型 (5.17) 的估计可通过极大化 (5.17) 的一个近似的似然函数来得到。在 $\mu_d^{(s)}(\cdot)$ (如 Huang, Wu 和 Zhou, 2004) 的一些温和光滑条件下, $\mu_d^{(s)}(\cdot)$ 可由 B-样条的展开式来逼近, 即

$$\mu_d^{(s)}(a_{id}) \approx \sum_{p=1}^{P_d} \gamma_p^{(d)} B_p^{(d)}(a_{id}) = \left(\gamma^{(d)} \right)^T \mathbf{B}^{(d)}(a_{id}), \quad (5.21)$$

其中, 对某个 P_d , $\{B_p^{(d)}(\cdot); 1 \leq p \leq P_d\}$ 是一个样条基,

$$\mathbf{B}^{(d)}(a_{id}) = \left(B_1^{(d)}(a_{id}), \dots, B_{P_d}^{(d)}(a_{id}) \right)^T$$

和 $\gamma^{(d)} = (\gamma_1^{(d)}, \dots, \gamma_{P_d}^{(d)})^T$ 为一组实值系数。由 (5.16) 得

$$S_i \approx \sum_{d=0}^{d_0} (\gamma^{(d)})^T \mathbf{B}^{(d)}(a_{id}) + \epsilon_i^{(s)}. \quad (5.22)$$

事实上, 只是用 $\sum_{d=0}^{d_0} (\gamma^{(d)})^T \mathbf{B}^{(d)}(a_{id})$ 替代 (5.16) 的 $\sum_{d=0}^{d_0} \mu_d^{(s)}(a_{id})$, 其参数为

$$\theta = \left(\alpha^T, \beta_1^T, \dots, \beta_{d_1}^T, \gamma^T \right)^T, \quad \gamma = \left((\gamma^{(0)})^T, \dots, (\gamma^{(d_0)})^T \right)^T, \quad \mathbf{V} = \left\{ \mathbf{V}_y, \mathbf{V}_a, \mathbf{V}_b, \sigma_s^2 \right\}.$$

令 $f_s^*(\cdot; \gamma, \sigma_s | \mathbf{a}_i)$ 为 $\sum_{d=0}^{d_0} \{(\gamma^{(d)})^T \mathbf{B}^{(d)}(a_{id})\} + \epsilon_i^{(s)}$ 在给定 \mathbf{a}_i 下的条件密度。若零均值的 ϵ_{ij} , $e_i^{(a)}$, $e_i^{(b)}$ 和 $e_i^{(s)}$ 的分布可由方差 \mathbf{V} 的参数向量参数化, 则密度 $f_s(S_i | \mathbf{a}_i)$ 可用 $f_s^*(S_i; \gamma, \sigma_s | \mathbf{a}_i)$ 来逼近, 且 (5.17) 的参数可通过极大化下面的给定 \mathcal{D}_i 下 $(\mathbf{Y}_i, \mathcal{S}_i^{(c)})$ 的近似对数似然函数

$$L_c^*(\phi) = \frac{1}{n} \sum_{i: \delta_i^{(c)}=0} \log f_{(y,s)}^*(\mathbf{Y}_i, S_i; \phi | \mathcal{D}_i) + \frac{1}{n} \sum_{l=1,2} \sum_{i: \delta_i^{(c)}=l} \log f_{(y,l)}^*(\mathbf{Y}_i; \phi | \mathcal{D}_i) \quad (5.23)$$

来得到, 其中 $\phi = (\theta, \mathbf{V})$, $f_{(y,s)}^*(\cdot | \mathcal{D}_i)$, $f_{(y,k)}^*(\cdot | \mathcal{D}_i)$, $k = 1, 2$, 是将 (5.20) 中的 $f_s(S_i | \mathbf{a}_i)$ 用 $f_s^*(S_i; \gamma, \sigma_s | \mathbf{a}_i)$ 替代后所得。如果 $L_d^*(\phi)$ 满足 ML 估计的正则条件, 近似的 ML 估计 $\hat{\phi}$ 满足 $L_c^*(\hat{\phi}) = \max_{\phi} L_c^*(\phi)$ 。

§5.5.4 两步估计方法

似然函数 (5.20) 和 (5.23) 包含了参数的非线性项。实际上, 对 (5.20) 和 (5.23) 关于 θ 和 \mathbf{V} 同时作全局极大化, 在计算上是不可行的。为了减轻计算负担, Wu, Tian 和 Jiang (2011) 结合了 REMLE 和 Newton-Raphson 算法, 建议采用如下两步算法:

