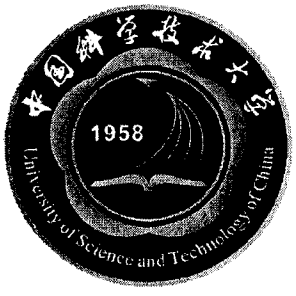
中国科学技术大学

硕士学位论文



纵向数据均值-协方差建模中的滑动平均因子模型

作者姓名：刘笑宇

学科专业：概率论与数理统计

导师姓名：张伟平 副教授

完成时间：二零一二年四月一日

University of Science and Technology of China

A dissertation for master degree

A moving average Cholesky factor model in joint mean-covariance modeling for longitudinal data

Author : Xiaoyu Liu

Speciality : Probability Theory and Math Statistics

Supervisor : Dr. Weiping Zhang

Finished Time : April 1st, 2012

中国科学技术大学学位论文原创性声明

本人声明所呈交的学位论文，是本人在导师指导下进行研究工作所取得的成果。除已特别加以标注和致谢的地方外，论文中不包含任何他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的贡献均已在论文中作了明确的说明。

作者签名：刘奘宇　　签字日期：2012.5.31

中国科学技术大学学位论文授权使用声明

作为申请学位的条件之一，学位论文著作权拥有者授权中国科学技术大学拥有学位论文的部分使用权，即：学校有权按有关规定向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅，可以将学位论文编入《中国学位论文全文数据库》等有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。本人提交的电子文档的内容和纸质论文的内容相一致。

保密的学位论文在解密后也遵守此规定。

公开　保密 _____ 年

作者签名：刘奘宇　　导师签名：张伟平

签字日期：2012.5.31　　签字日期：2012.5.31

摘 要

在纵向数据分析中，若要运用广义估计方程的方法有效估计均值参数，同时对均值和协方差建模是一个非常普遍的途径。在这篇文章中，我们提出了一种新的回归模型来对协方差结构进行参数化。我们引入一种新的且具有良好统计意义解释的 Cholesky 分解，其中两个分解因子是对角元为 1 的下三角阵，其非对角元素是滑动平均系数，另一个因子是对角元为误差项方差的对角阵，而这些参数都可以用协变量的线性函数来建模。我们可以证明由此得到的广义估计方程估计都具有相合性与渐近正态性。一组真实数据分析以及数值模拟研究也都验证了我们提出的估计方法能够高效估计均值参数和协方差结构，并且比以往的一些方法节省了很多工作量。

关键词： 滑动平均因子，广义估计方程，纵向数据，均值与协方差结构建模

原书空白页，
不缺内容

ABSTRACT

Modeling the mean and covariance simultaneously is a common strategy to efficiently estimate the mean parameters when using generalized estimating equation techniques to longitudinal data. In this article, we propose new regression models for parameterizing covariance structures. Using a novel Cholesky factor, the entries in this decomposition having moving average and log innovation interpretation and are modeled as linear functions of covariates. The resulting estimators for the regression coefficients in both the mean and covariance are shown to be consistent and asymptotically normally distributed. Simulation studies and a real data analysis show that the proposed approach yields highly efficient estimators for the parameters in the mean, and provides parsimonious estimation for the covariance structure.

Keywords: Moving average factor, Generalized estimating equation, Longitudinal data, Modeling of mean and covariance structures

原书空白页，
不缺内容

目 录

第一章　绪论

纵向数据由于其在实际中的广泛存在与应用，越来越受到理论与应用学者们的重视，已经发展成为统计学中一个非常重要的研究方向。纵向数据最为标志性的特点就是同一个个体在不同时间被重复观测若干次，因而即使我们可以假设不同个体间是相互独立的，也不能忽略同一个体的重复观测值间的相依关系。因此，纵向数据不但能够反映出同一时间个体间的差异，而且能够反映出观测值随时间的变化趋势，而后者往往也是我们所关心的。纵向数据不但广泛存在于很多科研领域，如经济学、流行病学、遗传学以及社会科学等等，在实际生活中也经常能够看到，例如很多支股票在一个月内的走势。应用各种模型与方法从这些复杂繁多的数据中提取我们想要的信息，被称为纵向数据分析。

1.1　纵向数据简介

在纵向数据分析中，最基本的问题是研究每组响应变量随时间的变化，以及这些变化在不同个体间的差异。在一些问题中，我们或许还需要由给定的数据对另外一些个体进行预测。在纵向数据中，如果每个个体的重复观测次数以及观测时间都一样，我们称这样的数据为“平衡”纵向数据。如若不然，称为“非平衡”纵向数据。后者也许在实际中更为常见，因为除非是严格设计的试验，往往我们不能保证每个个体的观测次数与时间都完全一致，即使在设计好的试验中，也有可能由于参与者提前离开或者记录的疏忽等原因造成不完整数据，从而产生非平衡数据。

在这里我们引入一些纵向数据分析中常用的简单记号。假设在一次试验中，有 N 个个体，第 i 个个体有 n_i 个观测值，记 y_{ij} 就是第 i 个个体的第 j 个观测值，其对应的观测时间是 t_{ij}，那么我们记个体 i 的观测值为 $y_i = (y_{i1}, \cdots, y_{in_i})'$。注意，如果 $\{n_i, i = 1, ..., N\}$ 以及 $\{t_{ij}, i = 1, ..., N, j = 1, ..., n_i\}$ 均与下标 i 无关，则该数据集就是平衡纵向数据。记 $\mu_{ij} = E(y_{ij})$ 为 y_{ij} 的均值，$\sigma_{ij}^2 = E(y_{ij} - \mu_{ij})^2$

为 y_{ij} 的方差。前面已经说到，同一个个体观测值间的相依性非常重要，因此我们记 $\sigma_{ijk} = E\{(y_{ij} - \mu_{ij})(y_{ik} - \mu_{ik})\}$ 为 y_{ij} 与 y_{ik} 的协方差。从而得到第 i 个响应变量 $y_i = (y_{i1}, \cdots, y_{in_i})'$ 的协方差矩阵为：

$$Cov\begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{pmatrix} = \begin{pmatrix} \sigma_{i1}^2 & \sigma_{i12} & \cdots & \sigma_{i1n_i} \\ \sigma_{i21} & \sigma_{i2}^2 & \cdots & \sigma_{i2n_i} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{in_i1} & \sigma_{in_i2} & \cdots & \sigma_{in_i}^2 \end{pmatrix}. \tag{1.1}$$

在实际中我们往往关心的是响应变量随时间的变化趋势，也就是数据的均值部分。近几十年来，学者们为此提出了参数模型，非参数模型以及两者结合的半参数模型等。下面我们就简单地介绍一下这些基本模型。

参数模型，也就是线性模型，主要是指响应变量的均值部分是参数的线性回归函数，即

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{pmatrix} = \begin{pmatrix} X_{i11} & X_{i12} & \cdots & X_{i1p} \\ X_{i21} & X_{i22} & \cdots & X_{i2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{in_i1} & X_{in_i2} & \cdots & X_{in_ip} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{in_i} \end{pmatrix}, \tag{1.2}$$

其中 $X_i = (X_{ijk})_{j=1,\ldots,n_i;k=1,\ldots,p}$ 是第 i 个个体对应的协变量矩阵，$\beta = (\beta_1, \ldots, \beta_p)'$ 为参数，而 $e_i = (e_{i1}, \ldots, e_{in_i})'$ 是第 i 个个体观测值的误差项。

我们常见的纵向数据参数模型就是多项式模型，如线性、二次甚至更高阶多项式。在这种模型中，出于研究响应变量均值随时间变化的趋势这一目的，我们一般都是将响应变量均值部分表示为时间的多项式函数，这样能够很轻易地处理高度非平衡数据。具体而言，在上述参数模型 (1.2) 中，协变量矩阵的元都是时间的多项式，例如 p 阶多项式模型：

$$X_i = \begin{pmatrix} 1 & t_{i1} & \cdots & t_{i1}^p \\ 1 & t_{i2} & \cdots & t_{i2}^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t_{in_i} & \cdots & t_{in_i}^p \end{pmatrix}.$$

在上述模型中，时间 t_{ij} 也可以换成任何与时间相关的变量，如何选取可以依赖于实际经验，最好能够有合理的理由作为支持。在本文中，我们就是引入时间多项式模型来拟合均值，而引入相应时间间隔的多项式来对表示前后测量值关系的滑动平均系数进行建模。

当我们假定均值部分是某个变量，比如时间 t 的函数时，我们就得到了所谓的非参数模型：

$$y_{ij} = f(t_{ij}) + e_{ij},$$

其中可以假定均值 $f(t_{ij})$ 与误差 e_{ij} 相互独立，$f(\cdot)$ 是满足某些条件的函数。此时，可以利用样条等非参数方法进行处理。推而广之，当均值部分同时包含参数项与非参数项时，称为半参数模型。

当我们对数据进行初步建模以后，就可以对分布进行假定，比如常见的多元正态分布。此时，我们可以引入一些常用的方法，比如极大似然、受限制的极大似然等进行统计估计或者检验。在参数情形下，还可以通过引入 BIC 等准则进行参数选择。若我们放宽对分布的假设，我们依然可以运用广义估计方程方法进行统计估计与推断。但是值得注意的是，在上述模型中我们都是只对均值部分进行了建模，而往往这些统计方法的运用都要求对协方差矩阵进行指定或者估计，例如 (1.1) 就给出了一个协方差矩阵最直观的结构。因此，尽管有时候我们的兴趣仅仅在于响应变量均值部分的变化趋势，但如果我们想运用一些依赖于协方差的统计方法，就要同时对均值与协方差进行建模，也就是所谓的均值 -协方差联合模型。

1.2　研究背景与意义

纵向数据广泛存在于很多应用领域，正如前面所提到的，其最主要的特征之一就是每个个体在不同时间被重复观测，因此同一个个体的观测值间存在着内在相依性。为了研究响应变量均值随时间的变化趋势等，以往的文献提出了很多种回归模型 (Diggle et al., 2002)。其中广义估计方程方法 (GEE, Liang & Zeger, 1986) 应用也比较广泛，这种方法主要是应用一种所谓的“工作”相关性模型来模拟内在的协方差，从而在协方差结构得到正确指定时得到回归系数的有效估计。Wang & Carey(2003) 告诉我们如果协方差结构被错误指定，估计的

效率可能会大大降低。

正因为很多统计方法的应用依赖于对协方差结构的选取，同时对均值与协方差进行建模被普遍认为是比较有吸引力的。然而，对协方差阵建模是一件非常有挑战性的工作。一方面，正如在 (1.1) 中所示，如果要无任何结构地参数化协方差阵，我们需要引入很多参数；另一方面，协防差阵必须是正定阵。正由于这两方面的原因，如何能够有效简便地对协方差结构建模成为一个非常热门的研究课题。Prentice & Zhao (1991) 提出了在均值参数的估计方程之外再引入另一组估计方差参数的估计方程。他们利用矩方法，并且通过某些工作相关矩阵来参数化协方差阵进而保证其正定性。特别是 Pourahmadi(1999, 2000) 首次提出了对协方差矩阵的逆进行 Cholesky 分解。这种分解最有吸引力的地方在于它能够在保证正定性的同时提供一种参数化协方差阵的好方法，并且这种分解有很好的统计意义的解释，对应了时间序列分析中的自回归模型。更进一步，回归模型可以用来参数化这些分解中的元素，包括均值、自回归系数等等，从而通过引入很少的无限制参数实现了均值 -协方差联合建模。进而一些统计方法可以应用其中，近期的一些工作可以参考 Pan & Mackenzie(2003)，Ye & Pan(2006)，Pourahmadi et al.(2007) 以及 Leng & Zhang(2010) 等。

