分类号 021 密级 公开
UDC 编号

雲南大學

博士研究生学位论文

题 目 响应变量缺失下的特征筛选和模型的参数估计

Title Feature screening and Model parameter estimation for missing response

学院（所、中心） 数学与统计学院

专业名称 概率论与数理统计

研究方向 缺失数据的统计分析

研究生姓名 李晓霞 学号 22015000115

导师姓名 巩馥洲 唐年胜 职称 教授

2019 年 12 月

摘　要

随着数据收集技术的快速发展, 很多领域的研究者用较低的成本可以获得大量的超高维数据, 在超高维数据分析中, 预测变量的个数 p 随着样本量 n 的增加呈指数增长, 但只有少数预测变量对响应变量有显著影响, 这一点已经众所周知. 为此, 统计学家们提出了许多边际特征筛选的方法. 但是在实际应用中, 由于各种原因, 缺失数据经常出现在经济学、社会学、生物医学、市场调研等很多领域中, 近年来, 缺失数据模型的统计推断引起了许多学者的关注. 很多经典的统计方法和理论都建立在完全观测数据分析的基础上, 不能直接应用到缺失数据模型的统计推断. 对缺失数据进行统计分析最常用的方法有基于完全观测数据的方法, 插补方法, 似然方法, 逆概率加权方法等. 如何在缺失数据的背景下将超高维数据中预测变量的维数大幅度降至中等规模, 然后建立模型并进行参数的估计是需要解决的一个重要问题. 为此, 在响应变量缺失框架下, 本文研究了超高维数据下的特征筛选, 广义部分线性单指标模型的参数估计以及数据存在异常值时的稳健估计. 其主要内容包括:

1. 讨论了响应变量带有随机缺失的超高维数据的特征筛选问题. 提出了一种新的非参数特征筛选方法, 通过条件边际插补 Spearman 秩相关关系来识别重要特征. 所提的非参数筛选方法不需要假定预测变量和响应变量之间任何的回归形式, 也不需要为缺失数据机制模型指定参数化模型, 而且对异常值和重尾数据具有稳健性. 在一些正则条件下, 证明了所提出的特征筛选方法具有确定筛选性质和秩相合性. 模拟研究和扩散性的大-B-淋巴瘤细胞实例数据分析都表明, 所提的非参数筛选过程优于现有的几种无模型筛选过程.

2. 研究了响应变量不可忽略缺失下广义部分线性单指标模型的参数估计. 考虑半参数逻辑回归模型作为响应变量缺失机制. 结合局部似然方法和倾向得分方法, 提出了一种基于截面加权估计方程 (WEE) 的广义部分线性单指标模型的参数和非参数估计方法. 基于截面原理, 利用核回归方法对非参数部分进行估计, 广义矩方法对参数进行估计, 并给出了估计量的渐近性质. 模拟研究验证了

所提方法的有效性和可行性.

3. 研究了在响应变量随机缺失且协变量和响应变量中都存在异常值时回归参数的有效估计. 首先, 构建一个加权的拟似然函数, 对缺失机制模型中参数进行稳健估计. 其次, 利用 Tukey's biweight 函数的一阶导函数, 基于逆概率加权和重新下降的思想, 建立了包含感兴趣未知参数的无偏估计方程组 (即能处理缺失数据还能处理异常值), 使用广义矩估计方法对感兴趣参数进行估计. 再次证明了估计量的相合性和渐近正态性. 模拟研究和实例分析都表明, 响应变量随机缺失且协变量和响应变量中都存在异常值时, 该方法具有良好的表现.

关键词： 特征筛选; 缺失数据; 边际 spearman 相关关系; 参数估计; 稳健估计

Abstract

With the rapid development of data collection technology, many researchers can obtain ultra-high dimensional data at low cost in many fields. In ultra-high dimensional data analysis, It is well known that the number of prediction variables p increases exponentially with the increase of sample size n, but only a few of prediction variables have a significant impact on the response variables. To this end, statisticians have proposed a number of marginal feature screening methods. But in actual application, due to various reasons, missing data often appear in economics, sociology, biomedicine, market research and many other fields, In recent years, statistical inference under missing data has attracted much attention. Classical statistical methods and theories are established on the basis of the complete data analysis, and cannot be directly applied to the missing data of statistical inference. The most commonly used method in missing data analysis has complete case analysis, impution method, the likelihood method, and inverse probability weighting method, etc. how to reduce the dimensionality of prediction variables to a medium scale in the context of missing data? and then model and estimate parameters are important problems to be solved. To this end, under the framework of missing data, this dissertation studies feature screening under ultra-high dimensional data, parameter estimation of generalized partial linear single-indicator model and robust estimation when abnormal points exist in data. The main content of this dissertation is as follows:

1. This chapter addresses the feature screening issue for ultrahigh-dimensional data with response missing at random. A novel nonparametric feature screening procedure is developed to identify the important features via the conditionally imputing marginal Spearman rank correlation. The proposed nonparametric screening approach doesn't have to assume any regression form of predictors on response variable and specify a parameterized model for the missing data mechanism model.

It is robust to outliers and heavy-tailed data. Under some regularity conditions, it is shown that the proposed feature screening procedure has the sure screening and ranking consistency properties. Simulation studies show that the proposed screening procedure outperforms several existing model-free screening procedures. An example taken from the microarray diffuse large-B-cell lymphoma study is used to illustrate the proposed methodologies.

2. This chapter studies parameter estimation for generalized partially linear single-index models when response variable is nonignorable missing. We consider the semi-parametric logistic regression model as response mechanism in this dissertation. Combining the local likelihood technique and the propensity score method, a profile weighted estimating equations (WEE)-based approach is proposed to estimate the parameters and non-parameters in generalized partially linear single-index models. Basing on profile principle, using a kernel-type estimator to estimate the nonparametric component, and then estimate the parameter of response mechanism by the generalized method of moments. Asymptotic properties of the proposed estimators are established. Simulation studies and real data applications are conducted to illustrate the effectiveness and feasibility of the estimators.

3. This chapter studies the parameter estimation of regression models with simultaneous missing response and outliers in both response and covariates. First, to be robust against outliers in both response and covariates, we define a weighted profile likelihood whose purpose is to obtain robust parameter estimates in the missing data mechanism model. Second, using the first derivative of the Tukey's biweight function, based on the inverse-probability weighted method and redescending technique, we construct a class of unbaised estimating equations containing the parameter involved in the regression model. Ihese new estimating equations not only deal with missing data but also eliminate the effect of outliers. Finally, we employ the generalized method of moments to estimate unknown parameters of interest, and then investigate the large sample theory of the proposed estimator. Expensive simulations and a real

data example have shown that the proposed method performs well when there exists missing response and outliers in both response and covariates.

Keywords: Feature screening; Missing data; Marginal Spearman rank correlation; Parameter estimation; Robust estimation

目　录

第一章 绪论

1.1 问题的提出

在经济学、社会学、生物医学、市场调研等许多应用领域中, 常常因为各种原因使得一些数据不能获得, 如一些被抽样的个体中途退出研究、不愿意回答调查问卷上的一些问题, 还有一些不可控制的因素导致信息丢失等等. 然而经典的统计方法与理论都建立在完全数据分析的基础上, 不能直接应用到缺失数据的统计推断. 缺失数据统计分析方法的有效性很大程度上依赖于缺失数据机制. 在过去十几年中, 基于三种不同缺失数据机制模型 (Little and Rubin, 2002), 各种统计方法被提出用来处理缺失数据问题. 三种缺失数据机制分别如下: 完全随机缺失 (MCAR), 其缺失数据不依赖于任何观察到的或未被观测的数据; 随机缺失 (MAR), 其缺失数据仅取决于观测数据; 非随机缺失 (MNAR), 其缺失数据与未观测到的数据有关, 可能也与观测到的数据有关, 这是缺失机制中最难处理的. 通常, MCAR 和 MAR 被称为可忽略缺失, 而 MNAR 被称为不可忽略缺失.

缺失数据分析中最常用的方法是基于完全观测数据进行统计推断. 然而, 当缺失数据机制不是 MCAR 时, 这种方法获得的参数估计可能存在偏差且方差较大. 另一种处理缺失数据的流行方法是插补方法, 包括单值插补; 例如热平台插补 (即, 缺失值由随机选取一个相似记录值插补), 冷平台插补; 均值插补 (即用该变量的所有其他观察到值的平均值来插补), 它对于单变量统计分析有一些吸引人的性质, 但是对于多变量分析, 它是有问题的, 因为它减弱了要插补的变量之间的相关性; 回归插补 (即, 利用所考虑的回归模型的拟合值进行插补缺失值)、多重插补 (MI). 特别地, Horvitz and Thompson (1952) 提出的逆概率加权 (IPW) 方法也是一种常用的方法, 但当缺失概率非常小时, 使用逆概率加权的方法存在不稳定性, 即使在 MAR 下是无偏的, 也可能由于数据遗漏而导致 IPW 估计有很大的方差，进而影响其有效性. 为了使用不完整数据个体中增加的额外信息, Rubins, Rotnitzky and Zhao (1994) 提出了增广逆概率加权 (AIPW) 方法. AIPW 方法具有双重稳健性, 即当缺失机制模型和回归模型中至少有一

个假定正确时, AIPW 估计是相合的. 同样地, AIPW 估计也易受缺失概率的影响. 与 AIPW 方法相关的大部分现有工作主要集中在缺失数据的 MAR 假设. 当缺失数据是 MNAR 时, 一般的处理 MAR 数据的方法针对不可忽略缺失数据可能是无效的, 由于在这种情况下, 缺失机制模型中缺失变量并不总是被观察到. 所以不可忽略缺失数据的统计分析方法是相当具有挑战性的. 为此, 许多学者在 PS 方法的基础上提出了许多处理缺失数据的方法. 例如 Hansen (1982), Kim and Shao (2013), Riddles (2013), Wang, Shao and Kim (2014), Riddles, Kim and Im (2016), Jiang, Zhao and Tang (2016). 此外基于模型的方法为处理缺失数据提供了额外的工具, 包括所考虑模型中检验缺失数据类型和估计参数. 广泛所使用的方法包括最大似然法 (ML) 或观测数据的拟似然法, 以及当似然函数或者拟似然函数未获取时, Dempster, Laird and Rubin (1977) 提出的从完全数据似然在观测数据下的期望的最大似然估计 (MLE) 的 EM 算法, 并将其推广到各种统计模型中. 例如, Laird and Ware (1982), Ibrahim (1990), Ibrahim, Chen and Lipsitz (2001), Lee and Zhu (2002), Zhao and Tang (2015), 和 Tang and Tang (2018). 此外 Owen (1988) 提出的经验似然 (EL) 方法, 由于其诱人的特性, 在缺失数据分析中也得到了广泛的应用. 例如, 它有与自助法相类似的抽样性质, 但不是重抽样. 关于缺失数据分析使用 EL 方法有相当多的的文献, 包括均值泛函和模型的估计参数. 例如, Cheng (1994), Wang and Rao (2002), Liang, Wang and Carroll (2007), Stute, Xue and Zhu (2007), Xue (2009), Qin, Zhang and Leung (2009), Wang and Chen (2009), Tang and Zhao (2013), Wang and Qin (2010), Chen and Kim (2017), Zhao, Tang and Tang (2013), Zhao, Zhao and Tang (2013), and Tang, Zhao and Zhu (2014).

随着现代科学技术的发展, 人类进入了大数据时代. 在超高维数据分析中, 预测变量的个数 p 随着样本量 n 的增加呈指数增长, 但只有少数预测变量对响应变量有显著影响, 这一点已经众所周知. 为此, 各种特征筛选过程被提出将预测变量的维数大幅度降至中等规模, 以至于经典统计推断方法可应用于其简化模型中. 有很多的文献研究了基于模型的和无模型的变量筛选方法, 例如, Fan and Lv (2008), Fan and Song (2010), Fan et al. (2011), Chang et al. (2013), Zhu et al. (2011), Li et al. (2012), He et al. (2013), Li, Zhong and Zhu (2012), Mai and

Zou (2015), Cui et al. (2015), Pan, Wang and Li (2016), Yan et al. (2018), Xie et al. (2019) 等等. 上述所有的特征筛选方法都集中于完全观测数据, 但当数据存在缺失时, 变量筛选的文献很少, 这就是本文研究的第一个问题, 响应变量随机缺失下的超高维数据的特征筛选. 当通过特征筛选的方法将预测变量的维数大幅度降低后, 接着就需要建立模型并进行统计推断, 而广义部分线性单指标模型 (GPLSIM) 是部分线性回归模型、单指标模型和广义线性模型的自然推广, 由于它既有参数模型的优点, 也有非参数模型的优点, 其灵活性能避免维度祸根且适应于连续响应变量和离散响应变量, 引起了许多学者的广泛关注. 有很多的文献针对随机缺失或不可忽略缺失的参数模型进行研究, 或针对随机缺失的单指标模型进行统计分析. 例如, Wang et al. (2010), He and Yi (2011), Lai and Wang (2011), Dong and Zhu (2013), Xue (2013), Lai and Wang (2014), Wang, Zhang and Hardle (2018), Jiang et al. (2016), Riddles et al. (2016), Zhao et al. (2017). 现有的方法在协变向量维数很高的情况下存在维数祸根, 且在缺失数据机制是不可忽略的情况下效率会明显降低. 受到上述问题的启发, 讨论了响应变量不可忽略缺失下广义部分线性单指标模型的参数估计问题. 这就是本文研究的第二个问题. 然而, 实例数据中不单单只存在缺失数据, 还可能存在异常值. 举一个最简单的例子: 在没有数据缺失下, 回归模型的最小二乘估计 (OLS) 对异常值或者潜在模型的正确性非常敏感. 实际中很多数据都包含异常值, 且这些异常值可能出现在响应变量和 (或) 协变量中, 如在收入调查数据中经常遇见. 稳健的统计推断目的是为了构造不受异常值影响的稳健估计或假设检验统计量. 例如, Hampel (1968), Huber (1973, 1981), Rousseeuw and Leroy (1987), Yohai and Zamar (1988), Konker and Basset (1978), Gervini and Yohai (2002), Koenker (2005), Zhu and Zhang (2004), Jiang et al (2018) 等等. 因此, 发展有效的统计推断方法使其对协变量和 (或) 响应变量中异常值都稳健显得越发重要. 在响应变量缺失且协变量和响应变量中都存在异常值时, 稳健估计尚未见相关文献, 这是本文研究的第三个问题.

1.2 国内外研究现状

1.2.1 变量筛选

随着数据收集技术的快速发展, 很多领域的研究者可以用较低的成本获得超高维数据, 例如生物医学成像, 肿瘤分类技术, 金融, 核磁共振成像等领域. 然而由于计算成本, 统计精度和算法的稳定性等问题, 许多变量选择和充分降维的方法不能有效地处理超高维数据的问题. 为此, 统计学家们提出了许多边际筛选的方法能够有效和快速地将超高维数据 $p = \exp\{O(n^{\xi})\}, \xi > 0$ 减少至一个相对低的维数 $d = O(n)$. 并且该方法能够在依概率 1 收敛情况下保证所有的重要变量被筛选出来. 继而充分降维和变量选择的方法可用于估计相对低的 d 维参数向量, 从而大大地减少计算量. 大多数变量筛选的方法分为两类: 基于模型和无模型的变量筛选方法.

目前, 基于模型变量筛选的方法有 Fan and Lv (2008) 针对线性回归模型基于边际皮尔逊相关关系提出了一种确定独立筛选 (SIS) 过程和迭代的确定独立筛选 (ISIS) 过程. Fan and Song (2010) 将 SIS 方法推广到广义线性模型中. Fan et al. (2011) 利用 B-样条基展开提出了针对可加模型的一种非参数边际筛选方法. Chang et al. (2013) 针对线性模型和广义线性模型提出了一种基于边际经验似然比的特征筛选方法. 上述基于模型的特征筛选过程只有在真实模型正确指定时, 结果才会令人满意, 但是当真实模型指定错误时, 结果可能不会表现良好. 众所周知, 在许多实际应用中, 给超高维数据指定一个正确的模型是相当具有挑战性的, 甚至几乎是不可能的. 为此, 近年来发展了一些无模型的特征筛选方法. 例如, Zhu et al. (2011) 给出了一个确定的独立秩筛选 (SIRS) 过程来识别重要的预测变量. Li, Zhong and Zhu (2012) 介绍了一种基于距离相关 (DC) 的 SIS 过程, 该过程对极值重尾数据不具有稳健性. Li, Peng, Zhang and Zhu (2012) 基于 Kendall τ 相关关系提出了一种稳健的特征筛选过程. Mai and Zou (2013) 针对二元分类问题利用 Kolmogorov－Smirnov 统计量提出一种确定特征筛选方法. He et al. (2013) 基于边际效用的样条近似值提出了一种自适应分位数的非线性独立筛选过程. Mai and Zou (2015) 研究了一种基于带了切片技术融合的 Kolmogorov filter

的确定特征筛选过程, 此过程在计算上非常耗时. Cui et al. (2015) 对超高维判别分析问题给出了一种均值方差筛选过程, 此方法仅仅适用于分类响应变量和连续的预测变量. Pan, Wang and Li (2016) 对带有发散维类和超高维预测变量的线性判别分析提出了两两确定筛选方法. Yan et al. (2018) 拓展了Cui et al. (2015) 的 MVS 过程, 通过引入切片技术使之适应于各种类型的预测变量, 连续型、离散型和分类的响应变量. Xie et al. (2019) 针对超高维异质性分类数据提出了一种自适应分类变量筛选过程. 然而, 上述所提到的无模型的特征筛选方法主要集中于完全观测数据.

在生物医学、社会学、临床试验、经济学、纵向研究等各个领域, 经常会出现数据缺失的情况 (Little and Rubin, 2002). 关于缺失数据回归模型的变量选择, 已有相当多的文献. 例如, 参见 Ibrahim et al. (2008), Garcia et al. (2010), Long and Johnson (2015), Fang and Shao (2016), Zhao et al. (2017), Tang and Tang (2018) 等. 然而, 上述文献关注的是缺失数据的低维回归模型. 最近, 人们已经认识到在超高维缺失数据分析中, 越来越需要发展一些特征筛选方法来识别重要的预测变量. 例如, Lai et al. (2017) 提出了一种采用逆概率加权 (IPW) 方法调整 (Zhu et al., 2011) 所提的确定独立秩筛选效用 (SIRS) 的无模型特征筛选方法.

然而, 此方法严重依赖于倾向得分函数的正确假定和估计, 对模型的错误假定不具有稳健性. 因此, 发展一种高效、计算可行、稳健、无模型的特征筛选方法来区分超高维缺失数据的重要预测变量和不重要预测变量是十分必要的. 为此, 本章针对超高维响应变量随机缺失数据, 提出了一种基于带有非参数插补的调整 Spearman 秩相关关系 (ASRC) 的新的特征筛选方法.

1.2.2　广义部分线性单指标模型

广义部分线性单指标模型 (GPLSIM) 是部分线性回归模型、单指标模型和广义线性模型的自然推广, 由于它允许一部分协变量线性建模和其他部分非参数建模的灵活性, 既有参数模型的优点, 也有非参数模型的优点, 能很好地避免维度祸根且对于响应变量是连续还是离散都适用, 因此得到了相当多的关注. 针对这个模型的参数估计有很多文献进行研究, 例如, Carroll et al. (1997) 用局部线性回归而不是简单的核回归方法, He and Yi (2011) 针对相关二元数据的成对

似然方法, Yi et al. (2009) 针对相依的二元响应变量提出了一种基于局部线性的截面最小二乘估计量, Boente and Rodriguez (2012) 的稳健估计, Huh and Park (2002) 对于没有部分线性项的广义单指数模型提出平均导数估计量, Poon and Wang (2013) 提出完全的贝叶斯方法, Yu et al. (2017) 的惩罚样条估计.

上述工作主要集中在完全观测数据上. 然而, 实际上, 由于各种原因, 例如药物的严重副作用, 一些抽样个人不愿意回答所需的信息, 由于无法控制的因素造成的信息丢失, 一些定期的访问者间歇性或退出研究 (Little and Rubin, 2002), 响应变量可能缺失. 关于在响应变量随机缺失下的单指标模型的统计推断已有相当多的研究成果. 譬如, Wang et al. (2010) 研究了在单指标模型的未知连接函数和方向参数的估计问题; He and Yi (2011) 提出了一种灵活的半参数方法来处理 GPLSIM 中的估计问题; Lai and Wang (2011) 借助插补方法研究了部分线性单指标模型的估计问题; Dong and Zhu (2013) 提供了一种带有缺失数据的单指标模型中方向参数估计的一般处理方法; Xue (2013) 针对协变量中数据缺失的单指标模型, 提出了一种经验似然方法; Lai and Wang (2014) 建立了响应变量随机缺失的异方差部分线性单指标模型的半参数有效界; Wang, Zhang and Hardle (2018) 考虑了响应变量随机缺失的广义单指标模型. 据我们所知, 带有非随机缺失的响应变量 (即, 不可忽略缺失) 下 GPLSIM 的统计推断尚未被研究.

不可忽略缺失数据的统计推断更具挑战性, 因为缺失概率模型往往是未知的, 而假定的缺失数据机制模型是不可验证的. 为此, 基于似然的方法和贝叶斯方法处理不可忽略缺失数据的统计分析研究, 许多学者已经付出了大量的努力. 例如 Ibrahim et al. (1999) 和 Lee and Tang (2006). 特别地, Kim and Yu (2011) 基于不可忽略缺失数据机制的指数倾斜模型, 提出了一种带有不可忽略缺失数据的均值估计的半参数方法. Tang, Zhao and Zhu (2014) 在具有不可忽略缺失数据的广义估计方程中提出了一种利用核回归插补缺失数据进行参数估计的经验似然方法.

最近, Horvitz and Thompson (1952) 提出的逆概率加权方法, 也称为倾向得分加权方法, 被广泛用于处理缺失数据问题. 例如, Qin et al. (2008) 提出了一种有效的回归插补方法, 当缺失机制是协变量相关的, 且倾向得分函数被正确指

定时, 该方法对回归模型的错误指定具有稳健性; Xue (2013) 利用逆概率加权方法, 建立了一种经验似然法, 用于协变量随机缺失的单指标模型中参数向量置信域的构造; Jiang et al. (2016) 针对一类协变量不可忽略缺失的线性模型, 提出了一种基于有效的稳健的回归过程的调整倾向得分参数估计; Riddles et al. (2016) 提出了一种以校准条件为辅助信息的倾向得分估计的最大似然方法; Zhao et al. (2017) 研究了协变量或响应变量不可忽略缺失时分位数回归中参数的几种逆概率加权估计. 上述工作建立在随机缺失或不可忽略缺失的参数模型或随机缺失的单指标模型的基础上. 因此, 现有的方法在协变向量维数很高的情况下存在维数祸根, 而在缺失数据机制是不可忽略的情况下效率会明显降低. 受到上述问题的启发, 本文讨论了响应变量不可忽略缺失下广义部分线性单指标模型的参数估计问题.

1.2.3　稳健估计

然而, 数据中不单单只存在缺失数据, 还可能存在异常值. 举一个最简单的例子: 在没有数据缺失下, 回归模型的最小二乘估计 (OLS) 对异常值或者潜在模型的正确性非常敏感. 实际中很多数据都包含异常值, 如在收入调查数据中经常遇见. 异常值的产生有很多原因, 如: 数据中一小部分的个体来自另外一个总体或者由于某些原因造成测量误差也会造成异常值的出现. 稳健的统计推断目的是为了构造不受异常值影响稳健的估计或者假设检验统计量. 另外, 数据集中包含异常值经常在统计分析中遇见, 且这些异常值可能出现在响应变量和 (或) 协变量中. 因此, 发展有效的统计推断方法使其对协变量和 (或) 响应变量中异常值都稳健显得越发重要. 稳健统计的参考文献请参看: Hampel (1968), Huber (1981), Rousseeuw and Leroy (1987), Zhu and Zhang (2004). 过去几十年, 稳健回归估计受到越来越多研究者的关注. Huber (1973) 首次提出最流行的稳健回归 M-type 估计量, 中位数回归模型、分位数回归模型 (Konker and Basset, 1978; Koenker, 2005), MM-估计 (Yohai, 1987), τ-估计 (Yohai and Zamar, 1988), 有效且稳健的加权最小二乘估计 (REWLS) (Gervini and Yohai, 2002), 扩展的带有新截尾的 Huber's 函数 (Huber-ESL) (Jiang et al. 2018). 这些估计量只对响应变量中的异常值稳健, 然而对协变量中的高杠杆点 (数据点脱离协变量空间主体) 却异

常敏感, 此时它们的渐近破坏点是零 (Yohai, 1987). 为了获得对协变量和响应变量都稳健的估计, 损失函数的一阶导数需要是重新下降的 (Rousseeuw and Yohai, 1984; Yohai, 1987). 最常见的满足这个性质的损失函数是 Tukey's biweight 函数 (Tukey, 1960). 综上, 可以发现: 在响应变量缺失且协变量和响应变量中都存在异常值时, 尚未见相关文献, 所以本文研究了在响应变量随机缺失且协变量和响应变量中都存在异常值时如何获得回归参数的有效估计.

1.3 本文主要工作及创新点

本论文研究了响应变量缺失下的特征筛选和模型的参数估计问题, 其主要研究工作包括:

首先, 针对超高维响应变量随机缺失数据, 提出了一种基于带有非参数插补的调整 Spearman 秩相关关系 (ASRC) 的新的特征筛选方法. 所提的非参数筛选方法有几个可取的优点. 第一, 它是非参数方法, 不需要为所考虑的响应变量和候选预测变量指定参数化模型,也不需要为缺失数据机制模型指定参数化模型. 第二, 此筛选可以研究响应变量和预测变量之间的非线性关系, 对异常值和重尾数据具有稳健性. 包括对缺失数据机制模型和假设回归模型的错误假定, 仍然表现良好. 第三, 在一些正则条件下, 证明了所提出的特征筛选方法具有确定筛选性质和秩相合性. 模拟研究和扩散性的大 B-淋巴瘤细胞实例数据分析都表明, 所提的非参数筛选过程优于现有的几种无模型筛选过程.

其次, 针对响应变量不可忽略缺失的 GPLSIM, 采用了倾向得分方法去估计模型中的未知参数和未知连接函数, 其主要思想就是对完全观测数据给定一个与响应变量被观测到的概率 (响应概率) 相关的权重. 响应概率模型采用了一种半参数逻辑回归模型 (SLRM). 特别地, 为了得到响应概率的相合估计, SLRM 中的非参数部分采用核回归方法估计, 参数部分基于工具变量构造的估计方程采用两步的广义矩估计方法 (Wang et al. 2014), 其中工具变量是与研究的响应变量有关, 但是与缺失数据机制无关的协变量. 基于局部线性方法, 给出了未知参数和未知非参数函数的估计, 提出了一种计算非参数函数和参数估计的数值迭代算法. 同时, 也系统地研究了这些估计量的渐近性质. 模拟研究验证了所提方法

的有效性和可行性.

最后, 详细地讨论在响应变量随机缺失且协变量和响应变量中都存在异常值时如何获得回归参数的有效估计. 首先, 对缺失机制模型, 构建一个加权的拟似然函数, 然后基于给定似然函数获得缺失机制模型中参数的稳健估计. 其次, 基于逆概率加权和重新下降的思想, 建立了包含感兴趣未知参数的无偏估计方程组 (即能处理缺失数据还能处理异常值). 最后, 使用广义矩估计方法对感兴趣参数的估计, 并证明了新提出估计量的相合性和渐近正态性. 模拟研究和实例分析都表明, 响应变量随机缺失且协变量和响应变量中都存在异常值时, 该方法具有良好的表现.

第二章 响应变量随机缺失下超高维数据的非参数特征筛选

2.1 引言

随着现代科学技术和数据收集技术的快速发展, 超高维数据经常出现在从基因组学, 健康科学到经济学, 机器学习等各个领域中. 在超高维数据分析中, 预测变量 p 随着样本量 n 的增加呈指数增长, 但只有少数预测变量对响应变量有显著影响, 这一点已经众所周知. 由于同时存在计算可行性, 统计准确性和算法稳定性的挑战 (Li, Zhong and Zhu, 2012), 针对超高维数据, 现有的正则化方法如 LASSO (Tibshirani, 1996)、平滑剪切绝对偏差 (Fan and Li, 2001) 和自适应 LASSO (Zou, 2006) 可能表现比较差. 为此, 各种特征筛选过程被提出将预测变量的维数大幅度降至中等规模中, 以至于经典统计推断方法可应用于其简化模型. 例如, Fan and Lv (2008) 针对线性回归模型基于边际皮尔逊相关关系提出了一种确定独立筛选 (SIS) 过程和迭代的确定独立筛选 (ISIS) 过程. Fan and Song (2010) 将 SIS 方法推广到广义线性模型中. Fan et al. (2011) 利用 B-样条基展开提出了针对可加模型的一种非参数边际筛选方法. Chang et al. (2013) 针对线性模型和广义线性模型提出了一种基于边际经验似然比的特征筛选方法. 上述基于模型的特征筛选过程只有在真实模型正确指定时, 结果才会令人满意, 但是当拟合模型指定错误时, 结果可能不会表现良好.

众所周知, 在许多实际应用中, 给超高维数据指定一个正确的模型是相当具有挑战性的, 甚至几乎是不可能的. 为此, 近年来发展了一些无模型的特征筛选方法. 例如, Zhu et al. (2011) 给出了一个确定的独立秩筛选 (SIRS) 过程来识别重要的预测变量. Li, Zhong and Zhu (2012) 介绍了一种基于距离相关 (DC) 的 SIS 过程, 该过程对极值重尾数据不具有稳健性. Li, Peng, Zhang and Zhu (2012) 基于 Kendall τ 相关关系提出了一种稳健的特征筛选过程. Mai and Zou (2013) 针对二元分类问题利用 Kolmogorov－Smirnov 统计量提出一种确定特征筛选方法. He et al. (2013) 基于边际效用的样条近似值提出了一种自适应分位数的非线性独立筛选过程. Mai and Zou (2015) 研究了一种基于带了切片技术融合的 Kolmogorov

filter 的确定特征筛选过程, 此过程在计算上非常耗时. Cui et al. (2015) 对超高维判别分析问题给出了一种均值方差筛选过程, 此方法仅仅适用于分类响应变量和连续的预测变量. Pan, Wang and Li (2016) 对带有发散维类和超高维预测变量的线性判别分析提出了两两确定筛选方法. Yan et al. (2018) 拓展了Cui et al. (2015) 的 MVS 过程, 通过引入切片技术使之适应于各种类型的预测变量, 连续型、离散型和分类的响应变量. Xie et al. (2019) 针对超高维异质性分类数据提出了一种自适应分类变量筛选过程. 然而, 上述所提到的无模型的特征筛选方法主要集中于完全观测数据.

在生物医学、社会学、临床试验、经济学、纵向研究等各个领域, 由于一些样本个体不愿意回答敏感性问题, 不可控制因素导致信息丢失, 一些定期的访问者间歇性地或退出研究等各种原因, 经常会出现数据缺失的情况 (Little and Rubin, 2002). 关于缺失数据的回归模型的变量选择, 已有相当多的文献. 例如, 参见 Ibrahim et al. (2008), Garcia et al. (2010), Long and Johnson (2015), Fang and Shao (2016), Zhao et al. (2017), Tang and Tang (2018) 等. 然而, 上述文献关注的是缺失数据的低维回归模型. 最近, 人们已经认识到在超高维缺失数据分析中, 越来越需要发展一些特征筛选方法来识别重要的预测变量. 例如, Song and Jeng (2014) 基于逆概率加权的方法将 Li et al. (2012) 的方法推广到删失数据中; Lai et al. (2017) 提出了一种采用逆概率加权 (IPW) 方法调整 (Zhu et al., 2011) 所提的确定独立秩筛选效用 (SIRS) 的无模型特征筛选方法. Tang et al. (2019) 基于截面边际估计方程和核插补技术, 针对响应变量随机缺失的超高维纵向数据部分线性模型提出了一种新的特征筛选方法. 然而, 这些方法严重依赖于倾向得分函数的正确假定和估计, 对模型的错误假定不具有稳健性. 因此, 发展一种高效、计算可行、稳健、无模型的特征筛选方法来区分超高维缺失数据的重要预测变量和不重要预测变量是十分必要的.

为此, 本章针对超高维响应变量随机缺失数据, 提出了一种基于带有非参数插补的调整 Spearman 秩相关关系 (ASRC) 的新的特征筛选方法. 与传统的参数化方法相比, 所提的非参数化方法具有以下优点: (i) 不需要为所考虑的响应变量和候选预测变量指定参数化模型, 也不需要为缺失数据机制模型指定参数化

模型; (ii) 即使存在异常值、重尾数据和模型错误假定, 包括对缺失数据机制模型和假设回归模型的错误假定, 仍然表现良好; (iii) 在活跃集和不活跃集的边际效用存在显著差异的情况下, 具有确定独立的筛选性质; (iv) 在一定的正则化条件下, 即使预测变量维数随样本量呈指数增长, 也具有秩相合性; (v) 可以研究响应变量和预测变量之间的非线性关系. (vi) 相对于直接推广 kendall's τ 相关关系, 此方法从理论和计算上都更容易实现.

本章的其余部分结构安排如下. 在第 2.2 节中, 介绍了 ASRC 在响应变量随机缺失情况下的筛选过程. 第 2.3 节在一定的正则条件下建立了确定的筛选和秩相合的理论性质. 第 2.4 节进行了模拟研究来说明所提筛选方法的表现. 第 2.5 节通过扩散性的大 B-细胞淋巴瘤实例进行举例说明. 第 2.6 节给出一些简要的讨论. 第 2.7 节理论证明.