- (S1) 假设 (5.15) 或 (5.17) 中的 $\{\epsilon_{ij}, \mathbf{a}_i, \mathbf{b}_i, S_i\}$ 为独立随机变量, 其协方差阵为 $\mathbf{V} = \{\mathbf{V}_y, \mathbf{V}_a, \mathbf{V}_b, \sigma_s^2\}$, 即假设简单混合效应模型 (5.1) 成立。用 REMLE 程序计算 \mathbf{V} 的估计 $\hat{\mathbf{V}}$;
- (S2) 用 $\hat{\mathbf{V}}$ 替换 \mathbf{V} , 然后用 Newton-Raphson 程序关于 θ 极大化 $L_c(\theta, \hat{\mathbf{V}})$, 其极大值点 $\hat{\theta} = \arg \max_{\theta} L_c(\theta, \hat{\mathbf{V}})$ 为 θ 的近似的 ML 估计。

由 $f_{(y,s)}(\mathbf{Y}_i, S_i | \mathcal{D}_i)$ 和 $f_{(y,k)}(\mathbf{Y}_i | \mathcal{D}_i)$ ($k = 1, 2$) 的表达式易知: 第二步 (S2) 中极大化 $L_c(\theta, \hat{\mathbf{V}})$ 的 Newton-Raphson 算法涉及到关于 \mathbf{a}_i 和 \mathbf{b}_i 的联合分布在 \mathbf{a}_i , \mathbf{b}_i 和 S_i 函数上的多维积分。虽然在 Newton-Raphson 算法中全部必须被涉及到的量, 包括对数似然函数以及其梯度和 Hessian 阵, 都能够通过蒙特卡罗模拟来计算, 但由于这里每次迭代都需要大量的蒙特卡罗样本来计算梯度和 Hessian 阵, 故完整运行 Newton-Raphson 算法一次就需要花费很长时间。然而如果可得到一个合适的初始估计, 该算法的计算就能被降到“一步” Newton-Raphson 程序 (Bickel, 1975)。Wu, Tian 和 Jiang (2011) 建议用 REMLE 程序计算的估计作为初始估计 $\hat{\theta}_0$, 并用可观测到 S_i (即 $\delta_i^{(c)} = 0$) 的个体拟合回归模型 $S_i = \gamma^T(1, \tilde{\mathbf{a}}_i^{pred}) + \epsilon_i^{(s)}$ 来计算 γ 的初始估计, 其中 $\tilde{\mathbf{a}}_i^{pred}$ 为 \mathbf{a}_i 的预测值。

§5.6 Bootstrap 置信区间

理论上, 当 n 较大且模型 (5.15) 服从一个已知的参数分布族时, 可用 ML 估计 $\hat{\phi}$ 的渐近分布构造参数向量 ϕ 的近似推断。在适当的正则条件 (Serfling, 1980, Ch. 4)

下, ML 估计的渐近正态性表明 $\hat{\phi}$ 渐近于正态分布 $\mathcal{N}(\phi, \text{Var}(\phi))$, 因此, ϕ 的线性组合 $\ell(\phi)$ 的渐近 $[100 \times (1 - \alpha)]$ 置信区间为 $\ell(\hat{\phi}) \pm Z_{\alpha/2} [\hat{\text{Var}}\{\ell(\phi)\}]^{1/2}$, 其中 $\hat{\text{Var}}\{\ell(\phi)\}$ 为方差估计, $Z_{\alpha/2}$ 为标准正态分布的上 $\alpha/2$ 分位点。然而对于出现了非参部分的可加模型 (5.17), (5.23) 中近似 ML 估计的渐近分布还未被研究。作为实用的另一种选择, bootstrap 程序是通过个体重复抽样产生 bootstrap 样本, 计算相应的 bootstrap 估计。从原始样本得到的估计自然可作为 bootstrap 样本的初始估计。基于百分位点的 $[100 \times (1 - \alpha)]$ 置信区间可通过 bootstrap 估计的相应的下、上 $[100 \times (\alpha/2)]$ 分位数 $(L_{\alpha/2}, U_{\alpha/2})$ 得到。Wu, Tian 和 Bang (2008) 与 Wu, Tian 和 Jiang (2011) 已用过这个 bootstrap 方法。此外, 也可由 bootstrap 样本计算 $\hat{\text{Var}}\{\ell(\phi)\}$, 然后采用近似置信区间 $\ell(\hat{\phi}) \pm Z_{\alpha/2} [\hat{\text{Var}}\{\ell(\phi)\}]^{1/2}$ 。

§5.7 ENRICH D 药物治疗法数据中的应用

§5.7.1 可观测到变点的个体中的应用

第 1.3 节的例 4 已给出了 ENRICH D 研究的一个简要介绍。该研究的目标是针对 6 个月认知行为疗法 (CBT) 治疗期接受药物治疗的病人, 评价药物治疗 (抗抑郁药物) 对由其 BDI 得分测得的抑郁严重性趋势的影响。因为使用抗抑郁药和开始使用的时间都是由病人或他们的主治医生决定的, 所以使用抗抑郁药的药物治疗是该试验的一个伴随干预。