另一方面，有时候协方差本身的结构也是研究兴趣所在 (Diggle & Verbyla, 1998)。而通过上述 Cholesky 分解建模得到的却是协方差的逆矩阵的结构，并不一定能够保留协方差阵的一些原始结构特点，比如稀疏性等。从而我们想对协方差本身进行 Cholesky 分解。此时分解的元素依然有非常好的统计意义，这种分解恰好对应了滑动平均模型 (Rothman et al., 2010)。类似地，我们可以引入回归模型来对相关元素进行参数化，再通过一些统计方法，比如极大似然估计、广义估计方程等，进行参数估计。这给我们提供了一种同时对均值与协方差阵建模的新途径。直观上来看，由于这种方法是直接对协方差本身进行分解并建模，它或许更能保留协方差的一些原有结构特点。

1.3 论文结构

本文将在第二章中提出基本模型与估计方法。不同于 Pourahmadi(1999) 等文献，我们对协方差本身进行 Cholesky 分解，即若有一个对称正定协方差阵

Σ，则其可以分解为 $\Sigma = LDL'$，其中 L 是一个对角元为 1 的下三角阵，而 D 是一个对角阵。我们可以发现，这种分解恰好对应了时间序列分析中另一个基本模型——滑动平均模型。在本文中，我们假设要处理的是非平衡的纵向数据，也就是每个个体的观测次数与时间都不完全一致。在第二章中，我们就利用这种新的 Cholesky 分解通过引入回归模型对协方差结构进行参数化，同时也对均值部分参数化。然后我们运用广义估计方程方法提出了估计方程，并且利用拟 Fisher 算法给出了一个比较快速有效的迭代算法来得到参数的估计。

在第三章中我们将以定理的形式说明我们在第二章中提出的估计具有相合性与渐近正态性。这些结果需要一些在实际中比较容易满足的正则条件，由这些正则条件我们就可以证明我们所列出的定理。此外基于定理的结果，我们进一步给出了一种“三明治”结构矩阵，用来估计参数估计值的协方差阵。

我们在第四章中对一组真实数据进行了分析，并且设计了三组数值模拟试验。通过真实数据分析，我们可以初步比较本文提出的方法与对应的 Ye & Pan (2006) 的方法在参数估计与数据拟合等方面的不同。第一组数值模拟研究是为了验证第三章中的理论结果，第二组模拟是为了探究本文提出的方法对于数据总体分布假设的稳健性，而最后一组模拟则可以比较在自回归与滑动平均两种数据结构下本文的方法与之前的方法的效果。

在最后一章，我们将对本文的结果进行总结，并且展望未来的工作。在附录里，我们将给出第三章中定理的证明，并且指出 Ye & Pan (2006) 理论部分中的一处错误。

原书空白页，
不缺内容

第二章　模型与估计方法

2.1　数据模型

假设我们共有 m 个个体，且 $y_i = (y_{i1}, \cdots, y_{im_i})'$ 是第 i 个个体 $(i = 1, \cdots, m)$ 在时间集 $t_i = (t_{i1}, \cdots, t_{im_i})'$ 上的 m_i 次重复测量值，从而总共有 $n = \sum_{i=1}^{m} m_i$ 个观测值。更一般的情况下，t_{ij} 并不一定是时间，也有可能是任何依赖于时间的非参数化的协变量。通过让 m_i 与 t_{ij} 都与个体 i 相关，我们的方法可以处理任意不规律观测时间和高度不平衡的数据集。我们假定 $E(y_{ij}|x_{ij}, t_{ij}) = \mu_{ij}$ 以及 $V(y_i|x_i, t_i) = \Sigma_i$，其中 x_{ij} 是一个 p 维协变量，而 $x_i = (x'_{i1}, x'_{i2}, ..., x'_{im_i})'$。

为了参数化 Σ_i，Pourahmadi (1999) 首次提出了将其分解为 $T_i \Sigma_i T_i' = D_i$，这种分解对应了自回归模型 $y_{ij} - \mu_{ij} = \sum_{k=1}^{j-1} \phi_{ijk}(y_{ik} - \mu_{ik}) + e_{ij}$。$T_i$ 是一个下三角阵，对角元为 1，次对角元 T_{ijk} 依次是上述自回归系数 ϕ_{ijk} 的相反数，而 D_i 是对角阵，其对角元依次是 $\sigma_{ij}^2 = Var(e_{ij})$。基于这种分解，Ye & Pan (2006) 提出了利用广义估计方程的方法对均值 -协方差进行建模。

通过令 $L_i = T_i^{-1}$，我们可以得到新的分解 $\Sigma_i = L_i D_i L_i'$。L_i 也是一个对角元为 1 的下三角阵，其非对角元 l_{ijk} 恰好是下述滑动平均模型中的滑动平均系数，

$$y_{ij} - \mu_{ij} = \sum_{k=1}^{j-1} l_{ijk} \epsilon_{ik} + \epsilon_{ij}, \tag{2.1}$$

其中 $\epsilon_{i1} = y_{i1} - \mu_{i1}$，且对 $\epsilon_i = (\epsilon_{i1}, \cdots, \epsilon_{im_i})'$ 有 $E\epsilon_i = 0$，$Var(\epsilon_i) = D_i$。注意到其中参数 l_{ijk} 与 $\log(\sigma_{ij}^2)$ 皆是非限制的。对于规则的平衡数据，Rothman et al. (2010) 基于这种分解提出了利用对分解的因子矩阵进行 “Banding” 来估计稀疏矩阵，并且指出该方法并不是完全有效的。他们并没有考虑构建回归模型。

鉴于我们的分解与 Pourahmadi (1999) 的分解最主要的区别在于是用到了 T_i 还是其逆 L_i，我们可以通过常见的协方差阵来进行简单对比。如果 Σ 是一个 $p \times p$ 阶对称阵 $\sigma^2\{(1-\rho)I + \rho J\}$，其中 J 是元素皆为 1 的矩阵，I 是单位

阵。那么根据 Pourahmadi (1999) 的分解方式有 $\phi_{jk} = \rho\{1 + (j-2)\rho\}^{-1}$，而我们的分解有 $l_{jk} = \rho\{1 + (k-1)\rho\}^{-1}$。再例如 $\Sigma = \sigma^2(\rho^{|i-j|})_{i,j=1}^{p}$ 是 $AR(1)$ 序列的协方差阵，则 Pourahmadi (1999) 的分解给出 $\phi_{j,j-1} = \rho$，$\phi_{jk} = 0(k < j-1)$，而我们的给出 $l_{jk} = \rho^{|j-k|}$。

为了参数化均值 -协方差结构且能够进一步减少参数维数，受 Pourahmadi (1999, 2000) 以及 Pan & MacKenzie (2003) 的启发，我们也引入如下线性回归模型，

$$g(\mu_{ij}) = x_{ij}'\beta, \qquad l_{ijk} = w_{ijk}'\gamma, \qquad log(\sigma_{ij}^2) = z_{ij}'\lambda, \tag{2.2}$$

在这里，μ_{ij}，l_{ijk} 与 $log(\sigma_{ij}^2)$ 分别被称为均值，滑动平均系数与误差项方差的对数。其中 $g(\cdot)$ 是单调可微的，被称为联系方程，x_{ij}，w_{ijk} 与 z_{ij} 分别是 $p \times 1$，$q \times 1$ 与 $d \times 1$ 维的协变量，β，γ 和 λ 是对应的回归系数。在这里我们假设 x_{ij} 与 z_{ij} 一般是各自观测时间或时间相关变量的多项式向量，而 w_{ijk} 是时间间隔 $t_{ij} - t_{ik}$ 的多项式向量。这样一来，观测时间的不同反映在了 w_{ijk} 上，而 w_{ijk} 正是滑动平均模型 (2.1) 中衡量 y_{ij} 和 y_{ik} 关系的滑动平均系数。引入参数 γ 使得度量很多个不同的滑动平均系数变得更加简便。在以后的行文中，我们将称我们引入的这种回归模型为滑动平均模型，而 Ye & Pan (2006) 中的模型将被称为自回归模型。

2.2　估计方程

首先注明一点，在本文中一个方程作用在一个向量上等于作用在这个向量的每个元素上，例如 $g(\mu_i) = (g(\mu_{i1}), \cdots, g(\mu_{im_i}))'$。利用 Liang & Zeger (1986) 提出的广义估计方程 (GEE) 方法，我们可以对三组参数 β，γ 与 λ 构造如下估计方程：

$$\begin{aligned} S_1(\beta) &= \sum_{i=1}^{m} X_i'\Delta_i\Sigma_i^{-1}(y_i - \mu_i(X_i\beta)) = 0, \\ S_2(\gamma) &= \sum_{i=1}^{m} \left(\frac{\partial \epsilon_i'}{\partial \gamma}\right) D_i^{-1}\epsilon_i = 0, \\ S_3(\lambda) &= \sum_{i=1}^{m} Z_i'D_i(Z_i\lambda)W_i^{-1}(\epsilon_i^2 - \sigma_i^2(Z_i\lambda)) = 0, \end{aligned} \tag{2.3}$$

其中 $\Delta_i = \Delta_i(X_i\beta) = diag\{\dot{g}^{-1}(x'_{ij}\beta), \cdots, \dot{g}^{-1}(x'_{im_i}\beta)\}$ 以及 $\dot{g}^{-1}(\cdot)$ 是反函数 $g^{-1}(\cdot)$ 的一次微分；注意到当 $j = 1$ 时，记求和式 $\sum_{k=1}^{0} \cdot$ 为零。$\epsilon_i = (\epsilon_{i1}, \cdots, \epsilon_{im_i})'$，其元素为 $\epsilon_{ij} = r_{ij} - \sum_{k=1}^{j-1} l_{ijk}\epsilon_{ik}$，其中 $r_{ij} = y_{ij} - \mu_{ij}$；$\partial\epsilon'_i/\partial\gamma$ 是一个 $q \times m_i$ 矩阵，它的第一列为零，第 $j(j > 2)$ 列为 $\partial\epsilon_{ij}/\partial\gamma = -\sum_{k=1}^{j-1}[\epsilon_{ik}w_{ijk} + l_{ijk}\partial\epsilon_{ik}/\partial\gamma]$；$Z_i = (z'_{i1}, ..., z'_{im_i})'$，$D_i = diag\{\sigma^2_{i1}, \cdots, \sigma^2_{im_i}\}$，$\sigma_i^2 = (\sigma^2_{i1}, \cdots, \sigma^2_{im_i})'$。不难看出 ϵ_{ij} 和 $\partial\epsilon_{ij}/\partial\gamma$ 都可以通过在滑动平均模型下递归计算得到。此外，W_i 是 ϵ_i^2 的协方差矩阵，也就是说 $W_i = Var(\epsilon_i^2)$。广义估计方程 (2.3) 的解 $\hat{\beta}$，$\hat{\gamma}$ 和 $\hat{\lambda}$ 就被称为待估参数 β，γ 与 λ 的广义估计方程估计。类似于 Ye & Pan (2006)，我们引入一种三明治“工作”协方差结构 $\hat{W}_i = A_i^{1/2}R_i(\rho)A_i^{1/2}$ 来近似真实的 W_i，其中 $A_i = 2diag\{\sigma^4_{i1}, \cdots, \sigma^4_{im_i}\}$，$R_i(\rho)$ 通过引入一个新参数 ρ 来表示 ϵ^2_{ij} 与 $\epsilon^2_{ik}(i \neq k)$ 之间的相关性。一般来说 $R_i(\rho)$ 最为典型的结构包括复合对称结构与 $AR(1)$ 结构。就好像在一般的均值广义估计方程中一样，参数 ρ 对 γ 和 λ 的估计也许几乎没有影响。在后文中我们的真实数据分析和模拟数值研究也都很好地证实了这一点。这就说明了即使我们错误地指定了 $R_i(\rho)$，我们提出的方法所得到的对均值，滑动平均系数以及误差项方差的估计都是比较有效且稳健的。