2.2　筛选方法

2.2.1　调整的 Spearman 秩相关效用

令 Y 是一个连续型的响应变量, $\boldsymbol{X} = (X_1, \ldots, X_p)^T$ 是一个 $p \times 1$ 连续型的预测向量. 假设 $\boldsymbol{X}$ 是完全观测, Y 可能存在缺失. 令 δ 是一个响应变量 Y 缺失的指标器, i.e., 如果 Y 缺失, $\delta = 0$; 如果 Y 被观测到, $\delta = 1$. 在这里假设 δ 的值仅依赖于 $\boldsymbol{X}$ 使得倾向得分的形式为 $\pi(\boldsymbol{X}) = \Pr(\delta = 1|\boldsymbol{X})$. 依据 Little and Rubin (2002) 的讨论, 上面定义的缺失数据机制是随机缺失 (MAR). 在这一章中, 假设预测变量的维数 p 可能随着样本量 n 呈指数级增长, i.e., 对某些常数 $\alpha > 0$, $\log(p) = O(n^{\alpha})$.

对于响应变量 Y, 不假定预测变量 $\boldsymbol{X}$ 的任何回归形式. 主要目的是发现一种有效的方法来区分超高维缺失数据问题中的活跃预测变量和非活跃预测变量. 现有的无模型特征筛选方法都是基于完全观测数据, 不能直接应用于上述缺失数据中. 为了解决这个问题, 首先回顾响应变量没有缺失的情况下广泛使用的秩相关筛选过程.

在响应变量 Y 没有缺失的情况下, Fan and Lv (2008) 利用绝对边际 Pearson 相关关系 $E_{F_k(x,y)}(X_k Y) - E_{F_k(x)}(X_k)E_{F(y)}(Y)$ 对 X_k 和 Y 之间的线性关系进行排序, 它

仅能测量响应变量和预测变量之间的线性相关性, 其筛选过程是基于模型的, 其中 X_k 是向量 $\boldsymbol{X}$ 的第 k 个分量 (i.e., p 个预测变量中的第 k 个), $F_k(x,y) = P(X_k \le x, Y \le y)$ 是 X_k 和 Y 的联合分布函数, $k = 1,\ldots,p$, $F(y) = P(Y \le y)$ 是 Y 的边际分布函数, E_F 是关于分布函数 F 的期望. 为了处理这个问题, Li et al. (2012a, b) 采用边际秩相关系数 $E_{F_k(x,y)}\{F_k(X_k,Y)\} - 1/4$, 即 Kendall's τ 统计量, 对 X_k 和 Y 的相关性进行排序. Song et al. (2014) 基于逆概率加权的方法将 Li et al. (2012a, b) 的方法推广到删失数据中. 然而, 对缺失数据直接推广 Li et al. (2012a, b) 可能不是很有效, 因为响应变量随机缺失的情况下, $E_{F_k(x,y)}\{F_k(X_k,Y)\}$ 不能可靠地被估计, 尤其当缺失率很高的时候. 注意到采用非参的方法很容易估计 $E\{F(Y)|X_k,\delta = 1\}$ (e.g., see Cheng, 1994; Cheng and Chu, 1996; Tang et al., 2014). 受上述两个边际筛选效用和一些事实的启发, 这儿使用下面这个量 $\omega_k^0 = E\{F_k(X_k)F(Y)\} - 1/4$ 来衡量 X_k 和 Y 的相关性. 因为 $F_k(x,y) = F_k(x)F(y)$ 当且仅当 X_k 和 Y 独立.

然而, 当响应变量 Y 存在随机缺失时, 有 $E\{F_k(X_k)F(Y)\} = E[F_k(X_k)E\{F(Y)|X_k,\delta\}]$. 受 Tang et al. (2014) 的启发, 给定 δ, X_k, 通过它的边际条件期望插补 $F(Y)$, 即, $E\{F(Y)|X_k,\delta\} = \delta F(Y) + (1-\delta)E\{F(Y)|X_k,\delta = 1\}$. 因此, 给定 $k = 1,\ldots,p$, 定义下面的指标

$$\omega_k = E\big[\delta F_k(X_k)F(Y) + (1-\delta)F_k(X_k)E\{F(Y)|X_k,\delta = 1\}\big] - 1/4 \tag{2.1}$$

来测量当存在缺失时 X_k 和 Y 之间的相关性. 很容易验证在响应变量 Y 在 MAR 假设下, ω_k=0 当且仅当 X_k 和 Y 的分布是统计不相关. 特别地, 如果 X_k 和 Y 独立时, 有 ω_k=0.

注 2.1 令 $\{(\boldsymbol{X}_i, Y_i), i = 1,\cdots,n\}$ 是从总体 $(\boldsymbol{X}, Y)$ 抽取的 n 个随机样本. 当响应变量完全观测时, ω_k 的样本估计形式为 $\widehat{\omega}_k = n^{-3}\sum_{i=1}^n\{\sum_{j=1}^n I(X_{jk} \le X_{ik})\}\{\sum_{j=1}^n I(Y_j \le Y_i)\} - 1/4$. 令 $R_{ik} = \sum_{j=1}^n I(X_{jk} \le X_{ik})$, $Q_i = \sum_{j=1}^n I(Y_j \le Y_i)$. 记 $\bar{R}_k = n^{-1}\sum_{i=1}^n R_{ik}$, $\bar{Q} = n^{-1}\sum_{i=1}^n Q_i$. 因此, $\widehat{\omega}_k$ 可以重新记为 $\widehat{\omega}_k = n^{-3}\sum_{i=1}^n R_{ik}Q_i - 1/4$.

而 Spearman's 秩相关系数能表示为

$$\rho_k = \frac{\sum_{i=1}^{n}(R_{ik}-\bar{R}_k)(Q_i-\bar{Q})}{\sqrt{\sum_{i=1}^{n}(R_{ik}-\bar{R}_k)^2\sum_{i=1}^{n}(Q_i-\bar{Q})^2}} = 12\left\{\frac{1}{n(n^2-1)}\sum_{i=1}^{n}R_{ik}Q_i - \frac{1}{4}\frac{n+1}{n-1}\right\},$$

这就蕴含了 ρ_k 是 $\widehat{\omega}_k$ 的一个函数, 渐近收敛到 $12\omega_k$. 因此, 上面定义的秩相关关系称为一个调整的 Spearman 秩相关关系 (ASRC).

2.2.2　ω_k 的估计

在很多应用中, 分布函数 $F_k(x)$ $(k=1,\ldots,p)$ 和 $F(y)$ 通常是未知的. 因此, 直接利用上述定义的指标 ω_k 是不可能的或者说是非常困难的. 为了解决这个问题, 下面考虑 ω_k 的估计问题.

令 $\{(\boldsymbol{X}_i, Y_i, \delta_i), i=1,\cdots,n\}$ 是从总体 $(\boldsymbol{X}, Y, \delta)$ 中抽取的 n 个随机样本. 为了估计 $\mathscr{F}_{k1}(x) = E\{F(Y)|X_k = x, \delta = 1\}$, 依据 Cheng (1994) 的论证, 由于不需要假定响应变量和缺失数据机制的参数回归模型, 考虑基于 Nadaraya-Watson 核回归估计量的非参数插补方法, 这蕴含它不涉及模型的可识别性问题, 避免了模型的错误假定的问题, 对异常点和重尾分布是不敏感的. 为此, 令 $K(\cdot)$ 是一个对称核函数满足 $\int K(t)dt = 1$, $h = h_n$ 是一个带宽序列满足 $h_n \to 0$ 和当 $n \to \infty$, 有 $nh_n \to \infty$. 在 MAR 的假设下, 有 $\mathscr{F}_{k1}(x) = E\{F(Y)|X_k = x, \delta = 1\} = E\{F(Y)|X_k = x, \delta = 0\} = E\{F(Y)|X_k = x, \delta\}$. 因此, 依据 Cheng (1994), 在数据集 $\{(\boldsymbol{X}_i, Y_i, \delta_i), i=1,\ldots,n\}$ 上, 可通过极小化下面目标函数获得 $\mathscr{F}_{k1}(x)$ 的一个非参数回归估计量 $\widetilde{\mathscr{F}}_{k1}(x)$:

$$\sum_{i=1}^{n} K_h(X_{ik}, x)\delta_i\{F(Y_i) - \mathscr{F}_{k1}(x)\}^2 \tag{2.2}$$

它可以表示为

$$\widetilde{\mathscr{F}}_{k1}(x) = \sum_{i=1}^{n} W_{ik}^0(x)F(Y_i), \quad W_{ik}^0(x) = \frac{\delta_i K_h(X_{ik}, x)}{\sum_{j=1}^{n}\delta_j K_h(X_{jk}, x)}, \tag{2.3}$$

其中 $K_h(u, x) = h^{-1}K\{(u-x)/h\}$, X_{ik} 是 $\boldsymbol{X}_i$ 的第 k 个分量.

利用 (2.3) 评估 $\mathscr{F}_{k1}(x)$ 的非参数估计 $\widetilde{\mathscr{F}}_{k1}(x)$, 需要分别估计 X_k 和 Y 的分布函数 $F_k(x)$, $F(y)$. 一般地, $F_k(x)$ 可以通过它的经验分布函数来估计:

$\widehat{F}_k(x) = n^{-1}\sum_{i=1}^n I(X_{ik} \le x)$, 其中 $I(\cdot)$ 是一个指示函数. 于此同时, $F(y)$ 的一个相合估计的形式为

$$\widehat{F}(y) = \frac{1}{n}\sum_{i=1}^n \left\{\delta_i I(Y_i \le y) + (1-\delta_i)\sum_{j=1}^n W_{jk}^0(X_{ik}) I(Y_j \le y)\right\}.$$

因此, 调整后的 Spearman 秩相关关系的样本形式可以表示为

$$\widehat{\omega}_k = \frac{1}{n}\sum_{i=1}^n \left\{\delta_i \widehat{F}_k(X_{ik})\widehat{F}(Y_i) + (1-\delta_i)\widehat{F}_k(X_{ik})\widehat{\mathscr{F}}_{k1}(X_{ik})\right\} - 1/4 \tag{2.4}$$

$k = 1, \ldots, p$, 其中 $\widehat{\mathscr{F}}_{k1}(x) = \sum_{i=1}^n W_{ik}^0(x)\widehat{F}(Y_i)$. 注意 $\widehat{\omega}_k$ 对 X_k 和 Y 的任何单调变换都是不变的.

对于依据 (2.4) 中 $\widehat{\omega}_k$ 的估计值, 定义预测变量的活跃集为

$$\widehat{\mathcal{A}} = \{k : |\widehat{\omega}_k| \ge cn^{-\tau}, 1 \le k \le p\}, \tag{2.5}$$

其中 c 和 τ 是定义在条件 (C1) 中事先给定的阈值. 然而, 根据经验, 指定 c 和 τ 的值是相当困难的. 为了解决这个问题, 考虑以下标准来选取预测变量的活跃集:

$$\widehat{\mathcal{A}}_{d_n} = \{k : |\widehat{\omega}_k| \text{ 是前} d_n \text{个最大的}\}, \tag{2.6}$$

其中 $d_n < n$ 是预先确定的正整数.

注 2.2 在估计 X_k 和 Y 的分布函数 $F_k(x)$, $F(y)$ 时, 用到边际条件独立性 (MCI) (i.e., $Y \perp \delta | X_k$, $k = 1, \ldots p$). He et al. (2013) 和 Zhou and Zhu (2017) 也采用了类似的假设. 这个 MCI 假设实际上比 $Y \perp \delta | \boldsymbol{X}$ 强, 它是在低维数据分析中一个广泛使用的条件独立性. 采用 MCI 假设来推导我们想要的理论. 更重要的是, 它可以避免在存在超高维数据的情况下出现“维数诅咒”问题.

注 2.3 带宽 h_n 和核函数 $K(\cdot)$ 的选择在估计 $F(y)$, $F_k(x)$ 和 $\mathscr{F}_{k1}(x)$ 中起着重要的作用. 带宽的经典最优速率是 $h_n = n^{-1/5}$ (Tang et al., 2014). 但是在假定随着 $n \to \infty$, $nh_n^2 \to 0$ 的情况下, 最优速率 $h_n = n^{-1/5}$ 是不合适的 (Tang et al., 2014). 为此, Tang et al. (2014) 建议了一个适当而简单的带宽 $h_n = c\hat{\sigma}_x n^{-\kappa}$, 其中 c 是某个正常数, $\kappa > 0.5$ 是一个常数, $\hat{\sigma}_x$ 是来自总体 X_k 的观测值 $\{X_{ik}, i = 1, \ldots, n\}$ 的标

准差. 一般地, 在假设核函数 $K(\cdot)$ 是对称的, 满足 $\int K(t)dt = 1$, 可以简单地选取 $K(\cdot)$ 为高斯核.

2.3　理论性质

在本节中, 研究了所提的 ASRC 特征筛选方法的确定筛选和秩相合的性质.

在真实稀疏模型中, 预测变量的活跃集定义为

$$\mathcal{A} = \{1 \le k \le p : \text{对 } y,\ \ F(y|\boldsymbol{X}) \text{ 依赖于 } X_k \}.$$

记 $\mathcal{A}^c = \{1, 2, \ldots, p\}\backslash\mathcal{A}$ 为预测变量的非活跃集, $|\mathcal{A}|$ 是集合 $\mathcal{A}$ 元素的个数. 在超高维数据分析中, 假设 $p \gg n$, $p \gg |\mathcal{A}|$. 在下文中, 将证明预测变量中估计的活跃集 $\widehat{\mathcal{A}}$ 包含真正的活跃集 $\mathcal{A}$ 的概率趋于 1. 从上述预测变量的活跃集和非活跃集的定义中, 很容易可以看出如果 $\{X_k : k \in \mathcal{A}\}$ 与 $\{X_k : k \in \mathcal{A}^c\}$ 独立, ω_k 是区分预测变量的活跃集和非活跃集的一个有效指标, 因为对 $k \in \mathcal{A}$, $\omega_k > 0$; 当 $k \in \mathcal{A}^c$ 时, $\omega_k = 0$. 显然, 由于 ω_k 只依赖于条件分布函数和无条件分布函数, 因此所提出的 ASRC 特征筛选过程是无模型的, 并且可以用于筛选响应变量 Y 和预测变量 X_k 之间的线性和非线性依赖关系.

为了研究提出的 ASRC 特征筛选方法的理论性质, 需要以下正则化条件.

(C1) 存在一个正常数 $c > 0$ 使得对于任意 $0 \le \tau < 1/2$, 有 $\min_{k\in\mathcal{A}} |\omega_k| > 2cn^{-\tau}$ 成立.

(C2) 存在一个正常数 $c_1 > 0$ 使得 $\lim_{p\to\infty} \inf\left(\min_{k\in\mathcal{A}} |\omega_k| - \max_{k\notin\mathcal{A}} |\omega_k|\right) \ge c_1$.

(C3) 核函数 $K(\cdot)$ 是一个概率密度函数使得 (i) 它有界且有紧的支撑; (ii) 对称且 $\sigma^2 = \int t^2 K(t)dt < \infty$; (iii) 在以 0 为中心的闭区间上, 对常数 d_1, 有 $K(\cdot) \ge d_1$; (iv) 随着 $n \to \infty$, 有 $nh^2 \to 0$.

(C4) 对于 $k = 1, \cdots, p$, X_k 的概率密度函数 $f_k(x)$ 及其相应的二阶导数均在 X_k 的支持集 $\mathcal{R}_{X_k}$ 的范围内一致地远离 0 和无穷大.

(C5) 对于 $k = 1, \cdots, p$, $\pi(X_k)$=Pr($\delta = 1 \mid X_k = x$) 的一阶、二阶导数和 $F_k(y \mid x)$=Pr($Y \le y \mid X_k = x$) 在它们相应的支撑集 $\mathcal{R}_{X_k}$ 和 $\mathcal{R}_Y$ 上是一致有限的.

条件 (C1) 允许真实信号的最小值不能太小, 但是随着样本量的增大退化为零. 这个条件在 Cui et al.(2015) 中也使用过. 条件 (C2) 表明当 $k \in \mathcal{A}$ 有 $|\omega_k| \ne 0$,

当 $k \notin \mathcal{A}$ 有 $|\omega_k| = 0$. 也就是说, 采用条件 (C2) 能保证所提的特征筛选指标能够较好地识别总体水平上的活跃预测变量和非活跃预测变量. 条件 (C3)-(C5) 分别给出了关于概率密度函数 $f_k(x)$, 缺失概率函数 $\pi_k(x)$, 条件分布 $F_k(y \mid x)$ 和核函数 $K(\cdot)$ 的正则化条件, 这是使用分布函数和核函数的广泛使用的条件. 基本上, 上述所提的正则化条件比带有缺失数据的特征筛选文献 (Lai et al., 2017) 中的条件相对弱些.

定理 2.1 (i) (确定筛选性质) 假如条件 (C1) 和 (C3)–(C5) 成立. 则, 存在一个依赖于 (C1) 中常数 c 的一个正常数 b 使得

$$\Pr\left(\max_{1\le k\le p} |\widehat{\omega}_k - \omega_k| \ge cn^{-\tau}\right) \le O\{p(n+4)\exp(-bn^{1-2\tau})\}.$$

特别地, 有$\Pr(\mathcal{A} \subset \widehat{\mathcal{A}}) \ge 1 - O\{|\mathcal{A}|(n+4)\exp(-bn^{1-2\tau})\}$.

(ii) (秩相合性) 假如条件 (C2) 和 (C3)–(C5) 成立. 如果 $\log(p)/n = o(1)$, 因此有

$$\liminf_{n\to\infty}\left(\min_{k\in\mathcal{A}} |\widehat{\omega}_k| - \max_{k\in\mathcal{A}^c} |\widehat{\omega}_k|\right) > 0. \ \ a.s.$$

定理 2.1 建立了上述所提的特征筛选过程的确定筛选和秩相合性质. 确定筛选性质所需的条件比 Fan and Lv (2008) 和 Li et al. (2012) 所给的条件相对温和一些, 表现在 (i) 与 SIS 方法 (Fan and Lv, 2008) 相比, 它没有假设预测变量 $\boldsymbol{X}$ 对响应变量 Y 的回归模型; (ii) 与 DCS 过程 (Li et al. , 2012) 相比, 没有对预测变量的矩作一些假设. 因此, 本章提出的非参数特征筛选方法对异常值、重尾分布以及模型和缺失数据机制模型的错误假定具有较强的稳健性. 而且, 定理 2.1 中的确定筛选性质适用于松弛条件 $c = O(n^{-\psi})$ 其中 $0 < \psi < 2\tau$. 除此之外, 所提出的非参数筛选方法可用于处理 NP-维数问题, 即, $\log(p) = O(n^{\zeta})$, 其中 $\zeta < 1 - 2\tau$, $0 \le \tau < 1/2$, 它取决于最小真实信号强度. 如果这样, 有

$$\Pr\left(\max_{1\le k\le p} |\widehat{\omega}_k - \omega_k| \le cn^{-\tau}\right) \ge 1 - O\left(\Lambda\exp\{-\Lambda n^{1-2\tau} + \log(n+4)\}\right), \tag{2.7}$$

其中 Λ 是一个常数. 秩相合性表明活跃预测变量 $|\widehat{\omega}_k|$ 的值以很高的概率要比非活跃预测变量的值大. 因此, 根据 (2.6) 给出的准则, 通过选取恰当的阈值 d_n 能分离出活跃预测变量和不活跃预测变量.

定理 2.2　(控制错误发生率) 假如条件 (C1)–(C5) 都成立. 对 $\varsigma > 0$, 如果 $\sum_{k=1}^{p} |\omega_k| = O(n^{\varsigma})$, 则存在两个正常数 b_0 , c' 使得

$$\Pr\left(|\widehat{\mathcal{A}}| \leq c' n^{\tau+\varsigma}\right) \geq 1 - O(n)p\exp(-b_0 n^{1-2\tau}).$$

定理 2.2 蕴含了可以用多项式单位来指定一个估计的模型大小. 然而, 这个控制假阳率的结论对于特征筛选是保守的, 因为假阳率越小, 假阴率越大. 因此, 在很多实际应用中, 经常采用定义在 (2.6) 中的选择准则通过预先指定的模型大小 d_n 来评估所估计的子模型 $\widehat{\mathcal{A}}_{d_n}$. 另一方面, 我们注意到, 当响应变量存在缺失, 缺失比例和缺失模型对所提的 ASRC 过程的确定筛选性质没有影响. 这个事实表明所得到的结论与没有缺失数据的情况下所得的结论相同. 定理 2.2 蕴含了如果基于样本 ASRC 的值按次序选取前 d_n 个变量, $d_n = \lfloor n^{\nu+\tau+\varsigma}/\log n \rfloor$, $\nu > 0$, 那么所有相关的变量以很高的概率被选取. 特别地, 取 $\nu = 1 - \tau - \varsigma$, 其中 $\varsigma < 1 - \tau$, 硬阈值减少为 $d_n = \lfloor n/\log(n) \rfloor$, 这个阈值被 Fan and Lv (2010), Mai and Zou (2015), Lai et al. (2017) 和 Ma et al. (2017) 使用过, 其中 $\lfloor a \rfloor$ 表示 a 的整数部分. 通常, d_n 的选取反映了研究人员对潜在的活跃预测变量个数或者预算限制的先验信息.

2.4　模拟研究

在本节中, 进行了三个模拟实验来研究所提出的调整的 Spearman 秩相关筛选方法 (简称“ASRC”) 在有限样本上的表现. 为了比较, 考虑下面的筛选方法: (A) 逆概率加权的确定独立秩筛选方法 (简称“IPW-SIRS”) (Lai et al., 2017); (B) 完全观测数据的融合均值-方差过滤筛选方法 (简称 “FMV”) (Yan et al., 2018); (C) 完全观测数据的 Spearman 秩相关筛选方法 (简称“SRC”); (D) 多重插补的确定独立筛选方法 (简称“MI-SIS”) (Fan et al., 2008); (E) 多重插补的最大边际似然估计 (简称“MI-MMLE”) (Fan et al., 2010). 这里使用估计的最小模型大小 $\mathcal{S}$ (即, 包含所有真实的活跃预测变量的所选活跃预测变量的最小模型个数) 衡量六种所考虑的筛选方法的有效性. 模拟给出了 200 次重复试验 $\mathcal{S}$ 的中位数、标准差 (SD) 和四分位间距 (IQR). 而且, 还给出了 200 次重复试验中前 $d_n = \lfloor n/\log(n) \rfloor$ 个估计的预测变量包含真实活跃集的经验覆盖概率.

实验 1 (线性回归模型). 考虑下面的线性回归模型:

$$Y_i = \boldsymbol{X}_i^T\boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n,$$

其中 $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})^T$ 是一个 $p \times 1$ 维的预测向量, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ 是一个 $p \times 1$ 维回归系数向量, ε_i 是测量误差, 与 $\boldsymbol{X}_i$ 独立. 在这个实验中, $\boldsymbol{X}_i$ 是由均值为 $\boldsymbol{\mu}$ (即, $\boldsymbol{\mu} = (0, \ldots, 0)^T$), 协方差矩阵为 $\boldsymbol{\Sigma} = (\sigma_{ij})_{p\times p}$ 的多元正态分布产生, 其中 $\sigma_{ij} = \rho^{|i-j|}$. $\boldsymbol{\beta}$ 的真值取 $\boldsymbol{\beta} = (1.5, 2.0, 2.5, 3.0, 0.0, \ldots, 0.0)^T$, 这表明前四个预测变量是活跃的, 后 $p-4$ 个预测变量是非活跃的. ρ 的真实值 $\rho = 0.2, 0.5, 0.8$ 分别代表低, 中, 高正相关. 设置 $(n, p) = (100, 1000)$, 这表明 $p \gg n$. 在这个实验中, 假设所有的 $\boldsymbol{X}_i$ 是完全观测的, 而 Y_i 容易缺失. 为了研究所提的特征筛选方法对异常值和重尾数据的稳健性, 考虑了以下三种模拟设置.

Case (A). $\varepsilon_i \sim \mathcal{N}(0, 1)$, $i = 1, \ldots, n$.

Case (B). $\varepsilon_i \sim \mathcal{N}(0, 1)$, $i = 1, \ldots, n$. 5% 的异常点的构造如下. 随机的从 100 个样本中抽取 5 个, $i = 1, \ldots, 5$, X_{i1} 和 X_{i3} 是由自由度为 3 的 t 分布中产生 (即, $X_{ik} \sim t(3)$ for $k = 1, 3$), 和 X_{i2} 是从 $t(3) + 1$ 中产生. 这个设置是为了研究所提的筛选过程对异常值的稳健性.

Case (C). ε_i 是从 t 分布中产生: $\varepsilon_i \sim t(3)$, 这表明响应变量服从重尾分布. 如此设置是为了研究该方法对重尾数据的稳健性.

为了产生缺失数据, 考虑下面四种倾向得分函数:

Case (M1). $\Pr(\delta_i = 1|\boldsymbol{X}_i = \boldsymbol{x}_i) = \exp(3 + 2x_{i1})/\{1 + \exp(3 + 2x_{i1})\}$;

Case (M2). 如果 $|x_{i1} - 1| \le 1$, $\Pr(\delta_i = 1|\boldsymbol{X}_i = \boldsymbol{x}_i) = 0.2 + 0.1|x_{i1} - 1|$, 否则 0.9;

Case (M3). $\Pr(\delta_i = 1|\boldsymbol{X}_i = \boldsymbol{x}_i) = \exp(1.2 + 0.8x_{i1} + 0.3x_{i2} + 0.1x_{ip})/\{1 + \exp(1.2 + 0.8x_{i1} + 0.3x_{i2} + 0.1x_{ip})\}$;

Case (M4). $\Pr(\delta_i = 1|\boldsymbol{X}_i = \boldsymbol{x}_i, Y_i = y_i) = \exp(1.5 + 1.2x_{i1} + 0.5x_{i5} + 0.001y_i)/\{1 + \exp(1.5 + 1.2x_{i1} + 0.5x_{i5} + 0.001y_i)\}$.

缺失机制模型 (M1) 和 (M3) 表示由 logistic 回归模型指定的 MAR 机制. 更重要的是, 缺失机制模型 (M3) 不仅取决于两个活跃的预测变量 X_1 和 X_2, 也取决于非活跃的预测变量 X_p, 这个设置为了研究在不满足边际条件独立性的情况下,

表 2.1: 实验 1 中四种缺失机制下的六种筛选方法的结果

			$\rho = 0.2$					$\rho = 0.5$				
PS	Case	Method	Median	IQR	Mean	SD	CP	Median	IQR	Mean	SD	CP
M1	A	FMV	8	27	43.02	100	0.67	4	1	4.98	3.23	0.98
		IPW-SIRS	6	9.25	33.31	87.6	0.81	4	0	4.68	3.21	0.99
		SRC	5	8	16.98	36.4	0.84	4	0	4.33	1.04	1.00
		ASRC	4	4.25	15.75	38.9	0.88	4	0	4.22	0.91	1.00
		MI-SIS	5	14	32.54	91.60	0.77	4	1	5.45	8.00	0.99
		MI-MMLE	5	11	27.69	78.79	0.82	4	0	4.97	5.11	0.98
	B	FMV	16	80	79.21	147	0.59	4	2	14.43	67.7	0.92
		IPW-SIRS	8	27.25	51.84	115	0.70	4	1	7.57	21.8	0.98
		SRC	6	9.25	23.59	50.6	0.79	4	1	5.84	11.9	0.98
		ASRC	5	8.25	20.13	51.2	0.83	4	0	5.77	9.17	0.98
		MI-SIS	5	10.25	32.85	94.72	0.82	4	1	6.90	18.82	0.98
		MI-MMLE	6	16	34.02	98.19	0.78	4	1	6.71	14.74	0.97
	C	FMV	11	41.25	55.82	117	0.64	4	1	7.38	18.9	0.98
		IPW-SIRS	6	17.25	32.08	67.8	0.76	4	1	5.47	7.86	0.98
		SRC	5	9	20.74	40.7	0.82	4	0	4.54	2.09	0.99
		ASRC	5	6	16.88	35.9	0.86	4	0	4.36	1.93	0.99
		MI-SIS	6	11.25	30.26	77.10	0.80	4	1	5.17	5.35	0.99
		MI-MMLE	6	12.25	27.33	61.53	0.80	4	0	4.49	2.12	1.00
M2	A	FMV	32.5	99	99.77	163	0.43	5	5	24.34	68.6	0.85
		IPW-SIRS	12	41.75	78.35	181	0.62	4	1	12.2	38.8	0.93
		SRC	11	25.25	36.92	67.4	0.66	4	1	9.76	21.2	0.93
		ASRC	7	17.25	32.05	77.3	0.76	4	0	6.48	12.3	0.97
		MI-SIS	12.5	63.25	74.10	153.51	0.61	4	1	7.08	13.36	0.97
		MI-MMLE	10	41.25	66.89	145.56	0.62	4	1	5.43	5.42	0.98
	B	FMV	27.5	86	89.6	156	0.48	5	7	23.76	83.7	0.82
		IPW-SIRS	13.5	43.75	62.72	133	0.6	4	2	19.32	87.4	0.93
		SRC	10	23	35.64	73.3	0.72	4	2	12.24	46.8	0.93
		ASRC	6	13.25	26.64	66	0.80	4	1	7.9	28.4	0.96
		MI-SIS	12	37.75	67.96	156.40	0.63	4	2	11.14	33.21	0.94
		MI-MMLE	15.5	58.75	75.86	159.20	0.57	5	3	12.83	37.87	0.91
	C	FMV	23	88.25	91.79	155	0.49	5	4	18.13	45.5	0.86
		IPW-SIRS	14.5	49.5	56.68	103	0.58	4	1	13.39	63	0.95
		SRC	8	26	38.50	88	0.69	4	1	7.33	10.8	0.96
		ASRC	6	16	25.71	61.9	0.76	4	0.25	5.76	6.25	0.97
		MI-SIS	16	65.25	73.52	138.75	0.57	4	2	11.73	31.32	0.93
		MI-MMLE	17	62	73.94	136.27	0.57	4	2	11.15	36.39	0.94

表 2.1: (续) 实验 1 中四种缺失机制下的六种筛选方法的结果

			$\rho = 0.2$					$\rho = 0.5$				
PS	Case	Method	Median	IQR	Mean	SD	CP	Median	IQR	Mean	SD	CP
M3	A	FMV	12	57.25	51.23	92.74	0.60	4	1.25	7.55	13.71	0.95
		IPW-SIRS	7	23.5	42.47	97.85	0.72	4	1	5.3	5.17	0.99
		SRC	6	13	21.18	43.26	0.80	4	1	4.83	3.04	0.99
		ASRC	6	12.25	19.96	36.96	0.80	4	1	4.83	2.82	1.00
		MI-SIS	7	21.5	40.4	104.91	0.73	4	1	6.22	11.37	0.98
		MI-MMLE	7	21.25	36.39	90.51	0.73	4	1	5.63	6.34	0.97
	B	FMV	22.5	73.5	88.35	162.37	0.50	4	2	15.2	54.46	0.93
		IPW-SIRS	10.5	41.25	61.01	134.25	0.65	4	2	7.15	19.10	0.97
		SRC	8	24	36.02	78.80	0.72	4	1	6.41	16.82	0.98
		ASRC	8	23.25	35.87	74.53	0.72	4	1	6.40	11.51	0.97
		MI-SIS	7	18	39.71	96.48	0.76	4	1	7.27	16.26	0.97
		MI-MMLE	8.5	25.25	43.14	98.77	0.70	4	2	8.92	24.26	0.96
	C	FMV	16	67	87.66	165.09	0.56	4	3.25	19.67	64.04	0.86
		IPW-SIRS	9	33.25	53.36	119.62	0.67	4	1	8.67	15.81	0.94
		SRC	6	16.25	32.75	65.10	0.77	4	1	6.74	10.16	0.97
		ASRC	6	14	29.31	58.14	0.80	4	1	6.00	7.21	0.97
		MI-SIS	8	24	39.41	98.42	0.72	4	1	9.87	29.00	0.95
		MI-MMLE	8	17.5	36.6	90.34	0.75	4	1	7.71	19.72	0.95
M4	A	FMV	17.5	64	73.62	128.05	0.54	4	1.25	9.58	48.64	0.98
		IPW-SIRS	9	30.25	40.61	78.07	0.71	4	1	5.03	4.17	0.99
		SRC	7	15.25	26.79	70.18	0.78	4	0	4.97	6.05	0.99
		ASRC	6	11	26.04	67.45	0.81	4	0	4.89	5.50	1.00
		MI-SIS	7	16	29.20	70.86	0.77	4	1	6.45	9.80	0.96
		MI-MMLE	7	17	24.94	48.45	0.78	4	1	6.1	10.34	0.98
	B	FMV	19.5	83	98.51	179.91	0.54	4	3	17.55	55.61	0.89
		IPW-SIRS	9	38.25	70.48	152.63	0.68	4	1	8.15	18.70	0.96
		SRC	7	25.25	42.06	110.83	0.72	4	1	8.22	20.19	0.96
		ASRC	6.5	16.25	41.78	98.27	0.76	4	1	7.6	16.5	0.96
		MI-SIS	7	23.25	43.12	98.10	0.72	4	1	9.53	26.21	0.94
		MI-MMLE	10	32.25	48.01	113.51	0.66	4	2	9.47	26.14	0.96
	C	FMV	17	77.25	86.14	163.50	0.55	4	2	9.63	21.26	0.92
		IPW-SIRS	10	37.25	51.25	119.31	0.67	4	1	6.33	13.63	0.99
		SRC	7	17	29.98	70.92	0.76	4	1	4.99	2.87	0.99
		ASRC	6	17	29.54	68.18	0.76	4	0	4.83	2.79	0.99
		MI-SIS	8	21.25	46.84	116.64	0.72	4	1	7.23	12.64	0.96
		MI-MMLE	7	22	43.98	107.36	0.72	4	1	6.18	8.93	0.98

表 2.1: (续) 实验 1 中四种缺失机制下的六种筛选方法的结果

			$\rho = 0.8$						$\rho = 0.8$				
PS	Case	Method	Median	IQR	Mean	SD	CP	PS	Median	IQR	Mean	SD	CP
M1	A	FMV	4	0	4.27	0.54	1.00	M3	4	1	4.33	0.57	1.00
		IPW-SIRS	4	0	4.24	0.48	1.00		4	0	4.29	0.56	1.00
		SRC	4	0	4.45	0.49	1.00		4	0	4.26	0.51	1.00
		ASRC	4	0	4.18	0.43	1.00		4	0	4.20	0.44	1.00
		MI-SIS	4	1	4.31	0.53	1.00		4	1	4.34	0.61	1.00
		MI-MMLE	4	0	4.18	0.40	1.00		4	0	4.18	0.40	1.00
	B	FMV	4	1	4.37	0.63	1.00		4	1	4.36	0.56	1.00
		IPW-SIRS	4	1	4.29	0.53	1.00		4	1	4.33	0.58	1.00
		SRC	4	1	4.33	0.53	1.00		4	1	4.32	0.52	1.00
		ASRC	4	0	4.21	0.45	1.00		4	0	4.26	0.51	1.00
		MI-SIS	4	1	4.33	0.60	1.00		4	1	4.40	0.72	1.00
		MI-MMLE	4	1	4.81	4.30	1.00		5	1	6.19	20.62	1.00
	C	FMV	4	1	4.28	0.47	1.00		4	1	4.33	0.59	1.00
		IPW-SIRS	4	1	4.3	0.48	1.00		4	1	4.34	0.60	1.00
		SRC	4	1	4.27	0.47	1.00		4	0	4.30	0.57	1.00
		ASRC	4	0	4.17	0.38	1.00		4	0	4.26	0.52	1.00
		MI-SIS	4	1	4.35	0.71	1.00		4	1	4.34	0.55	1.00
		MI-MMLE	4	0	4.14	0.37	1.00		4	0	4.23	0.47	1.00
M2	A	FMV	4	1	4.47	1.09	1.00	M4	4	0	4.28	0.55	1.00
		IPW-SIRS	4	0	4.2	0.57	1.00		4	0	4.25	0.50	1.00
		SRC	4	1	4.40	0.72	1.00		4	0.25	4.28	0.51	1.00
		ASRC	4	0	4.17	0.5	1.00		4	0	4.21	0.47	1.00
		MI-SIS	4	0.25	4.31	0.59	1.00		4	1	4.34	0.60	1.00
		MI-MMLE	4	0	4.23	0.45	1.00		4	0	4.17	0.40	1.00
	B	FMV	4	1	5.08	3.52	1.00		4	1	4.33	0.61	1.00
		IPW-SIRS	4	1	4.36	0.69	1.00		4	0	4.33	0.89	1.00
		SRC	4	1	4.72	1.16	1.00		4	0	4.27	0.56	1.00
		ASRC	4	1	4.31	0.7	1.00		4	0	4.25	0.48	1.00
		MI-SIS	4	1	4.46	0.84	1.00		4	0	7.27	41.70	1.00
		MI-MMLE	4	1	5.11	3.71	1.00		4	1	5.69	17.04	1.00
	C	FMV	4	1	4.53	1.59	1.00		4	1	4.34	0.57	1.00
		IPW-SIRS	4	0	4.18	0.53	1.00		4	1	4.31	0.56	1.00
		SRC	4	1	4.43	0.65	1.00		4	1	4.33	0.56	1.00
		ASRC	4	0	4.16	0.43	1.00		4	0	4.24	0.50	1.00
		MI-SIS	4	1	4.37	0.65	1.00		4	1	4.38	0.62	1.00
		MI-MMLE	4	0	4.26	0.47	1.00		4	0	4.24	0.46	1.00

所提出的筛选方法对不满足 MCI 假设下的稳健性. 缺失机制模型 (M2) 也是一种 MAR 机制, 由分段函数而不是参数回归模型来指定. 缺失机制模型 (M4) 是一种 NMAR 机制, 包括活动预测变量 X_1、非活动预测变量 X_5 和缺失响应变量 Y, 用于研究该方法对错误指定的倾向评分函数的稳健性. 与缺失机制模型 (M1), (M2), (M3), (M4) 相应的响应变量的平均缺失率分别为 13.2%, 41.5%, 26.5%, 24.2%.