采用第 5.4 节回归结构, 这个分析中包括了 CBT 组病人中在试验期接受药物治疗并有他们开始药物治疗时间清晰记录的 91 个病人, 共 1,446 个观测, 其中排除了一般护理组 (UC) 病人的数据, 因为这些病人的药物治疗开始的时间和重复测得 BDI 得分没有被准确的记录下来, 也排除了 CBT 组中在六个月 CBT 治疗期没有被记录药物治疗具体开始时间的病人数据。在这 91 个病人中, 有 43 个病人是在试验一开始就采用药物治疗, 另 48 人是在试验开始后的 7 到 172 天之间开始药物治疗的。这些病人回访的次数为 5 次 ~36 次, 其中位数为 16。

根据 5.2 节, Y_{ij} 、 T_{ij} 、 S_i 、 $R_{ij} = T_{ij} - S_i$ 和 $\delta_{ij} = 1_{[T_{ij} \geq S_i]}$ 分别为第 j 次回访时第 i 个个体的 BDI 得分、试验时间 (月)、药物治疗开始时间、已药物治疗的时间、以及药物治疗标记。用来评价 BDI 得分随时间 T_{ij} 的变化趋势的一个简单混合效应模型为线性混合效应模型

$$Y_{ij} = a_{0i} + a_{1i}T_{ij} + b_{0i}\delta_{ij} + b_{1i}\delta_{ij}R_{ij} + \epsilon_{ij}, \quad (5.24)$$

其中 $E(a_{0i}, a_{1i}, b_{0i}, b_{1i})^T = (\alpha_0, \alpha_1, \beta_0, \beta_1)^T$. 当 $\delta_{ij} = 1$ 和 $R_{ij} = r$ 时, $(\beta_0 + \beta_1 r)$ 表示从开始药物治疗起的第 r 月的平均药物治疗效应。

由于 (5.24) 忽视了 S_i 和药物治疗前的抑郁趋势之间可能存在的相关性, 因此该模型存在潜在的偏, 故将其推广为变系数混合效应模型

$$Y_{ij} = \alpha_0(S_i) + \alpha_1(S_i)T_{ij} + \beta_0\delta_{ij} + \beta_1\delta_{ij}R_{ij} + e_{ij}, \quad (5.25)$$

其中 $e_{ij} = a_{i0}^* + a_{i1}^*T_{ij} + b_{i0}^*\delta_{ij} + b_{i1}^*\delta_{ij}R_{ij} + \epsilon_{ij}$, $\alpha_0(S_i) = \gamma_{00} + \gamma_{01}S_i$ 和 $\alpha_1(S_i) = \gamma_{10} + \gamma_{11}S_i$. 在模型 (5.25) 中, 药物治疗前 BDI 趋势是通过截距 $\alpha_0(S_i)$ 和斜率 $\alpha_1(S_i)$ 与 S_i 产生关联的。从开始药物治疗起的第 r 月的平均药物治疗效应为 $\beta_0 + \beta_1 r$ 。 $\beta_0 + \beta_1 r$ 的负 (正) 值对应于有益 (有害) 于降低抑郁。

为数学上简单, 假设模型 (5.25) 中 $\beta_0(S_i) \equiv \beta_0$ 和 $\beta_1(S_i) \equiv \beta_1$, 这样药物治疗的效应仅依赖于使用抗抑郁药的时间长度。这种假设下, 尽管平均 BDI 得分有不同的药物治疗前的时间趋势, 但 β_0 和 β_1 在简单线性混合效应模型 (Näive LME) (5.24) 和共享参数线性混合效应模型 (SP-LME) (5.25) 下具有相同的解释。

表 4 概括了由带无结构的相关的限制极大似然 (REML) 程序得到的 β_0 和 β_1 的估计和它们相应的标准差 (SE)、95% 置信区间和 p- 值。在 Näive LME 和 SP-LME 模型下, (β_0, β_1) 的估计皆为负的, 这表明当分析中只包括 CBT 期有药物治疗变点时间的病人时, Näive LME 和 SP-LME 模型下都检测到药物治疗对这个病人总体有益的影响。

表 4 对简单线性混合效应模型 (Näive LME) (5.24) 和共享参数线性混合效应模型 (SP-LME) (5.25) 使用带无结构的相关的 REML 程序获得的 β_0 和 β_1 的参数估计和它们的标准差 (SE)、95% 置信区间 (CIs) 和 p- 值.