记 $\zeta = (\beta', \gamma', \lambda')'$，由估计方程 (2.3)，我们就可以计算拟 Fisher 信息矩阵 $\mathcal{I}_\zeta$：

$$\mathcal{I}_\zeta = -E\left\{\frac{\partial S'(\zeta)}{\partial\zeta}\right\} = -E\left\{\frac{\partial S'(S'_1(\beta), S'_2(\gamma), S'_3(\lambda))}{\partial(\beta', \gamma', \lambda')'}\right\}.$$

通过一些计算，我们看出拟 Fisher 信息矩阵 $\mathcal{I}_\zeta$ 是一个分块对角阵，也就是说 $\mathcal{I}_\zeta = diag(\mathcal{I}_\beta, \mathcal{I}_\gamma, \mathcal{I}_\lambda)$，其中

$$\begin{aligned}
\mathcal{I}_\beta &= \sum_{i=1}^{m} X_i'\Delta_i\Sigma_i^{-1}\Delta_i X_i, \\
\mathcal{I}_\gamma &= \sum_{i=1}^{m}\sum_{j=2}^{m_i} \frac{1}{\sigma^2_{ij}} \left(\sum_{k=1}^{j} a_{ijk}W_{ik}\right) D_i \left(\sum_{k=1}^{j} a_{ijk}W_{ik}\right)', \\
\mathcal{I}_\lambda &= \sum_{i=1}^{m} Z_i'D_i(Z_i\lambda)W_i^{-1}D_i'(Z_i\lambda)Z_i,
\end{aligned} \tag{2.4}$$

在这里，$W_{ij} = (w_{ij1}, \ldots, w_{ij(j-1)}, 0, \ldots, 0)$ 是一个 $q \times m_i$ 阶矩阵，a_{ijk} 是下三角阵 L_i^{-1} 的 (j,k) 元素，并且注意到 $d \times m_i$ 阶矩阵 $Z_{i1} = 0$。这样一来，我们就

可以独立地估计各个参数 β，γ 与 λ。

2.3 主要算法

我们想要求解广义估计方程，即满足广义估计方程 (2.3) 的解 β，γ 与 λ。我们可以通过固定其他两个，利用拟 Fisher 得分迭代算法迭代另一个直接得到广义估计方程 (2.3) 的数值解，即三组参数的估计。具体的迭代步骤如下：

1. 给三组参数赋予初始值，$\beta^{(0)}$，$\gamma^{(0)}$ 与 $\lambda^{(0)}$，并且令 k=0。

2. 利用 $\gamma^{(k)}$ 与 $\lambda^{(k)}$ 计算 Σ_i。更新 β：

$$\beta^{(k+1)} = \beta^{(k)} + \left\{ \mathcal{I}_\beta^{-1} \sum_{i=1}^{m} X_i' \Delta_i \Sigma_i^{-1} (y_i - \mu_i(X_i\beta)) \right\} \Big|_{\beta=\beta^{(k)}} \tag{2.5}$$

3. 固定 $\beta = \beta^{(k+1)}$ 以及 $\lambda = \lambda^{(k)}$，更新 γ：

$$\gamma^{(k+1)} = \gamma^{(k)} - \left\{ \mathcal{I}_\gamma^{-1} \sum_{i=1}^{m} \left(\frac{\partial \epsilon_i'}{\partial \gamma} \right) D_i^{-1} \epsilon_i \right\} \Big|_{\gamma=\gamma^{(k)}} \tag{2.6}$$

4. 固定 $\beta = \beta^{(k+1)}$ 以及 $\gamma = \gamma^{(k+1)}$，更新 λ：

$$\lambda^{(k+1)} = \lambda^{(k)} + \left\{ \mathcal{I}_\lambda^{-1} \sum_{i=1}^{m} Z_i' D_i(Z_i\lambda) W_i^{-1} (\epsilon_i^2 - \sigma_i^2(Z_i\lambda)) \right\} \Big|_{\lambda=\lambda^{(k)}} \tag{2.7}$$

5. 令 $k \leftarrow k+1$，重复上述步骤 2-5 直到达到我们预设的一个收敛准则。

很自然地，我们可以在 (2.5) 中令协方差阵 Σ_i 为单位阵求得一个 β 作为初始值。然后在 (2.6) 中令 $D_i = I_i$ 求解作为 γ 的初始值。并不难看出，这些初始值都是 $\sqrt{m}$ 相合的。从而这样的 Σ_i 的初始值保证了在均值部分的相合性，进而保证了在滑动平均系数以及误差项方差部分的参数估计的相合性。在下文的一些数值分析中，我们设定相继的估计间差的欧氏模小于 10^{-6} 为收敛准则，即直到某个 k 使得 $\|\zeta^{(k+1)} - \zeta^{(k)}\| < 10^{-6}$ 才停止迭代。我们经过数值分析发现，这个算法的收敛速度还是比较快的，通常在 10 步迭代以内就可以收敛。

第三章　渐近性质

3.1　正则条件

在本文中，对一个向量而言，$\|\cdot\|$ 表示它的欧氏模，而对任意的方阵 A 来说，$\|A\|$ 表示它最大奇异值的模。我们将 $S(\zeta)/\sqrt{m}=(S_1'(\beta),S_2'(\gamma),S_3'(\lambda))'/\sqrt{m}$ 的协方差阵记作 $V_m=(v_m^{kl})_{k,l=1,2,3}$，其中对任意 $k\neq l$ 有 $v_m^{kl}=m^{-1}Cov(S_k,S_l)$，而 $v_m^{kk}=m^{-1}Var(S_k)(k,l=1,2,3)$。我们假定在真实参数 ζ_0 处，协方差阵 V_m 是正定的，且有当 $m\to\infty$ 时，

$$V_m=\begin{pmatrix} v_m^{11} & v_m^{12} & v_m^{13}\\ v_m^{21} & v_m^{22} & v_m^{23}\\ v_m^{31} & v_m^{32} & v_m^{33}\end{pmatrix}\to V=\begin{pmatrix} v^{11} & v^{12} & v^{13}\\ v^{21} & v^{22} & v^{23}\\ v^{31} & v^{32} & v^{33}\end{pmatrix}, \tag{3.1}$$

其中常值矩阵 V 也假定为正定阵。

为了研究估计的渐近性质，我们假设有以下三个正则条件：

C1. 我们假设 (2.2) 中协变量 x_{ij}，w_{ijk} 与 z_{ij} 的维数 p，q 与 d，以及每个个体重复观测的次数 m_i 都是固定的。我们还假设响应变量的前四阶矩皆存在。

C2. 参数空间 Θ 是 R^{p+q+d} 上的紧集，且真实参数在参数空间 Θ 的内部。

C3. 各协变量 x_{ij}，w_{ijk} 和 z_{ij}，向量

$$\Delta_i=\Delta_i(X_i\beta)=diag\{\dot{g}^{-1}(x_{ij}'\beta),\cdots,\dot{g}^{-1}(x_{im_i}'\beta)\}$$

以及矩阵 W_i^{-1} 都是有界的，也就是说它们所有的元素都是有界的。

不难看出，在实际中这些正则条件都并不难满足的，也比较自然。由这些正则条件还可以推出以下一些必要条件：

N1. (2.3) 中的方程以及它们对参数 β，γ 与 λ 的一次，二次微分皆存在。

N2. (2.3) 中方程满足

$$ES_1(\beta)=0,$$
$$ES_2(\gamma)=0,$$
$$ES_3(\lambda)=0.$$

N3. 信息矩阵满足:

$$E\left[\left\{X_i'\Delta_i\Sigma_i^{-1}(y_i-\mu_i)\right\}\left\{X_i'\Delta_i\Sigma_i^{-1}(y_i-\mu_i)\right\}'\right]=-E\left[\frac{\partial}{\partial\beta}\left\{X_i'\Delta_i\Sigma_i^{-1}(y_i-\mu_i)\right\}'\right],$$

$$E\left[\left\{Z_i'D_iW_i^{-1}(\epsilon_i^2-\sigma_i^2)\right\}\left\{Z_i'D_iW_i^{-1}(\epsilon_i^2-\sigma_i^2)\right\}'\right]=-E\left[\frac{\partial}{\partial\lambda}\left\{Z_i'D_iW_i^{-1}(\epsilon_i^2-\sigma_i^2)\right\}'\right],$$

以及

$$-E\left[\frac{\partial}{\partial\gamma}\left\{\left(\frac{\partial\epsilon_i'}{\partial\gamma}\right)D_i^{-1}\epsilon_i\right\}'\right]$$

是有界的. 因此我们可以假定

$$\frac{1}{m}\mathcal{I}_\gamma\to w^{22},\ m\to\infty.$$

N4. 注意到 $\zeta=(\beta',\gamma',\lambda',)'$, $S(\zeta)=(S_1'(\beta),S_2'(\gamma),S_3'(\lambda))'$, 当 $m\to\infty$ 时, 我们有

$$\frac{1}{m}\frac{\partial S'(\zeta)}{\partial\zeta}-\frac{1}{m}E\left\{\frac{\partial S'(\zeta)}{\partial\zeta}\right\}_{\zeta=\zeta_0}\to 0,a.s.$$

N5. 当 $\zeta=\zeta_0$ 时, 有如下渐近结果: 当 $m\to\infty$ 时,

$$\frac{1}{\sqrt{m}}\begin{pmatrix}S_1(\beta)\\S_2(\gamma)\\S_3(\lambda)\end{pmatrix}\to N\left\{0,\begin{pmatrix}v^{11}&v^{12}&v^{13}\\v^{21}&v^{22}&v^{23}\\v^{31}&v^{32}&v^{33}\end{pmatrix}\right\}.$$