结合四种缺失数据机制和三种与 $\boldsymbol{X}_i$ 和 ε_i 相关指定的分布共有 12 种设置, 针对每一种设置重复 200 次的每一次, 采用本章所提的特征筛选过程和五种现有的特征筛选方法去识别活跃和非活跃预测变量. 为了实现所提的筛选过程, 取高斯核函数 $K(u) = \exp(-u^2/2)/(2\pi)^{1/2}$, 带宽 h 设置为 $h = 1.06\hat{\sigma}_x n^{-2/3}$, 其中 $\hat{\sigma}_x$ 是 X_k 的样本标准差.

结果见表 2.1. 观察表 2.1 可以得到 (i) 针对所有的设置, 所提的 ASRC 筛选方法在六种特征筛选过程中表现最好, 尤其是, 在出现重尾数据和异常点时, 它的中位数与真实的活跃预测变量的个数越接近, IQR, 均值和方差都几乎是最小的, CP 值也是所有六种特征筛选方法中最大的; (ii) 当预测变量的相关性很大 (例, $\rho = 0.8$), 缺失概率很小时 (例, M1), 即使存在异常点和重尾数据, 六种特征筛选方法都表现不错; (iii) 当预测变量的相关性很小 (例, $\rho = 0.2$), FMV, IPW-SIRS, SRC, MI-SIS 和 MI-MMLE 筛选方法都表现比较差, 对缺失数据机制很敏感; (iv) 当缺失机制错误指定时, 所提的 ASRC 筛选方法比其他五种筛选方法表现都要好. 上述发现的结果可以通过下面的事实解释, FMV 和 SRC 方法仅仅使用了完全观测数据; IPW-SIRS 方法包括了倾向得分的估计; MI-SIS 和 MI-MMLE 方法在给定 $X = X_k$ 的情况下, 需要从 Y 的条件分布估计中重复采样观察值, 并且严重依赖于指定的模型; 而 ASRC 方法是无模型的.

实验 2 (非线性回归模型). 考虑下面响应变量 Y_i 和预测变量 $\boldsymbol{X}_i$ 之间的非线性关系:

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 \sin(X_{i3}) + \beta_4 X_{i4}^3 + \beta_5 X_{i5} + \ldots + \beta_p X_{ip} + \varepsilon_i, \quad i = 1, \ldots, n,$$

其中 $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})^T$ 是一个 $p \times 1$ 维预测向量, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ 是一个

$p \times 1$ 维回归系数向量, ε_i 是测量误差, 与 $\boldsymbol{X}_i$ 独立. 上述模型假设前四个预测变量是活跃的, 其他预测变量是非活跃的. 在这个实验中, $\boldsymbol{X}_i$ 的产生与实验 1 相同, 其中 $\rho = 0.2, 0.5, 0.8$, $(n, p) = (100, 1000)$, $\boldsymbol{\beta}$ 的真值设置为 $\boldsymbol{\beta} = (2.0, 2.0, 3.0, 4.0, 0.0, \ldots, 0.0)^T$. 同样, 假设所有的 $\boldsymbol{X}_i$ 是完全观测的但 Y_i 有缺失. 为了研究所提筛选过程对异常值和重尾数据的稳健性, ε_i 采用与实验 1 相同的设置. 缺失响应机制仍采用与实验 1 相同的设置除了 (M3), (M4) 被下面 (M3'), (M4') 取代:

Case (M3') $\Pr(\delta_i = 1|\boldsymbol{X}_i = \boldsymbol{x}_i) = \exp(1.5 + 0.8x_{i1} + 0.5x_{i2} + 0.2x_{ip})/\{1 + \exp(1.5 + 0.8x_{i1} + 0.5x_{i2} + 0.2x_{ip})\}$;

Case (M4') $\Pr(\delta_i = 1|\boldsymbol{X}_i = \boldsymbol{x}_i, Y_i = y_i) = \exp(1.8 + 0.001y_i)/\{1 + \exp(1.8 + 0.001y_i)\}$.

缺失机制模型 (M3') 与实验 1 中 (M3) 有相同的解释. 而缺失机制模型 (M4') 与 (M4) 不同, 虽然仍然是不可忽略缺失但是仅仅依赖于响应变量. 与缺失机制模型 (M1), (M2), (M3'), (M4') 相应的响应变量的平均缺失率分别为 13.2%, 41.5%, 24.5%, 14.5%.

基于 200 次重复试验的六种特征筛选过程相应的结果见表 2.2. 仔细观察表 2.2 可以看出 (i) 无论预测变量的相关系数是低, 中还是高相关, 所提出的 ASRC 特征筛选过程对异常点和重尾数据是稳健的; (ii) 当相关系数 ρ 是低或者中相关, FMV, IPW-SIRS 和 SRC 特征筛选过程对异常点和重尾数据很敏感; (iii) 无论是否正确指定倾向得分函数, 所提出的 ASRC 特征筛选方法都优于其他五种方法, 因为前者的中位数比后者更接近真实活跃预测变量的个数, 前者的均值、方差和 IQR 值都均小于后者, 但前者的 CP 大于后者; (iv) IPW-SIRS 过程对于错误指定的倾向得分函数很敏感, 因为需要估计倾向得分函数; (v) 无论预测变量的相关系数 ρ 是高还是低, MI-SIS 和 MI-MMLE 方法都表现较差, 因为它们是参数模型, 它们的表现严重地依赖于假设的工作模型与真实模型接近的程度.

表 2.2: 实验 2 中四种缺失机制下的六种筛选方法的结果

			$\rho = 0.2$					$\rho = 0.5$				
PS	Case	Method	Median	IQR	Mean	SD	CP	Median	IQR	Mean	SD	CP
M1	A	FMV	9	35	44.63	91.38	0.67	4	1	7.67	13.09	0.94
		IPW-SIRS	321	561.5	388.1	317.05	0.08	15	66.75	83.72	159.83	0.57
		SRC	8	17	22.81	37.29	0.75	4	1	5.87	5.94	0.98
		ASRC	7	16	22.67	45.11	0.77	4	1	5.6	6.7	0.99
		MI-SIS	276	477	346.04	279.45	0.07	15	54.5	68.86	129.15	0.56
		MI-MMLE	280	471	346.46	282.13	0.06	14	54.25	63.62	124.19	0.62
	B	FMV	13	55.25	65.77	133.73	0.62	4	1.25	13.01	46.94	0.93
		IPW-SIRS	385	595.25	421.2	316.65	0.09	29	148	119.3	185.96	0.50
		SRC	8	20	28.76	58.03	0.74	4	1	7.22	12.22	0.97
		ASRC	7	15.25	25.52	55.17	0.78	4	1	6.7	11.21	0.97
		MI-SIS	238.5	348.25	299.98	259.85	0.11	16	66.25	76.59	135.91	0.57
		MI-MMLE	252.5	371.5	318.56	260.18	0.07	20	76.25	81.03	143.58	0.52
	C	FMV	11	40	59.21	121.77	0.65	4	1	6.92	15.34	0.97
		IPW-SIRS	274	526.75	360.6	301.34	0.08	15.5	62	81.12	153.55	0.57
		SRC	8	20	33.4	63.56	0.72	4	1	5.82	7.82	0.98
		ASRC	7	16.25	28.94	59.53	0.77	4	1	5.34	5.41	0.99
		MI-SIS	242	355.25	305.72	256.21	0.09	16	58.25	79.18	149.17	0.58
		MI-MMLE	252.5	374	301.70	249.41	0.09	14.5	44.5	75.53	147.71	0.6
M2	A	FMV	37	112	119.5	184.06	0.43	5	6	27.82	75.33	0.85
		IPW-SIRS	392.5	598.75	433.1	323.77	0.07	14	66.5	92.52	190.49	0.57
		SRC	20	65	59.31	91.56	0.53	5	3	14.64	35.51	0.89
		ASRC	13	40.75	49.84	89.73	0.61	4	1	9.75	30.46	0.95
		MI-SIS	378.5	611.5	418.84	317.26	0.04	21.5	105.25	100.54	172.50	0.51
		MI-MMLE	364	537.75	406.13	306.22	0.04	21	95	98.89	179.81	0.51
	B	FMV	50.5	88.25	109.3	161.64	0.35	5	11	29.15	86.51	0.82
		IPW-SIRS	356.5	578.5	437.4	317.56	0.05	21	105.75	105.5	194.73	0.51
		SRC	24.5	61.25	63.78	96.40	0.48	5	5	14.56	41.03	0.90
		ASRC	14	43	57.02	110	0.62	4	2	9.39	28.33	0.93
		MI-SIS	378.5	516.25	408.12	302.61	0.04	37	89.5	115.33	192.39	0.44
		MI-MMLE	364.5	578.75	427.06	318.19	0.03	40.5	119.5	128.05	202.81	0.38
	C	FMV	38	122.75	104.4	157.41	0.4	5	5.25	17.52	35.89	0.84
		IPW-SIRS	337	547.25	410.6	314.03	0.09	27	94.25	107.6	196.74	0.47
		SRC	21	60.25	58.59	98.32	0.54	5	4	10.27	17.13	0.9
		ASRC	17.5	37.25	50.64	86.13	0.57	4	2	8.13	13.50	0.94
		MI-SIS	365	547.5	411.54	308.13	0.05	29	102.75	106.09	178.18	0.45
		MI-MMLE	346.5	602.25	413.75	315.12	0.04	30.5	106	102.25	168.49	0.44

表 2.2: (续) 实验 2 中四种缺失机制下的六种筛选方法的结果

			$\rho = 0.2$					$\rho = 0.5$				
PS	Case	Method	Median	IQR	Mean	SD	CP	Median	IQR	Mean	SD	CP
M3’	A	FMV	14	43.25	50.37	90.52	0.61	4	2	16.39	66.00	0.90
		IPW-SIRS	372.5	553	407.79	308.69	0.06	5	1	5.34	1.53	1.00
		SRC	9	22	31.86	61.31	0.71	5	2	9.39	26.01	0.94
		ASRC	9	19	32.04	60.82	0.74	4	2	11.27	28.76	0.94
		MI-SIS	322	474	387.36	287.18	0.03	19	56.5	65.08	115.06	0.57
		MI-MMLE	315.5	469.25	385.58	287.30	0.03	16.5	51.5	66.63	124.07	0.57
	B	FMV	16	77.5	70.66	127.12	0.55	4	2	14.35	49.15	0.91
		IPW-SIRS	342	526.75	401.38	297.85	0.07	27.5	129.5	122.37	195.79	0.46
		SRC	14	36	49.51	97.63	0.62	4.5	2	6.94	8.38	0.96
		ASRC	12.5	37.25	47.74	86.21	0.63	4	2	6.72	7.43	0.96
		MI-SIS	272	446.25	332.24	267.17	0.05	21	86.25	111.99	196.61	0.51
		MI-MMLE	299.5	431.25	354.47	273.11	0.04	35.5	116.5	116.88	188.08	0.39
	C	FMV	21.5	60.25	68.78	129.65	0.51	4	2	12.54	31.92	0.90
		IPW-SIRS	363	530.75	406.12	304.55	0.07	15.5	68.25	85.08	158.62	0.56
		SRC	14	36.25	43.55	92.05	0.62	4	1	10.3	32.81	0.96
		ASRC	13	34	46.44	86.39	0.63	4	1	9.17	28.28	0.96
		MI-SIS	346.5	493.5	410.46	295.82	0.03	28	123.25	113.4	179.67	0.46
		MI-MMLE	366.5	453.75	409.97	286.62	0.06	27	108	107.27	174.34	0.48
M4’	A	FMV	7	26.25	34.25	63.33	0.71	4	0	4.84	3.53	0.99
		IPW-SIRS	250.5	568.75	358.28	313.62	0.14	7.5	27.75	66.30	156.99	0.70
		SRC	6	9	18.64	32.05	0.81	4	1	4.87	2.97	0.99
		ASRC	6	8	18.48	30.83	0.81	4	1	4.85	3.49	0.99
		MI-SIS	188	411	275.44	254.00	0.1	11	39	59.46	114.65	0.64
		MI-MMLE	182	384	273.81	254.21	0.09	10	39.25	57.31	112.24	0.63
	B	FMV	8.5	36.5	47.95	105.64	0.65	4	1	6.16	11.68	0.98
		IPW-SIRS	240	585.25	368.55	318.23	0.12	13.5	60.25	70.64	133.28	0.60
		SRC	7	16	25.07	48.88	0.76	4	1	5.03	3.23	0.99
		ASRC	6.5	15.25	25.85	51.51	0.76	4	1	5.03	3.22	0.99
		MI-SIS	237.5	417.25	323.88	287.11	0.13	10	29	64.99	137.63	0.69
		MI-MMLE	275.5	467.25	355.29	285.79	0.09	13	55.25	75.01	142.38	0.61
	C	FMV	7	31.75	41.72	91.88	0.71	4	1	5.60	5.94	0.98
		IPW-SIRS	243	477.5	339.33	300.99	0.12	14.5	62	74.74	147.95	0.59
		SRC	7	14	23.25	6.67	0.79	4	1	5.56	7.11	0.99
		ASRC	7	15	22.56	47.72	0.80	4	1	5.31	4.58	0.99
		MI-SIS	215.5	445.75	314.57	281.35	0.09	13	42	57.71	122.93	0.64
		MI-MMLE	202	444.5	317.53	289.33	0.07	12	35	54.74	116.41	0.62

表 2.2: (续) 实验 2 中四种缺失机制下的六种筛选方法的结果

			$\rho = 0.8$						$\rho = 0.8$				
PS	Case	Method	Median	IQR	Mean	SD	CP	PS	Median	IQR	Mean	SD	CP
M1	A	FMV	4	1	4.32	0.55	1.00	M3’	4	1	4.39	0.62	1.00
		IPW-SIRS	5	1	5.12	0.82	1.00		5	1	5.34	1.53	1.00
		SRC	4	1	4.39	0.59	1.00		4	1	4.48	0.70	1.00
		ASRC	4	0	4.27	0.53	1.00		4	1	4.37	0.59	1.00
		MI-SIS	5	2	5.18	1.24	1.00		5	2	5.41	1.35	1.00
		MI-MMLE	5	0	5.17	1.24	1.00		5	1	5.32	1.22	1.00
	B	FMV	4	1	4.50	0.67	1.00		4	1	4.41	0.63	1.00
		IPW-SIRS	5	1	5.81	2.78	0.99		5	1	5.43	1.51	1.00
		SRC	4	1	4.54	0.66	1.00		4	1	4.52	0.68	1.00
		ASRC	4	1	4.48	0.66	1.00		4	1	4.46	0.62	1.00
		MI-SIS	5	2	5.30	1.17	1.00		5	1	6.89	9.89	0.99
		MI-MMLE	5	1	7.07	8.17	0.97		6	2	14.39	53.74	0.94
	C	FMV	4	1	4.32	0.53	1.00		4	1	4.39	0.60	1.00
		IPW-SIRS	5	1	5.12	0.84	1.00		5	1.25	5.15	1.09	1.00
		SRC	4	1	4.4	0.57	1.00		4	1	4.49	0.64	1.00
		ASRC	4	1	4.33	0.54	1.00		4	1	4.41	0.61	1.00
		MI-SIS	5	2	5.30	1.75	1.00		5	1.25	5.55	2.83	0.99
		MI-MMLE	5	1	5.30	1.51	1.00		5	1	5.43	1.75	1.00
M2	A	FMV	4	1	4.51	0.82	1.00	M4’	4	0	4.26	0.54	1.00
		IPW-SIRS	4	1	4.75	1.10	1.00		5	1	4.99	1.05	1.00
		SRC	4	1	4.57	0.85	1.00		4	1	4.33	0.57	1.00
		ASRC	4	0	4.26	0.60	1.00		4	1	4.32	0.54	1.00
		MI-SIS	5	2	6.10	8.94	1.00		5	1	5.45	2.72	1.00
		MI-MMLE	5	2	5.77	5.04	1.00		5	0	5.28	2.11	1.00
	B	FMV	4	1	5.07	2.96	0.99		4	1	4.28	0.53	1.00
		IPW-SIRS	5	2	13.08	70.69	0.98		5	2	5.89	12.13	1.00
		SRC	5	1	5.06	2.08	0.99		4	1	4.39	0.56	1.00
		ASRC	4	1	4.64	1.98	0.99		4	1	4.34	0.62	1.00
		MI-SIS	5	2	8.00	15.74	0.97		5	1	5.66	2.05	1.00
		MI-MMLE	6	2	14.96	48.23	0.95		5	2	12.10	37.87	0.95
	C	FMV	4	1	4.59	1.16	1.00		4	0	4.27	0.50	1.00
		IPW-SIRS	4	1	4.79	1.75	1.00		5	1	5.2	1.56	1.00
		SRC	4	1	4.65	0.93	1.00		4	1	4.36	0.58	1.00
		ASRC	4	0	4.27	0.53	1.00		4	0	4.33	0.53	1.00
		MI-SIS	5	2	6.21	4.36	0.98		5	2	5.2	1.05	1.00
		MI-MMLE	5	1	5.97	4.23	0.99		5	1	5.18	0.95	1.00

实验 3 (单指标回归模型). 考虑下面响应变量 Y_i 和预测变量 $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})^T$ 的非单调关系:

$$Y_i = \exp(\boldsymbol{X}_i^T \boldsymbol{\beta} + 4)^3 + \varepsilon_i, \quad i = 1, \ldots, n,$$

其中 $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ 是一个 $p \times 1$ 维回归系数向量, ε_i 是测量误差, 与 $\boldsymbol{X}_i$ 独立. 在这个实验中, $\boldsymbol{X}_i$ 和 ε_i 产生与实验 1 相同, 考虑 $(n, p) = (100, 1000)$. $\boldsymbol{\beta}$ 的真实值取 $\boldsymbol{\beta} = (1.0, 1.2, 1.4, 1.6, 0.0, \ldots, 0.0)^T$, 这表明前四个预测变量是活跃变量, 后 $p - 4$ 个预测变量是非活跃变量. 同样, 假设所有的 $\boldsymbol{X}_i$ 是完全观测的但 Y_i 有缺失. 为了研究所提筛选过程对异常值和重尾数据的稳健性, ε_i 采用与实验 1 相同的设置. 缺失响应机制仍采用与实验 1 相同的设置除了 (M4) 被下面 (M4") 取代: $\Pr(\delta_i = 1|\boldsymbol{X}_i = \boldsymbol{x}_i, Y_i = y_i) = \exp(1.5 + 0.8x_{i1} + 0.00001y_i)/\{1 + \exp(1.5 + 0.8x_{i1} + 0.00001y_i)\}$, 而缺失机制模型 (M4') , 虽然仍然是不可忽略缺失, 与 (M4) 不同的是, (M4") 不包括非活跃变量 X_5. 与缺失机制模型 (M1), (M2), (M3), (M4") 相应的响应变量的平均缺失率分别为 13.2%, 41.5%, 26.5%, 12.5%.

基于 200 次重复试验的六种特征筛选过程相应的结果见表 2.3. 表 2.3 与实验 1 和实验 2 的观察结果相同. 也就是说, 即使预测变量和预测变量之间的关系是非单调的, 所提出的 ASRC 过程也表现的相当好.

2.5　实例分析

采用微阵列弥漫性大 B 细胞淋巴瘤 (DLBCL) 数据 (Rosenwald et al., 2002) 说明本章所提的特征筛选过程. 考虑到 DLBCL 是成人最常见的淋巴瘤类型, 标准化治疗后的存活率只有 35% 到 40%, 本研究的主要目的是识别影响生存结果的遗传因素. DLBCL 数据包含 $n = 240$ 患者的生存时间, 作为响应变量, 从每个患者的 cDNA 微阵列中获得的 $p = 7399$ 个基因的测量值被作为预测变量. 预测变量数量很大, 样本量较小, 由于计算的可行性、统计的准确性和算法的稳定性等方面的挑战, 很难直接利用正则化方法同时估计参数和选择重要基因 (Li et al., 2012). 为了解决这个问题, 在进行详细的数据分析之前, 有必要首先进行特征筛选过程来识别最相关的基因. 一般可以采用一些参数化的特征筛选方法, 但这些方法在很大程度上依赖于可能会错误指定的工作模型. 此外, 为上述数据集

表 2.3: 实验 3 中四种缺失机制下的六种筛选方法的结果

			$\rho = 0.2$					$\rho = 0.5$				
PS	Case	Method	Median	IQR	Mean	SD	CP	Median	IQR	Mean	SD	CP
M1	A	FMV	5	9	19.99	45.21	0.83	4	0	4.34	1.34	1.00
		IPW-SIRS	438	528.75	499.95	282.85	0.00	201.5	365.75	289.53	253.59	0.08
		SRC	4	1	9.85	27.54	0.93	4	0	4.10	0.41	1.00
		ASRC	4	1	8.62	24.22	0.95	4	0	4.08	0.37	1.00
		MI-SIS	455	342.5	492.28	241.09	0	274	307.5	329.57	233.34	0.01
		MI-MMLE	468.5	350.25	479.91	239.98	0	256	296	315.87	230.59	0.01
	B	FMV	5	12.25	38.6	115.27	0.78	4	0	6.6	20.62	0.99
		IPW-SIRS	435	408.5	465.43	262.77	0.01	253.5	343	338.04	283.33	0.02
		SRC	4	2	12.95	37.83	0.93	4	0	4.50	3.08	0.99
		ASRC	4	2	10.72	26.72	0.93	4	0	4.22	1.82	0.98
		MI-SIS	499.5	417.75	527.31	255.90	0	294	310.25	345.76	227.14	0
		MI-MMLE	537.5	417.75	538.05	260.36	0	310.5	287.5	358.9	227.14	0
	C	FMV	5	7	24.80	71.70	0.86	4	0	4.70	4.94	1.00
		IPW-SIRS	404.5	472	471.26	278.41	0.00	218	322	289.35	234.11	0.03
		SRC	4	1	8.50	14.65	0.93	4	0	4.2	1.32	1.00
		ASRC	4	1	7.43	12.95	0.96	4	0	4.14	1.22	1.00
		MI-SIS	417	390.5	457.97	244.40	0	283.5	351	351.62	259.62	0.01
		MI-MMLE	415.5	368.75	463.48	238.97	0	255.5	291	336.46	247.06	0.02
M2	A	FMV	12	38	56.91	1109.92	0.61	4	2	14.83	42.36	0.91
		IPW-SIRS	635.5	428.5	623.29	258.25	0.00	435	460.75	459.86	281.79	0.02
		SRC	6	12.25	21.24	43.62	0.81	4	1	6.84	10.96	0.96
		ASRC	5	5	18.77	52.45	0.86	4	0	5.35	6.69	0.98
		MI-SIS	604.5	406	589	251.21	0	271.5	426	363.44	287.37	0.01
		MI-MMLE	622	446.5	589.71	261.42	0	257	416	362.98	281.56	0.01
	B	FMV	14	58.5	72.06	145.26	0.59	4	2	11.95	27.00	0.92
		IPW-SIRS	609	391	595.06	250.65	0.00	363.5	434	430.28	277.09	0.01
		SRC	7	17.25	29.52	58.47	0.76	4	1	7.04	14.36	0.96
		ASRC	5	13	20.53	43.23	0.82	4	0	5.31	6.79	0.99
		MI-SIS	584	414.75	584.79	251.5	0	296.5	282.44	370.13	282.44	0.01
		MI-MMLE	608.5	431.5	590.52	254.56	0	287	405.25	366.80	267.10	0.01
	C	FMV	11	36.25	53.67	116.75	0.67	4	2	9.52	19.69	0.94
		IPW-SIRS	635	450.75	612.72	258.88	0.00	441	499	480.65	287.93	0.01
		SRC	6	9	21.52	57.59	0.81	4	1	5.54	5.72	0.98
		ASRC	4.5	4	12.28	29.17	0.91	4	0	4.44	2.12	1.00
		MI-SIS	557.5	435	567.41	249.84	0	255	379.25	338.95	271.27	0.01
		MI-MMLE	567	373.75	561.17	142.62	0	228.5	346.5	325.47	267.23	0.02

表 2.3: (续) 实验 3 中四种缺失机制下的六种筛选方法的结果

			$\rho = 0.2$					$\rho = 0.5$				
PS	Case	Method	Median	IQR	Mean	SD	CP	Median	IQR	Mean	SD	CP
M3	A	FMV	6	20.25	26.87	49.76	0.74	4	0	5.32	4.94	0.98
		IPW-SIRS	450.5	512	499.16	280.71	0.00	208	354.25	291.24	254.27	0.07
		SRC	4	4.25	10.93	24.85	0.91	4	0	4.38	1.55	1.00
		ASRC	4	4	9.38	15.51	0.92	4	0	4.31	1.09	1.00
		MI-SIS	546	431.75	534.69	248.81	0	310.5	290.5	361.04	227.62	0.01
		MI-MMLE	504.5	416.75	525.2	240.35	0	307.5	259.25	353.97	218.30	0
	B	FMV	8	30.75	57.89	128.21	0.67	4	1	9.42	46.97	0.97
		IPW-SIRS	491.5	471	500.14	272.86	0.01	269.5	377.5	343.65	260.95	0.03
		SRC	5	7	19.54	50.03	0.83	4	0	4.84	6.25	0.99
		ASRC	5	6	18.70	41.89	0.86	4	0	4.79	4.37	0.99
		MI-SIS	457.5	363.5	497.6	238.51	0	337.5	328.5	383.55	244.25	0
		MI-MMLE	502.5	373	519.32	229.42	0	341	305.5	392.82	224.65	0
	C	FMV	6	17.25	35.29	84.26	0.76	4	1	8.62	27.46	0.96
		IPW-SIRS	437.5	490.5	496.13	281.77	0.01	164.5	260.25	255.44	235.87	0.05
		SRC	4	4	13.28	29.10	0.90	4	0	4.80	4.06	1.00
		ASRC	4	4	12.54	27.05	0.90	4	0	4.49	1.73	1.00
		MI-SIS	537.5	368	520.07	239.70	0	294.5	298.5	345.92	215.48	0
		MI-MMLE	490.5	363.5	514.26	240.47	0	300	258	348.16	207.73	0
M4”	A	FMV	5	13	31.5	76.42	0.78	4	0	7.60	25.39	0.98
		IPW-SIRS	430.5	389.25	465.31	264.68	0.00	202	286.75	265.18	222.94	0.05
		SRC	4	3.25	10.94	21.6	0.89	4	0	4.73	5.15	0.99
		ASRC	4	4	10.18	19.40	0.90	4	0	4.59	4.10	0.99
		MI-SIS	514	345.25	522.24	226.89	0	291.5	270	363.44	241.40	0
		MI-MMLE	495	377.25	519.08	236.87	0	272.5	263.5	350.39	244.86	0
	B	FMV	8	25	51.74	113.79	0.72	4	1	6.12	9.50	0.97
		IPW-SIRS	455.5	487.25	484.75	284.49	0.01	259	374.5	332.55	254.58	0.04
		SRC	4.5	7	20.52	67.10	0.83	4	0	4.49	2.24	1.00
		ASRC	4	6	22.14	69.76	0.84	4	0	4.47	2.14	1.00
		MI-SIS	473	406.5	516.23	248.35	0	296.5	290	342.62	224.02	0.01
		MI-MMLE	483	356	520.14	232.91	0	310	314.75	348.56	216.10	0.01
	C	FMV	7	16	38.62	101.94	0.77	4	0	4.84	3.90	1.00
		IPW-SIRS	486	419.25	494.89	279.03	0.01	177	275.25	246.34	221.08	0.08
		SRC	5	4	11.22	22.95	0.93	4	0	4.22	0.73	1.00
		ASRC	5	4	12.16	23.80	0.93	4	0	4.15	0.50	1.00
		MI-SIS	461.5	407.75	494.45	249.14	0	262	237	307.96	184.21	0
		MI-MMLE	452.5	324.5	488.18	243.91	0	254	235.25	294.07	190.21	0.01

表 2.3: (续) 实验 3 中四种缺失机制下的六种筛选方法的结果

			$\rho = 0.8$						$\rho = 0.8$				
PS	Case	Method	Median	IQR	Mean	SD	CP	PS	Median	IQR	Mean	SD	CP
M1	A	FMV	4	0	4.17	0.40	1.00	M3	4	0	4.18	0.48	1.00
		IPW-SIRS	40	88.25	76.63	98.37	0.37		37.5	78	83.9	136.86	0.34
		SRC	4	0	4.14	0.38	1.00		4	0	4.14	0.40	1.00
		ASRC	4	0	4.10	0.30	1.00		4	0	4.12	0.36	1.00
		MI-SIS	102.5	121.5	146.57	133.65	0.07		144	153.25	170.33	121.19	0.03
		MI-MMLE	100.5	123.25	137.41	133.50	0.1		138	128.75	167.37	116.11	0.03
	B	FMV	4	0	4.18	0.44	1.00		4	0	4.26	0.66	1.00
		IPW-SIRS	57.5	116.5	97.33	110.45	0.28		61	130.75	114.72	158.93	0.27
		SRC	4	0	4.18	0.42	1.00		4	0	4.2	0.51	1.00
		ASRC	4	0	4.14	0.37	1.00		4	0	4.19	0.50	1.00
		MI-SIS	108.5	125.25	144.96	126.40	0.06		150.5	167.75	191.54	152.88	0.02
		MI-MMLE	138.5	156.25	172.57	138.84	0.06		177	184.25	220.9	161.30	0.02
	C	FMV	4	0	4.14	0.36	1.00		4	0	4.18	0.44	1.00
		IPW-SIRS	42.5	96.5	83.35	111.89	0.37		41	78.5	83.49	133.40	0.35
		SRC	4	0	4.11	0.32	1.00		4	1	4.15	0.39	1.00
		ASRC	4	0	4.07	0.26	1.00		4	0	4.10	0.30	1.00
		MI-SIS	89	98.5	122.42	117.64	0.08		137	127.25	166.77	119.95	0.02
		MI-MMLE	87.5	104.25	122.59	118.30	0.06		139	140	162.94	110.45	0.03
M2	A	FMV	4	0.25	4.38	0.85	1.00	M4”	4	0	4.11	0.31	1.00
		IPW-SIRS	88	151.5	140.47	153.38	0.16		39	84.5	90.48	139.24	0.35
		SRC	4	0.25	4.33	0.64	1.00		4	0	4.09	0.28	1.00
		ASRC	4	0	4.08	0.38	1.00		4	0	4.07	0.25	1.00
		MI-SIS	61	74.5	98.75	112.74	0.14		102.5	120.5	132.79	115.97	0.09
		MI-MMLE	59	88.5	89.90	96.62	0.20		101.5	119.75	124.83	107.50	0.12
	B	FMV	4	1	4.51	0.98	1.00		4	0	4.27	0.51	1.00
		IPW-SIRS	112.5	182.5	182.88	196.25	0.13		62	95.5	97.3	116.18	0.22
		SRC	4	1	4.38	0.68	1.00		4	0	4.24	0.49	1.00
		ASRC	4	0	4.16	0.42	1.00		4	0	4.20	0.48	1.00
		MI-SIS	61	94.5	114.05	154.71	0.19		111	142.5	144.85	128.90	0.10
		MI-MMLE	80.5	124	130.35	151.14	0.14		133	149.5	171.99	152.87	0.06
	C	FMV	4	1	4.40	0.99	1.00		4	0	4.24	0.48	1.00
		IPW-SIRS	99.5	175.75	153.4	170.27	0.17		34.5	80.25	80.04	121.8	0.37
		SRC	4	1	5.54	5.72	0.98		4	0	4.22	0.42	1.00
		ASRC	4	0	4.09	0.32	1.00		4	0	4.16	0.38	1.00
		MI-SIS	62	72.5	86.69	97.62	0.15		97.5	127.75	142.22	137.78	0.07
		MI-MMLE	59	75.25	86.99	96.20	0.18		97	131.5	139.8	129.86	0.07

指定正确的工作模型几乎是不可能. 因此, 本章的非参数特征筛选方法可能是一种较好的相关基因识别方法. 为了统一量纲, 对所选的预测变量和响应变量进行标准化.