模型	参数	估计	标准差	95% CI	p- 值
Näive LME	β_0	-3.410	0.994	(-5.399, -1.422)	0.0013
	β_1	-1.584	0.521	(-2.626, -0.542)	0.0039
SP-LME	β_0	-4.302	1.041	(-6.385, -2.220)	0.0001
	β_1	-2.062	0.773	(-3.608, -0.516)	0.0105

§5.7.2 应用于个体变点删失的情形

上面的分析仅用了 ENRICH D CBT 组中在 CBT 治疗期变点时间 S_i 被观测到的 91 个抑郁病人这个子样本。因此，抗抑郁药有利于降低 BDI 得分的结论忽视了 CBT 治疗期没有开始使用药物疗法的病人信息，下面用共享参数模型，基于 557 抑郁病人进行分析，这些病人有具体的开始使用抗抑郁药的时间记录，并在 6 个月的治疗期参加了 5 次或更多次 CBT 课程。为实际应用考虑，该分析排除了 UC 组的病人、没有记录开始服用抗抑郁药的时间的病人和对要求参加每周 CBT 课程的依从性差 (参加课程少于 5 次) 的病人。由于每个病人抗抑郁药的使用由研究精神病的医师单独监督和尽可能准确地记录的 (Taylor 等, 2005, 第 794 页)，因此，假设抗抑郁药开始日期记录的缺失为随机缺失是合理的。纵向样本包括的 11 个在试验前就是用抗抑郁药的病人，92 个在治疗期开始使用抗抑郁药的病人和 454 个在治疗前和治疗期都没有使用抗抑郁药的病人。这些病人的回访次数范围是 5 到 36 次，中位数为 12 次。

对第 5.7.1 节的记号稍加修改，令 Y_{ij} , T_{ij} , $S_i^{(c)}$, and $R_{ij} = T_{ij} - S_i^{(c)}$ 分别表示第 j 次回访中第 i 个病人的 BDI 得分，试验时间 (月)，使用抗抑郁药的开始时

间 (月)，和已使用抗抑郁药的时间 (月)。对所有的 $1 \leq i \leq n$, 被观测的 $(S_i^{(c)}, \delta_i^{(c)})$ 分别为 $(S_i^{(c)} = S_i, \delta_i^{(c)} = 0)$ 如果第 i 个病人在 CBT 期内使用抗抑郁药物； $(S_i^{(c)} = T_{in_i}, \delta_i^{(c)} = 1)$ 如果第 i 个病人在 CBT 期内没有使用抗抑郁药物； $(S_i = 0, \delta_i^{(c)} = 2)$ 如果第 i 个病人在 CBT 期前使用抗抑郁药物。当 $\lambda_{ij} = 1_{[S_i < T_{ij}]}$ 且采用线性模型 $\mu_0(T_{ij}; a_{i0}, a_{i1}) = a_{i0} + a_{i1}T_{ij}$ 和 $\mu_1(R_{ij}; b_{i0}, b_{i1}) = b_{i0} + b_{i1}R_{ij}$ 时， (a_{i0}, a_{i1}) 表示使用抗抑郁药物前第 i 个个体的 BDI 轨迹的截距和斜率， (b_{i0}, b_{i1}) 表示使用抗抑郁药物后第 i 个个体的 BDI 轨迹改变量的截距和斜率。

Wu, Tian 和 Jiang (2011, 网络附加材料附录 D.1) 的一系列初步分析表明：上面线性模型 $\mu_0(T_{ij}; a_{i0}, a_{i1})$ 和 $\mu_1(R_{ij}; b_{i0}, b_{i1})$ 可以被用来作为该研究中 BDI 趋势的精确近似。

与 (5.24) 类似，可应用 REMLE 程序来估计下列简单混合效应模型

$$\begin{cases} Y_{ij} = a_{i0} + a_{i1}T_{ij} + \lambda_{ij}(b_{i0} + b_{i1}R_{ij}) + \epsilon_{ij}, \\ (a_{i0}, a_{i1}, b_{i0}, b_{i1})^T \sim \mathcal{N}((\alpha_0, \alpha_1, \beta_0, \beta_1)^T, \Sigma) \end{cases} \quad (5.26)$$

的未知总体均值参数 $\alpha_0, \alpha_1, \beta_0$ 和 β_1 ，其中 $(\alpha_0, \alpha_1, \beta_0, \beta_1)^T$ 为未知的均值向量， Σ 为多元正态分布 $\mathcal{N}(\cdot, \cdot)$ 的无结构协方差阵。总体均值伴随干预效应为 β_0 和 β_1 ，分别表示使用抗抑郁药物后“调整项”的平均截距和斜率，故 $\beta_0 + \beta_1 R_{ij} = 0$ 表明抗抑郁药对平均 BDI 得分的影响可以被忽略。