上述收敛是依分布收敛, 且渐近协方差阵是一个正定矩阵。

不难看出, $N1-N3$ 是正则条件 $C1-C3$ 的直接推论, $N4$ 与 $N5$ 可能并不是那么显然, 我们将在附录里给出证明。

3.2 主要结果

有了上述假设与条件后, 我们可以证明由上一章算法得到的估计是相合的且具有渐近正态性, 也就是下面的定理。受篇幅局限, 定理的证明被放在了附录里。

定理 3.2.1. 在一些正则条件下, 广义估计方程估计 $(\hat\beta_m',\hat\gamma_m',\hat\lambda_m')'$ 是 $\sqrt{m}$ 相

合的，且满足渐近正态性。即当 $m \to \infty$ 时，

$$\sqrt{m}\begin{pmatrix}\hat{\beta}_m-\beta_0\\ \hat{\gamma}_m-\gamma_0\\ \hat{\lambda}_m-\lambda_0\end{pmatrix}$$
$$\to N\left\{0,\begin{pmatrix}v^{11}&0&0\\0&w^{22}&0\\0&0&v^{33}\end{pmatrix}^{-1}\begin{pmatrix}v^{11}&v^{12}&v^{13}\\v^{21}&v^{22}&v^{23}\\v^{31}&v^{32}&v^{33}\end{pmatrix}\begin{pmatrix}v^{11}&0&0\\0&w^{22}&0\\0&0&v^{33}\end{pmatrix}^{-1}\right\},$$

该收敛是依分布收敛，其中矩阵元素 v^{kl} 是当 $\zeta=\zeta_0(k,l=1,2,3)$ 时的协方差。

通过定理可以看出正态分布响应变量的渐近方差是一个对角阵。在以后的推断中，我们将引用下面一种结构的估计作为估计值 $\hat{\beta}_m$ 的协方差，

$$V(\hat{\beta}_m)=M_0^{-1}M_1M_0^{-1}, \tag{3.2}$$

其中 $M_0=\sum_{i=1}^{m}X_i'\hat{\Delta}_i\hat{\Sigma}_i^{-1}\hat{\Delta}_iX_i$，$M_1=\sum_{i=1}^{m}X_i'\hat{\Delta}_i\hat{\Sigma}_i^{-1}(y_i-\hat{\mu}_i)(y_i-\hat{\mu}_i)'\hat{\Sigma}_i^{-1}\hat{\Delta}_iX_i$。对于另外两个参数的估计值 $\hat{\gamma}_m$ 和 $\hat{\lambda}_m$，它们的协方差也用类似的“三明治”结构去估计。至于这种估计的优良性如何，我们将在数值模拟研究中，将其与样本方差进行比较。

原书空白页，
不缺内容

第四章　数值研究

4.1　CD4+ 细胞数据的分析

在这一节，我们将应用本文提出的方法重新分析一下在很多纵向数据文献中都会用到的一组真实数据——CD4+ 细胞数据。CD4+ 细胞数据 (Diggle et al.，2002) 来自于 369 位 HIV 病毒感染者，在大约 8 年半的时间里，试验组织者在不同的时间测量每个感染者体内 CD4+ 细胞数目，共计 2376 个数据。每个个体的重复观测次数并不一样，最少的只有 1 次，最多的达到 12 次，并且测量的时间也不是等间距的。所以 CD4+ 细胞数据是一组高度不平衡的纵向数据实例，这也是为什么很多文献都会用这组数据来检验他们提出的方法或观点。如果需要关于这次医学实验的设计或者意义的详细内容，可以查阅文献 Diggle et al.(2002)。

类似于 Ye & Pan (2006)，我们也采用同样的多项式模型来对均值与协方差结构进行建模：

$$
\begin{aligned}
&y_{ij} = \beta_0 + \beta_1 t_{ij} + \cdots + \beta_p t_{ij}^{p-1},\\
&l_{ijk} = \gamma_0 + \gamma_1 (t_{ij} - t_{ik}) + \cdots + \gamma_q (t_{ij} - t_{ik})^{q-1},\\
&log(\sigma_{ij}^2) = \lambda_0 + \lambda_1 t_{ij} + \cdots + \lambda_d t_{ij}^{d-1}.
\end{aligned}
$$

基于 Ye & Pan (2006) 中提出的模型选择结果也为了更直观地进行比较，我们也令 $p = 7$，$q = d = 4$。注意到在数据集中有大概 54 个个体的重复观测次数要小于 q 与 d，也就是说共有大概 15% 的协变量 w_{ijk} 或者 z_{ij} 的维数要大于它们相应的重复观测次数。

对于第三个广义估计方程中 ϵ_i^2 的“工作”协方差阵 $W_i = A_i^{\frac{1}{2}} R_i(\rho) A_i^{\frac{1}{2}}$ 中的 $R_i(\rho)$，我们考虑两种相关性结构：复合对称结构与 $AR(1)$ 结构。对于每种情况，我们让参数 ρ 取 4 个不同的值，即 $\rho = 0$，0.2，0.5 以及 0.8，这样我们可以观察错误指定 $R_i(\rho)$ 对广义估计方程估计 $\hat{\beta}$，$\hat{\gamma}$ 与 $\hat{\lambda}$ 的影响。表 4.1 记录了

$R_i(\rho)$ 为 $AR(1)$ 结构时参数的估计以及相应的估计标准差，在复合对称结构下的结果也非常类似。在表 4.1 中，我们可以看到不同的参数 ρ 对于参数 β，γ 与 λ 的估计影响很小，这就意味着我们的估计相对于 $R_i(\rho)$ 的选取呈现了比较好的稳健性。并且我们发现 β 的估计与 Ye & Pan (2006) 中相应参数的估计基本吻合。这是由于我们两篇文章的方法的不同之处在于如何对协方差建模，在均值部分建模一致，从而参数估计也一致。此外，我们的估计值 $\hat{\gamma}_3$ 与 $\hat{\gamma}_4$ 并不是非常显著，意味着滑动平均系数与时间间隔呈现一种近似于线性的关系，比 Ye & Pan (2006) 中的模型更加简单一些。

表 4.1 CD4+ 细胞数据。$R_i(\rho)$ 为 $AR(1)$ 结构时各参数的估计，括号里是参数估计相应的估计标准差

	$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
β_1	867.01(16.87)	867.15(16.90)	867.64(16.89)	868.76(16.74)
β_2	-203.32(12.94)	-203.73(12.95)	-204.73(12.97)	-206.68(13.00)
β_3	-17.98(8.28)	-18.05(8.37)	-18.41(8.53)	-19.42(8.74)
β_4	28.72(4.64)	28.89(4.66)	29.35(4.71)	30.35(4.80)
β_5	-1.42(0.88)	-1.43(0.90)	-1.43(0.92)	-1.39(0.96)
β_6	-1.56(0.47)	-1.57(0.47)	-1.62(0.48)	-1.72(0.49)
β_7	0.21(0.06)	0.21(0.06)	0.21(0.06)	0.23(0.06)
γ_1	0.55(0.04)	0.55(0.04)	0.55(0.04)	0.55(0.04)
γ_2	-0.19(0.06)	-0.19(0.06)	-0.19(0.06)	-0.19(0.06)
γ_3	4.9×10^{-2}	4.9×10^{-2}	4.9×10^{-2}	5.0×10^{-2}
	(2.7×10^{-2})	(2.7×10^{-2})	(2.7×10^{-2})	(2.7×10^{-2})
γ_4	-4.8×10^{-3}	-4.8×10^{-3}	-4.8×10^{-3}	-4.9×10^{-3}
	(3.3×10^{-3})	(3.3×10^{-3})	(3.3×10^{-3})	(3.3×10^{-3})
λ_1	11.54(0.04)	11.54(0.05)	11.52(0.06)	11.49(0.07)
λ_2	-0.35(0.03)	-0.35(0.03)	-0.36(0.03)	-0.37(0.03)
λ_3	-0.05(0.01)	-0.05(0.01)	-0.03(0.01)	-0.01(0.01)
λ_4	1.7×10^{-2}	1.6×10^{-2}	1.5×10^{-2}	1.2×10^{-2}
	(3.2×10^{-3})	(3.5×10^{-3})	(3.4×10^{-3})	(2.6×10^{-3})

图 4.1 展示了当 $\rho = 0.2$ 且 $R_i(\rho)$ 是 $AR(1)$ 结构时均值，滑动平均系数，误差项方差的对数三组变量的拟合曲线。图中虚线表示的是拟合值的 95% 渐近置信区间。通过对比可以发现，均值与误差项方差的对数两者的拟合曲线与 Ye &

Pan (2006) 相吻合。不同的是，拟合的滑动平均系数曲线更接近于线性，而在使用自回归分解时，自回归系数拟合曲线更接近于一条三次函数曲线。

为了进一步比较我们的方法与 Ye & Pan (2006) 的不同，我们分别采用两种方法进行建模，然后采用逐一剔除的交叉验证方法比较两者对于数据拟合以及协方差估计的效果。也就是说每次建模，我们剔除一个个体的观测数据，利用其余 368 组数据分别用两种方法建模，得到参数的估计，然后预测之前被剔除的那组数据并计算预测误差。经过计算，我们发现两种方法对于数据本身的预测误差 $\sum_{i=1}^{m}\|y_i-\hat{y}_{(i)}\|^2/m$ 基本相同（自回归模型 768.72，滑动平均模型 768.04），而对于样本方差与预测方差之间的差别 $\sum_{i=1}^{m}\|(y_i-\hat{y}_i)(y_i-\hat{y}_i)'-\hat{\Sigma}_{(i)}\|^2/m$ 而言，滑动平均模型方法是 8.48×10^5，比自回归模型方法（8.59×10^5）要稍微小一些。我们注意到两种方法的预测误差都比较大，这也许是因为在建模的时候我们只选取了时间和时间间隔的多项式作为协变量，如果能考虑引入试验数据中记录的其他更加相关的协变量的话，效果或许会有提高。但是在这里，我们只需要比较一下两种方法的区别，对如何更好地针对这组数据选择合适的协变量并不是本文最关心的问题。

4.2 数值模拟研究

我们设计了三组模拟研究，来检验文中提出的估计方法以及定理 3.2.1 后提出的方差的估计表达式 (3.2) 的效果，估计方法相对于总体分布假设的稳健性，以及比较滑动平均模型与自回归模型。在每个模拟研究中，我们基本都是运用 Monte Carlo 方法，每种情况下的模拟都进行 1000 次重复抽样。

4.2.1 模拟一

该模拟研究主要是为了观察在实际中我们的估计方法是否真的满足定理 3.2.1 中的相合性与渐近正态性以及 (3.2) 是否合适。我们按照如下模型产生数据集：

$$y_{ij}=\beta_0+x_{ij1}\beta_1+x_{ij2}\beta_2+x_{ij3}\beta_3+e_{ij},\ (i=1,\ldots,m;\quad j=1,\ldots,m_i),$$