作为一个例证, 与 Zhu et al. (2011) 类似, 将数据集分为 $n_1 = 160$ 个病人的训练集和 $n_2 = 80$ 个病人的测试集. 虽然在训练集中没有缺失数据, 但是包含响应变量删失的数据, 可以认为是缺失数据. 缺失率大约 41.25%. 对于训练数据集, 利用所提出的 ASRC 方法识别与生存时间相关的基因 $d_n = [n_1/\log(n_1)] = 31$. 与实验 1 类似, 取高斯核函数 $K(u) = \exp(-u^2/2)/(2\pi)^{1/2}$, 带宽 h 设置为 $h = 1.06\hat{\sigma}_x n^{-1/5}$, 其中 $\hat{\sigma}_x$ 是 X_k 的样本标准差. 为了比较, 仍然考虑其他五种筛选方法(例. IPW-SIRS, FMV, SRC, MI-SIS, MI-MMLE). 在表 2.4 中给出了各个方法筛选出来的前 31 个基因序号. 针对所考虑的六种方法, 为了评价所筛选出来的 31 个基因对响应变量的影响, 与 Zhu et al. (2011) 类似, 首先对训练数据集进行 Cox 比例风险模型的拟合, 然后计算训练和测试数据集的风险得分, 基于测试数据集中的风险评分, 将患者分为低风险和高风险组, 其中临界值由训练集的估计得分的中位数决定.

图 2.1 表示测试组中患者的两个风险组的 6 种特征筛选方法的 Kaplan-Meier 生存曲线估计值. 观察图 2.1 结果表明, 在比较的五种筛选方法中, ASRC 方法得到的两条曲线分离效果最好, log-rank 检验得到 p -值为0.0066, 表明拟合模型具有较好的预测效果; 而 IPW-SIRS, FMV, SRC, MI-SIS, MI-MMLE 方法相应的 p-值分别为 0.0559, 0.0133, 0.0891, 0.6966 and 0.5465. 这说明用无模型的筛选方法 (例, IPW-SIRS, FMV, SRC) 得到的拟合模型比有模型的筛选方法 (例, MI-SIS, MI-MMLE) 具有较好的预测效果.

表 **2.4:** 微阵列弥漫性大 **B** 细胞淋巴瘤 **(DLBCL)** 数据的前 **31** 个筛选变量

order	FMV	IPW-SIRS	SRC	ASRC	MI-SIS	MI-MMLE
1	6827	7394	3787	3787	5027	5027
2	4714	6304	5051	5025	1393	4748
3	5052	4847	5025	3578	4748	2105
4	5051	6732	5050	5352	4400	1393
5	4648	4447	5352	5050	5044	4400
6	5025	4959	6827	4268	2105	2516
7	3580	6130	3578	6827	5028	5044
8	3787	3715	3580	3580	5046	907
9	6125	5653	7177	5046	3142	3575
10	5050	3665	4785	6722	5969	5879
11	5444	6685	3593	7177	6685	6685
12	6551	6325	5260	5044	2576	2576
13	4637	5938	4648	7188	2516	3715
14	3578	1994	7188	6474	5851	6743
15	3922	95	7184	4785	5026	6365
16	7188	1494	5254	5051	907	7354
17	5364	7042	6722	1484	2103	4712
18	5442	1349	3377	5260	1346	3513
19	5352	4075	5296	4266	7354	5851
20	6722	4321	6909	3593	1296	1440
21	3593	7365	3592	5283	2104	3283
22	5071	4452	4259	3377	1789	1671
23	5296	3377	5024	4648	7279	368
24	4375	1995	4266	1483	1269	3142
25	3924	1355	1484	4259	4324	3074
26	3377	3714	5046	5254	5852	2104
27	5488	96	5444	4269	1484	3578
28	5254	6712	4268	3592	2588	5969
29	4268	4250	6474	5296	4750	2103
30	3921	3283	3998	4794	6744	2588
31	5488	5045	6248	5629	1383	3684

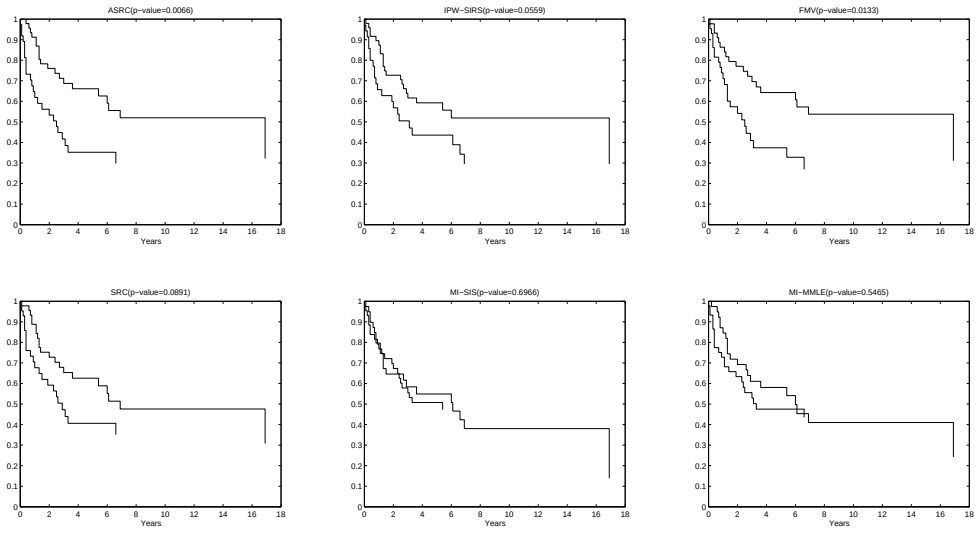

图 2-1: 两个风险组的Kaplan-Meier 生存曲线估计和测试组病人log-rank 检验的p 值.

2.6　定理证明

引理 2.1　假如条件 (C3)–(C4) 成立. 对于任意的 $\epsilon \in (0,1)$ 和某个常数 s, 对 $k = 1, \cdots, p$, 一致地有

$$\Pr\left\{\sup_{x\in\mathcal{R}_{X_k}}\left|\frac{1}{n}\sum_{i=1}^{n}\delta_i K_h(X_{ik},x) - f_k(x)\pi_k(x)\right| > \epsilon\right\} \le 2(1-s\epsilon/4)^n, \tag{2.8}$$

$$\Pr\left\{\sup_{x\in\mathcal{R}_{X_k},y\in\mathcal{R}_Y}\left|\frac{1}{n}\sum_{i=1}^{n}\delta_i K_h(X_{ik},x)I(Y_i\le y) - f_k(x)\pi_k(x)F_k(y\mid x)\right| > \epsilon\right\} \le 2(1-s\epsilon/4)^n, \tag{2.9}$$

其中 $\pi_k(x) = \Pr(\delta = 1|X_k = x)$, $F_k(y|x) = \Pr(Y \le y|X_k = x)$, $f_k(x)$ 是 X_k 的概率密度函数, $\mathcal{R}_{X_k}$ 和 $\mathcal{R}_Y$ 分别表示 X_k, Y 的支撑.

证明. 因为 (2.8) 和 (2.9) 证明方法相同, 为了节约空间, 在这里仅仅证明 (2.9). 记 $\xi_n = n^{-1}\sum_{i=1}^{n}\delta_i K_h(X_{ik},x)I(Y_i \le y)$. 则有

$$\Pr\left(|\xi_n - f_k(x)\pi_k(x)F_k(y|x)| > \epsilon\right) \le T_1 + T_2,$$

其中 $T_1 = \Pr\left(\xi_n - f_k(x)\pi_k(x)F_k(y|x) > \epsilon\right)$, $T_2 = \Pr\left(\xi_n - f_k(x)\pi_k(x)F_k(y|x) < -\epsilon\right)$. 对

任意的 $t>0$, 使用 Markov 不等式可以得到

$$\begin{aligned} T_1 &\leq \Pr\left(\exp[t\{\xi_n - f_k(x)\pi_k(x)F_k(y|x)\}] > \exp(t\epsilon)\right) \\ &\leq \exp(-t\epsilon)\exp\{-tf_k(x)\pi_k(x)F_k(y|x)\}E\{\exp(t\xi_n)\}. \end{aligned}$$

对于 i.i.d 随机样本, 有

$$E\exp(t\xi_n) = \left[E\exp\{t\delta_i K_h(X_{ik},x)I(Y_i\leq y)/n\}\right]^n.$$

令 $\phi(s) = E\exp\{s\delta_i K_h(X_{ik},x)I(Y_i\leq y)\}$. 对正常数 s, 取 $t=ns$ 可得

$$T_1 \leq [\exp(-s\epsilon)\exp\{-sf_k(x)\pi_k(x)F_k(y|x)\}\phi(s)]^n.$$

由 $\phi(s)$ 的定义可以得到

$$\begin{aligned} \exp\{-sf_k(x)\pi_k(x)F_k(y|x)\}\phi(s) &= E\exp\left[s\{\delta_i K_h(X_{ik},x)I(Y_i\leq y) - f_k(x)\pi_k(x)F_k(y|x)\}\right] \\ &= \mathcal{I}_1(y,x)\mathcal{I}_2(y,x), \end{aligned}$$

其中 $\mathcal{I}_1(y,x) = \exp\left(s[E\{\delta_i K_h(X_{ik},x)I(Y_i\leq y)\} - f_k(x)\pi_k(x)F_k(y|x)]\right)$, $\mathcal{I}_2(y,x) = E\exp\left(s[\delta_i K_h(X_{ik},x)I(Y_i\leq y) - E\{\delta_i K_h(X_{ik},x)I(Y_i\leq y)\}]\right)$.

对足够小的 s, 由不等式 $\exp(u)\leq 1+2|u|$ 可得

$$\mathcal{I}_1(y,x) \leq 1 + 2s|E\{\delta_i K_h(X_{ik},x)I(Y_i\leq y)\} - f_k(x)\pi_k(x)F_k(y|x)|.$$

记 $\Delta(y,x) = E\{\delta_i K_h(X_{ik},x)I(Y_i\leq y)\} - f_k(x)\pi_k(x)F_k(y|x) = E\{K_h(X_{1k},x)\pi_k(x)F_k(y|x)\} - f_k(x)\pi_k(x)F_k(y|x)$, 可以重新记为 $\Delta(y,x) = \int\{f_k(x+th) - f_k(x)\}\pi_k(x)F_k(y|x)K(t)dt$.

因为 $K(t)$ 是对称的, 有 $\int tK(t)dt = 0$. 同时, 在条件 (C3), 有 $\lim_{h\to 0}\{f_k(x+th) - f_k(x)\}/h = tf_k'(x)$. 结合上述不等式可得 $\mathcal{I}_1(y,x)\leq 1+Csh$.

当 $n\to\infty$ 有 $nh^2\to 0$, 那么对于足够大的 n, 有 $h<\epsilon/(16C)$. 因此, 有 $\sup_{x\in\mathcal{R}_{X_k},y\in\mathcal{R}_Y}\mathcal{I}_1(y,x) < 1+\epsilon s/16$. 类似地, 有 $\sup_{x\in\mathcal{R}_{X_k},y\in\mathcal{R}_Y}\mathcal{I}_2(y,x) < 1+\epsilon s/16$.

对于足够小的 ϵ 和 s, 有 $\exp(-s\epsilon)\leq 1 - s\epsilon + s^2\epsilon^2/2 \leq 1 - s\epsilon + s\epsilon/2 = 1 - s\epsilon/2$. 结合上述不等式和 $\exp(-s\epsilon)\leq 1-s\epsilon/2$ 得到 $T_1\leq(1-\epsilon s/4)^n$.

现在考虑 T_2 的上界. 对任意的 $t>0$, 根据 Markov 不等式可得

$$\begin{aligned} T_2 &= \Pr\left(f_k(x)\pi_k(x)F_k(y|x)-\xi_n>\epsilon\right) \\ &\le \Pr\left(\exp[t\{f_k(x)\pi_k(x)F_k(y|x)-\xi_n\}]>\exp(t\epsilon)\right) \\ &\le \exp(-t\epsilon)\exp\{tf_k(x)\pi_k(x)F_k(y|x)\}E\{\exp(-t\xi_n)\} \end{aligned}$$

类似于上述 T_1 的讨论, 可得 $T_2\le(1-\epsilon s/4)^n$. 结合上述的结果 (2.9) 得证. 类似地, 可以证明 (2.8).

引理 2.2 假如条件 (C3)–(C4) 成立. 如果 $\epsilon\in(0,1)$, 有

$$\Pr\left\{\sup_{y\in\mathcal{R}_Y}|\widehat{F}(y)-F(y)|>4\epsilon\right\}\le 4(n+3)\exp(-2n\epsilon^2), \tag{2.10}$$

$$\Pr\left\{\sup_{x\in\mathcal{R}_{X_k}}\left|\widehat{\mathscr{F}}_{k0}(x)-\mathscr{F}_{k0}(x)\right|>16\epsilon\right\}\le 6(n+4)\exp(-2n\epsilon^2). \tag{2.11}$$

证明. 为了证明 (2.10), 注意到

$$\begin{aligned} |\widehat{F}(y)-F(y)| &= \Big|\widehat{F}(y)-\frac{1}{n}\sum_{i=1}^n E\{I(Y_i\le y)|X_{ik},\delta_i\} \\ &\quad +\frac{1}{n}\sum_{i=1}^n E\{I(Y_i\le y)|X_{ik},\delta_i\}-E\{I(Y\le y)\}\Big| \\ &\le \Big|\widehat{F}(y)-\frac{1}{n}\sum_{i=1}^n E\{I(Y_i\le y)|X_{ik},\delta_i\}\Big| \\ &\quad +\Big|\frac{1}{n}\sum_{i=1}^n E\{I(Y_i\le y)|X_{ik},\delta_i\}-E\{I(Y\le y)\}\Big| = |\mathcal{H}_1|+|\mathcal{H}_2|. \end{aligned}$$

现在证明 $\mathcal{H}_1$ 的概率上界. 存在常数 s, 根据 (2.8) 可得

$$\begin{aligned} &\Pr\Big(\sup_{y\in\mathcal{R}_Y}\Big|\widehat{F}(y)-\frac{1}{n}\sum_{i=1}^n E\{I(Y_i\le y)|X_{ik},\delta_i\}\Big|>2\epsilon\Big) \\ =\ &\Pr\Big(\sup_{y\in\mathcal{R}_Y}\Big|\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n(1-\delta_i)W_{jk}^0(X_{ik})I(Y_j\le y)-\frac{1}{n}\sum_{i=1}^n E\{I(Y_i\le y)|X_{ik},\delta_i=0\}\Big|>2\epsilon\Big) \\ \le\ &\Pr\Big(\sup_{x\in\mathcal{R}_{X_k},y\in\mathcal{R}_Y}\Big|\frac{1}{n}\sum_{j=1}^n\frac{\delta_jK_h(X_{jk},x)}{f_k(x)\pi_k(x)}I(Y_j\le y)-E\{I(Y\le y)|x,\delta=1\}\Big|>\epsilon\Big) \end{aligned}$$

$$+2(1-s\epsilon/4)^n,$$

其中 $W_{ik}^0(x) = \delta_i K_h(X_{ik},x)/\sum_{j=1}^n \delta_j K_h(X_{jk},x)$. 对于某些 s 有 $(1-s\epsilon/4)^n \le 2\exp(-2n\epsilon^2)$, 由 Hoeffding's 不等式可得

$$\Pr\left(\sup_{y\in\mathcal{R}_Y}|\mathcal{H}_1|>2\epsilon\right)\le 2(n+3)\exp(-2n\epsilon^2).$$

对 $\mathcal{H}_2$, 由于 $E\{I(Y_i\le y)|X_{ik},\delta_i\}$ 是独立有界的随机变量, 有有限界, 由 Hoeffding's 不等式和经验过程理论可得 $\Pr(\sup_{y\in\mathcal{R}_Y}|\mathcal{H}_2|>2\epsilon)\le 2(n+1)\exp(-8n\epsilon^2)$. 结合上述不等式可得

$$\begin{aligned}\Pr\left\{\sup_{y\in\mathcal{R}_Y}|\widehat{F}(y)-F(y)|>4\epsilon\right\} &\le \Pr\left(\sup_{y\in\mathcal{R}_Y}|\mathcal{H}_1|>2\epsilon\right)+\Pr\left(\sup_{y\in\mathcal{R}_Y}|\mathcal{H}_2|>2\epsilon\right)\\ &= 4(n+3)\exp(-2n\epsilon^2).\end{aligned}$$

现在证明 (2.11). 注意到 $|\widehat{\mathscr{F}}_{k1}(x)-\mathscr{F}_{k1}(x)|$ 有下面的分解:

$$\begin{aligned}|\widehat{\mathscr{F}}_{k1}(x)-\mathscr{F}_{k1}(x)| &= \left|\sum_{i=1}^n W_{ik}^0(x)\{\widehat{F}(Y_i)-F(Y_i)\}+\sum_{i=1}^n W_{ik}^0(x)F(Y_i)-E\{F(Y)|x\}\right|\\ &= |\mathcal{H}_3+\mathcal{H}_4|.\end{aligned}$$

对 $\mathcal{H}_3$, 对某些 s, 由 (2.9) 可得

$$\begin{aligned}&\Pr\Big\{\sup_{x\in\mathcal{R}_{X_k}}\Big|\sum_{i=1}^n W_{ik}^0(x)\{\widehat{F}(Y_i)-F(Y_i)\}\Big|>8\epsilon\Big\}\\ \le\ &\Pr\Big\{\sup_{x\in\mathcal{R}_{X_k}}\Big|\frac{1}{n}\sum_{i=1}^n\frac{\delta_i K_h(X_{ik},x)}{f_k(x)\pi_k(x)}\{\widehat{F}(Y_i)-F(Y_i)\}\Big|>8\epsilon\Big\}\\ \le\ &\Pr\Big\{\sup_{y\in\mathcal{R}_Y}\Big|\widehat{F}(y)-F(y)\Big|>4\epsilon\Big\}+2(1-s\epsilon/2)^n\\ \le\ &4(n+4)\exp(-2n\epsilon^2).\end{aligned}$$

类似地, 对 $\mathcal{H}_4$, 有

$$\begin{aligned}&\Pr\Big\{\sup_{x\in\mathcal{R}_{X_k}}\Big|\sum_{i=1}^n W_{ik}^0(x)F(Y_i)-E\{F(Y)|X_k=x\}\Big|>8\epsilon\Big\}\\ \le\ &\Pr\Big\{\sup_{x\in\mathcal{R}_{X_k}}\Big|\frac{1}{n}\sum_{i=1}^n\frac{\delta_i K_h(X_{ik},x)}{f_k(x)\pi_k(x)}F(Y_i)-E\{F(Y)|X_k=x,\delta=1\}\Big|>4\epsilon\Big\}+2(1-s\epsilon)^n\end{aligned}$$

$$\leq \quad 2(n+3)\exp(-2n\epsilon^2).$$

因此, 结合上面的结果可得

$$\begin{aligned}\Pr\left(\sup_{x\in\mathcal{R}_{X_k}}|\widehat{\mathscr{F}}_{k1}(x)-\mathscr{F}_{k1}(x)|>16\epsilon\right) &\leq \Pr\left(\sup_{x\in\mathcal{R}_{X_k}}|\mathcal{H}_3|>8\epsilon\right)+\Pr\left(\sup_{x\in\mathcal{R}_{X_k}}|\mathcal{H}_4|>8\epsilon\right)\\ &\leq 6(n+4)\exp(-2n\epsilon^2).\end{aligned}$$

引理 2.3 假如条件 (C3) 和 (C4) 成立. 如果 $\epsilon\in(0,1)$, 有

$$\Pr\left(\max_k\left|\widehat{\omega}_k-\omega_k\right|>96\epsilon\right)\leq 30p(n+4)\exp(-2n\epsilon^2).$$

证明. 记 $\omega_k = E\{F_k(X_k)F(Y)\}-1/4$, $\widehat{\omega}_k = n^{-1}\sum_{i=1}^n[\widehat{F}_k(X_{ik})\{\delta_i\widehat{F}(Y_i)+(1-\delta_i)\widehat{\mathscr{F}}_{k1}(X_{ik})\}]-1/4$, $\tilde{\omega}_k = n^{-1}\sum_{i=1}^n[F_k(X_{ik})\{\delta_iF(Y_i)+(1-\delta_i)\mathscr{F}_{k1}(X_{ik})\}]-1/4$, $\bar{\omega}_k = n^{-1}\sum_{i=1}^n[\widehat{F}_k(X_{ik})\{\delta_iF(Y_i)+(1-\delta_i)\mathscr{F}_{k1}(X_{ik})\}]-1/4$. 因此, $\widehat{\omega}_k-\omega_k$ 有以下分解:

$$\begin{aligned}\widehat{\omega}_k-\omega_k &= \widehat{\omega}_k-\bar{\omega}_k+\bar{\omega}_k-\tilde{\omega}_k+\tilde{\omega}_k-\omega_k\\ &= \frac{1}{n}\sum_{i=1}^n\widehat{F}_k(X_{ik})[\delta_i\{\widehat{F}(Y_i)-F(Y_i)\}+(1-\delta_i)\{\widehat{\mathscr{F}}_{k1}(X_{ik})-\mathscr{F}_{k1}(X_{ik})\}]\\ &\quad+\frac{1}{n}\sum_{i=1}^n\{\delta_iF(Y_i)+(1-\delta_i)\mathscr{F}_{k1}(X_{ik})\}\{\widehat{F}_k(X_{ik})-F_k(X_{ik})\}\\ &\quad+\frac{1}{n}\sum_{i=1}^n[F_k(X_{ik})\{\delta_iF(Y_i)+(1-\delta_i)\mathscr{F}_{k1}(X_{ik})\}]-E\{F_k(x)F(y)\}\\ &\triangleq J_1+J_2+J_3.\end{aligned}$$

对 J_1, 因为 $\left|\widehat{F}(Y_j)-F(Y_j)\right|\leq 1$ 和 $\left|\widehat{\mathscr{F}}_{k1}(X_{ik})-\mathscr{F}_{k1}(X_{ik})\right|\leq\left|\widehat{F}(Y_i|X_{ik})-F(Y_i|X_{ik})\right|\leq 1$ 满足 Cui et al. (2015) 的 Hoeffding's 不等式的条件, 有 $|\widehat{F}_k(X_{jk})[\delta_i\{\widehat{F}(Y_j)-F(Y_j)\}+(1-\delta_i)\{\widehat{\mathscr{F}}_{k1}(X_{ik})-\mathscr{F}_{k1}(X_{ik})\}]|\leq|\widehat{F}(Y_j)-F(Y_j)|+|\widehat{\mathscr{F}}_{k1}(X_{ik})-\mathscr{F}_{k1}(X_{ik})|$. 结合上述等式和 (2.10), (2.11) 可得

$$\begin{aligned}\Pr\left(|\widehat{\omega}_k-\bar{\omega}_k|>32\epsilon\right) &\leq \Pr\left(\left|\frac{1}{n}\sum_{i=1}^n\{\widehat{F}(Y_i)-F(Y_i)\}\right|>16\epsilon\right)\\ &\quad+\Pr\left(\left|\frac{1}{n}\sum_{i=1}^n\{\widehat{\mathscr{F}}_{k1}(X_{ik})-\mathscr{F}_{k1}(X_{ik})\}\right|>16\epsilon\right)\end{aligned}$$

$$\begin{aligned} &\leq \Pr\Big(\sup_{y\in\mathcal{R}_Y}\big|\widehat{F}(y)-F(y)\big|>4\epsilon\Big)\\ &\quad+\Pr\Big(\sup_{x\in\mathcal{R}_{X_k}}\big|\widehat{\mathscr{F}}_{k1}(x)-\mathscr{F}_{k1}(x)\big|>16\epsilon\Big)\\ &\leq 10(n+4)\exp(-2n\epsilon^2). \end{aligned}$$

对 J_2, 由 $|\{\delta_i F(Y_i)+(1-\delta_i)\mathscr{F}_{k1}(X_{ik})\}\{\widehat{F}_k(X_{jk})-F_k(X_{jk})\}| \leq 2|\{\widehat{F}_k(X_{jk})-F_k(X_{jk})\}| \leq 2$, 因此有

$$\begin{aligned} \Pr(|\bar{\omega}_k-\tilde{\omega}_k|>32\epsilon) &= \Pr\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}\{\delta_i F(Y_i)+(1-\delta_i)\mathscr{F}_{k1}(X_{ik})\}\{\widehat{F}_k(X_{jk})-F_k(X_{jk})\}\Big|>32\epsilon\Big)\\ &\leq \Pr\Big(\sup_{x\in\mathcal{R}_x}\big|\widehat{F}_k(x)-F_k(x)\big|>16\epsilon\Big)\\ &\leq 2(n+1)\exp(-2n\epsilon^2). \end{aligned}$$

对 J_3, 因为 $E[F_k(X_{ik})\{\delta_i F(Y_i)+(1-\delta_i)\mathscr{F}_{k1}(X_{ik})\}] = E[E\{F_k(X_{ik})F(Y_i)|X_{ik},\delta_i\}] = E\{F_k(X_k)F(Y)\}$, 有

$$\begin{aligned} \Pr(|\tilde{\omega}_k-\omega_k|>32\epsilon) &= \Pr\Big(\sup_x\sup_{y\in\mathcal{R}_Y}\Big|\frac{1}{n}\sum_{i=1}^{n}[F_k(X_{ik})\{\delta_i F(Y_i^*)+(1-\delta_i)\mathscr{F}_{k1}(X_{ik})\}]\\ &\quad -E\{F_k(x)F(y)\}\Big|>32\epsilon\Big)\\ &\leq 2(n+1)\exp(-2n\epsilon^2). \end{aligned}$$

由于 $\big|\widehat{\omega}_k-\omega_k\big| \leq \big|\widehat{\omega}_k-\bar{\omega}_k\big|+\big|\bar{\omega}_k-\tilde{\omega}_k\big|+\big|\tilde{\omega}_k-\omega_k\big|$, 有

$$\begin{aligned} \Pr(\big|\widehat{\omega}_k-\omega_k\big|>96\epsilon) &\leq \Pr(|J_1|>32\epsilon)+\Pr(|J_2|>32\epsilon)+\Pr(|J_3|>32\epsilon)\\ &\leq 30(n+4)\exp(-2n\epsilon^2). \end{aligned}$$

结合上述不等式可以得到

$$\Pr(\max_k|\widehat{\omega}_k-\omega_k|>96\epsilon) \leq \sum_{k=1}^{p}\Pr(\big|\widehat{\omega}_k-\omega_k\big|>96\epsilon) \leq 30p(n+4)\exp(-2n\epsilon^2).$$

定理 2.1 的证明. 首先证明确定筛选性质. 在条件 (C1) 下, 根据条件 $\max_{k\in\mathcal{A}}|\widehat{\omega}_k-\omega_k|<cn^{-\tau}$ 可得 $\mathcal{A}\subset\widehat{\mathcal{A}}$. 事实上, 对 $k\in\mathcal{A}$, 根据 $\max_{k\in\mathcal{A}}|\widehat{\omega}_k-\omega_k|<cn^{-\tau}$ 可得 $|\widehat{\omega}_k|>|\omega_k|-cn^{-\tau}\geq cn^{-\tau}$. 因此, 由引理 2.3 可得

$$\Pr(\mathcal{A}\subset\widehat{\mathcal{A}}) \geq \Pr\Big(\max_{k\in\mathcal{A}}|\widehat{\omega}_k-\omega_k|<cn^{-\tau}\Big)=1-\Pr\Big(\max_{k\in\mathcal{A}}|\widehat{\omega}_k-\omega_k|\geq cn^{-\tau}\Big)$$

$$\geq \quad 1 - O\{|\mathcal{A}|(n+4)\exp(-bn^{1-2\tau})\},$$

其中 $b = c^2/4608$.

接着证明秩的相合性. 对常数 c_3, 根据条件 (C2) 和引理 2.3 可得

$$\begin{aligned}
\Pr\Big(\min_{k\in\mathcal{A}}|\widehat{\omega}_k| - \max_{k\notin\mathcal{A}}|\widehat{\omega}_k| < \tfrac{c_1}{2}\Big) &\leq \Pr\Big\{\Big(\min_{k\in\mathcal{A}}|\widehat{\omega}_k| - \max_{k\notin\mathcal{A}}|\widehat{\omega}_k|\Big) \\
&\quad -\Big(\min_{k\in\mathcal{A}}|\omega_k| - \max_{k\notin\mathcal{A}}|\omega_k|\Big) \leq -\tfrac{c_1}{2}\Big\} \\
&\leq \Pr\Big\{\Big|\Big(\min_{k\in\mathcal{A}}|\widehat{\omega}_k| - \max_{k\notin\mathcal{A}}|\widehat{\omega}_k|\Big) \\
&\quad -\Big(\min_{k\in\mathcal{A}}|\omega_k| - \max_{k\notin\mathcal{A}}|\omega_k|\Big)\Big| \geq \tfrac{c_1}{2}\Big\} \\
&\leq \Pr\Big(2\max_{1\leq k\leq p}|\widehat{\omega}_k - \omega_k| \geq \tfrac{c_1}{2}\Big) \\
&\leq O(n)p\exp(-c_3 n).
\end{aligned}$$

此外, 附加条件意味着存在常数 n_0, 当 $n > n_0$, 有 $p \leq \exp(c_3 n/2)$. 结合上述事实和 $c_3 n/2 \geq 3\log(n)$ 得

$$\begin{aligned}
\sum_{n=n_0}^{\infty} np\exp(-c_3 n) &\leq \sum_{n=n_0}^{\infty}\exp\{-c_3 n + c_3 n/2 + \log(n)\} \\
&\leq \sum_{n=n_0}^{\infty}\exp\{-3\log(n) + \log(n)\} \\
&= \sum_{n=n_0}^{\infty} n^{-2} < \infty.
\end{aligned}$$

根据 Borel-Cantelli 引理得

$$\liminf_{n\to\infty}\Big(\min_{k\in\mathcal{A}}|\widehat{\omega}_k| - \max_{k\notin\mathcal{A}}|\widehat{\omega}_k|\Big) > 0. \ \ a.s.$$

定理 2.2 的证明. 注意到 $\{k : |\omega_k| \geq cn^{-\tau}/2\}$ 的个数少于 $2n^{\tau}\sum_k|\omega_k|/c$. 记事件 $\mathcal{G} = \{\max_k|\widehat{\omega}_k - \omega_k| \leq cn^{-\tau}/2\}$. 因此, $\{k : |\widehat{\omega}_k| \geq cn^{-\tau}\}$ 的个数小于 $\{k : |\omega_k| \geq cn^{-\tau}/2\}$ 的个数, 它的界是 $2n^{\tau}\sum_k|\omega_k|/c$. 由于 $|\omega_k| \leq 1/2$ 和 $\sum_k|\omega_k| = O(n^{\varsigma})$, 存在常数 c' 使得 $\Pr(|\widehat{\mathcal{A}}| \leq c'n^{\tau+\varsigma}) \geq \Pr(\mathcal{G}) \geq 1 - O(n)p\exp(-b_0 n^{1-2\tau})$, 其中 b_0 是常数. 因此, 定理 2.2 证毕.

2.7 本章小结

在超高维数据分析中, 在不存在缺失数据的情况下一些特征筛选方法被提出识别活跃特征, 如 SIRS, RCS, DCS, FMV. 当响应变量存在缺失, 现有的筛选方法不能直接使用. 为了解决这个问题, 本章在响应变量随机缺失的情况下提出一种非参数特征筛选过程. 所提的特征筛选方法有以下几个优点. 首先, 它不需要假定任何模型. 其次, 它对异常值、重尾数据、相关协变量、模型的错误指定和缺失数据机制的错误指定都有较强的稳健性. 最后, 它对于响应变量和协变量的单调变换也是不变的. 在一些正则化的条件下, 建立所提筛选方法的确定筛选和秩相合的性质. 模拟研究和实例分析都表明所提的特征筛选过程比现有的特征筛选方法表现都好.