为了考虑药物疗法变点时间和用抗抑郁药前的 BDI 趋势之间可能存在的关联，直接将 (5.26) 推广为一个共享参数模型

$$\begin{cases} Y_{ij} = a_{i0} + a_{i1}T_{ij} + \lambda_{ij}(b_{i0} + b_{i1}R_{ij}) + \epsilon_{ij}, & (a_{i0}, a_{i1})^T = (\alpha_0, \alpha_1)^T + \epsilon_i^{(a)}, \\ b_{i0} = \beta_0 + \epsilon_{i0}^{(b)}, & b_{i1} = \beta_1 + \epsilon_{i1}^{(b)}, & S_i = \gamma_0 + \gamma_1 a_{i0} + \epsilon_i^{(s)}, \end{cases} \quad (5.27)$$

其中 $\epsilon_i^{(a)}$ 和 $\epsilon_i^{(b)} = (\epsilon_{i0}^{(b)}, \epsilon_{i1}^{(b)})^T$ 是均值为零、无结构协方差阵分别为 $\Sigma^{(a)}$ 和 $\Sigma^{(b)}$ 的二元正态随机向量， $\epsilon_i^{(s)}$ 是均值为零、方差为 σ_s^2 的随机变量， $\epsilon_i^{(a)}$, $\epsilon_i^{(b)}$ 和 $\epsilon_i^{(s)}$ 独立。

(5.27) 中总体均值参数 $\alpha_0, \alpha_1, \beta_0$ 和 β_1 与 (5.26) 中相应部分具有相同的解释。

表 5 ENRICHHD 针对药物治疗数据, 分别基于简单混合效应模型 (Näive-LMEM) (5.26) 和线性共享参数模型 (LSPM)(5.27) 计算的参数估计, 标准差 (SE) 和 95% 置信区间 (CI).

效应 参数	Näive-LMEM			LSPM		
	估计	标准差	95% CI	估计	标准差	95% CI
α_0	14.454	0.312	(13.842, 15.066)	15.867	0.375	(15.216, 16.771)
α_1	-1.887	0.067	(-2.018, -1.756)	-1.816	0.070	(-1.989, -1.714)
β_0	3.579	0.825	(1.962, 5.196)	-6.646	0.915	(-8.239, -4.990)
β_1	0.036	0.227	(-0.409, 0.481)	-0.453	0.284	(-0.962, 0.122)
γ_0	—	—	—	15.162	1.134	(13.497, 17.652)
γ_1	—	—	—	0.480	0.050	(-0.593, -0.392)

表 5 显示简单混合效应模型 (Näive-LMEM) (5.26) 下采用无结构协方差阵 Σ 的 REMLE 程序和线性共享参数模型 (LSPM)(5.27) 下采用两步 ML 程序的 10 次 Newton-Raphson 迭代计算的 α_0 , α_1 , β_0 和 β_1 的估计, 它们的标准差 (SE) 和 95% 置信区间 (CI)。在 Näive-LMEM, 负估计 $\tilde{\alpha}_1 = -1.887$ 表明从开始 CBT 课程起, 这些病人的平均 BDI 得分倾向于随着试验时间降低, 而正的 $\tilde{\beta}_0 = 3.579$ 和它的 95% 置信区间似乎表明使用抗抑郁药物增加了病人的平均 BDI 得分。由于在 (5.26) 中没有考虑作为伴随干预的使用抗抑郁药物的“自选性”, 因此, 在该模型下 β_0 的正估计不能反映药物治疗对抑郁严重性的真实效应。另一方面, 在 LSPM (5.27) 下, 负估计 $\hat{\beta}_0 = -6.646$ 和 $\hat{\beta}_1 = -0.453$ 表明平均意义下药物治疗对降低病人的抑郁严重性有好的影响。由 (5.27) 得到的 γ_0 和 γ_1 的 95% 置信区间表明: 对这个数据集而言, 模型 (5.26) 可能被错误指定。

§5.8 小结与讨论

本节主要结果表明结果自适应协变量, 如伴随干预, 一般情况下不应该被当作是简单混合效应模型下的时间依赖协变量; 对于带一个伴随干预的简单情形, 可以

考虑共享参数模型来减少估计的偏和调整伴随干预的自选性。这里提供的方法只是针对纵向临床试验中仅有一个伴随干预的情形。伴随干预通常也出现在其它情形，如在流行病学研究中，当研究个体的血压在研究期呈现不理想的趋势或在一个不可容忍的区域时，他们可以服用抗高血压药。在 ENRICH 药物治疗数据中，药物治疗法作为伴随干预是在很模糊指导下给出的，选择线性共享参数模型看起来是有合理的。然而当改变干预选择机制时，这个模型就不适合了，在一些情形，整个共享参数方法可能不得不重新评估。