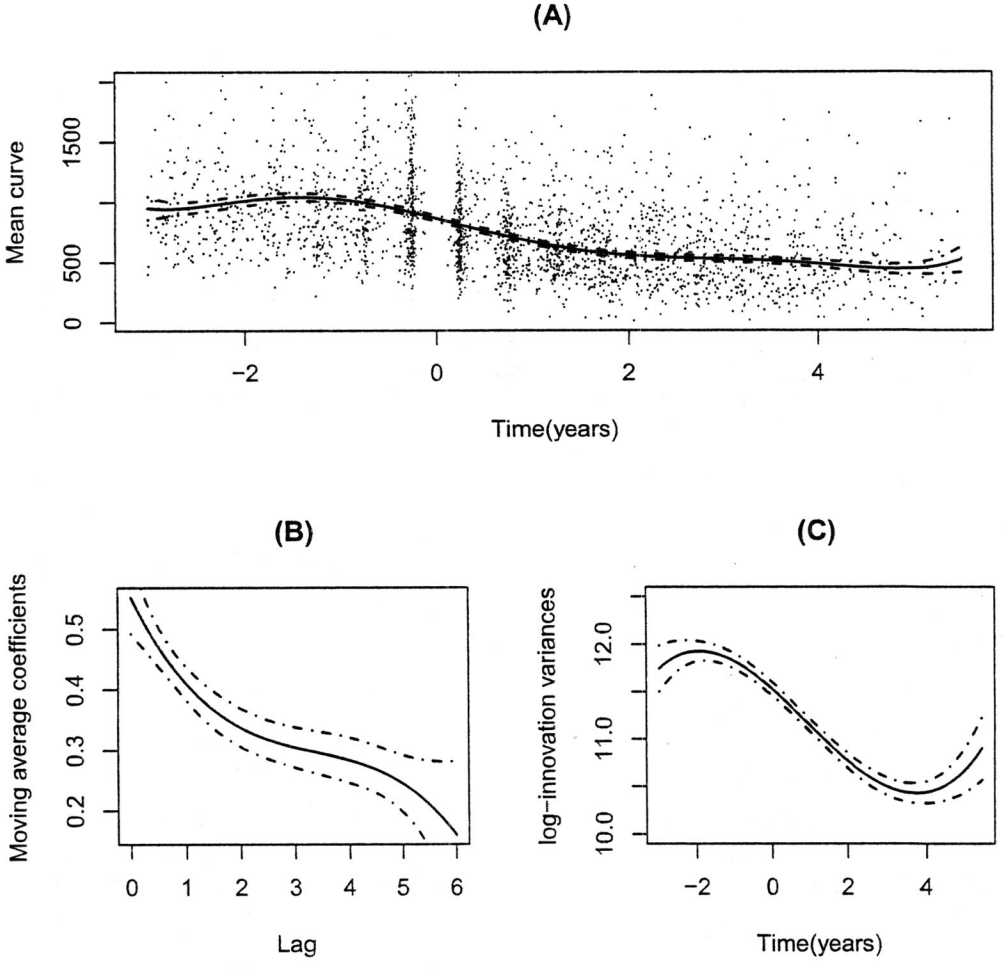


图 4.1 CD4+ 细胞数据。当 $\rho = 0.2$ 且 $R_i(\rho)$ 采用 $AR(1)$ 结构时，(A) 均值 -时间拟合曲线，(B) 滑动平均系数 -时间间隔拟合曲线，(C) 误差项方差对数 -时间拟合曲线。虚线表示 95% 渐近置信区间。

而滑动平均系数与误差项方差对数由下面的模型产生：

$$l_{ijk} = \gamma_0 + z_{ijk1}\gamma_1 + z_{ijk2}\gamma_2 + z_{ijk3}\gamma_3,\ \log(\sigma_{ij}^2) = \lambda_0 + h_{ij1}\lambda_1 + h_{ij2}\lambda_2 + h_{ij3}\lambda_3,$$

由此，当误差 e_{ij} 服从正态分布时，样本 y_i 即是从多元正态分布 $N(\mu_i, \Sigma_i)$ 中抽取的随机样本，其中 μ_i 与 Σ_i 分别由上述线性模型得到。我们分别研究当样本量 $m = 100$ 以及 $m = 200$ 时的结果。每个个体假设被测量了 m_i 次，其中，$m_i - 1 \sim Binomial(11, 0.8)$，而观测时间 t_{ij} 由标准均匀分布产生。这样就产生了每个个体观测次数不同，观测时间也不尽相同的非平衡数据。协变量 $x_{ij} = (x_{ij1}, x_{ij2}, x_{ij3})'$ 由多元正态分布随机产生，其均值为零，协方差对角元为 1，非对角元为 0.5。与 CD4+ 细胞数据中的假定类似，我们取 $h_{ij} = x_{ij}$，$z_{ijk} = (1, t_{ij} - t_{ik}, (t_{ij} - t_{ik})^2, (t_{ij} - t_{ik})^3)'$。对每组模拟数据，我们在 $W_i = A_i^{1/2} R_i(\rho) A_i^{1/2}$ 中令 $R_i(\rho)$ 为 $AR(1)$ 结构，且取参数 $\rho = 0.2$，0.5，0.8，以此来检验估计对于参数 ρ 的稳健性。

表 4.2 模拟一（样本量 $m = 100$）。表中 SD、SE 与 Std 均为乘以 10^2 后的数值.

		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$	
	True	$\hat{\zeta}_{SD}$	SE_{Std}	$\hat{\zeta}_{SD}$	SE_{Std}	$\hat{\zeta}_{SD}$	SE_{Std}
β_0	1.00	$1.00_{0.75}$	$0.69_{0.058}$	$1.00_{0.81}$	$0.72_{0.067}$	$1.00_{0.85}$	$0.79_{0.071}$
β_1	0.50	$0.50_{2.38}$	$2.14_{0.28}$	$0.50_{2.43}$	$2.21_{0.32}$	$0.50_{2.65}$	$2.38_{0.33}$
β_2	0.10	$0.10_{2.32}$	$2.14_{0.29}$	$0.10_{2.49}$	$2.21_{0.28}$	$0.10_{2.59}$	$2.38_{0.32}$
β_3	−0.20	$-0.20_{2.30}$	$2.12_{0.28}$	$-0.20_{2.43}$	$2.22_{0.31}$	$-0.20_{2.75}$	$2.39_{0.33}$
γ_0	−0.30	$-0.30_{0.95}$	$0.85_{0.13}$	$-0.30_{0.97}$	$0.88_{0.13}$	$-0.30_{1.03}$	$0.96_{0.15}$
γ_1	0.10	$0.10_{4.45}$	$4.21_{0.63}$	$0.10_{4.87}$	$4.38_{0.61}$	$0.10_{5.23}$	$4.73_{0.68}$
γ_2	0.20	$0.20_{4.61}$	$4.25_{0.70}$	$0.20_{4.75}$	$4.41_{0.70}$	$0.20_{5.23}$	$4.82_{0.74}$
γ_3	−0.15	$-0.15_{10.73}$	$10.10_{1.67}$	$-0.15_{11.77}$	$10.49_{1.63}$	$-0.16_{12.37}$	$11.33_{1.79}$
λ_0	0.20	$0.18_{4.51}$	$4.39_{0.35}$	$0.19_{5.04}$	$4.66_{0.39}$	$0.19_{6.06}$	$6.06_{0.54}$
λ_1	−0.05	$-0.05_{5.93}$	$5.51_{0.60}$	$-0.06_{6.93}$	$6.22_{0.72}$	$-0.05_{7.10}$	$6.88_{0.93}$
λ_2	0.15	$0.16_{5.67}$	$5.53_{0.59}$	$0.15_{6.64}$	$6.23_{0.73}$	$0.15_{6.98}$	$6.87_{0.94}$
λ_3	−0.10	$-0.10_{5.38}$	$5.50_{0.61}$	$-0.09_{6.43}$	$6.21_{0.79}$	$-0.10_{7.30}$	$6.81_{0.92}$

表 4.2 是样本量为 100 时的估计结果，从中我们可以看出我们提出的方法得到的参数估计是基本无偏的，不同的 ρ 对参数估计的影响微乎其微，也就是说估计对于参数 ρ 表现出了一定的稳健性。表中不但给出了参数的估计值，还

给出了估计的标准差，其中 SD 表示 1000 组参数估计值的样本标准差，可以被看作参数估计的真实标准差，而 SE 是 1000 组由公式 (3.2) 计算得到的标准差的估计的平均值，可以看做是参数估计的标准差的估计，而 Std 代表了这 1000 组标准差的估计的标准差。通过观察可以看出，对于每个参数 β，γ，λ 来说，它们的 SD 与 SE 都非常吻合，这也就是说我们提出的标准差估计公式 (3.2) 也是一个比较好的估计。

表 4.3 是样本量为 200 时的估计结果，我们可以看出该表反映的结果与上一个表基本一致。不过，相对于表 4.2，表 4.3 中参数的估计与真实参数更加吻合，只有 4 个估计稍微偏离真实参数，并且得到的估计的标准差也要更小。不同样本量 m 下的比较也印证了我们在上一章中给出的渐近结果。

表 4.3 模拟一（样本量 $m = 200$）。表中 SD、SE 与 Std 均为乘以 10^2 后的数值.

		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$	
	True	$\hat{\zeta}_{SD}$	SE_{Std}	$\hat{\zeta}_{SD}$	SE_{Std}	$\hat{\zeta}_{SD}$	SE_{Std}
β_0	1.00	$1.00_{0.61}$	$0.55_{0.034}$	$1.00_{0.61}$	$0.55_{0.035}$	$1.00_{0.58}$	$0.55_{0.034}$
β_1	0.50	$0.50_{1.74}$	$1.67_{0.16}$	$0.50_{1.74}$	$1.67_{0.16}$	$0.50_{1.74}$	$1.67_{0.16}$
β_2	0.10	$0.10_{1.86}$	$1.67_{0.16}$	$0.10_{1.86}$	$1.67_{0.16}$	$0.10_{1.75}$	$1.67_{0.16}$
β_3	−0.20	$-0.20_{1.66}$	$1.68_{0.16}$	$-0.20_{1.66}$	$1.68_{0.16}$	$-0.20_{1.72}$	$1.67_{0.17}$
γ_0	−0.30	$-0.30_{0.70}$	$0.67_{0.076}$	$-0.30_{0.70}$	$0.67_{0.076}$	$-0.30_{0.72}$	$0.67_{0.076}$
γ_1	0.10	$0.10_{3.43}$	$3.31_{0.37}$	$0.10_{3.43}$	$3.31_{0.37}$	$0.10_{3.40}$	$3.28_{0.36}$
γ_2	0.20	$0.20_{3.55}$	$3.34_{0.41}$	$0.20_{3.56}$	$3.34_{0.41}$	$0.20_{3.64}$	$3.30_{0.40}$
γ_3	−0.15	$-0.15_{8.07}$	$7.90_{0.99}$	$-0.15_{8.08}$	$7.90_{1.00}$	$-0.16_{8.07}$	$7.81_{0.97}$
λ_0	0.20	$0.19_{3.19}$	$3.21_{0.17}$	$0.19_{3.38}$	$3.37_{0.19}$	$0.19_{4.11}$	$4.26_{0.29}$
λ_1	−0.05	$-0.05_{4.16}$	$4.05_{0.33}$	$-0.05_{4.60}$	$4.50_{0.38}$	$-0.05_{4.85}$	$4.84_{0.44}$
λ_2	0.15	$0.15_{4.06}$	$4.04_{0.34}$	$0.15_{4.54}$	$4.50_{0.40}$	$0.15_{4.83}$	$4.82_{0.43}$
λ_3	−0.10	$-0.10_{4.16}$	$4.05_{0.35}$	$-0.10_{4.59}$	$4.48_{0.40}$	$-0.10_{4.84}$	$4.87_{0.45}$