第三章 响应变量不可忽略缺失下广义线性单指标模型的倾向得分方法

3.1 引言

半参数模型是一类很重要的回归模型, 不仅具有参数模型的优点, 而且具有非参数模型的优点. 广义部分线性单指标模型 (GPLSIM) 是典型的半参数模型, 其灵活性很强, 是部分线性回归模型、单指标模型和广义线性模型的自然推广. 此外, 这类模型能很好地避免维数祸根问题, 近年来得到了很多研究者的关注, 例如, Carroll et al. (1997) 用局部线性回归解决了完全观测数据下广义部分线性单指标模型的参数估计问题, 并给出了其大样本理论; He and Yi (2011) 针对相关二元数据提出成对似然方法; Yi et al. (2009) 针对相依的二元响应变量提出了一种基于局部线性的截面最小二乘估计量; Boente and Rodriguez (2012) 的稳健估计; Huh and Park (2002) 对于没有部分线性项的广义单指数模型提出的平均导数估计量; Poon and Wang (2013) 提出完全的贝叶斯方法, Yu et al. (2017) 的惩罚样条估计.

上述工作主要集中在完全观测数据上. 然而, 实际上, 由于药物的严重副作用, 一些抽样个人不愿意回答所需的信息, 由于无法控制的因素造成的信息丢失, 一些定期的访问者间歇性或退出研究 (Little and Rubin, 2002) 等各种原因, 导致响应变量可能缺失. 关于在响应变量随机缺失下的单指标模型的统计推断已有相当多的研究成果. 譬如, Wang et al. (2010) 研究了在单指标模型的未知连接函数和方向参数的估计问题; He and Yi (2011) 提出了一种灵活的半参数方法来处理 GPLSIM 中的估计问题; Lai and Wang (2011) 借助插补方法研究了部分线性单指标模型的估计问题; Dong and Zhu (2013) 提供了一种带有缺失数据的单指标模型中方向参数估计的一般处理方法; Xue (2013) 针对协变量中数据缺失的单指标模型, 提出了一种经验似然方法; Lai and Wang (2014) 建立了响应变量随机缺失的异方差部分线性单指标模型的半参数有效界; Wang, Zhang and Hardle (2018) 考虑了响应变量随机缺失的扩展的单指标模型. 据我们所知, 带有非随机

缺失的响应变量 (即, 不可忽略缺失) 的 GPLSIM 的统计推断尚未被研究.

不可忽略缺失数据的统计推断更具挑战性, 因为缺失概率模型往往是未知的, 而假定的缺失数据机制模型是不可验证的. 为此, 基于似然的方法和贝叶斯方法处理不可忽略缺失数据已经付出了大量的努力. 例如, 见 Ibrahim et al. (1999) 和 Lee and Tang (2006). 特别地, Kim and Yu (2011) 基于不可忽略缺失数据机制的指数倾斜模型, 提出了一种带有不可忽略缺失数据的均值估计的半参数方法. Tang, Zhao and Zhu (2014) 在具有不可忽略缺失数据的广义估计方程中提出了一种利用核回归插补缺失数据进行参数估计的经验似然方法.

最近, Horvitz and Thompson (1952) 提出的逆概率加权方法, 也称为倾向得分加权方法, 被广泛用于处理缺失数据问题. 例如, Qin et al. (2008) 提出了一种有效的回归插补方法, 当缺失机制是协变量相关的, 且倾向得分函数被正确指定时, 该方法对回归模型的错误指定具有稳健性; Xue (2013) 利用逆概率加权方法, 建立了一种经验似然法, 用于协变量随机缺失的单指标模型中参数向量置信域的构造; Jiang et al. (2016) 针对一类协变量不可忽略缺失的线性模型, 提出了一种基于有效的稳健的回归过程的调整倾向分数参数估计; Riddles et al. (2016) 提出了一种以校准条件为辅助信息的倾向得分估计的最大似然方法; Zhao et al. (2017) 研究了协变量或响应变量不可忽略缺失时分位数回归中参数的几种逆概率加权估计. 上述工作建立在随机缺失或不可忽略缺失的参数模型或随机缺失的单指标模型的基础上. 因此, 现有的过程在协变向量维数很高的情况下存在维数祸根, 而在缺失数据机制是不可忽略的情况下效率会明显降低.

受到上述问题的启发, 针对响应变量不可忽略缺失的 GPLSIM, 本章采用了倾向得分方法去估计模型中的未知参数和未知连接函数, 其主要思想就是对完全观测数据给定一个与响应变量被观测到的概率 (响应概率) 相关的权重. 为了得到响应概率的相合估计, 响应概率模型采用了一种半参数逻辑回归模型 (SLRM). 特别地, SLRM 中的非参数部分采用核回归方法估计, 参数部分基于工具变量构造的估计方程采用两步的广义矩估计方法 (Wang et al. 2014), 其中工具变量是与研究的响应变量有关, 但是与缺失数据机制无关的协变量. 采用局部线性方法, 本文研究了未知参数和未知非参数函数的估计, 提出了一种计算非参数

函数和参数估计的数值迭代算法. 同时, 本文也系统地研究了这些估计量的渐近性质.

本章的其余部分构建如下. 第 3.2 节首先介绍了具有非随机缺失响应变量的 GPLSIM, 然后提出了一种基于拟似然的估计方法. 第 3.3 节研究了在某些正则条件下所提估计量的渐近性质. 第 3.4 节进行了大量的模拟研究. 第 3.5 节通过一个实例进行阐述. 第 3.6 节给出定理的证明.

3.2　模型及估计方法

3.2.1　广义线性单指标模型 (GPLSIM)

考虑一个容量为 n 随机样本 $\{(Y_i;\boldsymbol{X}_i,\boldsymbol{Z}_i)\}_{i=1}^n$, 且 Y_i 为一维响应变量, $\boldsymbol{X}_i=(X_{i1},\ldots,X_{ip})^\top, \boldsymbol{Z}_i=(Z_{i1},\ldots,Z_{iq})^\top$ 是 $p\times 1$ 和 $q\times 1$ 维的协向量. 在本章中, 假设所有的 $\boldsymbol{X}_i$ 和 $\boldsymbol{Z}_i$ 是完全观测的, 而 Y_i 可能存在缺失. 假设 Y_i 服从指数族分布, 其密度函数为

$$f(Y_i|\boldsymbol{X}_i,\boldsymbol{Z}_i,\phi)=\exp\{\phi^{-1}[Y_i\zeta_i-\mathcal{B}(\zeta_i)]+c(Y_i,\phi)\},\quad i=1,\ldots,n, \tag{3.1}$$

其中 ζ_i 是自然参数, $\mathcal{B}(\cdot)$ 和 $c(\cdot)$ 是两个已知的函数, ϕ 是一个尺度参数或已知或未知. 为了简便, 在本章中假设 ϕ 是已知的. 因此, 给定协变量 X_i 和 Z_i, Y_i 的条件均值和方差分别是 $\mu_i=E(Y_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)=\dot{\mathcal{B}}(\zeta_i)$ 和 $\mathrm{var}(Y_i|\boldsymbol{X}_i,\boldsymbol{Z}_i)=V(\mu_i)=\phi\ddot{\mathcal{B}}(\zeta_i)$, 其中 $\dot{\mathcal{B}}(\zeta)=\partial\mathcal{B}(\zeta)/\partial\zeta$, $\ddot{\mathcal{B}}(\zeta)=\partial^2\mathcal{B}(\zeta)/\partial\zeta^2$. 密度函数 (3.1) 包括很多分布, 其特殊的分布有正态分布、二项分布、泊松分布等. 将自然参数与协变量 $\boldsymbol{X}_i,\boldsymbol{Z}_i$ 联系起来的广义部分线性单指标模型可表示为

$$G(\mu_i)=\boldsymbol{Z}_i^\top\boldsymbol{\beta}+g(\boldsymbol{X}_i^\top\boldsymbol{\alpha}), \tag{3.2}$$

其中 $G(\cdot)$ 是一个已知单调的连接函数, $g(\cdot)$ 是一个未知的可微函数, $\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_p)^\top$ 和 $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_q)^\top$ 分别是 $p\times 1$ 和 $q\times 1$ 维感兴趣的未知参数向量. 为了可识别, 假设 $\|\boldsymbol{\alpha}\|=1$ 和第一个分量 $\alpha_1>0$, 其中 $\|\cdot\|$ 为欧几里德范数. 参数 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 的真实分别记为 $\boldsymbol{\alpha}_0$ 和 $\boldsymbol{\beta}_0$.

当 Y_i 的分布为正态分布时, 上述定义的 GPLSIM 变为一个部分线性单指标模型 (Dong et al., 2016), 是模型 (3.1) 的一种特殊情况. 当 $\boldsymbol{\beta}=0$, 上述定义的

GPLSIM 简化为一个单指标模型. 当 X_i 是一个一维变量且 $\alpha = 1$ 时, 上述定义的 GPLSIM 简化为一个部分线性模型. 因此, 上述定义的 GPLSIM 是一个部分线性单指标模型, 单指标模型和部分线性模型的自然推广. 当没有缺失的数据时, Carroll et al. (1997) 提出了一种估计 GPLSIM 中参数和非参数函数的局部线性方法.

当缺失数据存在时, 假设 δ_i 是表示 Y_i 是否缺失的示性变量, 即, 如果 Y_i 可观测, 则 $\delta_i = 1$; 否则, 如果 Y_i 缺失, $\delta_i = 0$. 因此, 完全数据集由观测数据 $\{(Y_i, \boldsymbol{X}_i, \boldsymbol{Z}_i, \delta_i) : i = 1, \ldots, n\}$ 组成. 假如 $(p+q)\times 1$ 维随机向量 $(\boldsymbol{X}_i^\top, \boldsymbol{Z}_i^\top)^\top$ 可以分解为 $(\boldsymbol{X}_i^\top, \boldsymbol{Z}_i^\top)^\top = (\boldsymbol{U}_i^\top, \boldsymbol{S}_i^\top)^\top (i = 1, \ldots, n)$, 其中 $\boldsymbol{U}_i = (U_{i1}, \ldots, U_{id_1})^\top$, $\boldsymbol{S}_i = (S_{i1}, \ldots, S_{id_2})^\top$ 分别是 $d_1 \times 1$ 和 $d_2 \times 1$ 维协向量. 假设 $\delta_i | (Y_i, \boldsymbol{U}_i) \sim \text{Bernoulli}(\pi_i)$, 其中 $\pi_i = \pi(Y_i, \boldsymbol{U}_i)$, 对任意 $i \neq j$, δ_i 和 δ_j 独立. 如果 π_i 不依赖于 Y_i 的值, 上述考虑的缺失数据机制模型就是随机缺失 (MAR)的. 如果 π_i 依赖于 Y_i 的值, 上述考虑的缺失数据机制模型是不可忽略缺失的. 类似于 Kim and Yu (2011) 和 Tang et al. (2014), 假设响应概率模型为半参数模型, 其形式为:

$$\pi_i = \pi(\boldsymbol{U}_i, Y_i; \gamma) = \Pr(\delta_i = 1 | Y_i, \boldsymbol{U}_i) = \frac{\exp\{m(\boldsymbol{U}_i) - \gamma Y_i\}}{1 + \exp\{m(\boldsymbol{U}_i) - \gamma Y_i\}}, \tag{3.3}$$

这说明缺失数据机制是不可忽略的, 其中 $m(\boldsymbol{U}_i)$ 是 $\boldsymbol{U}_i$ 的未知函数, γ 是倾斜参数. 当 $\gamma = 0$, 模型 (3.3) 简化为一个 MAR 模型 (Little and Rubin, 2002).

3.2.2 参数和非参数函数的估计

针对 GPLSIM, 采用剖面估计方法对未知参数 $\boldsymbol{\alpha}$、$\boldsymbol{\beta}$ 和未知非参数函数 $g(\cdot)$ 进行估计. 具体过程如下: (1) 给定 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$, 结合局部线性核函数方法和加权的倾向得分对数似然 (PS-WL) 方法估计 $g(\cdot)$; (2) 给定预估的 $\hat{g}(\cdot)$, 可以使用剖面的 PS-WL 方法估计 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$.

令 $Q(\boldsymbol{\mu}, \boldsymbol{Y}) = \sum_{i=1}^n Q(\mu_i, Y_i) = \sum_{i=1}^n \{(Y_i\zeta_i - \mathcal{B}(\zeta_i))/\phi + c(Y_i, \phi)\}$ 是 $Y_1, \ldots, Y_n$ 的对数似然函数, 其中 $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_i)^\top$, $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$. 这很容易证明 $\partial Q(\boldsymbol{\mu}, Y)/\partial \boldsymbol{\mu} = \sum_{i=1}^n \partial Q(\mu_i, Y_i)/\partial \mu_i = \sum_{i=1}^n (Y_i - \mu_i)/V(\mu_i)$. 因此, 当没有缺失的数据时, 根据 Carroll et al. (1997) 的讨论, 给定 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 的值, 通过最大化与 a, b 有关的局部线性核加权

对数似然函数

$$\sum_{i=1}^{n} Q(\mu_i, Y_i) K_h(\boldsymbol{X}_i^\top \boldsymbol{\alpha} - t),$$

可以得到 $g(t) = a$ 及其导数 $\dot{g}(t) = b$ 的局部线性估计量, 其中 $\mu_i = G^{-1}(a + b(\boldsymbol{X}_i^\top \boldsymbol{\alpha} - t) + \boldsymbol{Z}_i^\top \boldsymbol{\beta})$, $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ 是核函数, h 是带宽. 对于缺失数据的情况, 响应概率 π_i 通常情况下是未知的, 需要进行相合估计 (Tang et al., 2014). 根据倾向得分方法的思想 (Qin et al., 2008), 关于 a 和 b 的调整的倾向得分局部线性估计量可通过极大化下面的 kernel-PS-weighted 对数似然获得：

$$\mathcal{H}(a, b, t) = \sum_{i=1}^{n} \frac{\delta_i}{\pi_i} Q(\mu_i, Y_i) K_h(\boldsymbol{U}_i^\top \boldsymbol{\alpha} - t). \tag{3.4}$$

记它们相应的估计量分别为 $\hat{a} = \hat{g}(t; \boldsymbol{\alpha}, \boldsymbol{\beta})$, $\hat{b} = \hat{\dot{g}}(t; \boldsymbol{\alpha}, \boldsymbol{\beta})$. 因此, 非参函数 $g(t)$ 的 PS-LLE 估计通过 $\hat{g}(t; \boldsymbol{\alpha}, \boldsymbol{\beta}) = \hat{a}$ 给出.

给定非参函数 $g(v)$ 的 PS-LL 估计量 $\hat{g}(v; \boldsymbol{\alpha}, \boldsymbol{\beta})$, 参数 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 的 PS-加权估计量通过极大化下面与 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 有关的剖面 PS-加权对数似然可以得到：

$$\mathcal{B}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{\delta_i}{\pi_i} Q(\mu_i, Y_i), \tag{3.5}$$

$\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 相应的 PS-加权估计量记为 $\hat{\boldsymbol{\alpha}}$ 和 $\hat{\boldsymbol{\beta}}$.

参数 $\boldsymbol{\alpha}, \boldsymbol{\beta}$ 和非参函数 $g(\cdot)$ 的估计过程可以总结如下：

步骤 1 . 基于数据集 $\{(Y_i, \boldsymbol{X}_i, \boldsymbol{Z}_i, \delta_i) : i = 1, \ldots, n\}$, 通过第 3.2.3 节中介绍的方法计算 $\pi_i = \pi(\boldsymbol{U}_i, Y_i)$ 的估计量 $\hat{\pi}_i = \hat{\pi}(\boldsymbol{U}_i, Y_i)$.

步骤 2 . 基于完全观测数据 $\{(Y_i, \boldsymbol{X}_i, \boldsymbol{Z}_i, \delta_i = 1) : i = 1, \ldots, n\}$, 拟合一个广义的参数线性模型, 且满足限制条件 $\|\boldsymbol{\alpha}\| = 1$ 和 $\alpha_1 > 0$ 获得 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 的初始值, 分别记为 $\hat{\boldsymbol{\alpha}}^0$ 和 $\hat{\boldsymbol{\beta}}^0$.

步骤 3 . 给定 $\hat{\boldsymbol{\alpha}}^0$ 和 $\hat{\boldsymbol{\beta}}^0$, 通过下列目标函数求解 $\hat{g}(\boldsymbol{X}_i^\top \hat{\boldsymbol{\alpha}}^0; \hat{\boldsymbol{\alpha}}^0, \hat{\boldsymbol{\beta}}^0) = \hat{a}_i$ 和 $\hat{\dot{g}}(\boldsymbol{X}_i^\top \hat{\boldsymbol{\alpha}}^0; \hat{\boldsymbol{\alpha}}^0, \hat{\boldsymbol{\beta}}^0) = \hat{b}_i$,

$$\sum_{j=1}^{n} \frac{\delta_j}{\hat{\pi}_j} K_h(\boldsymbol{X}_j^\top \hat{\boldsymbol{\alpha}}^0 - \boldsymbol{X}_i^\top \hat{\boldsymbol{\alpha}}^0) \{-Y_j \zeta_j + \mathcal{B}(\zeta_j)\} \tag{3.6}$$

其中, $\zeta_j = a_i + b_i(\boldsymbol{X}_j^\top \hat{\boldsymbol{\alpha}}^0 - \boldsymbol{X}_i^\top \hat{\boldsymbol{\alpha}}^0) + \boldsymbol{Z}_j^\top \hat{\boldsymbol{\beta}}^0$.

步骤 4 . 给定 $\hat{g}(\boldsymbol{X}_i^{\top}\hat{\boldsymbol{\alpha}}^0;\hat{\boldsymbol{\alpha}}^0,\hat{\boldsymbol{\beta}}^0)=\hat{a}_i$ 和$\hat{\dot{g}}(\boldsymbol{X}_i^{\top}\hat{\boldsymbol{\alpha}}^0;\hat{\boldsymbol{\alpha}}^0,\hat{\boldsymbol{\beta}}^0)=\hat{b}_i$, 更新 $(\boldsymbol{\alpha},\boldsymbol{\beta})$ 的估计通过解

$$\min_{\boldsymbol{\alpha}\boldsymbol{\beta}}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}_i}\{-Y_i\zeta_i^*+\mathcal{B}(\zeta_i^*)\},$$
$$s.t.\|\boldsymbol{\alpha}\|=1,\alpha_1>0 \tag{3.7}$$

得到, 其中 $\zeta_i^*=\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}+\hat{a}_i+\hat{b}_i\boldsymbol{X}_i^{\top}(\boldsymbol{\alpha}-\hat{\boldsymbol{\alpha}}^0)$.

步骤 5 . 重复步骤 3 和步骤 4, 直到收敛. 记 $(\boldsymbol{\alpha},\boldsymbol{\beta})$ 最后的估计为 $(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})$.

步骤 6 . 固定 $(\boldsymbol{\alpha},\boldsymbol{\beta})$ 为步骤 5 所得的估计值. $g(t)$ 最后的估计值是 $\hat{g}(t;\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})=\hat{a}_(t)$, 其中 $(\hat{a},\hat{b})$ 通过求解方程 (3.6) 获得.

注记 3.1 在数值研究中, 使用 Matlab 命令“fmincon”中的 sqp 算法得到 (3.7) 的解.

3.2.3 倾向得分估计

在本节中, 使用众所周知的 GMM 方法估计倾向得分函数 $\pi_i=\pi(\boldsymbol{U}_i,Y_i)$. 当 π_i 仅仅依赖于协变量 $\boldsymbol{U}_i$, 即, 缺失数据机制是 MAR, 使用加权最小二乘法很容易估计 π_i. 但是当缺失数据机制是不可忽略时, 估计 π_i 是相当困难的. 在不可忽略缺失数据机制的假设下, 有几种估计 π_i 的方法. 例如, 见 Kim and Yu (2011) 的核实样本方法, 其优点具有较好的稳健性; Qin et al. (2002) 和 Zhao et al. (2017) 的半参数经验似然方法; Wang et al. (2014), Jiang et al. (2016) 和 Zhao et al. (2017) 提出了基于估计方程的 GMM 方法. 在下文中, 采用基于估计方程的 GMM 方法对 π_i 进行估计.

基于方程 (3.3) 和 Tang et al. (2014) 文章的思想, 有

$$\exp\{-m(\boldsymbol{U}_i)\}=\frac{E\{(1-\delta_i)|\boldsymbol{U}_i\}}{E\{\delta_i\exp(\gamma Y_i)|\boldsymbol{U}_i\}}.$$

由核回归方法可知 $W(\boldsymbol{U}_i)=\exp\{-m(\boldsymbol{U}_i)\}$ 可由

$$\hat{W}(\boldsymbol{U}_i,\gamma)=\exp\{-\hat{m}(\boldsymbol{U}_i)\}=\frac{\sum_{j=1}^{n}(1-\delta_j)\bar{K}_{\bar{h}}(\boldsymbol{U}_i-\boldsymbol{U}_j)}{\sum_{j=1}^{n}\delta_j\exp(\gamma Y_j)\bar{K}_{\bar{h}}(\boldsymbol{U}_i-\boldsymbol{U}_j)} \tag{3.8}$$

进行估计, 其中 $\bar{K}_{\bar{h}}(\boldsymbol{u}) = \bar{h}^{-d_1}\bar{K}(\boldsymbol{u}/\bar{h})$ 是一个 d_1-维核函数, $\bar{h} = \bar{h}_n$ 是一个带宽序列. 因此, 当 $\gamma = \gamma_0$ 已知, 给出 π_i 的一个非参数估计 $\hat{\pi}_i(\gamma) = \{1 + \hat{W}(\boldsymbol{U}_i, \gamma)\exp(\gamma Y_i)\}^{-1}$.

然而, 在很多实际应用中, 参数 γ 通常是未知的. 注意 $\boldsymbol{S}_i$ 是与响应变量 Y_i 均值有关的协变量, 但它们没有出现在方程中 (3.3), 这样的协变量 $\boldsymbol{S}_i$ 通常称为响应变量的工具变量 (Wang et al., 2014). 令 $\mathbb{S} = (\boldsymbol{U}, \boldsymbol{S}^*)$, $\boldsymbol{S}^*$ 是 $\boldsymbol{S}$ 的 $d_3 \times 1$ 分量.

在不可忽略缺失数据机制假设下, 如果 $\pi(\boldsymbol{U}_i, Y_i)$ 被正确指定, 有

$$E\{M(\mathbb{S}, Y, \delta, m(\boldsymbol{U}, \gamma), \gamma)\} = E\left\{\left(\frac{\delta}{\pi(\boldsymbol{U}, Y)} - 1\right)\begin{pmatrix} 1 \\ \mathbb{S} \end{pmatrix}\right\} = 0,$$

这说明可以得到 γ 的一个相合估计值, 通过求解下面等式

$$\frac{1}{n}\sum_{i=1}^{n} M_k(\mathbb{S}_i, Y_i, \delta_i, \hat{m}_\gamma, \gamma) = 0, \quad k = 1, \ldots, \mathbb{L}.$$

其中 $\hat{m}_\gamma = \hat{m}(\boldsymbol{U}, \gamma)$.

记

$$\bar{M}_n(\hat{m}_\gamma, \gamma) = \frac{1}{n}\sum_{i=1}^{n} M(\mathbb{S}_i, Y_i, \delta_i, \hat{m}_\gamma, \gamma),$$

其中 $M(\mathbb{S}_i, Y_i, \delta_i, \hat{m}_\gamma, \gamma) = (M_1(\mathbb{S}_i, Y_i, \delta_i, \hat{m}_\gamma), \ldots, M_{\mathbb{L}}(\mathbb{S}_i, Y_i, \delta_i, \hat{m}_\gamma, \gamma))^\top$. 因此, 通过求解方程 $\bar{M}_n(\hat{m}_\gamma, \gamma) = 0$ 可以得到 γ 的一个估计量 $\hat{\gamma}$. 显然, $\bar{M}_n(\hat{m}_\gamma, \gamma) = 0$ 这个等式是过度识别, 这意味着直接解方程是不可能得到 γ 的估计量的. 为了解决这个问题, 采用了以下两步 GMM 过程 (Wang et al., 2014 and Shao, 2016).

步骤 R1 . 通过最小化 $\bar{M}_n(\hat{m}_\gamma, \gamma)^\top\bar{M}_n(\hat{m}_\gamma, \gamma)$ 求出 γ 的最初始的估计 $\hat{\gamma}^{(1)}$.

步骤 R2 . 记 $\hat{W}_n = \frac{1}{n}\sum_{i=1}^{n} M(\mathbb{S}_i, Y_i, \delta_i, \hat{m}_{\gamma^{(1)}}, \hat{\gamma}^{(1)})M(\mathbb{S}_i, Y_i, \delta_i, \hat{m}_{\gamma^{(1)}}, \hat{\gamma}^{(1)})^\top$. γ 的一种有效的两步广义矩估计可以通过求解

$$\hat{\gamma} = \arg\min_{\gamma} \tilde{Q}_n(\hat{m}_\gamma, \gamma)$$

得到, 其中 $\tilde{Q}_n(\gamma) = \bar{M}_n(\hat{m}_\gamma, \gamma)^\top\hat{W}_n\bar{M}_n(\hat{m}_\gamma, \gamma)$.

3.3 渐近性质

在本节中, 首先研究在步骤 R2 中半参数两步广义矩估计 γ 的相合性和渐近正态性.

定理 3.1 假设条件 (A1)–(A3) 成立. 如果 γ_0 是 $M_0(\gamma) = E\{M(\mathbb{S}, Y, \delta, \gamma)\}$ 的唯一解和 $\sup_{\gamma\in\Upsilon} \|M_0(\gamma)\| < \infty$. 则, 当 $n \to \infty$, 有 $\hat{\gamma} \xrightarrow{\mathcal{P}} \gamma$, $\hat{\gamma}$ 有下列渐近展开:

$$\sqrt{n}(\hat{\gamma} - \gamma_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{G}(\mathbb{S}_i, Y_i, \gamma_0) + o_p(1),$$

其中 $\mathcal{G}(\mathbb{S}_i, Y_i, \gamma_0) = -2\{\nabla^{(2)}_{\gamma_0} \tilde{Q}_n(\hat{m}_{\tilde{\gamma}}, \tilde{\gamma})\}^{-1}\{\nabla^{(1)}_{\gamma_0} \bar{M}_n(\hat{m}_{\gamma_0}, \gamma_0)\}^{\top} \hat{W}_n\{n^{1/2} \bar{M}_n(\hat{m}_{\gamma_0}, \gamma_0)\}$, $\tilde{\gamma}$ 介于 $\hat{\gamma}$ 和 γ_0 之间, $\nabla^{(1)}_{\gamma} \mathcal{F}(\gamma)$ 和 $\nabla^{(2)}_{\gamma} \mathcal{F}(\gamma)$ 表示函数 $\mathcal{F}(\gamma)$ 关于 γ 的一阶和二阶偏导数, $\xrightarrow{\mathcal{P}}$ 表示依概率收敛.

定理 3.2 假设条件 (A1)–(A6) 成立. 则, 当 $n \to \infty$, 有 $\sqrt{n}(\hat{\gamma} - \gamma_0) \xrightarrow{\mathcal{D}} N(0, \Sigma_\gamma)$, 其中 $\Sigma_\gamma = (\Phi^{\top} W \Phi)^{-1} \Phi^{\top} W \Omega_\gamma W \Phi (\Phi^{\top} W \Phi)^{-1}$, $\Omega_\gamma = \mathrm{var}[M(\mathbb{S}, Y, \delta, m_{\gamma_0}, \gamma_0) + \upsilon(\mathbb{S}, Y, \delta, m_{\gamma_0}, \gamma_0)]$, $\upsilon(\mathbb{S}, Y, \delta, m_{\gamma_0}, \gamma_0) = \omega(\boldsymbol{U})(1 - \delta, \delta \exp(\gamma_0 Y))^{\top} - E\{\omega(\boldsymbol{U})(1 - \delta, \delta \exp(\gamma_0 Y))^{\top}\}$, Φ 是条件 A5 给出.

接下来, 研究 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 的相合性和渐近正态性. 为了方便陈述, 引入一些记号. 令

$$\nu_j = \int u^j K^2(u) du, \quad \kappa_j = \int u^j K(u) du, \quad j = 0, 1, 2, \ldots$$

和

$$q_l(t, Y) = (\partial^l / \partial t^l) Q\{G^{-1}(t, Y)\}, l = 1, 2.$$

则, 有

$$q_1(t, Y) = \{Y - G^{-1}(t)\}\rho_1(t)$$

$$q_2(t, Y) = \{Y - G^{-1}(t)\}\rho_1'(t) - \rho_2(t),$$

其中 $\rho_l = \{dh^{-1}(t)/dt\}^l V^{-1}\{h^{-1}(t)\}$.

定理 3.3 假设条件 A 和 B 成立. 当 γ_0 已知时, 随着 $n \to \infty$, 如果 $nh^4 \to \infty$ 和 $nh^6 \to 0$ 成立, 则

$$\begin{aligned}\hat{g}(t;\boldsymbol{\alpha},\boldsymbol{\beta})-g(t)=&\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}\frac{K_h(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t)\varepsilon_i}{f(t)E[\rho_2\{g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0=t]}\\&-\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i-\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}{\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}\frac{E[K_h(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t)\varepsilon_i|\boldsymbol{U}_i,\delta_i=0]}{f(t)E[\rho_2\{g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0=t]}\\&-\begin{pmatrix}\frac{\dot{g}(t)E[\boldsymbol{X}\rho_2\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0=t]}{E[\rho_2\{g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0=t]}\\-\frac{E[\boldsymbol{Z}\rho_2\{g(\boldsymbol{Z}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0=t]}{E[\rho_2\{g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0=t]}\end{pmatrix}^{\top}\begin{pmatrix}\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\\\boldsymbol{\beta}-\boldsymbol{\beta}_0\end{pmatrix}\\&+\frac{1}{2}\ddot{g}(t)h^2\kappa_2+o_p(n^{-1/2})+O_p\left(h^2+\sqrt{\frac{\log n}{nh}}\right),\end{aligned}$$

$$\begin{aligned}\hat{\dot{g}}(t;\boldsymbol{\alpha},\boldsymbol{\beta})-\dot{g}(t)=&\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}\frac{K_h(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t)[(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t)/h]\varepsilon_i}{f(t)E[\rho_2\{g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0=t]}\\&-\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i-\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}{\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}\frac{E[K_h(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t)[(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t)/h]\varepsilon_i|\boldsymbol{U}_i,\delta_i=0]}{f(t)E[\rho_2\{g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0=t]}\\&+o_p(n^{-1/2})+O_p\left(h^2+\sqrt{\frac{\log n}{nh}}\right),\end{aligned}$$

其中 $\varepsilon_i=[Y_i-G^{-1}\{g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0\}]\rho_1\{g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0\}$.

定理 3.4 假设条件 A 和 B 成立. 当 γ_0 已知时, 随着 $n\to\infty$, 如果 $nh^4\to\infty$ 和 $nh^6\to 0$ 成立, 则

$$\sqrt{n}\begin{pmatrix}\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0\\\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\end{pmatrix}\xrightarrow{\mathcal{D}}N(0,\mathcal{A}_0^{-1}\Sigma_1\mathcal{A}_0^{-1}),$$

其中

$$\begin{aligned}\mathcal{A}_0=&E\left\{\rho_2\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}\begin{pmatrix}\dot{g}(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)\boldsymbol{X}\\\boldsymbol{Z}\end{pmatrix}\begin{pmatrix}\dot{g}(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)\boldsymbol{X}\\\boldsymbol{Z}\end{pmatrix}^{\top}\right\}\\&-E\left\{\rho_2\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}\begin{pmatrix}\dot{g}(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)\boldsymbol{X}\\\boldsymbol{Z}\end{pmatrix}\begin{pmatrix}\frac{E[\boldsymbol{X}\dot{g}(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)\rho_2\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0]}\\\frac{E[\boldsymbol{Z}\rho_2\{g(\boldsymbol{Z}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0]}\end{pmatrix}^{T}\right\},\end{aligned}$$

$$\Omega=\left[\begin{pmatrix}\dot{g}(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)\boldsymbol{X}\\\boldsymbol{Z}\end{pmatrix}-\begin{pmatrix}\frac{E[\boldsymbol{X}\dot{g}(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)\rho_2\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0]}\\\frac{E[\boldsymbol{Z}\rho_2\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0]}\end{pmatrix}\right],$$

$\Sigma_1 = Var\left\{\frac{\delta}{\pi(\boldsymbol{U},Y)}q_1\{\eta_0, Y\}\Omega - \frac{\delta-\pi(\boldsymbol{U},Y)}{\pi(\boldsymbol{U},Y)}E[q_1\{\eta_0, Y\}|\boldsymbol{U}, \delta = 0]\Omega\right\}$

和$\eta_0 = g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0) + \boldsymbol{Z}^\top\boldsymbol{\beta}_0$.

3.4 模拟研究

对具有三种缺失数据机制的正态部分线性单指标模型和泊松部分线性单指标模型进行了模拟研究, 给出了所提方法的有限样本表现. 核函数取具有两个特定带宽 $\tau = 1.5\text{std}(U)n^{-1/3}$ 和 $\tau = 1.06\text{std}(X)n^{-1/5}$ 的高斯核函数, 其中 std(·) 表示样本标准差.