作为共享参数模型的一种特殊情形，变系数混合效应模型之所以研究中被考虑主要是因为，在仅有一个伴随干预、全部个体的变点时间都能被观测到这种简单情形下，该模型有简单清楚的生物学解释。与共享参数模型相比，变系数混合效应模型下基于最小二乘的估计方法不要求分布函数为已知的参数形式。另一方面，共享参数模型可应用于变点时间带双向删失的伴随干预情形，但它们的估计需要应用计算量很大的 ML 和近似 ML 算法。

在这个领域，有以下几方面的推广值得进一步研究。第一，在纵向研究中的个体可能有单一或不同时间点开始或停止的多重伴随干预，这种情形，需要更一般的共享参数模型来适应多重干预和 / 或 多重变点的可能性；第二，本节所研究的全部共享参数模型都是通过线性函数来描述干预前、后的时间趋势，但对某些疾病结果线性响应曲线可能是不适合的，在实际应用中，非线性响应曲线模型更合适，更应该被研究；第三，这部分的估计依赖于经典的频率学派体系的 B- 样条方法，在不同的背景下，Fahrmeir 和 Lang (2000) 显示一个基于 Markov 随机场先验的广义可加混合效应模型的前景更好的贝叶斯推断程序。本节共享参数模型下的类似的贝叶斯估计和推断该方法可以导出计算更简单的估计和推断方法。第四，对于 ML 和近似 ML 估计的大样本性质，如收敛速度和渐近分布，仍没有很好的了解，应该进一步系统研究，并为这些估计的合理性提供的理论依据。最后，由于能不总是清楚

的知道是否一个干预是伴随干预, 因此, 评估共享参数模型合适性的模型诊断方法将是一个很值得研究的问题。

参考文献

- [1] AKAIKE, H. (1970). Statistical predictor identification. *Ann. Inst. Statist. Math.* **22**, 203–217.
- [2] ALTMAN, N. S. (1990). Kernel smoothing of data with correlated errors. *J. Amer. Statist. Assoc.* **85**, 749–759.
- [3] BANG, H. and ROBINS, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61**, 962–972.
- [4] BATES, D. M. and PINHEIRO, J. C. (1999) *Mixed Effects Models in S*. Springer-Verlag, New York.
- [5] BICKEL, P. J. (1975). One-step Huber estimates in linear models. *Journal of the American Statistical Association* **34**, 584653.
- [6] CHENG, S. C., WEI, L. J. and YING, Z. (1995). Analysis of transformation models with censored data. *Biometrika* **82**, 835–845.
- [7] CHENG, S. C., WEI, L. J. and YING, Z. (1997). Predicting survival probabilities with semi-parametric transformation models. *Journal of the American Statistical Association* **92**, 227–235.
- [8] CHIANG, C.-T., RICE, J. A. and WU, C. O. (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variable. *J. Amer. Statist. Assoc.* **96**(454):605-619.
- [9] DANIELS, S. R., MCMAHON, R. P., OBARZANEK, E., WACLAWIW, M. A., SIMILO, S. L., BIRO, F. M, SCHREIBER, G. B., KIMM, S. Y. S., MORRISON, J. A. and BARTON, B. A. (1998). Longitudinal correlates of change in blood pressure in adolescent girls. *Hypertension* **31**, 97–103.
- [10] DAVIDIAN, M. and GILTINAN, D.M. (1995). *Nonlinear Models for Repeated Measurement Data*. London; New York: Chapman Hall.
- [11] DE BOOR, C. (1978). *A Practical Guide to Splines*. Springer-Verlag, New York.
- [12] DIGGLE, P. J. (1988). An approach to the analysis of repeated measurements. *Biometrics* **44**, 959-971.

- [13] DIGGLE, P. J., HEAGERTY, P., LIANG, K.-Y. and ZEGER, S. L. (2002). *Analysis of Longitudinal Data.*, 2nd ed. Oxford: Oxford University Press, England.
- [14] EUBANK, R. L. and SPECKMAN, P. L. (1993). Confidence bands in nonparametric regression. *J. Amer. Statist. Assoc.* **88**, 1287-1301.
- [15] ENRICHD INVESTIGATORS. (2001). Enhancing recovery in coronary heart disease patients (ENRICHD): study intervention rationale and design *Psychosomatic Medicine* **63**, 747-755.
- [16] ENRICHD INVESTIGATORS. (2003). Enhancing recovery in coronary heart disease patients (ENRICHD): the effects of treating depression and low perceived social support on clinical events after myocardial infarction. *Journal of the American Medical Association* **289**, 3106-3116.
- [17] FAHRMEIR, L. and LANG, S. (2000). Bayesian inference for generalised additive mixed models based on Markov random field priors. *Applied Statistics*, **50**, 201-220.
- [18] FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modeling and Its Applications*. London: Chapman and Hall.
- [19] FAN, J. and MARRON, J. S. (1994). Fast implementations of nonparametric curves estimators. *J. Comput. Graph. Statist.*, **3**, 35-56.
- [20] FAN, J. Q. and ZHANG, J.-T. (2000). Functional linear models for longitudinal data. *J. Roy. Statist. Soc. B* **62**, 303-322.
- [21] FOLLMANN, D. and WU, M. C. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics* **51**, 151-168.
- [22] HALL, P., and MÜLLER, H.G. (2003). Order-preserving Nonparametric regression, with applications to conditional distributions and quantile function estimation. *J. Amer. Statist. Assoc.* **98**, 598-608.
- [23] HALL, P. and TITTERINGTON, D. M. (1988). On confidence bands in nonparametric density estimation and regression. *J. Multi. Anal.* **27**, 228-254.
- [24] HALL, P., WOLFF, R. C. L. and YAO, Q. (1999). Methods for estimating a conditional distribution function. *J. Amer. Statist. Assoc.* **94**, 154-163.
- [25] HÄRDLE, W. (1990). *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, U.K.
- [26] HÄRDLE, W. and MARRON, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* **19**, 778-796.