4.2.2　模拟二

模拟二是为了研究我们提出的方法相对于总体分布假设的稳健性，也就是说如果总体分布偏离正态分布，我们的估计会有什么相应的变化。为了达到这

个目的，我们沿用 Ye & Pan (2006) 中提出的混合正态模型：

$$F_{m_i} = \pi N(\mu_i(1+\tau), \Sigma_i) + (1-\pi)N(\mu_i, \Sigma_i),$$

其中 μ_i 与 Σ_i 同模拟一中的假设，π 是混合权重参数，而 τ 是均值飘移参数。为了更好地观察在不同程度的混合下估计的效果，我们分别令 $\tau = 1/10$，$1/5$，$1/3$ 以及 $\pi = 0.25$，0.5，0.75，这样我们总共有九种不同的混合组合。注意到混合后的真实均值与方差分别为 $\tilde{\mu}_i = \mu_i(1+\pi\tau)$ 以及 $\tilde{\Sigma}_i = \Sigma_i + \pi(1-\pi)\tau^2\mu_i\mu_i'$。我们定义相对误差 $\mathrm{err}(\hat{\mu}_i)=\|\hat{\mu}_i - \tilde{\mu}_i\|/\|\tilde{\mu}_i\|$，$\mathrm{err}(\hat{\Sigma}_i)=\|\hat{\Sigma}_i - \tilde{\Sigma}_i\|/\|\tilde{\Sigma}_i\|$ 来衡量我们提出的方法得到的估计的误差大小。同前，我们每组依然产生 1000 组数据集。

表 4.4 模拟二。假设 $R_i(\rho)$ 基于 $AR(1)$ 结构，且 $\rho = 0.2$，不同混合下产生 1000 组抽自于混合正态分布的随机样本，得到的相对误差 $\mathrm{err}(\hat{\mu})=\sum_{i=1}^{m}\mathrm{err}(\hat{\mu}_i)/m$ 与 $\mathrm{err}(\hat{\Sigma})=\sum_{i=1}^{m}\mathrm{err}(\hat{\Sigma}_i)/m$ 的平均值

	m=100		m=200	
(π,τ)	$\mathrm{err}(\hat{\mu})$	$\mathrm{err}(\hat{\Sigma})$	$\mathrm{err}(\hat{\mu})$	$\mathrm{err}(\hat{\Sigma})$
$(0.25, 1/10)$	3.45×10^{-2}	0.11	2.34×10^{-2}	7.28×10^{-2}
$(0.25, 1/5)$	4.31×10^{-2}	0.14	3.08×10^{-2}	0.11
$(0.25, 1/3)$	5.27×10^{-2}	0.21	3.92×10^{-2}	0.20
$(0.50, 1/10)$	3.54×10^{-2}	0.11	2.32×10^{-2}	7.49×10^{-2}
$(0.50, 1/5)$	4.33×10^{-2}	0.16	3.04×10^{-2}	0.13
$(0.50, 1/3)$	5.03×10^{-2}	0.23	3.86×10^{-2}	0.24
$(0.75, 1/10)$	3.21×10^{-2}	0.11	2.29×10^{-2}	7.20×10^{-2}
$(0.75, 1/5)$	3.73×10^{-2}	0.14	2.70×10^{-2}	0.11
$(0.75, 1/3)$	4.49×10^{-2}	0.22	3.40×10^{-2}	0.21

表 4.4 给出了在不同程度的混合下相对误差的均值。可以看出，不管是在何种混合下，相对误差都比较小，也就是说我们的估计值都比较接近于真实值。具体地说，均值的相对误差随着均值漂移参数 τ 与混合权重 π 的增大而增大，而协方差阵的估计值的平均相对误差 $\mathrm{err}(\hat{\Sigma}_i)$ 随着均值漂移参数 τ 的增大而增大，但是与混合权重 π 好像关系不大，这都比较符合理论结果。我们注意到有些情况下 $\mathrm{err}(\hat{\Sigma}_i)$ 达到了 20% 左右，但是我们认为这是可以接受的，因为 $\|\hat{\Sigma}_i - \tilde{\Sigma}_i\|$ 是基于维数大至 12×12 的矩阵计算得到的。并且我们可以观察到随着样本量的增加，相对误差有明显减小的趋势。总体来说，我们提出的方法对

于混合正态数据依然表现良好，具有一定的稳健性。

4.2.3 模拟三

在这个模拟试验中，我们是为了比较两种不同的分解方式所引出的估计方法在不同的数据结构下的表现如何。我们主要比较拟合均值和协方差阵分别与真实均值和协方差阵的差距，即定义相对误差 $err(\hat{\mu}_i) = \|\hat{\mu}_i - \mu_i\|/\|\mu_i\|$ 与 $err(\hat{\Sigma}_i) = \|\hat{\Sigma}_i - \Sigma_i\|/\|\Sigma_i\|$。首先，我们按照 Ye & Pan (2006) 的分解方式产生样本量为 m（m=100 或 200）的随机数据集，也就是自回归结构数据，分别用我们提出的方法（记作 MA 方法）与 Ye & Pan (2006) 提出的方法（记作 AR 方法）进行建模与估计，并计算相对误差。如此重复 1000 次，记录均值与协方差阵相对误差的均值。同样的，再按照我们的分解方式产生滑动平均结构数据，分别用两种方法估计并记录相对误差。由此，得到了表 4.5。

表 4.5 模拟三。相对误差 $err(\hat{\mu}) = \sum_{l=1}^{m} err(\hat{\mu}_l)/m$ 与 $err(\hat{\Sigma}) = \sum_{l=1}^{m} err(\hat{\Sigma}_l)/m$ 的均值

	Fit	MA		AR	
True	m	$err(\hat{\mu})$	$err(\hat{\Sigma})$	$err(\hat{\mu})$	$err(\hat{\Sigma})$
MA	100	2.97×10^{-2}	0.10	5.90×10^{-2}	0.38
	200	1.93×10^{-2}	0.07	4.23×10^{-2}	0.38
AR	100	4.99×10^{-2}	0.24	4.79×10^{-2}	0.11
	200	3.38×10^{-2}	0.22	3.23×10^{-2}	0.08

MA: 滑动平均分解方法　AR: 自回归分解方法 (Ye & Pan, 2006)

首先，通过表 4.5 可以看出当样本量 m=200 时，相对误差的均值要明显小于 m=100 时的相对误差均值，再一次验证了估计的渐近性质。其次，当真实的数据产自于滑动平均模型时，我们的方法（MA 方法）要明显优于 AR 方法，而当数据拥有自回归结构时，AR 方法的估计效果要更好一些。另外，我们注意到对于均值来说，不管在哪种数据结构下用何种方法，相对误差都在 10^{-2} 量级，这也正好说明了我们两种方法得到的关于均值的估计都是相合的。但是对于协方差阵来说，如果用的方法不当，相对误差还是比较大的，甚至可以达到将近 40%，这也正是我们需要提出新方法的原因。并且就我们的模拟结果来说，我们所提出的方法（MA 方法）即使在错误判断数据模型情况下，相对误差也

要小于 AR 方法在 MA 数据情况下的相对误差。也就是说，MA 方法在某些情况下，可能更稳健一些。

原书空白页，
不缺内容

第五章　结论

在这篇文章中，我们对协方差矩阵本身提出了一种新的 Cholesky 分解，恰好对应于滑动平均模型。通过引入参数模型以及广义估计方程方法，我们可以对均值与协方差结构同时建模。理论结果显示，参数的估计都是相合且渐近正态的。通过真实数据分析以及数值模拟研究，我们也进一步从数值角度验证了这一很好的理论性质。通过引入广义估计方程，我们放松了数据满足正态分布的假设，并且通过在不同程度的混合正态下的模拟可以看出，我们的估计具有较高的稳健性。同时在与 Ye & Pan (2006) 提出的自回归广义估计方程模型比较的时候，我们发现了如果真实数据是滑动平均结构，用我们的方法会得到更加准确的估计，反之亦然。

因此在处理纵向数据均值 -协方差联合建模时，除了 Ye & Pan (2006) 提出的自回归模型，我们提出的基于滑动平均 Cholesky 分解的这种新的广义估计方程模型也是一个比较有效的途径。至于如何选择这两种方法，要根据数据的情况。在实际应用中，我们也许可以引入一些图形工具，如 Pourahmadi (1999) 中的适用于平衡数据集的回归图，如果样本呈现出某种明显的趋势，我们可以采取对应的因子分解。我们也可以采用数值方法来选择合适的分解与参数化方式，例如我们可以应用交叉验证的方法来比较不同模型下均值与协方差矩阵的预测偏差，进而选择预测更加准确的那个模型。

至此，滑动平均与自回归两种模型都可以用来参数化协方差矩阵，很自然地我们想到是否可以把这两种模型结合起来，也就是所谓的自回归滑动平均混合模型 (ARMA 模型)。如果能够建立混合 ARMA 模型，那么我们提出的滑动平均模型与 Pourahmadi 等人提出的自回归模型都是其特殊情况。除此以外，我们可以研究非线性情况，此时可以引入半参数的均值 -协方差模型，如 Fan et al. (2007), Fan & Wu (2008) 以及 Leng et al. (2010) 等工作。

原书空白页，
不缺内容

参考文献

[1] Diggle, P. J., Heagerty P. J., Liang, K. Y., Zeger, S. L. Analysis of Longitudinal Data. Oxford University Press, 2002.

[2] Diggle, P. J. and Verbyla, A. P. Nonparametric estimation of covariance structure in longitudinal data. Biometrics, 1998, 54:403-415.

[3] Fan, J., Huang, T., Li, R. Analysis of longitudinal data with semiparametric estimation of covariance function. J. Am. Statist. Assoc., 2007, 102:632-641.

[4] Fan, J., Wu, Y. Semiparametric estimation of covariance matrices for longitudinal data. J. Am. Statist. Assoc., 2008, 103:1520-1533.

[5] Leng, C., Zhang, W., Pan, J. Semiparametric mean-covariance regression analysis for longitudinal data. J. Am. Statist. Assoc., 2010, 105:181-193.

[6] Liang, K. Y., Zeger, S. L. Longitudinal data analysis using generalised linear models. Biometrika, 1986, 73:13-22.

[7] McGullagh, P. Quasi-likelihood functions. Ann. Statist., 1983, 11:59-67.

[8] Pan, J., Mackenzie, G. Model Selection for Joint Mean-Covariance Structures in Longitudinal Studies. Biometrika, 2003, 90:239-244.

[9] Prentice, R. L., Zhao, L. P. Estimating Equations for Parameters in Means and Covariances of Multivariate Discrete and Continuous Responses. Biometrics, 1991, 47:825-839.

[10] Pourahmadi, M. Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. Biometrika, 1999, 86:677-690.

[11] Pourahmadi, M. Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. Biometrika, 2000, 87:425-435.

[12] Rothman, A. J., Levina, E., Zhu, J. A new approach to Cholesky-based covariance regularization in high dimensions. Biometrika, 2010, 97:539–550.