实验1 (恒等连接函数). 依据下面模型产生数据:

$$Y_i = 4\sin\left\{\frac{\pi}{2}\boldsymbol{X}_i^\top\boldsymbol{\alpha}\right\} + \boldsymbol{Z}_i^\top\boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n,$$

其中 $\boldsymbol{X}_i = (X_{i1}, X_{i2}, X_{i3})^\top$ 是三个变量独立产生于均匀分布 $U(0, 1)$, 协变量 $\boldsymbol{Z}_i = (Z_{i1}, Z_{i2})^\top$ 独立产生于一个二元正态分布, 均值为 0, 方差为 $\Sigma_Z = (\sigma_{zjk})_{2\times2}$ 其中 $\sigma_{z11} = \sigma_{z22} = 1$ 和 $\sigma_{z12} = \sigma_{z21} = 0.5$. 为了产生重尾分布的协变量, Z_{i2} 独立产生于自由度为 3 的 t -分布(即, $t(3)$), 和 ε_i 服从均值为 0, 方差为 0.2 的正态分布, 即 $\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0, 0.2)$. 参数 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 的真值为 $\boldsymbol{\alpha}^\top = (1, 1, 1)/\sqrt{3}$ 和 $\boldsymbol{\beta} = (0.8, 0.9)$. 假设 Y_i 有可能存在缺失, $\boldsymbol{X}_i$ 和 $\boldsymbol{Z}_i$ 是完全被观测的. 为了得到 Y_i 的缺失数据, 假设缺失指标 δ_i 产生于一个响应概率为 $\pi(\boldsymbol{X}_i, \boldsymbol{Z}_i, Y_i)$ 的 Bernoulli 分布. 这里考虑下面的关于 $\pi(\boldsymbol{X}_i, \boldsymbol{Z}_i, Y_i)$ 的不可忽略缺失数据机制为:

$$\text{logit}\{\pi(\boldsymbol{X}_i, Z_i, Y_i)\} = \gamma_0 + X_{1i}\gamma_1 + Z_{1i}\gamma_2 - \gamma Y_i$$

其中 $(\gamma_0, \gamma_1, \gamma_2, \gamma) = (1.0, 0.1, 0.1, 0.2)$.

考虑样本量 $n = 100$, 模拟 500 次. 平均缺失率为 43%. 为了比较, 计算了缺失机制在 MAR 下相应的估计. 表 3.1 给出可参数 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 的三个评价指标, 其中 'Bias' 是基于 500 次重复模拟估计值的平均和真值之间的差距, 'RMS' 基于 500 次重复模拟估计值和真值之间的均方根, 'SD' 基于 500 次重复模拟估计值的标准差.

观察表 3.1 可以看出在 NMAR 方法中参数 $\alpha_1, \alpha_2, \alpha_3, \beta_1$ 和 β_2 的偏差都非常的小且小于 0.09, SD 和 RMS 的值也都很小且小于 0.095. 在 MAR 方法中, 即在

等式 (3.3) 中 $\gamma = 0$, 参数 α_1, α_2 和 β_2 的 PS-WL 估计值是不准确的, 因为它们的偏差和 RMS 都相当的大, 这就暗示了缺失数据机制的错误识别会导致不合理的甚至错误的结论.

表3.1: 实验 1 中 PS-LL 估计量的表现

	NMAR			MAR		
Par.	Bias	SD	RMS	Bias	SD	RMS
α_1	-0.00161	0.04544	0.04544	-0.00193	0.04791	0.04792
α_2	-0.0068	0.08311	0.08334	-0.01061	0.11162	0.11207
α_3	-0.00367	0.07016	0.07022	-0.00433	0.06784	0.06795
β_1	0.08637	0.03626	0.09367	0.08000	0.03694	0.08811
β_2	-0.00114	0.02271	0.02272	-0.00570	0.02247	0.02317

实验 2 (泊松回归模型). 数据集 $\{(Y_i, X_i, Z_i) : i = 1, \ldots, n\}$ 生成如下. 对 $i = 1, \ldots, n$, 协变量 $X_i = (X_{1i}, X_{2i})^\top$ 独立产生于均匀分布 $U(0, 1)$; 协变量 Z_i 独立产生于一个二元正态分布, 均值为 0, 方差为 $\Sigma_Z = (\sigma_{zjk})_{2\times 2}$ 其中 $\sigma_{z11} = \sigma_{z22} = 1$ 和 $\sigma_{z12} = \sigma_{z21} = 0.5$; Y_i 从参数为 μ_i 的泊松分布中产生, 其中 $\eta_i = \log(\mu_i) = g(\boldsymbol{X}_i^\top\boldsymbol{\alpha}) + \boldsymbol{Z}_i^\top\boldsymbol{\beta}$, $g(t) = t^2$. 参数 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 的真值取 $\boldsymbol{\alpha} = (1/\sqrt{2}, 1/\sqrt{2})^\top$, 这满足 $\|\boldsymbol{\alpha}\| = 1$, 和 $\boldsymbol{\beta} = (0.8, 0.9)^\top$. 考虑下面的缺失数据机制:

$$\text{logit}\{\pi(\boldsymbol{X}_i, Z_i, Y_i)\} = \gamma_0 + X_{1i}\gamma_1 + Z_{1i}\gamma_2 - \gamma Y_i$$

其中 $(\gamma_0, \gamma_1, \gamma_2, \gamma) = (0.1, -0.1, -0.1, -0.2)$.

同样地, 考虑样本量 $n = 100$, 进行 500 次重复模拟. 平均缺失概率为 34%. 为了比较, 同样计算在 MAR 下相应的估计量. 表 3.2 给出了在以上考虑的设置下参数 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 的 Bias, RMS, SD 值. 观察表 3.2 同样可以看出在 NMAR 方法中参数 $\alpha_1, \alpha_2, \beta_1$ 和 β_2 的偏差都非常的小且小于 0.03, SD 和 RMS 的值也都很小且小于 0.12. 在 MAR 方法中, 即在等式 (3.3) 中 $\gamma = 0$, 参数 α_1, β_1 和 β_2 的 PS-WL 估计值是不准确的, 因为它们的偏差和 RMS 都相当的大, 这就暗示了缺失数据机制的错误识别会导致不合理的甚至错误的结论.

表3.2: 实验 2 中 PS-LL 估计量的表现

	NMAR			MAR		
Par.	Bias	SD	RMS	Bias	SD	RMS
α_1	-0.00256	0.11731	0.11722	-0.00326	0.11731	0.11724
α_2	-0.01723	0.11826	0.11939	-0.01631	0.11704	0.11806
β_1	-0.01532	0.06606	0.06775	-0.03312	0.06182	0.07008
β_2	-0.02752	0.07303	0.07798	-0.04701	0.07086	0.08498

3.5 定理证明

定理证明需要以下条件.

条件 A . 正则化假设:

(A1) 在 $\boldsymbol{U}$ 的支撑集上, $\boldsymbol{U}$ 的边际概率密度函数 $f(\mathbf{u})$ 远离 ∞, $E\{\exp(4\gamma Y)\}$ 矩是有限的, 函数 $E\{\exp(4\gamma Y)|\boldsymbol{U}\}f(\mathbf{u})$ 是有界的, 在一个包含 $\boldsymbol{U}$ 的支撑集的上, $m(\boldsymbol{U})$ 的真实函数是连续可微且有界的.

(A2) $\boldsymbol{K}(\cdot)$ 是一个 d_1-维核函数, 即, $\boldsymbol{K}(\cdot) = \prod_{i=1}^{d_1} K(\cdot)$, 其中 d_1 是的 $\boldsymbol{U}$ 的维数. 核函数 $K(\cdot)$ 是概率密度函数满足 (a1) 它是有界的, 并且具有紧支撑; (a2) 关于 τ 是有界导数; (a3) 它是对称的, 直到 $m-1$ 阶矩都是 0, m 阶矩非零.

(A3) 当 $n \to \infty$, 带宽 $h = h_{1n}$ 满足 $h_{1n} \to \infty$, $nh_{1n}^{d_1} \to \infty$, $\sqrt{n}h_{1n}^{d_1+2\tau}/\log n \to \infty$ 和 $nh_{1n}^{2m} \to 0$.

(A4) 有一个泛函的向量 $\mathcal{T}(Y, \boldsymbol{U}, \delta, \boldsymbol{\vartheta})$, 它对 $\boldsymbol{\vartheta} = (\vartheta_1, \vartheta_2)^T$ 是线性的, 使得:

(I) 对于足够小的 $\| \boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0 \|$, $\| \tilde{\mathrm{M}}(Y, \boldsymbol{U}, \delta, \boldsymbol{\vartheta}, \gamma_0) - \tilde{\mathrm{M}}(Y, \boldsymbol{U}, \delta, \boldsymbol{\vartheta}_0, \gamma_0) - \mathcal{T}(Y, \boldsymbol{U}, \delta, \boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) \| \leq c(Y, \boldsymbol{\vartheta}, \delta)(\|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0\|)^2$, 其中 $\tilde{\mathrm{M}}(Y, \boldsymbol{U}, \delta, \boldsymbol{\vartheta}, \gamma) = \{\delta[1 + \exp(\gamma Y)\boldsymbol{\vartheta}_1(\boldsymbol{U})/\boldsymbol{\vartheta}_2(\boldsymbol{U})] - 1\}(1, \boldsymbol{S}^T)^T$, $\boldsymbol{\vartheta}_0 = [E(1-\delta|\boldsymbol{U}), E\{\delta \exp(\gamma_0 Y)|\boldsymbol{U}\}]^T$ and $E[c(Y, \boldsymbol{U}, \delta)] < \infty$;

(II) $\|\mathcal{T}(Y, \boldsymbol{U}, \delta, \boldsymbol{\vartheta})\| \leq b(Y, \boldsymbol{U}, \delta)\|\boldsymbol{\vartheta}\|$ 和 $E[b(Y, \boldsymbol{U}, \delta)^2] < \infty$;

(III) 存在一个几乎处处连续函数 $\omega(\boldsymbol{U})$ 有 $\int \|\omega(\boldsymbol{U})\| d\boldsymbol{U} < \infty$, $E[\mathcal{T}(Y, \boldsymbol{U}, \delta, \boldsymbol{\vartheta})] = \int \boldsymbol{\vartheta}(\boldsymbol{U})\omega(\boldsymbol{U}) d\boldsymbol{U}$ 对所有的 $\|\boldsymbol{\vartheta}\| \leq \infty$, 对 $\varpi > 0$, 有 $E\left[\sup_{\|\boldsymbol{\varsigma}\| \leq \varpi} \|\omega(\boldsymbol{U} + \boldsymbol{\varsigma})\|^4\right] < \infty$.

(A5) 对于足够小的 $\|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0\|$, 在 γ_0 的邻域中的 γ, $\tilde{\mathrm{M}}(Y, \boldsymbol{U}, \delta, \boldsymbol{\vartheta}, \gamma)$ 是连续可微的, 对 $\epsilon > 0$, 带有 $E[t(Y, \boldsymbol{U}, \delta)] < \infty$ 的 $t(Y, \boldsymbol{U}, \delta)$ 使得 $\|\nabla_\gamma \tilde{\mathrm{M}}(Y, \boldsymbol{U}, \delta, \boldsymbol{\vartheta}, \gamma) -$

$\nabla_\gamma \tilde{\mathrm{M}}(Y,\boldsymbol{U},\delta,\boldsymbol{\vartheta}_0,\gamma_0)\| \le t(Y,\boldsymbol{U},\delta)(\|\gamma-\gamma_0\|^\epsilon + \|\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0\|^\epsilon)$, 和 $\Phi = E[\nabla_\gamma \tilde{\mathrm{M}}(Y,\boldsymbol{U},\delta,\boldsymbol{\vartheta}_0,\gamma_0)]$ 存在而且是满秩的.

(A6) 对于某个正常数 c_0, 概率函数 $\pi(\boldsymbol{U},Y;\gamma)$ 几乎处处满足 $\min_{1\le i\le n}\pi(\boldsymbol{U},Y;\gamma) \ge c_0 > 0$.

条件 B . 正则化假设:

(B1) $\boldsymbol{\alpha}^T\boldsymbol{X}_i$ 的密度函数 $f(\cdot)$ 是正的且在 $\mathcal{U} = \{\boldsymbol{\alpha}^T\boldsymbol{X} : \boldsymbol{X}\in\mathcal{X}, \alpha\in\Theta\}$ 上有一个连续的二阶导数, 其中 Θ 是关于 α 的一个紧的参数空间和 $\mathcal{X}$ 是一个 $\boldsymbol{X}_i$ 的紧支撑. 除此之外, $f(\cdot)$ 远离 0 , 而且在 $\boldsymbol{\alpha}_0$ 的邻域内 $\boldsymbol{\alpha}$ 是一致连续的.

(B2) 函数 $g(\cdot)$ 和 $V(\cdot)$ 都是二阶连续可微的, $g'''(\cdot)$ 是连续的, 其中 g''' 表示 g 的三阶导数.

(B3) 对于 $\eta_0 = g(\boldsymbol{\alpha}_0^T\boldsymbol{X}) + \boldsymbol{\beta}_0^T\boldsymbol{Z}$, $E[q_1^2(\eta_0,Y)|\boldsymbol{\alpha}_0^T\boldsymbol{X}=t]$, $E[q_1^2(\eta_0,Y)\boldsymbol{X}|\boldsymbol{\alpha}_0^T\boldsymbol{X}=t]$, $E[q_1^2(\eta_0,Y)\boldsymbol{X}\boldsymbol{X}^T|\boldsymbol{\alpha}_0^T\boldsymbol{X}=t]$, $E[q_1^2(\eta_0,Y)\boldsymbol{Z}|\boldsymbol{\alpha}_0^T\boldsymbol{X}=t]$ 和 $E[q_1^2(\eta_0,Y)\boldsymbol{Z}\boldsymbol{Z}^T|\boldsymbol{\alpha}_0^T\boldsymbol{X}=u]$ 关于 t 是连续函数. 而且对 $\varsigma\ge 2$ 有 $E[q_2^2(\eta_0,Y)] < \infty$ 和 $E[q_1^{2+\varsigma}(\eta_0,Y)] < \infty$.

(B4) 在本章中定义的矩阵 $\Sigma_i (i=1,2)$ 是正定的.

(B5) 核函数 K 是一个有界的对称概率密度函数, 并且是二次连续可微的, 满足

$$\int_{-\infty}^{\infty} u^2K(u)du \ne 0, \qquad \int_{-\infty}^{\infty}|u^i|K(u)du < \infty, \quad i=1,2,\cdots$$

(B6) 对响应变量范围内的 y, 函数 $q_2(x,y) < 0$.

注记 3.2 条件 A 是半参数两步估计文献中的一般假设, 同样见于 Shao (2016). 条件 (A6) 是缺失数据文献中常用的条件 (Tang et al. (2014)). 条件 B 是广义部分线性单指标文献中的一般假设, 见文献 Carroll et al. (1997) 和 Peng et al. (2014).

定理 3.1 的证明 定理 3.1 的证明可以按照 Shao and Wang (2016) 的类似思路得到, 因此在此处省略了细节.

定理 3.2 的证明 定理 3.2 的证明可以按照 Shao and Wang (2016)的类似思路得到, 因此在此处省略了细节.

定理 3.3 的证明 对于给定的 $\boldsymbol{\alpha},\boldsymbol{\beta}$ 和 $t = \mathbf{x}^\top\boldsymbol{\alpha}$, 令 $\hat{a} = \hat{g}(t;\boldsymbol{\alpha},\boldsymbol{\beta})$ 和 $\hat{b} = \hat{\dot{g}}(t;\boldsymbol{\alpha},\boldsymbol{\beta})$,

是下面有关 (a,b) 的目标函数的最大值点:

$$\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}_i(\gamma_0)}Q(G^{-1}\{a+b(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}-t)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}\},Y_i)K_h(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}-t).$$

进一步定义

$$\hat{\boldsymbol{\xi}}=\begin{pmatrix}\hat{a}-g(t)\\ h(\hat{b}-\dot{g}(t))\end{pmatrix},\quad \tilde{\boldsymbol{X}}_i=\begin{pmatrix}1\\ (\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}-t)/h\end{pmatrix}$$

和 $\eta_i=g(t)+\dot{g}(t)(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}-t)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}$. 则, 很容易验证 $\hat{\boldsymbol{\xi}}$ 也是下面有关 $\boldsymbol{\xi}$ 问题的最大值点

$$\mathcal{L}_n(\boldsymbol{\xi})=h\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}_i(\gamma_0)}\{Q(G^{-1}\{\tilde{\boldsymbol{X}}_i^{\top}\boldsymbol{\xi}+\eta_i\},Y_i)-Q(G^{-1}\{\eta_i\},Y_i)\}K_h(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}-t).$$

类似于 Carroll et al. (1997) 定理 1 的证明, 对函数 $Q(G^{-1}\{\cdot\},Y)$ 进行泰勒展开, 可以得到

$$\mathcal{L}_n(\boldsymbol{\xi})=B_n^T\boldsymbol{\xi}+\frac{1}{2}\boldsymbol{\xi}^TA_n\boldsymbol{\xi}\{1+o_p(1)\}\tag{3.9}$$

其中

$$B_n=h\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i,Y_i;\gamma_0)}q_1(\eta_i,Y_i)\tilde{\boldsymbol{X}}_iK_h(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}-t)$$

和

$$A_n=h\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i,Y_i;\gamma_0)}q_2(\eta_i,Y_i)\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{X}}_i^TK_h(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}-t).$$

令

$$\boldsymbol{D}_0=\begin{pmatrix}1 & \frac{\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t}{h}\\ \frac{\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t}{h} & \frac{(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t)^2}{h}\end{pmatrix},\quad \boldsymbol{D}=\begin{pmatrix}1 & \frac{\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}-t}{h}\\ \frac{\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}-t}{h} & \frac{(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}-t)^2}{h}\end{pmatrix},$$

$\eta_{0i} = g(\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0) + \boldsymbol{Z}_i^\top \boldsymbol{\beta}_0$, 我们有

$$\begin{aligned}
\frac{A_n}{nh} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}_i(\boldsymbol{U}_i, Y_i; \gamma_0)} K_h(\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t)\boldsymbol{D}_0 q_2(\eta_{0i}, Y_i) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}_i(\boldsymbol{U}_i, Y_i; \gamma_0)} K_h(\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t)\boldsymbol{D}_0 [q_2(\eta_i, Y_i) - q_2(\eta_{0i}, Y_i)] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}_i(\boldsymbol{U}_i, Y_i; \gamma_0)} K_h(\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t)[\boldsymbol{D} - \boldsymbol{D}_0] q_2(\eta_{0i}, Y_i) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}_i(\boldsymbol{U}_i, Y_i; \gamma_0)} K_h(\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t)[\boldsymbol{D}_0 - \boldsymbol{D}][q_2(\eta_{0i}, Y_i) - q_2(\eta_i, Y_i)] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}_i(\boldsymbol{U}_i, Y_i; \gamma_0)} [K_h(\boldsymbol{X}_i^\top \boldsymbol{\alpha} - t) - K_h(\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t)]\boldsymbol{D}_0 q_2(\eta_{0i}, Y_i) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}_i(\boldsymbol{U}_i, Y_i; \gamma_0)} [K_h(\boldsymbol{X}_i^\top \boldsymbol{\alpha} - t) - K_h(\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t)]\boldsymbol{D}_0 [q_2(\eta_i, Y_i) - q_2(\eta_{0i}, Y_i)] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}_i(\boldsymbol{U}_i, Y_i; \gamma_0)} [K_h(\boldsymbol{X}_i^\top \boldsymbol{\alpha} - t) - K_h(\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t)][\boldsymbol{D} - \boldsymbol{D}_0] q_2(\eta_{0i}, Y_i) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}_i(\boldsymbol{U}_i, Y_i; \gamma_0)} [K_h(\boldsymbol{X}_i^\top \boldsymbol{\alpha} - t) - K_h(\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t)][\boldsymbol{D} - \boldsymbol{D}_0][q_2(\eta_i, Y_i) - q_2(\eta_{0i}, Y_i)] \\
&=: A_{n1} + A_{n2} + A_{n3} + A_{n4} + A_{n5} + A_{n6} + A_{n7} + A_{n8}
\end{aligned} \tag{3.10}$$

对于 A_{n1}, 可以得到

$$\begin{aligned}
A_{n1} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i(\boldsymbol{U}_i, Y_i; \gamma_0)} K_\tau(\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t)\boldsymbol{D}_0 q_2(\eta_{0i}, Y_i) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\delta_i \exp(\gamma_0 Y_i)[\hat{W}(\boldsymbol{U}_i, \gamma_0) - W(\boldsymbol{U}_i, \gamma_0)] K_h(\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t)\boldsymbol{D}_0 q_2(\eta_{0i}, Y_i) \\
&:= \begin{pmatrix} E_{11} & E_{12} \\ E_{12} & E_{13} \end{pmatrix} + \begin{pmatrix} E_{21} & E_{22} \\ E_{22} & E_{23} \end{pmatrix}
\end{aligned} \tag{3.11}$$

其中

$$E_{11} = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i(\boldsymbol{U}_i, Y_i; \gamma_0)} K_h(\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t) q_2(\eta_{0i}, Y_i),$$

$$E_{12}=\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i(\boldsymbol{U}_i,Y_i;\gamma_0)}K_h(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t)\frac{\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t_0}{h}q_2(\eta_{0i},Y_i),$$

$$E_{13}=\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i(\boldsymbol{U}_i,Y_i;\gamma_0)}K_h(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t)\left(\frac{\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t}{h}\right)^2q_2(\eta_{0i},Y_i)$$

$$E_{21}=\frac{1}{n}\sum_{i=1}^{n}\delta_i\exp(\gamma_0Y_i)[\hat{W}(\boldsymbol{U}_i,\gamma_0)-W(\boldsymbol{U}_i,\gamma_0)]K_h(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t)q_2(\eta_{0i},Y_i),$$

$$E_{22}=\frac{1}{n}\sum_{i=1}^{n}\delta_i\exp(\gamma_0Y_i)[\hat{W}(\boldsymbol{U}_i,\gamma_0)-W(\boldsymbol{U}_i,\gamma_0)]K_h(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t)\frac{\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t}{h}q_2(\eta_{0i},Y_i),$$

$$E_{23}=\frac{1}{n}\sum_{i=1}^{n}\delta_i\exp(\gamma_0Y_i)[\hat{W}(\boldsymbol{U}_i,\gamma_0)-W(\boldsymbol{U}_i,\gamma_0)]K_h(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t)\left(\frac{\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t}{h}\right)^2q_2(\eta_{0i},Y_i)$$

直接计算可以得到

$$E_{11}=-f(t)E[\rho_2\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0=t]+O_p(h^2+\sqrt{\log n/nh}),$$

$$E_{12}=O_p(h^2+\sqrt{\log n/nh}),$$

$$E_{13}=-f(t)\kappa_2E[\rho_2\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0=t]+O_p(h^2+\sqrt{\log n/nh}).$$

根据核估计的一个标准推导, 可以看出条件 (A2) 暗含了

$$\max_{1\le i\le n}\mid\hat{W}(\boldsymbol{U}_i,\gamma_0)-W(\boldsymbol{U}_i,\gamma_0)\mid=o_p(n^{-1/4})$$

直接计算可以得到

$$\frac{1}{n}\sum_{i=1}^{n}K_h(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t)\left(\frac{\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t}{h}\right)^{\upsilon}q_2(\eta_{0i},Y_i)\xrightarrow{\mathcal{P}}f(t)\kappa_{\upsilon}E\{q_2(\eta_0,Y)|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0=t\}<\infty,$$

其中 $\kappa_{\upsilon}=\int u^{\upsilon}K(u)du,\upsilon=0,1,2$. 因此, 有

$$|E_{21}|<Co_p(n^{-1/4})=O_p(n^{-1/4}),|E_{22}|<O_p(n^{-1/4}),|E_{23}|<O_p(n^{-1/4}),$$

则 $E_{2i}\xrightarrow{\mathcal{P}}0,i=1,2,3.$

结合上面的结论可以得到

$$A_{n1}=-f(t)\begin{pmatrix}1&0\\0&\kappa_2\end{pmatrix}E[\rho_2\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0=t]+O_p(h^2+\sqrt{\log n/nh})+O_p(n^{-1/4})$$

对 A_{n2}, 注意到

$$
\begin{aligned}
&q_2(\eta_i, Y_i) - q_2(\eta_{0i}, Y_i)\\
&= \dot{q}_2(\eta_{0i}, Y_i)[\boldsymbol{Z}_i^\top\boldsymbol{\beta} + g(t) + \dot{g}(t)(\boldsymbol{X}_i^\top\boldsymbol{\alpha} - t) - g(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0) - \boldsymbol{Z}_i^\top\boldsymbol{\beta}_0] + o_p(1)\\
&= \dot{q}_2(\eta_{0i}, Y_i)\{g(\boldsymbol{X}_i^\top\boldsymbol{\alpha}) - g(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0) - [g(\boldsymbol{X}_i^\top\boldsymbol{\alpha}) - g(t) - \dot{g}(t)(\boldsymbol{X}_i^\top\boldsymbol{\alpha} - t)] + \boldsymbol{Z}_i^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\} + o_p(1)\\
&= \dot{q}_2(\eta_{0i}, Y_i)[g(\boldsymbol{X}_i^\top\boldsymbol{\alpha}) - g(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0)] - \dot{q}_2(\eta_{0i}, Y_i)[g(\boldsymbol{X}_i^\top\boldsymbol{\alpha}) - g(t) - \dot{g}(t)(\boldsymbol{X}_i^\top\boldsymbol{\alpha} - t)]\\
&+ \dot{q}_2(\eta_{0i}, Y_i)\boldsymbol{Z}_i^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o_p(1)\\
&=: I_{1i} + I_{2i} + I_{3i}
\end{aligned}
$$

有

$$
\begin{aligned}
A_{n2} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}_i(\boldsymbol{U}_i, Y_i; \gamma_0)}K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)\boldsymbol{D}_0 I_{1i}\\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}_i(\boldsymbol{U}_i, Y_i; \gamma_0)}K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)\boldsymbol{D}_0 I_{2i}\\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}_i(\boldsymbol{U}_i, Y_i; \gamma_0)}K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)\boldsymbol{D}_0 I_{3i} + o_p(1)\\
&=: A_{n21} + A_{n22} + A_{n23}
\end{aligned}
\tag{3.12}
$$

类似于 A_{n1}, 也可以得到

$$
\begin{aligned}
A_{n21} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i(\boldsymbol{U}_i, Y_i; \gamma_0)}K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)\boldsymbol{D}_0 I_{1i}\\
&+ \frac{1}{nh}\sum_{i=1}^{n}\delta_i\exp(\gamma_0 Y_i)[\hat{W}(\boldsymbol{U}_i, \gamma_0) - W(\boldsymbol{U}_i, \gamma_0)]K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)\boldsymbol{D}_0 I_{1i}\\
&=: \begin{pmatrix} S_{11} & S_{12}\\ S_{12} & S_{13}\end{pmatrix} + \begin{pmatrix} S_{21} & S_{22}\\ S_{22} & S_{23}\end{pmatrix}
\end{aligned}
$$

其中

$$
\begin{aligned}
S_{11} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i(\boldsymbol{U}_i, Y_i; \gamma_0)}K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)I_{1i},\\
S_{12} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i(\boldsymbol{U}_i, Y_i; \gamma_0)}K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)\frac{\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t}{h}I_{1i},
\end{aligned}
$$

$$S_{13} = \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{\pi_i(\boldsymbol{U}_i, Y_i; \gamma_0)} K_h(\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t)\left(\frac{\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t}{h}\right)^2 I_{1i}$$

$$S_{21} = \frac{1}{n}\sum_{i=1}^{n} \delta_i \exp(\gamma_0 Y_i)[\hat{W}(\boldsymbol{U}_i, \gamma_0) - W(\boldsymbol{U}_i, \gamma_0)] K_h(\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t) I_{1i},$$

$$S_{22} = \frac{1}{n}\sum_{i=1}^{n} \delta_i \exp(\gamma_0 Y_i)[\hat{W}(\boldsymbol{U}_i, \gamma_0) - W(\boldsymbol{U}_i, \gamma_0)] K_h(\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t)\frac{\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t}{h} I_{1i},$$

$$S_{23} = \frac{1}{n}\sum_{i=1}^{n} \delta_i \exp(\gamma_0 Y_i)[\hat{W}(\boldsymbol{U}_i, \gamma_0) - W(\boldsymbol{U}_i, \gamma_0)] K_h(\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t)\left(\frac{\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t}{h}\right)^2 I_{1i}.$$

直接计算可以得到

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{\pi_i(\boldsymbol{U}_i, Y_i; \gamma_0)} K_h(\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t)\left(\frac{\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t}{h}\right)^\upsilon \dot{q}_2(\eta_{0i}, Y_i)$$
$$\xrightarrow{\mathcal{P}} f(t)\kappa_\upsilon E\{\dot{q}_2(\eta_0, Y)|\boldsymbol{X}^\top \boldsymbol{\alpha}_0 = t\} < \infty$$

其中 $\kappa_\upsilon = \int u^\upsilon K(u)du, \upsilon = 0, 1, 2.$

注意

$$g(\boldsymbol{X}_i^\top \boldsymbol{\alpha}) - g(\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0) = \dot{g}(\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0)(\boldsymbol{X}_i^\top \boldsymbol{\alpha} - \boldsymbol{X}_i^\top \boldsymbol{\alpha}_0) + O_p(n^{-1/2}),$$

因此, 有

$$|S_{1i}| < Cn^{-1/2}, i = 1, 2, 3.$$

因为 $n^{1/2} \to \infty$, 有 $S_{1i} \xrightarrow{\mathcal{P}} 0, i = 1, 2, 3.$ 类似于 A_{n1} 和 S_{11}, 可以得到 $S_{2i} \xrightarrow{\mathcal{P}} 0, i = 1, 2, 3.$ 所以, 有 $A_{n21} = o_p(1), A_{n22} = A_{n23} = o_p(1).$ 类似地, 有 $A_{nj} \xrightarrow{\mathcal{P}} 0, j = 2, \ldots, 8.$ 因此, 很容易可以得到

$$A_n = -A + O_p\{h^2 + \sqrt{\log n/(nh)}\} + O_p(n^{-1/4}) \tag{3.13}$$

其中 $A = f(t)\begin{pmatrix} 1 & 0 \\ 0 & \kappa_2 \end{pmatrix} E[\rho_2\{g(\boldsymbol{X}^\top \boldsymbol{\alpha}_0) + \boldsymbol{Z}^\top \boldsymbol{\beta}_0\}|\boldsymbol{X}^\top \boldsymbol{\alpha}_0 = t].$

令

$$\boldsymbol{D}_0^* = \begin{pmatrix} 1 \\ \frac{\boldsymbol{X}_i^\top \boldsymbol{\alpha}_0 - t}{h} \end{pmatrix}, \quad \boldsymbol{D}^* = \begin{pmatrix} 1 \\ \frac{\boldsymbol{X}_i^\top \boldsymbol{\alpha} - t}{h} \end{pmatrix},$$

那么有

$$\begin{aligned}
\frac{B_n}{hn} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i, Y_i; \gamma_0)} K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)\boldsymbol{D}_0^* q_1(\eta_{0i}, Y_i) \\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i, Y_i; \gamma_0)} K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)\boldsymbol{D}_0^*[q_1(\eta_i, Y_i) - q_1(\eta_{0i}, Y_i)] \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i, Y_i; \gamma_0)} K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)[\boldsymbol{D}^* - \boldsymbol{D}_0^*] q_1(\eta_{0i}, Y_i) \\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i, Y_i; \gamma_0)} K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)[\boldsymbol{D}_0^* - \boldsymbol{D}^*][q_1(\eta_{0i}, Y_i) - q_1(\eta_i, Y_i)] \\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i, Y_i; \gamma_0)} [K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha} - t) - K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)]\boldsymbol{D}_0^* q_1(\eta_{0i}, Y_i) \\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i, Y_i; \gamma_0)} [K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha} - t) - K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)]\boldsymbol{D}_0^*[q_1(\eta_i, Y_i) - q_1(\eta_{0i}, Y_i)] \\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i, Y_i; \gamma_0)} [K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha} - t) - K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)][\boldsymbol{D}^* - \boldsymbol{D}_0^*] q_1(\eta_{0i}, Y_i) \\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i, Y_i; \gamma_0)} [K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha} - t) - K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)][\boldsymbol{D}^* - \boldsymbol{D}_0^*][q_1(\eta_i, Y_i) - q_1(\eta_{0i}, Y_i)] \\
&=: B_{n1} + B_{n2} + B_{n3} + B_{n4} + B_{n5} + B_{n6} + B_{n7} + B_{n8}
\end{aligned} \tag{3.14}$$

对于 B_{n1}, 有

$$\begin{aligned}
B_{n1} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi(\boldsymbol{U}_i, Y_i; \gamma_0)}\{\varphi_0^* - E[\varphi_0^*|\boldsymbol{U}_i, \delta_i]\} + \frac{1}{n}\sum_{i=1}^{n} E\{\varphi_0^*|\boldsymbol{U}_i, \delta_i\} \\
&+ \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i, Y_i; \gamma_0)} - \frac{\delta_i}{\pi(\boldsymbol{U}_i, Y_i; \gamma_0)}\right\}\{\varphi_0^* - E[\varphi_0^*|\boldsymbol{U}_i, \delta_i]\} \\
&+ \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i, Y_i; \gamma_0)} - 1\right\}\{E[\varphi_0^*|\boldsymbol{U}_i, \delta_i]\} \\
&=: B_{n11} + B_{n12} + B_{n13} + B_{n14}
\end{aligned} \tag{3.15}$$

其中 $\varphi_0^* = K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)\boldsymbol{D}_0^* q_1(\eta_{0i}, Y_i)$. 依据 Shao and Wang (2016) 和 Zhao, Tang and Tang (2014) 类似的思路, 有 $B_{n13} = o_p(1)$ 和 $B_{n14} = o_p(1)$.