- [27] HART, T. D. (1991). Kernel regression estimation with time series errors. *J. Roy. Statist. Soc., Ser. B* **53**, 173-187.
- [28] HART, T. D. and WEHRLY, T. E. (1986). Kernel regression estimation using repeated measurements data. *J. Amer. Statist. Assoc.* **81**, 1080-1088.
- [29] HARVILLE, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika* **61**, 383-385.
- [30] HASTIE, T. J. and TIBSHIRANI, R. J. (1993). Varying-coefficient models. *J. Roy. Statist. Soc. B* **55**, 757-796.
- [31] HOOVER, D. R., RICE, J. A., WU, C. O. and YANG, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85**, 809-822.
- [32] HUANG, J., WU, C. O. and ZHOU, L. (2002). Varying coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89**, 111-128.
- [33] HUANG, J. Z., WU, C. O. and ZHOU, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica* **14**, 763-88.
- [34] JONES, R. H. and ACKERSON, L. M. (1990). Serial correlation in unequally spaced longitudinal data. *Biometrika* **77**, 721-731.
- [35] JONES, R. H. and BOADI-BOTENG, F. (1991). Unequally spaced longitudinal data with serial correlation. *Biometrics* **47**, 161-175.
- [36] KASLOW, R. A., OSTROW, D. G., DETELS, R., PHAIR, J. P., POLK, B. F. and RINALDO, C. R. (1987). The Multicenter AIDS Cohort Study: Rationale, Organization and Selected Characteristics of the Participants. *American Journal of Epidemiology* **126**, 310-318.
- [37] KIMM, S. Y., BARTON, B. A., OBARZANEK, E., MCMAHON, R. P., SABRY, Z. I., WACLAWIW, M. A., et al. (2001). Racial divergence in adiposity during adolescence: the NHLBI Growth and Health Study. *Pediatrics* 2001;107:E34.
- [38] KIMM, S. Y., BARTON, B. A., OBARZANEK, E., MCMAHON, R. P., KRONBERG, S. S., WACLAWIW, M. A., et al. (2002). Obesity development during adolescence in a biracial cohort: the NHLBI Growth and Health Study. *Pediatrics* 2002;110:E54.
- [39] KIMM, S. Y., GLYNN, N. W., KRISKA, A. M., FITZGERALD, S. L., AARON, D. J., SIMILO, S. L., et al. (2000). Longitudinal changes in physical activity in a biracial cohort during adolescence. *Med Sci Sports Exerc* 2000;**32**:1445-54.
- [40] KNAFL, G., SACKS, J. and YLVISAKER, D. (1985). Confidence bands for regression functions. *J. Amer. Statist. Assoc.* **80**, 683-691.