[13] Wang, Y. G., Carey, V. Working correlation structure misspecification, estimation and covariate design: Implications for generalised estimating equations performance. Biometrika, 2003, 90:29-41.

[14] Ye, H., Pan, J. Modelling of covariance structures in generalized estimating equations for longitudinal data. Biometrika, 2006, 93:927-941.

附录 A

在前面受篇幅限制，我们没有给定理 3.2.1 以及 N4-N5 的证明。我们现在将给出简明的证明过程。我们首先给出相合性的证明，再给出渐近正态性的证明，最后给出 N4-N5 的证明。然后我们会计算一下第二个估计方程的 Fisher 信息阵。在附录的最后我将指出文献 Ye & Pan (2006) 中的一处错误。

A.1 相合性的证明

在这里我们只给出参数 β 的相合性证明，即 $\hat{\beta}_m \to \beta_0$，a.s.，对于另外两个参数 $\hat{\gamma}_m$ 与 $\hat{\lambda}_m$，其证明都是类似的。

根据 McCullagh (1983)，我们有

$$\hat{\beta}_m - \beta_0 = \left\{ \frac{1}{m} \sum_{i=1}^{m} X_i' \Delta_i \Sigma_i^{-1} \Delta_i X_i \right\}^{-1} \left\{ \frac{1}{m} \sum_{i=1}^{m} X_i' \Delta_i \Sigma_i^{-1} (y_i - \mu_i) \right\} \Big|_{\beta=\beta_0} + o_p(m^{-\frac{1}{2}}). \tag{A.1}$$

另一方面，$U_i \equiv X_i' \Delta_i \Sigma_i^{-1} (y_i - \mu_i(X_i\beta))$ 的期望与方差分别是 $E(U_i|_{\beta=\beta_0}) = 0$，

$$Var(U_i|_{\beta=\beta_0}) = X_i' \Delta_i \Sigma_i^{-1} \Delta_i X_i|_{\beta=\beta_0}. \tag{A.2}$$

由于 $\Sigma_i = L_i D_i L_i'$，则 $(A.2)$ 中的协方差阵可以进一步写成

$$Var(U_i|_{\beta=\beta_0}) = X_i' \Delta_i L_i'^{-1} D_i^{-1} L_i^{-1} \Delta_i X_i|_{\beta=\beta_0}.$$

正则条件 C3 表示存在一个常数 η_0 使得对任意的 i 以及 $\zeta \in \Theta$ 有 $Var(U_i|_{\beta=\beta_0}) \leqslant \eta_0 1_{p\times p}$，其中 $1_{p\times p}$ 是一个元素都是 1 的 $p \times p$ 阶方阵，也就是说 $Var(U_i|_{\beta=\beta_0})$ 中所有的元素都被 η_0 控制住。从而有 $\sum_{i=1}^{\infty} Var(U_i|_{\beta=\beta_0})/i^2 < \infty$。由 Kolmogorov 强大数律知，当 $m \to \infty$ 时有

$$\left\{ \frac{1}{m} \sum_{i=1}^{m} X_i' \Delta_i \Sigma_i^{-1} (y_i - \mu_i(X_i\beta)) \right\}_{\beta=\beta_0} \to E(U_i|_{\beta=\beta_0}) = 0, \quad a.s. \tag{A.3}$$

类似地可以知道

$$\left\{\frac{1}{m}\sum_{i=1}^{m}X_i'\Delta_i\Sigma_i^{-1}\Delta_i X_i\right\}_{\beta=\beta_0}$$

也是一个有界矩阵。进而由 $(A.3)-(A.3)$ 知当 $m\to\infty$ 时有 $\hat{\beta}_m-\beta_0\to 0, a.s.$。

A.2 渐近正态性的证明

根据正则条件 C1-C3 及其必要条件 N1-N5，我们可以证明 $(\hat{\beta}_m', \hat{\gamma}_m', \hat{\lambda}_m')'$ 是渐近正态的。

由于 $\hat{\zeta}_m=(\hat{\beta}_m', \hat{\gamma}_m', \hat{\lambda}_m')'$ 以概率 1 是方程 $S(\hat{\zeta}_m)=0$ 的根，则在 ζ 的真实值 $\zeta_0=(\beta_0', \gamma_0', \lambda_0')'$ 的某个小邻域 $J_0=\{(\beta',\gamma',\lambda')': (\|\beta-\beta_0\|^2+\|\gamma-\gamma_0\|^2+\|\lambda-\lambda_0\|^2)^{1/2}\leqslant\epsilon\}$ 内，我们有 $S_1(\hat{\beta}_m)=0$，$S_2(\hat{\gamma}_m)=0$ 以及 $S_3(\hat{\lambda}_m)=0$。定义

$$\Psi(\xi)=(S_1'(\beta), S_2'(\gamma), S_3'(\lambda))'|_{\zeta_\xi},$$

其中 $\zeta_\xi\triangleq(\beta_0'+\xi(\hat{\beta}_m-\beta_0)', \gamma_0'+\xi(\hat{\gamma}_m-\gamma_0)', \lambda_0'+\xi(\hat{\lambda}_m-\lambda_0)')'$，$\xi\in[0,1]$。我们有

$$\Psi(1)-\Psi(0)=\int_0^1\Psi'(\xi)d\xi.$$

由此可得

$$\frac{1}{\sqrt{m}}\begin{pmatrix}S_1(\beta)\\ S_2(\gamma)\\ S_3(\lambda)\end{pmatrix}_{\beta_0,\gamma_0,\lambda_0}=-\omega_m\left\{\sqrt{m}\begin{pmatrix}\hat{\beta}_n-\beta_0\\ \hat{\gamma}_n-\gamma_0\\ \hat{\lambda}_n-\lambda_0\end{pmatrix}\right\}, \tag{A.4}$$

其中

$$\omega_m=\int_0^1\frac{1}{m}\left\{\frac{\partial S'(\zeta)}{\partial\zeta}\right\}_{\zeta_\xi}d\xi=\begin{pmatrix}\omega_m^{11} & \omega_m^{12} & \omega_m^{13}\\ \omega_m^{21} & \omega_m^{22} & \omega_m^{23}\\ \omega_m^{31} & \omega_m^{32} & \omega_m^{33}\end{pmatrix}, \tag{A.5}$$

在这里

$$\omega_m^{11}=\int_0^1\frac{1}{m}\left\{\frac{\partial S'(\beta)}{\partial\beta}\right\}_{\zeta_\xi}d\xi,$$

其余的 $\omega_m^{k,l}(k,l=1,2,3)$ 也是类似的。由条件 $N2-N4$ 知，

$$-\omega_m \to \left(\begin{array}{ccc} v^{11} & 0 & 0 \\ 0 & v^{22} & 0 \\ 0 & 0 & v^{33} \end{array} \right)_{\beta_0,\gamma_0,\lambda_0}.$$

因此，再由条件 $N5$ 以及 $(A.4)(A.5)$ 可知，渐近正态性成立。

A.3 N4-N5 的证明

$N4$ 的证明：

证明类似于前面的相合性证明，我们也只给出 $\frac{\partial S_1'(\beta)}{\partial \beta}$ 部分的证明，其余的都是一样的。首先，我们定义 $T_i \triangleq \frac{\partial}{\partial \beta}\left\{X_i'\Delta_i\Sigma_i^{-1}(y_i-\mu_i(X_i\beta))\right\}'$。因此有

$$\frac{\partial S_1'(\beta)}{\partial \beta} = \sum_{i=1}^{m} T_i.$$

由正则条件我们知道对任意的 $i=1,2,...,m$，$E(T_i|_{\zeta=\zeta_0})<\infty$ 以及

$$Var(T_i) = X_i'\dot{\Delta}_i\Sigma_i^{-1}\dot{\Delta}_i X_i$$

是有界的，从而可知 $\sum\limits_{i=1}^{\infty} Var(T_i)/i^2 < \infty$。最后由 Kolmogorov 强大数律可知，

$$\frac{1}{m}\sum_{i=1}^{m} T_i \to \frac{1}{m}E(\sum_{i=1}^{m} T_i),$$

也就是说当 $m\to\infty$ 时有

$$\frac{1}{m}\frac{\partial S_1'(\beta)}{\partial \beta} - \frac{1}{m}E\left\{\frac{\partial S_1'(\beta)}{\partial \beta}\right\}_{\zeta=\zeta_0} \to 0,\ a.s.$$

$N5$ 的证明:

正则条件 $C3$ 意味着对任意的 $\phi\in R^p$，$\varphi\in R^q$ 与 $\psi\in R^d$，存在一个与 i 无关的常数 M 使得

$$E\left[\left|\phi' X_i'\Delta_i\Sigma_i^{-1}(y_i-\mu_i)+\varphi'\left(\frac{\partial \epsilon_i'}{\partial \gamma}\right)D_i^{-1}\epsilon_i+\psi' Z_i' D_i W_i^{-1}(\epsilon_i^2-\sigma_i^2)\right|^3\right]_{\zeta_0} \leqslant M.$$

在 $\zeta=\zeta_0$ 处，由在 (3.1) 中 V 的正定性可知，

$$\frac{1}{m}\sum_{i=1}^{m}V\left[\phi' X_i'\Delta_i\Sigma_i^{-1}(y_i-\mu_i)+\varphi'\left(\frac{\partial\epsilon_i'}{\partial\gamma}\right)D_i^{-1}\epsilon_i+\psi' Z_i' D_i W_i^{-1}(\epsilon_i^2-\sigma_i^2)\right]_{\zeta_0}$$

$$=(\phi',\varphi',\psi')\begin{pmatrix} v_m^{11} & v_m^{12} & v_m^{13}\\ v_m^{21} & v_m^{22} & v_m^{23}\\ v_m^{31} & v_m^{32} & v_m^{33}\end{pmatrix}\begin{pmatrix}\phi\\ \varphi\\ \psi\end{pmatrix}$$

$$\to(\phi',\varphi',\psi')\begin{pmatrix} v^{11} & v^{12} & v^{13}\\ v^{21} & v^{22} & v^{23}\\ v^{31} & v^{32} & v^{33}\end{pmatrix}\begin{pmatrix}\phi\\ \varphi\\ \psi\end{pmatrix}>0,$$

因此由多元形式的 Liapounov 中心极限定理可知，$N5$ 成立。

A.4 $\mathcal{I}_\gamma$ 的计算

Fisher 信息阵 $\mathcal{I}_\beta$ 与 $\mathcal{I}_\lambda$ 的计算都很直观，在这里就不赘述了。下面我们具体计算一下 $\mathcal{I}_\gamma$。由于

$$\frac{\partial S_2}{\partial\gamma'}=\sum_{i=1}^{m}\sum_{j=1}^{m_i}\frac{1}{\sigma_{ij}^2}\left[\frac{\partial\epsilon_{ij}}{\partial\gamma}\left(\frac{\partial\epsilon_{ij}}{\partial\gamma}\right)'+\frac{\partial^2\epsilon_{ij}}{\partial\gamma\partial\gamma'}\epsilon_{ij}\right],$$