对于 B_2, 可以得到

$$\begin{aligned}B_{n2} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi(\boldsymbol{U}_j, Y_j; \gamma_0)}K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)\boldsymbol{D}_0^*[q_1(\eta_i, Y_i) - q_1(\eta_{0i}, Y_i)] \\ &+ \frac{1}{n}\sum_{i=1}^{n}\delta_i \exp(\gamma_0 Y_i)[\hat{W}(\boldsymbol{U}_i;\gamma_0) - W(\boldsymbol{U}_i;\gamma_0)]K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)\boldsymbol{D}_0^*[q_1(\eta_i, Y_i) - q_1(\eta_{0i}, Y_i)] \\ &=: B_{21} + B_{22}\end{aligned} \tag{3.16}$$

类似于 A_{n2}, 有

$$\begin{aligned}B_{n2} &= -f(t)\begin{pmatrix}1\\0\end{pmatrix}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^\top\dot{g}(t)E[\boldsymbol{X}\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0) + \boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}^\top\boldsymbol{\alpha}_0 = t] \\ &+ f(t)h^2\begin{pmatrix}\kappa_2\\0\end{pmatrix}\ddot{g}(t)E[\boldsymbol{X}\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0) + \boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}^\top\boldsymbol{\alpha}_0 = t] \\ &- f(t)\begin{pmatrix}1\\0\end{pmatrix}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top E[\boldsymbol{Z}\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0) + \boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}^\top\boldsymbol{\alpha}_0 = t] \\ &+ o_p(n^{-1/2}) + O_p\{h^2 + \sqrt{\log n/(nh)}\}.\end{aligned} \tag{3.17}$$

类似地可以证明 $B_{nj} \xrightarrow{\mathcal{P}} 0, j = 3, \ldots, 8.$

综合以上结果可以得到

$$\begin{aligned}\frac{B_n}{nh} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi(\boldsymbol{U}_i, Y_i; \gamma_0)}K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)\begin{pmatrix}1\\ \frac{\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t}{h}\end{pmatrix}q_1(\eta_{0i}, Y_i) \\ &+ \frac{1}{n}\sum_{i=1}^{n}\left[1 - \frac{\delta_i}{\pi(\boldsymbol{U}_i, Y_i; \gamma_0)}\right]\begin{pmatrix}E[K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)q_1(\eta_{0i}, Y_i)|\boldsymbol{U}_i, \delta_i = 0] \\ E[K_h(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)[(\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0 - t)/h]q_1(\eta_{0i}, Y_i)|\boldsymbol{U}_i, \delta_i = 0]\end{pmatrix} \\ &- f(t)\begin{pmatrix}1\\0\end{pmatrix}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^\top\dot{g}(t)E[\boldsymbol{X}\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0) + \boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}^\top\boldsymbol{\alpha}_0 = t] \\ &+ f(t)h^2\begin{pmatrix}\kappa_2\\0\end{pmatrix}\ddot{g}(t)E[\boldsymbol{X}\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0) + \boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}^\top\boldsymbol{\alpha}_0 = t]\end{aligned}$$

$$- f(t)\begin{pmatrix}1\\0\end{pmatrix}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^{\top}E[\boldsymbol{Z}\rho_2\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0=t]$$

$$+ o_p(n^{-1/2}h^{-1/2}) + o_p(n^{-1/2}) + O_p\{h^2 + \sqrt{\log n/(nh)}\}. \tag{3.18}$$

结合上面的结果, 有

$$\begin{aligned}\hat{g}(t;\boldsymbol{\alpha},\boldsymbol{\beta}) - g(t) = {}& \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}\frac{K_h(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t)\varepsilon_i}{f(t)E[\rho_2\{g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0=t]}\\ & - \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i-\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}{\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}\frac{E[K_h(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t)\varepsilon_i|\boldsymbol{U}_i,\delta_i=0]}{f(t)E[\rho_2\{g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0=t]}\\ & - \begin{pmatrix}\frac{\dot{g}(t)E[\boldsymbol{X}\rho_2\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0=t]}{E[\rho_2\{g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0=t]}\\ \frac{E[\boldsymbol{Z}\rho_2\{g(\boldsymbol{Z}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0=t]}{E[\rho_2\{g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0=t]}\end{pmatrix}^{\top}\begin{pmatrix}\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\\ \boldsymbol{\beta}-\boldsymbol{\beta}_0\end{pmatrix}\\ & + \frac{1}{2}\ddot{g}(t)h^2\kappa_2 + o_p(n^{-1/2}) + O_p\left(h^2+\sqrt{\frac{\log n}{nh}}\right)\end{aligned}$$

$$\begin{aligned}\hat{\dot{g}}(t;\boldsymbol{\alpha},\boldsymbol{\beta}) - \dot{g}(t) = {}& \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}\frac{K_h(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t)[(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t)/h]\varepsilon_i}{f(t)E[\rho_2\{g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0=t]}\\ & - \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i-\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}{\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}\frac{E[K_h(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t)[(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-t)/h]\varepsilon_i|\boldsymbol{U}_i,\delta_i=0]}{f(t)E[\rho_2\{g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0=t]}\\ & + o_p(n^{-1/2}) + O_p\left(h^2+\sqrt{\frac{\log n}{nh}}\right)\end{aligned}$$

其中 $\varepsilon_i = [Y_i - G^{-1}\{g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0\}]\rho_1\{g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0\}$.

定理 3.4 的证明 使用与定理 3.3 的证明中类似的记号, 记 $\hat{\alpha}^1$ 和 $\hat{\beta}^1$ 是 α_0 和 β_0 从第 3.2 节提出的算法经过一次迭代所获得的估计量. 则, $\hat{\alpha}^1$ 和 $\hat{\beta}^1$ 是以下损失函数的最大值点:

$$\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i,Y_i;\gamma_0)}Q(G^{-1}\{\hat{g}(\boldsymbol{X}_i^{\top}\hat{\boldsymbol{\alpha}}^0;\hat{\boldsymbol{\alpha}}^0,\hat{\boldsymbol{\beta}}^0)+\hat{\dot{g}}(\boldsymbol{X}_i^{\top}\hat{\boldsymbol{\alpha}}^0;\hat{\boldsymbol{\alpha}}^0,\hat{\boldsymbol{\beta}}^0)(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}-\boldsymbol{X}_i^{\top}\hat{\boldsymbol{\alpha}}^0)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}\}, Y_i).$$

令 $\eta_{0i} = g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0$, $\eta_i^* = \hat{g}(\boldsymbol{X}_i^{\top}\hat{\boldsymbol{\alpha}}^0;\hat{\boldsymbol{\alpha}}^0,\hat{\boldsymbol{\beta}}^0)+\hat{\dot{g}}(\boldsymbol{X}_i^{\top}\hat{\boldsymbol{\alpha}}^0;\hat{\boldsymbol{\alpha}}^0,\hat{\boldsymbol{\beta}}^0)(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0-\boldsymbol{X}_i^{\top}\hat{\boldsymbol{\alpha}}^0)+\boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0$,

$$\hat{\boldsymbol{\theta}} = \sqrt{n}\begin{pmatrix}\hat{\boldsymbol{\alpha}}^1-\boldsymbol{\alpha}_0\\ \hat{\boldsymbol{\beta}}^1-\boldsymbol{\beta}_0\end{pmatrix},\quad \Delta_i = \begin{pmatrix}\hat{\dot{g}}(\boldsymbol{X}_i^{\top}\hat{\boldsymbol{\alpha}}^0;\hat{\boldsymbol{\alpha}}^0,\hat{\boldsymbol{\beta}}^0)\boldsymbol{X}_i\\ \boldsymbol{Z}_i\end{pmatrix},\quad \Delta_{0i} = \begin{pmatrix}\dot{g}(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0)\boldsymbol{X}_i\\ \boldsymbol{Z}_i\end{pmatrix}.$$

则, $\hat{\boldsymbol{\theta}}$ 是下面损失函数的最大值点:

$$\mathcal{L}_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i, Y_i; \gamma_0)} Q(G^{-1}\{\eta_i^* + \Delta_i^\top \boldsymbol{\theta}/\sqrt{n}\}, Y_i) - \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i, Y_i; \gamma_0)} Q(G^{-1}\{\eta_i^*\}, Y_i)$$

依据泰勒展开, 有

$$\mathcal{L}_n(\boldsymbol{\theta}) = \mathbb{B}_n^\top \boldsymbol{\theta} + \frac{1}{2}\boldsymbol{\theta}^T \mathbb{A}_n \boldsymbol{\theta}\{1 + o_p(1)\} \tag{3.19}$$

其中

$$\mathbb{B}_n = n^{-1/2} \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i, Y_i; \gamma_0)} q_1\{\eta_i^*, Y_i\}\Delta_i$$

和

$$\mathbb{A}_n = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i, Y_i; \gamma_0)} q_2\{\eta_i^*, Y_i\}\Delta_i\Delta_i^\top$$

因为 $q_2\{\eta_i^*, Y_i\} = q_2\{\eta_{0i}, Y_i\} + o_p(1), \Delta_i = \Delta_{0i} + o_p(1)$, 有

$$\begin{aligned}
\mathbb{A}_n &= \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i, Y_i; \gamma_0)} q_2\{\eta_{0i}, Y_i\}\Delta_{0i}\Delta_{0i}^\top + o_p(1) \\
&= \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{\pi_i(\boldsymbol{U}_i, Y_i; \gamma_0)} q_2(\eta_{0i}, Y_i)\Delta_{0i}\Delta_{0i}^\top + o_p(1) \\
&+ \frac{1}{nh} \sum_{i=1}^{n} \delta_i \exp(\gamma_0 Y_i)[\hat{W}(\boldsymbol{U}_i, \gamma_0) - W(\boldsymbol{U}_i, \gamma_0)] q_2(\eta_{0i}, Y_i)\Delta_{0i}\Delta_{0i}^\top + o_p(1) \\
&=: \mathbb{A}_{n1} + \mathbb{A}_{n2}
\end{aligned} \tag{3.20}$$

类似于定理 3.3, 有 $\mathbb{A}_{n2} \xrightarrow{\mathcal{P}} 0$ 和 $\mathbb{A}_{n1} = -f(t)E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0) + \boldsymbol{Z}^\top\boldsymbol{\beta}_0\}\Delta_0\Delta_0^\top|\boldsymbol{X}^\top\boldsymbol{\alpha}_0 = t_0] + O_p(h^2 + \sqrt{\log n/(nh)})$. 因此, 有

$$\begin{aligned}
\mathbb{A}_n &= -f(t)E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0) + \boldsymbol{Z}^\top\boldsymbol{\beta}_0\}\Delta_0\Delta_0^\top|\boldsymbol{X}^\top\boldsymbol{\alpha}_0 = t_0] + O_p(h^2 + \sqrt{\log n/(nh)}) \\
&= -\mathbb{A} + o_p(h^2 + \sqrt{\log n/(nh)})
\end{aligned} \tag{3.21}$$

其中 $\mathbb{A} = f(t)E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0) + \boldsymbol{Z}^\top\boldsymbol{\beta}_0\}\Delta_0\Delta_0^\top|\boldsymbol{X}^\top\boldsymbol{\alpha}_0 = t]$. 则, 等式 (3.19) 简化为

$$\mathcal{L}_n(\boldsymbol{\theta}) = \mathbb{B}_n^\top \boldsymbol{\theta} - \frac{1}{2}\boldsymbol{\theta}^T \mathbb{A}\boldsymbol{\theta} + o_p(1).$$

利用凸函数的性质, 有

$$\hat{\boldsymbol{\theta}} = \mathbb{A}^{-1}\mathbb{B}_n^{\top} + o_p(1) \tag{3.22}$$

对于 $\mathbb{B}_n$, 根据泰勒展开, 有

$$\begin{aligned}\mathbb{B}_n &= n^{-1/2}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i, Y_i; \gamma_0)}q_1\{\eta_{0i}, Y_i\}\Delta_i \\ &\quad + n^{-1/2}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i, Y_i; \gamma_0)}q_2\{\eta_{0i}, Y_i\}[\eta_i^* - \eta_{0i}]\Delta_i + o_p(1) \\ &=: \mathbb{B}_{n1} + \mathbb{B}_{n2} + o_p(1)\end{aligned} \tag{3.23}$$

类似于定理 3.3, 有

$$\begin{aligned}\mathbb{B}_{n1} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i(\boldsymbol{U}_i, Y_i; \gamma_0)}q_1(\eta_{0i}, Y_i)\Delta_{0i} \\ &\quad - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i - \pi_i(\boldsymbol{U}_i, Y_i; \gamma_0)}{\pi_i(\boldsymbol{U}_i, Y_i; \gamma_0)}E\{q_1(\eta_0, Y)\Delta_0|\boldsymbol{U}_i, \delta_i = 0\} + o_p(1)\end{aligned}$$

因为

$$\begin{aligned}&\eta_i^* - \eta_{0i} = \hat{g}(\boldsymbol{X}_i^{\top}\hat{\boldsymbol{\alpha}}^0; \hat{\boldsymbol{\alpha}}^0, \hat{\boldsymbol{\beta}}^0) + \dot{\hat{g}}(\boldsymbol{X}_i^{\top}\hat{\boldsymbol{\alpha}}^0; \hat{\boldsymbol{\alpha}}^0, \hat{\boldsymbol{\beta}}^0)(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0 - \boldsymbol{X}_i^{\top}\hat{\boldsymbol{\alpha}}^0) - g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0) \\ &= \hat{g}(\boldsymbol{X}_i^{\top}\hat{\boldsymbol{\alpha}}^0; \hat{\boldsymbol{\alpha}}^0, \hat{\boldsymbol{\beta}}^0) - g(\boldsymbol{X}_i^{\top}\hat{\boldsymbol{\alpha}}^0) + [\dot{\hat{g}}(\boldsymbol{X}_i^{\top}\hat{\boldsymbol{\alpha}}^0; \hat{\boldsymbol{\alpha}}^0, \hat{\boldsymbol{\beta}}^0) - \dot{g}(\boldsymbol{X}_i^{\top}\hat{\boldsymbol{\alpha}}^0)](\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0 - \boldsymbol{X}_i^{\top}\hat{\boldsymbol{\alpha}}^0) + o_p(n^{-1/2})\end{aligned}$$

根据定理 3.3的结论, 同样有

$$\begin{aligned}\eta_i^* - \eta_{0i} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi(\boldsymbol{U}_i, Y_i; \gamma_0)}\frac{K_h(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0 - t_0)\varepsilon_i}{f(t_0)E[\rho_2\{g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0) + \boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0 = t_0]} \\ &\quad - \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i - \pi(\boldsymbol{U}_i, Y_i; \gamma_0)}{\pi(\boldsymbol{U}_i, Y_i; \gamma_0)}\frac{E[K_h(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0 - t_0)\varepsilon_i|\boldsymbol{U}_i, \delta_i = 0]}{f(t_0)E[\rho_2\{g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0) + \boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0 = t_0]} \\ &\quad - \begin{pmatrix} \frac{\dot{g}(t)E[\boldsymbol{X}\rho_2\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0) + \boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0 = t_0]}{E[\rho_2\{g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0) + \boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0 = t_0]} \\ \frac{E[\boldsymbol{Z}\rho_2\{g(\boldsymbol{Z}^{\top}\boldsymbol{\alpha}_0) + \boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0 = t_0]}{E[\rho_2\{g(\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0) + \boldsymbol{Z}_i^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^{\top}\boldsymbol{\alpha}_0 = t_0]} \end{pmatrix}^{\top}\begin{pmatrix}\hat{\boldsymbol{\alpha}}^0 - \boldsymbol{\alpha}_0 \\ \hat{\boldsymbol{\beta}}^0 - \boldsymbol{\beta}_0\end{pmatrix} + o_p(1)\end{aligned} \tag{3.24}$$

对于 $\mathbb{B}_{n2}$, 类似于 $\mathbb{A}_{n2}$, 有

$$
\begin{aligned}
\mathbb{B}_{n2} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\hat{\pi}(\boldsymbol{U}_i,Y_i;\gamma_0)}q_2\{\eta_{0i},Y_i\}[\eta_i^*-\eta_{0i}]\Delta_{0i}+o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i q_2\{\eta_{0i},Y_i\}\Delta_{0i}}{\hat{\pi}(\boldsymbol{U}_i,Y_i;\gamma_0)}\frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j}{\pi(\boldsymbol{U}_j,Y_j;\gamma_0)}\frac{K_h(\boldsymbol{X}_j^\top\boldsymbol{\alpha}_0-X_i^\top\boldsymbol{\alpha}_0)\varepsilon_j}{f(t_0)E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|X_i^\top\boldsymbol{\alpha}_0]}\\
&- \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i q_2\{\eta_{0i},Y_i\}\Delta_{0i}}{\hat{\pi}(\boldsymbol{U}_i,Y_i;\gamma_0)}\frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j-\pi(\boldsymbol{U}_j,Y_j;\gamma_0)}{\pi(\boldsymbol{U}_j,Y_j;\gamma_0)}\frac{E[K_h(\boldsymbol{X}_j^\top\boldsymbol{\alpha}_0-X_i^\top\boldsymbol{\alpha}_0)\varepsilon|\boldsymbol{U}_j,\delta_j=0]}{f(t_0)E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0]}\\
&- \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i q_2\{\eta_{0i},Y_i\}\Delta_{0i}}{\hat{\pi}(\boldsymbol{U}_i,Y_i;\gamma_0)}\begin{pmatrix}\frac{E[\dot{g}(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)\boldsymbol{X}\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0]}\\ \frac{E[\boldsymbol{Z}\rho_2\{g(\boldsymbol{Z}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0]}\end{pmatrix}^\top\begin{pmatrix}\hat{\boldsymbol{\alpha}}^0-\boldsymbol{\alpha}_0\\ \hat{\boldsymbol{\beta}}^0-\boldsymbol{\beta}_0\end{pmatrix}+o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}\varepsilon_i\frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j q_2\{\eta_{0j},Y_j\}\Delta_{0j}}{\hat{\pi}(\boldsymbol{U}_j,Y_j;\gamma_0)}\frac{K_h(\boldsymbol{X}_j^\top\boldsymbol{\alpha}_0-X_i^\top\boldsymbol{\alpha}_0)}{f(t_0)E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|X_j^\top\boldsymbol{\alpha}_0]}\\
&- \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i-\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}{\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}\frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j q_2\{\eta_{0j},Y_j\}\Delta_{0j}}{\hat{\pi}(\boldsymbol{U}_j,Y_j;\gamma_0)}\frac{K_h(\boldsymbol{X}_j^\top\boldsymbol{\alpha}_0-X_i^\top\boldsymbol{\alpha}_0)E[\varepsilon|\boldsymbol{U}_i,\delta_i=0]}{f(t_0)E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}_j^\top\boldsymbol{\alpha}_0]}\\
&- \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i q_2\{\eta_{0i},Y_i\}\Delta_{0i}}{\hat{\pi}(\boldsymbol{U}_i,Y_i;\gamma_0)}\begin{pmatrix}\frac{E[\dot{g}(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)\boldsymbol{X}\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0]}\\ \frac{E[\boldsymbol{Z}\rho_2\{g(\boldsymbol{Z}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0]}\end{pmatrix}^\top\begin{pmatrix}\hat{\boldsymbol{\alpha}}^0-\boldsymbol{\alpha}_0\\ \hat{\boldsymbol{\beta}}^0-\boldsymbol{\beta}_0\end{pmatrix}+o_p(1)\\
&=: \mathbb{B}_{n21}+\mathbb{B}_{n22}+\mathbb{B}_{n23}+o_p(1)
\end{aligned}
\tag{3.25}
$$

类似于定理 3.3, 利用非参数回归, 有

$$
\mathbb{B}_{n21} = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}\frac{\varepsilon_i E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}\Delta_0|X_i^\top\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|X_i^\top\boldsymbol{\alpha}_0]}+o_p(1),
$$

和

$$
\mathbb{B}_{n22} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i-\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}{\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}\frac{E[\varepsilon|\boldsymbol{U}_i,\delta_i=0]E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}\Delta_0|X_i^\top\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|X_i^\top\boldsymbol{\alpha}_0]}+o_p(1),
$$

因此, 得到

$$\begin{aligned}\mathbb{B}_{n2} = &-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}\frac{\varepsilon_i E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}\Delta_0|X_i^\top\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|X_i^\top\boldsymbol{\alpha}_0]}\\ &+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i-\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}{\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}\frac{E[\varepsilon|\boldsymbol{U}_i,\delta_i=0]E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}\Delta_0|X_i^\top\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|X_i^\top\boldsymbol{\alpha}_0]}\\ &-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i q_2\{\eta_{0i},Y_i\}\Delta_{0i}}{\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}\begin{pmatrix}\frac{E[\dot{g}(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)\boldsymbol{X}\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0]}\\ \frac{E[\boldsymbol{Z}\rho_2\{g(\boldsymbol{Z}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0]}\end{pmatrix}^\top\begin{pmatrix}\hat{\boldsymbol{\alpha}}^0-\boldsymbol{\alpha}_0\\ \hat{\boldsymbol{\beta}}^0-\boldsymbol{\beta}_0\end{pmatrix}+o_p(1)\end{aligned}\tag{3.26}$$

结合上面的结果, 等式 (3.22) 可以重新写为

$$\mathcal{L}_n(\boldsymbol{\theta}) = (\mathbb{B}_{n1}^* - \mathbb{B}_{n2}^* + \mathbb{B}_0\boldsymbol{\theta}^*)^\top\boldsymbol{\theta} - \frac{1}{2}\boldsymbol{\theta}^T\mathbb{A}\boldsymbol{\theta} + o_p(1)\tag{3.27}$$

和

$$\hat{\boldsymbol{\theta}} = \mathbb{A}^{-1}(\mathbb{B}_{1n}^* - \mathbb{B}_{2n}^*) + \sqrt{n}\mathbb{A}^{-1}\mathbb{B}_0\begin{pmatrix}\hat{\boldsymbol{\alpha}}^0-\boldsymbol{\alpha}_0\\ \hat{\boldsymbol{\beta}}^0-\boldsymbol{\beta}_0\end{pmatrix}+o_p(1)\tag{3.28}$$

其中

$$\begin{aligned}\mathbb{B}_{n1}^* &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i(\boldsymbol{U}_i,Y_i;\gamma_0)}q_1(\eta_{0i},Y_i)\left[\Delta_{0i}-\frac{E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}\Delta_0|X_i^\top\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|X_i^\top\boldsymbol{\alpha}_0]}\right]\\ \mathbb{B}_{n2}^* &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i-\pi_i(\boldsymbol{U}_i,Y_i;\gamma_0)}{\pi_i(\boldsymbol{U}_i,Y_i;\gamma_0)}E\{q_1(\eta_0,Y)\Delta_0|\boldsymbol{U}_i,\delta_i=0\}\\ &\quad-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i-\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}{\pi(\boldsymbol{U}_i,Y_i;\gamma_0)}\frac{E[q_1(\eta_0,Y)|\boldsymbol{U}_i,\delta_i=0]E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}\Delta_0|X_i^\top\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|X_i^\top\boldsymbol{\alpha}_0]}\\ \mathbb{B}_0 &= E\left\{\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}\Delta_0\begin{pmatrix}\frac{E[\dot{g}(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)\boldsymbol{X}\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0]}\\ \frac{E[\boldsymbol{Z}\rho_2\{g(\boldsymbol{Z}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}_i^\top\boldsymbol{\alpha}_0]}\end{pmatrix}^\top\right\}\begin{pmatrix}\hat{\boldsymbol{\alpha}}^0-\boldsymbol{\alpha}_0\\ \hat{\boldsymbol{\beta}}^0-\boldsymbol{\beta}_0\end{pmatrix}\end{aligned}$$

回顾 $\hat{\boldsymbol{\theta}}$ 的定义, 有

$$\begin{pmatrix}\hat{\boldsymbol{\alpha}}^1-\boldsymbol{\alpha}_0\\ \hat{\boldsymbol{\beta}}^1-\boldsymbol{\beta}_0\end{pmatrix} = \frac{1}{\sqrt{n}}\mathbb{A}^{-1}(\mathbb{B}_{1n}^* - \mathbb{B}_{2n}^*) + \mathbb{A}^{-1}\mathbb{B}_0\begin{pmatrix}\hat{\boldsymbol{\alpha}}^0-\boldsymbol{\alpha}_0\\ \hat{\boldsymbol{\beta}}^0-\boldsymbol{\beta}_0\end{pmatrix}+o_p(n^{-1/2})\tag{3.29}$$

记 $\mathbb{A}^* = \mathbb{A}^{-1/2}\mathbb{B}_0\mathbb{A}^{-1/2}$ 和 $(\hat{\boldsymbol{\alpha}}^k,\hat{\boldsymbol{\beta}}^k)$ 是 $(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)$ 是第 k 次迭代所获得的估计值. 分别用 $(\hat{\boldsymbol{\alpha}}^{k+1},\hat{\boldsymbol{\beta}}^{k+1})$ 和 $(\hat{\boldsymbol{\alpha}}^k,\hat{\boldsymbol{\beta}}^k)$ 取代 $(\hat{\boldsymbol{\alpha}}^1,\hat{\boldsymbol{\beta}}^1)$ 和$(\hat{\boldsymbol{\alpha}}^0,\hat{\boldsymbol{\beta}}^0)$. 等式 (3.29) 仍然成立.

令

$$\vartheta^k = \mathbb{A}^{1/2}\begin{pmatrix}\hat{\boldsymbol{\alpha}}^k - \boldsymbol{\alpha}_0\\ \hat{\boldsymbol{\beta}}^k - \boldsymbol{\beta}_0\end{pmatrix},$$

有

$$\vartheta^{k+1} = \frac{1}{\sqrt{n}}\mathbb{A}^{-1/2}(\mathbb{B}^*_{1n} - \mathbb{B}^*_{2n}) + \mathbb{A}^*\vartheta^k + o_p(n^{-1/2})$$

则, 类似于 Liu et al. (2014), 可以得到所提出算法的收敛性. 对于足够大的 k, 有

$$\mathbb{A}^{1/2}\begin{pmatrix}\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\end{pmatrix} = \frac{1}{\sqrt{n}}\mathbb{A}^{-1/2}(\mathbb{B}^*_{1n} - \mathbb{B}^*_{2n}) + \mathbb{A}^*\mathbb{A}^{1/2}\begin{pmatrix}\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\end{pmatrix} + o_p(n^{-1/2}).$$

则,

$$(\mathbb{A} - \mathbb{A}^{1/2}\mathbb{A}^*\mathbb{A}^{1/2})\begin{pmatrix}\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\end{pmatrix} = \frac{1}{\sqrt{n}}(\mathbb{B}^*_{1n} - \mathbb{B}^*_{2n}) + o_p(n^{-1/2}). \tag{3.30}$$

回顾 $\mathbb{A}^*$ 的定义, 等式 (3.30) 可以写为

$$\sqrt{n}\mathcal{A}_0\begin{pmatrix}\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\end{pmatrix} = (\mathbb{B}^*_{1n} - \mathbb{B}^*_{2n}) + o_p(1), \tag{3.31}$$

其中

$$\begin{aligned}\mathcal{A}_0 &= \mathbb{A} - \mathbb{B}_0\\ &= f(t)E\left\{\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0) + \boldsymbol{Z}^\top\boldsymbol{\beta}_0\}\begin{pmatrix}\dot{g}(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)\boldsymbol{X}\\ \boldsymbol{Z}\end{pmatrix}\begin{pmatrix}\dot{g}(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)\boldsymbol{X}\\ \boldsymbol{Z}\end{pmatrix}^\top\right\}\\ &\quad - f(t)E\left\{\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0) + \boldsymbol{Z}^\top\boldsymbol{\beta}_0\}\begin{pmatrix}\dot{g}(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)\boldsymbol{X}\\ \boldsymbol{Z}\end{pmatrix}\begin{pmatrix}\frac{E[\boldsymbol{X}\dot{g}(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}^\top\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}^\top\boldsymbol{\alpha}_0]}\\ \frac{E[\boldsymbol{Z}\rho_2\{g(\boldsymbol{Z}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}^\top\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^\top\boldsymbol{\alpha}_0)+\boldsymbol{Z}^\top\boldsymbol{\beta}_0\}|\boldsymbol{X}^\top\boldsymbol{\alpha}_0]}\end{pmatrix}^T\right\}.\end{aligned}$$

因此, 有

$$\sqrt{n}\begin{pmatrix}\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\end{pmatrix} \xrightarrow{\mathcal{D}} N(0, \mathcal{A}_0^{-1}\Sigma_1\mathcal{A}_0^{-1})$$

其中

$$\Omega = \left[\begin{pmatrix} \dot{g}(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)\boldsymbol{X} \\ \boldsymbol{Z} \end{pmatrix} - \begin{pmatrix} \frac{E[\boldsymbol{X}\dot{g}(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)\rho_2\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0]} \\ \frac{E[\boldsymbol{Z}\rho_2\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0]}{E[\rho_2\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0]} \end{pmatrix}\right]$$

$$\Sigma_1 = Var\left\{\frac{\delta}{\pi(\boldsymbol{U},Y)}q_1\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}\Omega - \frac{\delta-\pi(\boldsymbol{U},Y)}{\pi(\boldsymbol{U},Y)}E[q_1\{g(\boldsymbol{X}^{\top}\boldsymbol{\alpha}_0)+\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0\}|\boldsymbol{U},\delta=0]\Omega\right\}$$

3.6　本章小结

在本章中, 针对具有不可忽略的响应变量的广义部分线性单指标模型, 提出了一种参数估计和未知连接函数的 PS-WEE 估计方法. 虽然已有文献研究了 GPLSM, 但所有的研究都只关注完全观测数据. 本文中允许考虑不可忽略缺失响应变量的信息, 提出的方法推广了现有的方法. 假设响应概率模型为半参数逻辑回归模型, 其中非参数部分采用非参数回归方法估计, 参数部分采用 GMM 方法估计. 而且系统地研究了所得估计量的理论性质. 模拟研究表明了该方法的优越性.

第四章 响应变量随机缺失的回归模型的稳健估计

4.1 引言

缺失数据经常存在于各种领域中, 譬如: 生物医学, 社会科学, 金融学等. 在过去的几十年中, 已经存在很多的统计方法用来处理缺失数据问题, 包括响应变量缺失、协变量缺失及二者都缺失. 现存的工作主要考虑三种数据缺失机制: 完全随机缺失 (MCAR), 即缺失不依赖于观测值和缺失值, 随机缺失 (MAR), 即缺失仅仅依赖于可观测数据; 非随机缺失 (MNAR), 即缺失依赖于未观测的数据本身, 该缺失机制最处理. 解决缺失数据问题方法的综述, 请参看 Little and Rubin (2002).

处理缺失数据最常用的方法是先将缺失数据删除, 再用回归或者似然方法去分析剩下的数据. 但如果缺失数据不是 MCAR, 则这种处理方法将获得有偏的结果. 为了得到无偏估计, Horvitz and Thompson (1952) 首先提出逆概率加权方法 (IPW) 去校准完全可观测数据的权重, 继而获得感兴趣未知的估计量的无偏估计量. 然而, 逆概率加权方法不具有很好的有效性, 因为它没有考虑未观测数据的额外信息. 为了获得更有效的估计量, Robins, Rotnitzky and Zhao (1994) 提出了一种增广的逆概率加权估计方法, 该方法不仅使用额外的数据信息, 还具有双重稳健性 (即回归模型或者倾向得分模型选择正确则可以获得相合的估计). 基于以上方法, 在 MAR 下处理缺失数据的成果, 详见: Robins, Rotnitzky, and Zhao (1995), Rotnitzky and Robins (1995), van der Laan and Robins (2003), Tsiatis (2006), Wang and Rao (2001, 2002), Rotnitzky (2008), Wang Q, Linton and Hardle (2004), Zhou and Liang (2005), Xue (2009a, 2009b), Qin, Zhang and Leung (2009), Tan (2010), Rotnitzky et al. (2012), Han and Wang (2013), Tang and Zhao (2014), Han (2014, 2018) 等等.

然而, 数据中不单单只存在缺失数据, 还可能存在异常值. 举一个最简单的例子: 在没有数据缺失下, 回归模型的最小二乘估计 (OLS) 对异常值或者潜在模型的正确性非常敏感. 实际中很多数据都包含异常值, 如在收入调查数据中

经常遇见. 异常值的产生有很多原因, 如: 数据中一小部分个体来自另外一个总体或者由于某些原因造成测量误差也会造成异常值的出现. 稳健的统计推断目的是为了构造不受异常值影响稳健的估计或者假设检验统计量. 另外, 数据集中包含异常值经常在统计分析中遇见, 且这些异常值可能出现在响应变量和 (或) 协变量中. 因此, 发展有效的统计推断方法使其对协变量和 (或) 响应变量中异常值都稳健显得越发重要. 稳健统计的参考文献请参看: Hampel (1968), Huber (1981), Rousseeuw and Leroy (1987), Zhu and Zhang (2004). 过去几十年, 稳健回归估计受到越来越多研究者的关注. Huber (1973) 首次提出最流行的稳健回归 M-type 估计量, 中位数回归模型、分位数回归模型 (Konker and Basset 1978; Koenker 2005), MM-估计 (Yohai, 1987), τ- 估计 (Yohai and Zamar, 1988), 有效且稳健的加权最小二乘估计 (REWLS) (Gervini and Yohai 2002), 扩展的带有新截尾的 Huber's 函数 (Huber-ESL) (Jiang et al. 2018). 这些估计量只对响应变量中的异常值稳健, 然而对协变量中的高杠杆点 (数据点脱离协变量空间主体) 却异常敏感, 此时它们的渐近破坏点是零 (Yohai, 1987). 为了获得对协变量和响应变量都稳健的估计, 损失函数的一阶导数需要是重新下降的 (Rousseeuw and Yohai 1984; Yohai 1987). 最常见的满足这个性质的损失函数是 Tukey's biweight 函数 (Tukey, 1960). 综上, 可以发现: 在响应变量缺失且协变量和响应变量中都存在异常值时, 尚未见相关文献, 这是本章的出发点.

在本章中, 详细地讨论在响应变量随机缺失且协变量和响应变量中都存在异常值时如何获得回归参数的有效估计. 首先, 对缺失机制模型, 构建一个加权的拟似然函数, 然后基于给定似然函数获得参数的稳健估计. 其次, 基于逆概率加权 (Horvitz and Thompson,1952) 和重新下降 (Tukey, 1960) 的思想, 建立了包含感兴趣未知参数的无偏估计方程组 (即能处理缺失数据还能处理异常值). 最后, 使用广义矩估计方法对感兴趣参数进行估计, 并证明了所得估计量的相合性和渐近正态性.

结构安排如下: 在 4.2 节中简单地回顾缺失数据下回归模型的参数估计. 在 4.3 节中, 当数据中存在响应变量随机缺失且协变量和响应变量中都存在异常值时, 给出一个新的稳健且无偏的参数估计量. 在 4.4 节中, 研究了新提出估计量

的大样本性质. 在 4.5 和 4.6 节中, 用数值模拟实验和实际数据来研究新提出的估计方法的表现. 4.7 节本章理论证明, 4.8 节本章小结.