- [41] LAIRD, N. M. and WARE, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963-974.
- [42] LIANG, K.-Y. and ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.
- [43] LIN, X. and CARROLL, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *J. Am. Statist. Assoc.* **95**, 520-534.
- [44] LIN, X. and ZHANG, D. (1999). Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society: Series B*, **61**(2), 381-400.
- [45] LU, W. and TSIATIS, A. A. (2006). Semiparametric transformation models for the case-cohort study. *Biometrika* **93**, 207-214.
- [46] LU, W. and YING, Z. (2004). On semiparametric transformation cure models. *Biometrika* **91**, 331-343.
- [47] MOLENBERGHS, G. and VERBEKE, G. (2005). *Models for Discrete Longitudinal Data*. Springer: New York, NY.
- [48] MOYEED, R. A. and DIGGLE, P. J. (1994). Rates of convergence in semiparametric modeling of longitudinal data. *Austral. J. Statist.* **36**, 75-93.
- [49] MÜLLER, H.-G. (1988). Nonparametric Regression Analysis of Longitudinal Data. *Lecture Notes in Statistics*, **46**. Springer-Verlag, Berlin.
- [50] NATIONAL HIGH BLOOD PRESSURE EDUCATION PROGRAM WORKING GROUP ON HIGH BLOOD PRESSURE IN CHILDREN AND ADOLESCENTS (NHBPEP Working Group) (2004). The fourth report on the diagnosis, evaluation, and treatment of high blood pressure in children and adolescents. *Pediatrics* **114**, 555-576.
- [51] NATIONAL HEART, LUNG, AND BLOOD INSTITUTE GROWTH AND HEALTH RESEARCH GROUP (NGHSRG) (1992). Obesity and cardiovascular disease risk factors in black and white girls: the NHLBI Growth and Health Study. *Am J. Public Health* **82** 1613-1620.
- [52] OBARZANEK, E., WU, C. O., CUTLER, J. A., KAVEY, R. W., PEARSON, G. D. and DANIELS, S. R. (2010). Prevalence and incidence of hypertension in adolescent girls. *The Journal of Pediatrics* **157**(3), 461-467.
- [53] PANTULA, S. G. and POLLOCK, K. H. (1985). Nested analysis of variance with autocorrelated errors. *Biometrics*, **41**, 909-920.
- [54] PATTERSON, H. D. and THOMPSON, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545-554.

- [55] PEPE, M. S. and ANDERSON, G. (1994). A cautionary note on inference of rmarginal regression models with longitudinal data and general correlated response data. *Communications in Statistics: Simulation and Computation*, **23**, 939-951.
- [56] RICE, J. A. and SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B* **53**, 233-243.
- [57] RICE, J. A. and WU, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57**, 253-259.
- [58] SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6**, 461-464.
- [59] SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New York: Wiley & Sons.
- [60] SHIBATA, R. (1981). An optimal selection of regression variables. *Biometrika* **68**, 45-54.
- [61] TAYLOR, C. B., YOUNGBLOOD, M. E., CATELLIER, D., VEITH, R. C., CARNEY, R. M., BURG, M. M., KAUFMANN, P., SHUSTER, J., MELLMAN, T., BLUMENTHAL, J. A., KRISHNAN, R. and JAFFE, A. S. (2005). Effects of antidepressant medication on morbidity and mortality in depressed patients after myocardial infarction. *Archives of General Psychiatry*, **62**, 792-798.
- [62] THOMPSON, D. R., OBARZANEK, E., FRANKO, D. L., BARTON, B. A., MORRISON, J., BIRO, F. M., DANIELS, S. R. and STRIEGEL-MOORE, R. H. (2007). Childhood overweight and cardiovascular disease risk factors: The National Heart, Lung, and Blood Institute Growth and Health Study. *Journal of Pediatrics* **150**, 18-25.
- [63] VERBEKE, G. and MOLENBERGHS, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer, New York.
- [64] VONESH, E. F. and CHINCHILLI, V. M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. Marcel Dekker, New York.
- [65] WAHBA, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.
- [66] WARE, J. H. (1985). Linear models for the analysis of longitudinal studies. *The American Statistician*, **39**, 95-101.
- [67] WU, C. O. and CHIANG, C.-T. (2000). Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Statistica Sinica* **10**, 433-456.
- [68] WU, C. O., CHIANG, C.-T. and HOOVER, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *J. Amer. Statist. Assoc.* **93**, 1388-1402.

- [69] WU, C. O., TIAN, X. and BANG, H. (2008). A varying-coefficient model for the evaluation of time-varying concomitant intervention effects in longitudinal studies. *Statistics in Medicine* **27**, 3042–3056.
- [70] WU, C. O., TIAN, X. and JIANG, W. (2011). A shared parameter model for the estimation of longitudinal concomitant intervention effects. *Biostatistics* **12**(4):737–749.
- [71] WU, C. O., TIAN, X. and YU, J. (2010). Nonparametric estimation for time-varying transformation models with longitudinal data. *Journal of Nonparametric Statistics* **22**, 133–147.
- [72] WU, C. O., YU, K. F. and CHIANG, C.-T. (2000). A two-step smoothing method for varying-coefficient models with repeated measurements. *Ann. Inst. Statist. Math.* **52**, 519–543.
- [73] WU, C. O., YU, K. F. and YUAN, V. W. S. (2000). Large sample properties and confidence bands for component-wise varying-coefficient regression with longitudinal dependent variable. *Commun. Statist.–Theory Meth.* **29**, 1017–1037.
- [74] ZEGER, S. L. and DIGGLE, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* **50**, 689–699.
- [75] ZEGER, S. L., LIANG, K.-Y. and ALBERT, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics* **44**, 1049–1060.
- [76] ZENG, D. and LIN, D. Y. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika* **93**, 627–640.