其中 $\frac{\partial^2\epsilon_{ij}}{\partial\gamma\partial\gamma'}=-\sum_{k=1}^{j-1}\left[w_{ijk}\frac{\partial\epsilon_{ik}}{\partial\gamma'}+\frac{\partial\epsilon_{ik}}{\partial\gamma}w_{ijk}'+l_{ijk}\frac{\partial^2\epsilon_{ik}}{\partial\gamma\partial\gamma'}\right]$，易得

$$G_i(\gamma)=E\frac{\partial S_2}{\partial\gamma}=\sum_{j=1}^{m_i}\frac{1}{\sigma_{ij}^2}E\frac{\partial\epsilon_{ij}}{\partial\gamma}\left(\frac{\partial\epsilon_{ij}}{\partial\gamma}\right)'. \tag{A.6}$$

注意到 $\partial\epsilon_{ij}/\partial\gamma=-\sum_{k=1}^{j-1}[\epsilon_{ik}w_{ijk}+l_{ijk}\partial\epsilon_{ik}/\partial\gamma]$ 可以重新表示为

$$\frac{\partial\epsilon_{ij}}{\partial\gamma}=-W_{ij}\epsilon_i-\sum_{k=1}^{j-1}a_{ijk}W_{ik}\epsilon_i,\quad j=2,\cdots,m_i \tag{A.7}$$

其中 $\epsilon_i=(\epsilon_{i1},\cdots,\epsilon_{im_i})^T$，$W_{ij}=(w_{ij1},\cdots,w_{ij(j-1)},0,\cdots,0)$ 是一个 $d\times m_i$ 矩

阵，并且 a_{ijk} 是下三角阵 L_i^{-1} 的 (j,k) 元。从而容易看出

$$\mathcal{I}_\gamma = \sum_{i=1}^{m}\sum_{j=2}^{m_i}\frac{1}{\sigma_{ij}^2}\left[W_{ij}+\sum_{k=1}^{j-1}a_{ijk}W_{ik}\right]D_i\left[W_{ij}+\sum_{k=1}^{j-1}a_{ijk}W_{ik}\right]^T. \tag{A.8}$$

A.5 Ye & Pan (2006) 中的一处错误

文献 Ye & Pan (2006) 是在自回归因子分解模型中引入广义估计方程方法进行均值 -协方差联合建模。简单来说，记 $r_{ij}=y_{ij}-\mu_{ij}$，有如下自回归模型：

$$r_{ij}=\sum_{k=1}^{j-1}\phi_{ijk}r_{ik}+\varepsilon_{ij},\ i=1,...,m,\ j=1,...,m_i.$$

模型中关于误差项分布的假设与我们的一致，即假设 $Cov(\varepsilon_i)=D_i=diag\{\sigma_{i1}^2,...,\sigma_{im_i}^2\}$。类似于我们的工作，他们对均值、自回归系数和误差项方差对数分别引入线性模型，

$$g(\mu_{ij})=x_{ij}'\beta,\qquad \phi_{ijk}=z_{ijk}'\gamma,\qquad log(\sigma_{ij}^2)=z_{ij}'\lambda,$$

其中 $g(\cdot)$ 是单调可微的，被称为联系方程，x_{ij}，w_{ijk} 与 z_{ij} 分别是 $p\times 1$，$q\times 1$ 与 $d\times 1$ 维的协变量，β，γ 和 λ 是对应的回归系数。

令 $\hat{r}_{ij}=\sum_{k=1}^{j-1}\phi_{ijk}r_{ik}$ 有广义估计方程：

$$\begin{aligned}
S_1(\beta)&=\sum_{i=1}^{m}X_i'\Delta_i\Sigma_i^{-1}(y_i-\mu_i(X_i\beta))=0,\\
S_2(\gamma)&=\sum_{i=1}^{m}\left(\frac{\partial\hat{r}_i'}{\partial\gamma}\right)D_i^{-1}(r_i-\hat{r}_i)=0,\\
S_3(\lambda)&=\sum_{i=1}^{m}Z_i'D_i(Z_i\lambda)W_i^{-1}(\varepsilon_i^2-\sigma_i^2(Z_i\lambda))=0,
\end{aligned}\tag{A.9}$$

其中 $r_i-\hat{r}_i=\varepsilon_i$，且 $\partial\hat{r}_i/\partial\gamma$ 是一个 $q\times m_i$ 矩阵，它的第一列为零，第 $j(j>2)$ 列为 $\partial\hat{r}_{ij}/\partial\gamma=-\sum_{k=1}^{j-1}r_{ik}z_{ijk}$。

在一些相似的条件下，他们指出广义估计方程解 $(\hat{\beta}_m,\hat{\gamma}_m,\hat{\lambda}_m)'$ 满足当

$m \to \infty$ 时有，

$$\sqrt{m}\begin{pmatrix} \hat{\beta}_m - \beta_0 \\ \hat{\gamma}_m - \gamma_0 \\ \hat{\lambda}_m - \lambda_0 \end{pmatrix}$$
$$\to N\left\{0, \begin{pmatrix} v^{11} & 0 & 0 \\ 0 & v^{22} & 0 \\ 0 & 0 & v^{33} \end{pmatrix}^{-1} \begin{pmatrix} v^{11} & v^{12} & v^{13} \\ v^{21} & v^{22} & v^{23} \\ v^{31} & v^{32} & v^{33} \end{pmatrix} \begin{pmatrix} v^{11} & 0 & 0 \\ 0 & v^{22} & 0 \\ 0 & 0 & v^{33} \end{pmatrix}^{-1}\right\}.$$

注意到，该渐近协方差阵也是一个“三明治”形式的矩阵，其两边的“面包”阵中的第二个对角块恰好与中间的“肉”矩阵中的第二个对角块一样，这与我们的结果不一致。但是我们在研读文献过程中发现他们的结果是不正确的，因为他们之所以能得到这样的一个结果，依赖于如下等式 (Ye & Pan (2006) 中附录 Condition A6)：

$$\begin{aligned} &E\left[\left\{\left(\frac{\partial \hat{r}_i'}{\partial \gamma}\right) D_i^{-1}(r_i - \hat{r}_i)\right\}\left\{\left(\frac{\partial \hat{r}_i'}{\partial \gamma}\right) D_i^{-1}(r_i - \hat{r}_i)\right\}'\right] \\ &= -E\left[\frac{\partial}{\partial \gamma}\left\{\left(\frac{\partial \hat{r}_i'}{\partial \gamma}\right) D_i^{-1}(r_i - \hat{r}_i)\right\}'\right], \end{aligned} \tag{A.10}$$

但是由于他们对误差项的假设是不相关但不一定独立，所以我们可以说明这个等式并不成立，为此我们在这里给出一个简单的反例。

考虑二维自回归模型 $r_i = (r_{i1}, r_{i2})'$，其中

$$r_{i1} = \varepsilon_{i1}, \quad r_{i2} = \phi_{i21} r_{i1} + \varepsilon_{i2}.$$

设 $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2})' \sim F = \pi N(0, \Sigma_1) + (1-\pi) N(0, \Sigma_2)$，其中混合权重 $\pi = 1/3$，且

$$\Sigma_1 = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}.$$

从而我们有

$$E(\varepsilon_i) = 0, \; Cov(\varepsilon_i) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

即 ε_{i1} 与 ε_{i2} 不相关。

我们可以算得 ε_i 的密度函数为

$$f_{\varepsilon_i}(x_1,x_2)=\frac{1}{12\pi}\exp\left\{-\frac{1}{2}(\frac{1}{2}x_1^2-x_1x_2+x_2^2)\right\}+\frac{1}{2\pi}\exp\left\{-\frac{1}{2}(2x_1^2+2x_1x_2+x_2^2)\right\}.$$

进而我们可以得到 ε_{i2} 的边际密度函数 $f_{\varepsilon_{i2}}(x_2)=\int_{-\infty}^{\infty}f_{\varepsilon_i}(x_1,x_2)dx_1$ 为

$$f_{\varepsilon_{i2}}(x_2)=\frac{1}{2\sqrt{\pi}}\exp\left\{-\frac{1}{4}x_2^2\right\}.$$

接着我们可以计算给定 ε_{i2} 时 ε_{i1} 的条件密度函数，例如

$$f_{\varepsilon_{i1}}(x_1|\varepsilon_{i2}=0)=\frac{f_{\varepsilon_i}(x_1,x_2)}{f_{\varepsilon_{i2}}(x_2)}\Big|_{x_2=0}=\frac{1}{6\sqrt{\pi}}\exp\left\{-\frac{1}{4}x_1^2\right\}+\frac{2}{3\sqrt{\pi}}\exp\left\{-x_1^2\right\}.$$

由此可得 $E(\varepsilon_{i1}|\varepsilon_{i2}=0)=0$，$Var(\varepsilon_{i1}|\varepsilon_{i2}=0)=1\neq Var(\varepsilon_{i1})=2$ 以及

$$E(\varepsilon_{i1}^2\varepsilon_{i2}^2)=\int\int x_1^2x_2^2f_{\varepsilon_i}(x_1,x_2)dx_1x_2=28/3,$$

因此，

$$\begin{aligned}&E\left[\left\{\left(\frac{\partial\hat{r}_i'}{\partial\gamma}\right)D_i^{-1}(r_i-\hat{r}_i)\right\}\left\{\left(\frac{\partial\hat{r}_i'}{\partial\gamma}\right)D_i^{-1}(r_i-\hat{r}_i)\right\}'\right]\\&=E\left[\left(\frac{1}{\sigma_{i2}^2}\right)^2\varepsilon_{i1}^2\varepsilon_{i2}^2z_{ij1}z_{ij1}'\right]=\frac{7}{3}z_{ij1}z_{ij1}',\end{aligned}$$

但是

$$-E\left[\frac{\partial}{\partial\gamma}\left\{\left(\frac{\partial\hat{r}_i'}{\partial\gamma}\right)D_i^{-1}(r_i-\hat{r}_i)\right\}'\right]=E\left[\frac{1}{\sigma_{i2}^2}\varepsilon_{i1}^2z_{ij1}z_{ij1}'\right]=z_{ij1}z_{ij1}'.$$

这也就是说等式 $(A.10)$ 不成立。

原书空白页，
不缺内容

致　谢

转眼间我即将在中国科学技术大学完成本科和硕士学业，在这段学习和研究经历中，我始终得到了来自导师以及学校其他老师和同学的谆谆教导和热忱帮助。

在论文完成之际，首先我要感谢我的导师张伟平副教授长久以来的指导和教诲。张老师严谨的研究态度、勤恳的工作精神以及渊博的学识，都是我学习的榜样。

感谢在本科以及研究生阶段所有的任课老师，他们的指导给我研究生阶段的学习工作打下了基础，使我终生受益。

感谢来自五湖四海的同学们，与你们的相处让我感受到了集体的温暖，与你们的讨论也使我受益良多。感谢你们，我们一同走过这段难忘而愉快的青春岁月。

最后，我要特别感谢家人对我一贯的鼓励与支持，焉得谖草，言树之背，养育之恩，无以回报。你们健康快乐才是我最大的心愿！

刘笑宇

2012 年 4 月

纵向数据均值-协方差建模中的滑动平均因子模型

作者： 刘笑宇
学位授予单位： 中国科学技术大学