4.2　缺失数据下回归模型的参数估计

考虑如下回归模型

$$y_i = \mu(\boldsymbol{x}_i, \boldsymbol{\beta}) + \epsilon_i \; i = 1, \ldots, n, \tag{4.1}$$

其中 y_i 是响应变量, $\boldsymbol{x}_i$ 是 p_1 维协变量, $\boldsymbol{\beta}$ 是 p 维参数向量, ϵ_i 是误差项, 且 $\mu(\cdot,\cdot)$ 是一个关于 $\boldsymbol{x}$ 和 $\boldsymbol{\beta}$ 的已知函数. 假定 $\mathbb{E}[\epsilon_i|\boldsymbol{x}_i] = 0$ 和 $\mathbb{E}[\epsilon_i^2|\boldsymbol{x}_i] = \sigma^2$, 则 $\mathbb{E}[y_i|\boldsymbol{x}_i] = \mu(\boldsymbol{x}_i, \boldsymbol{\beta})$. 一般地, 若 $\mu(\boldsymbol{x}_i, \boldsymbol{\beta}) = \boldsymbol{x}_i^\top\boldsymbol{\beta}$, 则 (4.1) 变成线性模型; 若 $\mu(\boldsymbol{x}_i, \boldsymbol{\beta}) = \exp(\boldsymbol{x}_i^\top\boldsymbol{\beta})$, 则 (4.1) 变成非线性回归模型.

假定从模型 (4.1) 中获得 n 个独立样本, 其中 n_1 个样本完全观测到 $\{y_i, \boldsymbol{x}_i\}$, 剩余的 $n - n_1$ 个样本只观测到 $\boldsymbol{x}_i$. 令 δ_i 是描述第 i 个观测中 y_i 是否被观测到的示性函数, 则若 y_i 观测, 则 $\delta_i = 1$. 反之, 若 y_i 缺失, 则 $\delta_i = 0$. 为了简单起见, 记 n 个观测到的数据集为 $\{y_i, \delta_i, \boldsymbol{x}_i : i = 1, \ldots, n\}$.

假定缺失机制服从 $\delta_i \sim \text{Bernoulli}(\pi_i)$, 其中

$$\pi_i = P(\delta_i = 1|y_i, \boldsymbol{x}_i) = P(\delta_i = 1|\boldsymbol{x}_i) = \omega(\boldsymbol{x}_i, \boldsymbol{\eta}), \tag{4.2}$$

其中 $\omega(\boldsymbol{x}_i, \boldsymbol{\eta})$ 是一个给定的概率分布函数, $\boldsymbol{\eta}$ 是 m 维未知参数向量. 特别地, $\omega(\boldsymbol{x}_i, \boldsymbol{\eta}) = \frac{1}{1+\exp(-\boldsymbol{x}_i^\top\boldsymbol{\eta})}$, 则缺失机制模型就简化成 Logistic 回归模型. 在模型 (4.2) 下, 缺失数据机制是 MAR.

注意到目标是估计回归系数 $\boldsymbol{\beta}$. 令 $U(y_i, \boldsymbol{x}_i, \boldsymbol{\beta})$ 是 $\boldsymbol{\beta}$ 的一组 q 维无偏估计方程, 则 $\mathbb{E}[U(y_i, \boldsymbol{x}_i, \boldsymbol{\beta})] = 0$. 在没有缺失数据时 (即 $\delta_i = 1, i = 1, \ldots, n$), 可以通过解 $\sum_{i=1}^n U(y_i, \boldsymbol{x}_i, \boldsymbol{\beta}) = 0$ 获得 $\boldsymbol{\beta}$ 的估计. 一般地, $U(y_i, \boldsymbol{x}_i, \boldsymbol{\beta})$ 可以取 $\mathbf{a}(\boldsymbol{x}_i, \boldsymbol{\beta})\{y_i - \mu(\boldsymbol{x}_i, \boldsymbol{\beta})\}$, 其中 $\mathbf{a}(\boldsymbol{x}_i, \boldsymbol{\beta})$ 是一个大于等于 $\boldsymbol{\beta}$ 维数的已知函数向量. 例如: 当 $\mu(\boldsymbol{x}_i, \boldsymbol{\beta}) = \boldsymbol{x}_i^\top\boldsymbol{\beta}$ 时, 则 $\mathbf{a}(\boldsymbol{x}_i, \boldsymbol{\beta}) = \boldsymbol{x}_i$, 这时 $U(y_i, \boldsymbol{x}_i, \boldsymbol{\beta})$ 就变成最小二乘估计的最优 KTT 条件. 类似地, 对不同的 $\mu(\boldsymbol{x}_i, \boldsymbol{\beta})$, 可以找不同维数的 $\mathbf{a}(\boldsymbol{x}_i, \boldsymbol{\beta})$ 来实参数 $\boldsymbol{\beta}$ 的可识别性和估计.

当缺失数据为 MAR 时, 最常用的 $\boldsymbol{\beta}$ 的 HT 估计 (记为 $\hat{\boldsymbol{\beta}}_{\text{HT}}$) 满足如下方程组

$$\sum_{i=1}^n \mathbf{g}(y_i, \delta_i, \boldsymbol{x}_i, \boldsymbol{\beta}, \hat{\boldsymbol{\eta}}) = \sum_{i=1}^n \frac{\delta_i U(y_i, \boldsymbol{x}_i, \boldsymbol{\beta})}{\omega(\boldsymbol{x}_i, \hat{\boldsymbol{\eta}})} = 0, \tag{4.3}$$

其中 $\hat{\boldsymbol{\eta}}$ 是在缺失机制模型 (4.2) 下基于数据集 $\{y_i, \boldsymbol{x}_i, \delta_i : i = 1, \ldots, n\}$ 上 $\boldsymbol{\eta}$ 的相合估计, 也就是, $\hat{\boldsymbol{\eta}}$ 通过极大化如下目标函数 $L(\boldsymbol{\eta})$ 获得

$$L(\boldsymbol{\eta}) = \sum_{i=1}^{n} \delta_i \log \omega(\boldsymbol{x}_i, \boldsymbol{\eta}) + (1 - \delta_i) \log\{1 - \omega(\boldsymbol{x}_i, \boldsymbol{\eta})\}. \tag{4.4}$$

然而, 当 $\boldsymbol{x}_i$ 和 y_i 中存在异常值时, (4.3) 得到的 $\boldsymbol{\beta}$ 的估计将不再相合且不具有稳健性, 理由如下:

(i) 模型 (4.1) 中的残差 $y_i - \mu(\boldsymbol{x}_i, \boldsymbol{\beta})$ 对异常值异常敏感, 也就是 $U(y_i, \boldsymbol{x}_i, \boldsymbol{\beta})$ 不具备重复下降的性质;

(ii) 当 $\boldsymbol{x}_i$ 中存在异常值时, 极大化 (4.4) 获得的 $\hat{\boldsymbol{\eta}}$ 将会远离其真实值且不相合.

因此, 考虑存在缺失数据且有异常值时的有效估计是一个非常重要且必要的研究课题.

4.3 缺失数据下参数的稳健估计

根据 4.2 节末尾的分析, 同时存在缺失数据和异常值时, 为了获得参数的有效估计, 需要解决两个方面的问题:

(i) 如何在带有异常值的条件下获得 $\boldsymbol{\eta}$ 的相合估计;

(ii) 如何基于模型 (4.1) 构建 $\boldsymbol{\beta}$ 的无偏估计方程.

下面我们分两步来解决以上问题.

4.3.1 缺失机制模型参数的稳健估计

考虑 $\boldsymbol{\eta}$ 的稳健估计. 众所周知, 当数据中存在异常值时, 极大化 (4.4) 将不能获得 $\boldsymbol{\eta}$ 的相合估计. 为了获得 $\boldsymbol{\eta}$ 的相合估计, 首先要消除 $\boldsymbol{x}_i$ 中异常值点对 $\boldsymbol{\eta}$ 估计的影响. 根据 Carroll and Pederson (1993), 先考虑如下加权似然函数

$$\widetilde{L}(\boldsymbol{\eta}) = \sum_{i=1}^{n} w\{h(\boldsymbol{x}_i)\}\Big[\delta_i \log \omega(\boldsymbol{x}_i, \boldsymbol{\eta}) + (1 - \delta_i) \log\{1 - \omega(\boldsymbol{x}_i, \boldsymbol{\eta})\}\Big], \tag{4.5}$$

其中 $h(\boldsymbol{x}_i)$ 是 $\boldsymbol{x}_i$ 的已知函数; $w\{h(\boldsymbol{x}_i)\}$ 是一个关于 $\boldsymbol{x}_i$ 的加权函数, 其目的是当 $\boldsymbol{x}_i$ 中存在异常值时, $w\{h(\boldsymbol{x}_i)\} \to 0$; 当 $\boldsymbol{x}_i$ 中没有异常值时, $w\{h(\boldsymbol{x}_i)\} \to 1$.

不失一般性, 根据 Maronna, Martin and Yohai (2006), 取

$$w(z)=\begin{cases}\dfrac{\psi(z)}{z}, & z\neq 0\\ 0, & z=0\end{cases} \tag{4.6}$$

其中 $\psi(z)$ 是给定的稳健的 ψ 函数. 例如, Huber 函数的 ψ 函数为

$$\psi(z)=\begin{cases}z, & |z|\leq c\\ sgn(z)c, & |z|>c\end{cases}$$

其中 c 是 Huber 函数中用来控制有效性的调谐参数.

现考虑 $h(\boldsymbol{x}_i)$ 的取法, 这里给出两种常用的取法:

(i) $h(\boldsymbol{x}_i)=[(\boldsymbol{x}_i-\hat{\mu})^{\top}\hat{\Sigma}^{-1}(\boldsymbol{x}_i-\hat{\mu})/p_1]^{1/2}$, 其中 $\hat{\mu}$ 和 $\hat{\Sigma}$ 是 $\boldsymbol{x}$ 的均值和协方差矩阵的稳健估计, p_1 是协变量的维数;

(ii) $h(\boldsymbol{x}_i)=\dfrac{\|\boldsymbol{x}_i-\mathrm{med}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)\|_2}{\mathrm{median}_{1\leq j\leq n}\|\boldsymbol{x}_j-\mathrm{med}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)\|_2}$,

其中 $\mathrm{med}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)=\arg\min_{\boldsymbol{\mu}}\sum_{i=1}^{n}\|\boldsymbol{x}_i-\boldsymbol{\mu}\|$, $\mathrm{median}(a_1,\ldots,a_n)$ 表示 $a_1,\ldots,a_n$ 的中位数.

综上, 可以看到极大化 (4.5) 可以获得 $\boldsymbol{\eta}$ 的稳健估计 $\hat{\boldsymbol{\eta}}_r$. 即, $\hat{\boldsymbol{\eta}}_r$ 是如下正则方程的解

$$\widetilde{S}_n(\boldsymbol{\eta})=\frac{\partial\widetilde{L}(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}=\frac{1}{n}\sum_{i=1}^{n}w\{h(\boldsymbol{x}_i)\}\frac{\delta_i-\omega(\boldsymbol{x}_i,\boldsymbol{\eta})}{\omega(\boldsymbol{x}_i,\boldsymbol{\eta})(1-\omega(\boldsymbol{x}_i,\boldsymbol{\eta}))}\frac{\partial\omega(\boldsymbol{x}_i,\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}. \tag{4.7}$$

4.3.2　回归模型参数的稳健估计

在本小节中, 使用逆概率加权和重复下降的思想, 来构建关于 $\boldsymbol{\beta}$ 的无偏估计方程. 首先, 记 Tukey's biweight 函数的一阶导函数为 $\psi_1(z,c)=z[1-(\frac{z}{c})^2]^2I(|z|\leq c)$, 其中 $I(\cdot)$ 是示性函数, c 是用来控制有效性的调谐参数. 为获得当 y_i 存在缺失且 y_i 和 $\boldsymbol{x}_i$ 都存在异常值, $\boldsymbol{\beta}$ 的稳健且无偏估计. 令

$$\boldsymbol{g}_1(y_i,\boldsymbol{x}_i,\delta_i,\boldsymbol{\beta},\sigma,\boldsymbol{\eta})=\begin{pmatrix}\frac{\delta_i}{\omega(\boldsymbol{x}_i,\boldsymbol{\eta})}\mathbf{a}(\boldsymbol{x}_i,\boldsymbol{\beta})\psi_1(\frac{y_i-\mu(\boldsymbol{x}_i,\boldsymbol{\beta})}{\sigma},c)\\ \frac{\delta_i}{\omega(\boldsymbol{x}_i,\boldsymbol{\eta})}w_1(\frac{y_i-\mu(\boldsymbol{x}_i,\boldsymbol{\beta})}{\sigma},c)\{(y_i-\mu(\boldsymbol{x}_i,\boldsymbol{\beta}))^2-\sigma^2\}\end{pmatrix}, \tag{4.8}$$

其中

$$w_1(z,c)=\begin{cases}\dfrac{\psi_1(z,c)}{z}, & z\neq 0\\ 0, & z=0\end{cases}$$

因此, $\mathbb{E}[\boldsymbol{g}_1(y_i,\boldsymbol{x}_i,\delta_i,\boldsymbol{\beta},\sigma,\boldsymbol{\eta})]=0$. 特别地, 为了实现 $\boldsymbol{\beta}$ 的稳健估计, 引入一个需要估计的尺度参数 σ. 为简单起见, 假定 σ 已知, 则 (4.8) 中的第二个式子可以省略. 下面简要地说明 (4.8) 的好处:

(i) $\delta_i/\omega(\boldsymbol{x}_i,\boldsymbol{\eta})$ 保证了缺失数据下的无偏性;

(ii) 当 y_i 和 $\boldsymbol{x}_i$ 中存在异常值时, w_1 和 ψ_1 能很好地消除异常值对估计的影响. 在这里, 需要说明的是当 $\boldsymbol{x}_i$ 和 y_i 中存在异常值且 y_i 未缺失时, 上述估计方程 $\boldsymbol{g}_1(y_i,\boldsymbol{x}_i,\delta_i,\boldsymbol{\beta},\sigma,\boldsymbol{\eta})$ 的期望仍然为零, 其原因是 w_1 和 ψ_1 自动将期望截断为零.

记 $\boldsymbol{\theta}=(\boldsymbol{\beta}^{\top},\sigma)^{\top}$, 定义 $\bar{\boldsymbol{g}}_1(\boldsymbol{\theta})=n^{-1}\sum_{i=1}^{n}\boldsymbol{g}_1(y_i,\boldsymbol{x}_i,\delta_i,\boldsymbol{\beta},\sigma,\hat{\boldsymbol{\eta}}_r)$, 其中 $\hat{\boldsymbol{\eta}}_r$ 是 $\boldsymbol{\eta}$ 的稳健相合估计, 考虑 $\boldsymbol{\theta}$ 的两步最优广义矩估计. 根据 Hanse (1982), 广义矩估计 (记 $\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}}$) 为

$$\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}}=\arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\bar{\boldsymbol{g}}_1(\boldsymbol{\theta})^{\top}W\bar{\boldsymbol{g}}_1(\boldsymbol{\theta}),\tag{4.9}$$

其中 W 是正定且对称的矩阵.

4.4 大样本性质

下面介绍参数估计的大样本性质. 记 $\boldsymbol{\eta}$ 和 $\boldsymbol{\theta}$ 的真实值分别为 $\boldsymbol{\eta}_0$ 和 $\boldsymbol{\theta}_0$. 为此, 作如下记号:

$$\begin{aligned}
&\omega_1(\boldsymbol{x}_i,\boldsymbol{\eta})=\partial\omega(\boldsymbol{x}_i,\boldsymbol{\eta})/\partial\boldsymbol{\eta},\\
&S_{11}(\boldsymbol{\eta})=\mathbb{E}\Big\{w\{h(\boldsymbol{x}_i)\}\frac{\omega_1(\boldsymbol{x}_i,\boldsymbol{\eta})\omega_1(\boldsymbol{x}_i,\boldsymbol{\eta})^{\top}}{\omega(\boldsymbol{x}_i,\boldsymbol{\eta})(1-\omega(\boldsymbol{x}_i,\boldsymbol{\eta}))}\Big\},\\
&S_{11}=S_{11}(\boldsymbol{\eta}_0),\\
&\widetilde{S}_n(\boldsymbol{\eta})=\frac{\partial\widetilde{L}(\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}=\frac{1}{n}\sum_{i=1}^{n}w\{h(\boldsymbol{x}_i)\}\frac{\delta_i-\omega(\boldsymbol{x}_i,\boldsymbol{\eta})}{\omega(\boldsymbol{x}_i,\boldsymbol{\eta})(1-\omega(\boldsymbol{x}_i,\boldsymbol{\eta}))}\frac{\partial\omega(\boldsymbol{x}_i,\boldsymbol{\eta})}{\partial\boldsymbol{\eta}},\\
&\widetilde{S}_n=\widetilde{S}_n(\boldsymbol{\eta}_0),\\
&\boldsymbol{g}_1(\boldsymbol{z}_i,\boldsymbol{\theta},\boldsymbol{\eta})=\boldsymbol{g}_1(y_i,\boldsymbol{x}_i,\delta_i,\boldsymbol{\beta},\sigma,\boldsymbol{\eta}),\ \boldsymbol{z}_i=(y_i,\boldsymbol{x}_i,\delta_i),\\
&\mathbf{B}_{\boldsymbol{\theta}}=\mathbb{E}\{\nabla_{\boldsymbol{\theta}}\boldsymbol{g}_1(y_i,\boldsymbol{x}_i,\delta_i,\boldsymbol{\beta}_0,\sigma_0,\boldsymbol{\eta}_0)\},
\end{aligned}$$

$$\begin{aligned}
&\mathbf{B}_{\eta} = \mathbb{E}\{\nabla_{\eta} \boldsymbol{g}_1(y_i, \boldsymbol{x}_i, \delta_i, \boldsymbol{\beta}_0, \sigma_0, \boldsymbol{\eta}_0)\},\\
&\psi(\boldsymbol{x}_i, \boldsymbol{\eta}) = -S_{11}^{-1}\left\{w\{h(\boldsymbol{x}_i)\}\frac{\delta_i - \omega(\boldsymbol{x}_i, \boldsymbol{\eta})}{\omega(\boldsymbol{x}_i, \boldsymbol{\eta})(1-\omega(\boldsymbol{x}_i, \boldsymbol{\eta}))}\frac{\partial \omega(\boldsymbol{x}_i, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right\},\\
&\bar{\boldsymbol{g}}_1(\boldsymbol{\theta}, \boldsymbol{\eta}) = n^{-1}\sum_{i=1}^{n} \boldsymbol{g}_1(y_i, \boldsymbol{x}_i, \delta_i, \boldsymbol{\beta}, \sigma, \boldsymbol{\eta})
\end{aligned}$$

首先证明 $\hat{\boldsymbol{\eta}}_r$ 的相合性和渐近正态性.

定理 4.1. 假设条件 (A1)-(A4) 成立, 则

$$\hat{\boldsymbol{\eta}}_r - \boldsymbol{\eta}_0 = S_{11}^{-1}\widetilde{S}_n + o_p(n^{-1/2}),$$

$$\sqrt{n}(\hat{\boldsymbol{\eta}}_r - \boldsymbol{\eta}_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, S_{11}^{-1}),$$

其中 $\xrightarrow{\mathcal{L}}$ 表示依分布收敛.

定理 4.1 表明当协变量存在异常值时, 本章所提出的估计仍然是相合且渐近有效的, 这与常规的极大似然估计没有区别. 唯一的区别在于本文估计的 Fisher 信息矩阵中含有一个权重 $w\{h(\boldsymbol{x}_i)\}$. 可以很容易地看到: 当数据中没有异常值点时, 所有的 $w\{h(\boldsymbol{x}_i)\}$ 都为 1, 此时就变成常规的极大似然估计.

其次, 考虑 $\hat{\boldsymbol{\theta}}$ 的相合性和渐近正态性.

定理 4.2. 假定缺失机制模型 $\omega(\boldsymbol{x}_i, \boldsymbol{\eta})$ 被正确指定. 若假设条件 (A1)-(A4), (B1)-(B5) 成立, $\bar{\boldsymbol{g}}_1(\hat{\boldsymbol{\theta}})^{\top} W \bar{\boldsymbol{g}}_1(\hat{\boldsymbol{\theta}}) \leq \inf_{\theta\in\Theta} \bar{\boldsymbol{g}}_1(\boldsymbol{\theta}, \boldsymbol{\eta}_0)^{\top} W \bar{\boldsymbol{g}}_1(\boldsymbol{\theta}, \boldsymbol{\eta}_0) + o_p(n^{-1})$, $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$, 则

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_1),$$

其中 $\Sigma_1 = (\mathbf{B}_{\boldsymbol{\theta}}^{\top} W \mathbf{B}_{\boldsymbol{\theta}})^{-1} \mathbf{B}_{\boldsymbol{\theta}}^{\top} W \Sigma_2 W \mathbf{B}_{\boldsymbol{\theta}} (\mathbf{B}_{\boldsymbol{\theta}}^{\top} W \mathbf{B}_{\boldsymbol{\theta}})^{-1}$, Σ_2 将在 4.7 节中给出, $\xrightarrow{p}$ 表示依概率收敛.

对渐近方差矩阵 Σ_1, 最优权重矩阵 $W = \Sigma_2$. 在选取最优权重矩阵后, Σ_1 就变成 $(\mathbf{B}_{\boldsymbol{\theta}}^{\top} \Sigma_2 \mathbf{B}_{\boldsymbol{\theta}})^{-1}$, 且 $\Sigma_1 - (\mathbf{B}_{\boldsymbol{\theta}}^{\top} \Sigma_2 \mathbf{B}_{\boldsymbol{\theta}})^{-1}$ 是半正定矩阵. 这说明选取最优权重矩阵可以获得最小的协方差估计.

4.5 模拟研究

考虑如下线性回归模型：

$$y_i = \beta_0 + \boldsymbol{x}_i^{\top}\boldsymbol{\beta} + \epsilon_i, i = 1, \ldots, n, \tag{4.10}$$

其中 β_0 是截距项, $\boldsymbol{\beta}$ 是 p 维未知参数, $\boldsymbol{x}_i$ 是来自多元正态分布 $N(0, \Sigma)$ 的观测样本, Σ 的每个元素为 $\sigma_{ij} = 0.5^{|i-j|}$. 这里取 $\beta_0 = 2, \boldsymbol{\beta} = (3, 3)^{\top}$. 为了使数据中存在异常值, 我们考虑以下几种情况:

(1) ϵ_i 来自标准正态分布;

(2) ϵ_i 来自一个 π -污染分布, 即 $\epsilon_i \sim (1-\pi)N(0,1) + \pi N(10, 10^2)$, 其中 $\pi = 0.1$;

(3) ϵ_i 来自自由度为 3 的学生分布, 即 $\epsilon_i \sim t(3)$;

(4) ϵ_i 来自标准正态分布, 协变量中含有 10% 的高杠杆值异常点且满足 $\boldsymbol{x}_i = (x_{1i}, x_{2i})^{\top} = (10, 10)^{\top}$;

(5) $\epsilon_i \sim (1-\pi)N(0,1) + \pi N(10, 10^2)$, 其中 $\pi = 0.05$, 协变量中含有 10% 的高杠杆值异常点且满足$\boldsymbol{x}_i = (x_{1i}, x_{2i})^{\top} = (10, 10)^{\top}$.

情形 (1) 考虑在误差项服从正态分布时, 研究在有限样本的情况下不同估计方法与最小二乘估计的表现; 情形 (2) 和 (3) 考虑响应变量中存在异常值的情形; 情形 (4) 研究协变量中存在异常值时不同估计方法的差异; 情形 (5) 主要是为了研究在响应变量和协变量都存在异常值时所提出估计方法的有效性.

下面我们考虑缺失数据机制模型. 假设协变量 $\boldsymbol{x}_i = (x_{1i}, x_{2i})^{\top}$ 全部观测, 响应变量 y_i 存在缺失, y_i 的缺失示性函数 δ_i 由伯努利分布产生, 其概率为 $Pr(\delta|y_i, \boldsymbol{x}_i) = Pr(\delta|\boldsymbol{x}_i) = \omega(\boldsymbol{x}_i, \boldsymbol{\eta})$. 特别地, 本章考虑如下两种缺失概率模型:

(a) $\omega(\boldsymbol{x}_i, \boldsymbol{\eta}) = \{1 + \exp(-\eta_1 - \eta_2 x_{1i} - \eta_3 x_{2i})\}^{-1}$, 其中 $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3)^{\top} = (1.25, 0.2, 0.2)^{\top}$;

(b) $\omega(\boldsymbol{x}_i, \boldsymbol{\eta}) = \Phi(\eta_1 + \eta_2 x_{1i} + \eta_3 x_{2i})$, 其中 $\Phi(\cdot)$ 是标准正态分布的累积分布函数, $\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3)^{\top} = (0.75, 0.35, 0.35)^{\top}$.

由于现有的稳健估计方法不能直接对缺失数据使用, 导致可以比较的方法较少. 为了比较估计的有效性, 本章将采用完全可观测数据 (即 $\delta_i = 1, i = 1, \ldots, n$) 拟合常见的稳健估计方法. 特别地, 基于完全可观测数据, 计算以下 6 种估计量: (i) 最小二乘估计, 记为LS_{cc}; (ii) Rousseeuw (1984) 的最小中位数二乘估计, 记为LMS_{cc}; (iii) Yohai (1987) 的 MM 估计, 记为 MM_{cc}; (iv) Huber (1981) 的 Huber 估计, 记为 $Huber_{cc}$; (v) S估计, 记为 S_{cc}; (vi) 中位数回归估计, 记为 LAD_{cc}. 另外, 还计算基于逆概率加权的最小二乘估计 (记为 LS_{ipw}) 和基于逆概率加权的最小一乘估计 (记为 LAD_{ipw}), 其中缺失机制中的参数 $\boldsymbol{\eta}$ 都通过极大化 (4.4) 获得. 对所有估计方法, 重复试验 1000 次, 且考虑 $n = 200$ 和 $n = 400$ 的情形.

为了说明估计方法在参数估计中的具体表现. 对每一次试验, 计算参数估计与真实参数值之间的差异. 具体地, 对每一种估计方法, “Bias” 代表 1000 个估计值的平均值与真实值之间的偏差, “SD” 代表 1000 个估计值的标准偏差, “RMSE”代表 1000 个估计值与真实值之间的均方根. 同时, 为了检验方法是否有较好的预测值, 此外还计算基于完全可观测数据的预测值与真实值之间差异的绝对中位数值 (median absolute prediction error, MAPE), “MMAPE” 表示基于 1000 次重复试验的 MAPE 的平均值. 相关结果将在表 4.1, 表 4.2, 表 4.3, 表 4.4, 表 4.5中给出.

分析表 4.1, 表 4.2, 表 4.3, 表 4.4, 表 4.5 的结果, 有下面的发现: (i) 当数据中不存在异常值时, 所提出的估计方法与其他稳健估计方法之间的差异较小; (ii) 当异常值只存在响应变量中时, 相比其他方法, $Huber_{cc}$ 表现最好, 因为它拥有最小的 RMSE 值; (iii) 当数据存在缺失时, 只使用完全可观测数据获得的稳健估计表现较差, 其原因是丢掉没观测数据的协变量信息会降低估计的有效性; (iv) LAD_{ipw} 和 LS_{ipw} 的结果说明协变量中存在高杠杆点对缺失机制模型参数的估计影响很大, 进而影响回归模型参数的估计; (v) 当响应变量存在缺失且响应变量和协变量中都存在异常值时, 所提出的估计方法在所有方法中表现最佳; (vi) 所提出方法的 SD 和 RMSE 很接近且 RMSE 值很小, 说明我们的方法表现良好; (vii) 增大样本量可降低估计的 SD 和 RMSE, 即样本量越大估计越有效; (viii) 当数据中存在异常值且响应变量缺失时, 所提出的方法有较小的平均 MAPE 值,

表 **4.1 :**当 n = 200 时基于 **logistic** 回归缺失模型下不同估计方法的模拟结果**.**

		β_1			β_2			β_3		
Case	Method	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
1	LS_{cc}	-0.0036	0.0818	0.0818	0.0012	0.0927	0.0926	0.0036	0.0949	0.0949
	LMS_{cc}	-0.0070	0.6202	0.6199	0.0278	0.8444	0.8444	-0.0325	0.9018	0.9019
	MM_{cc}	0.0085	0.2442	0.2443	-0.0121	0.3040	0.3040	0.0030	0.2970	0.2968
	S_{cc}	0.00089	0.1527	0.1527	-0.0021	0.1784	0.1783	0.0148	0.1851	0.1856
	$Huber_{cc}$	-0.0022	0.0844	0.0844	0.0014	0.0951	0.0951	0.0037	0.0973	0.0973
	LAD_{cc}	-0.0001	0.1001	0.1001	0.0013	0.1182	0.1181	0.0076	0.1185	0.1187
	LAD_{ipw}	-0.00034	0.1001	0.1000	0.0003	0.1204	0.1203	0.0083	0.1200	0.1202
	LS_{ipw}	-0.0035	0.0819	0.0819	0.00067	0.0935	0.0934	0.0038	0.0949	0.0950
	Proposed	-0.0062	0.1150	0.1152	0.0091	0.1773	0.1774	0.0049	0.1855	0.1854
2	LS_{cc}	0.9966	0.3585	1.0590	0.0023	0.4307	0.4305	-0.0113	0.4139	0.4139
	LMS_{cc}	-0.0083	0.8278	0.8274	-0.0662	1.1383	1.1397	-0.0311	101445	1.1444
	MM_{cc}	-0.0019	0.2271	0.2270	0.0116	0.3063	0.3064	-0.0015	0.3078	0.3077
	S_{cc}	-0.0059	0.1501	0.1501	-0.0013	0.1697	0.1696	0.0063	0.1742	0.1742
	$Huber_{cc}$	0.1287	0.1027	0.1647	0.0012	0.1139	0.1139	0.00011	0.1170	0.1169
	LAD_{cc}	0.0942	0.1125	0.1467	0.0015	0.1264	0.1263	0.0019	0.1326	0.1325
	LAD_{ipw}	0.0933	0.1123	0.1459	0.0017	0.1287	0.1286	0.0023	0.1343	0.1343
	LS_{ipw}	0.9966	0.3589	1.0592	0.0033	0.4352	0.4350	-0.0117	0.4232	0.4231
	Proposed	0.0031	0.1240	0.1240	0.00006	0.1949	0.1948	0.0003	0.2063	0.2062
3	LS_{cc}	-0.0097	0.3941	0.3941	-2.6146	0.4247	2.6488	-2.6061	0.4251	2.6405
	LMS_{cc}	-0.0166	0.8803	0.8800	-0.0631	1.1989	1.2000	0.00037	1.2672	0.2666
	MM_{cc}	0.0055	0.1953	0.1953	0.0113	0.2445	0.2446	-0.0330	0.2545	0.2544
	S_{cc}	-0.0024	0.1493	0.1492	-0.0093	0.2633	0.2633	-0.0238	0.3286	0.3293
	$Huber_{cc}$	-0.0084	0.2889	0.2889	-1.4422	0.7692	1.6343	-1.4180	0.7410	1.5998
	LAD_{cc}	-0.0047	0.2947	0.2946	-1.2477	0.8156	1.4904	-1.2201	0.7829	1.4494
	LAD_{ipw}	-0.0416	0.2040	0.2081	-0.7835	0.3552	0.8602	-0.7686	0.3280	0.8357
	LS_{ipw}	-0.1000	0.3882	0.4007	-2.5833	0.4169	2.6167	-2.5744	0.4179	2.6081
	Proposed	0.0047	0.1415	0.1415	0.0008	0.2190	0.2189	0.0025	0.2270	0.2269
4	LS_{cc}	-0.0291	0.4175	0.4183	-2.8268	0.4313	2.8595	-2.7572	0.4352	2.7913
	LMS_{cc}	-0.0832	0.8461	0.8498	-0.1752	1.1783	1.1907	-0.0823	1.1121	1.1146
	MM_{cc}	-0.0015	0.2172	0.2171	-0.0129	0.2971	0.2973	0.0039	0.2656	0.2655
	S_{cc}	0.00044	0.1910	0.1910	-0.0539	0.4172	0.4205	-0.0661	0.5136	0.5175
	$Huber_{cc}$	-0.0255	0.4843	0.4847	-2.7719	0.4919	2.8151	-2.7271	0.5019	2.7729
	LAD_{cc}	-0.0335	0.5092	0.5100	-2.7660	0.5315	2.8166	-2.7283	0.5401	2.7812
	LAD_{ipw}	-0.0682	0.5044	0.5087	-2.7549	0.5155	2.8027	-2.7230	0.5269	2.7734
	LS_{ipw}	-0.0672	0.4136	0.4188	-2.8182	0.4304	2.8508	-2.7484	0.4341	2.7824
	Proposed	0.0011	0.1206	0.1205	0.0004	0.1677	0.1676	0.0038	0.1713	0.1712

表 **4.2 :**当 $n = 400$ 时基于 **logistic** 回归缺失模型下不同估计方法的模拟结果**.**

		β_1			β_2			β_3		
Case	Method	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
1	LS_{cc}	0.0015	0.0865	0.0864	0.0041	0.0953	0.0954	-0.0049	0.0981	0.0982
	LMS_{cc}	0.0205	0.6638	0.6638	-0.0161	0.8755	0.8753	0.0184	0.9342	0.9339
	MM_{cc}	0.0046	0.2906	0.2905	0.0015	0.3368	0.3366	-0.0090	0.3516	0.3516
	S_{cc}	0.0017	0.1638	0.1637	0.0027	0.1866	0.1865	0.0057	0.1892	0.1892
	$Huber_{cc}$	0.0019	0.0885	0.0884	0.0048	0.0964	0.0964	-0.0046	0.0997	0.0998
	LAD_{cc}	0.00039	0.1070	0.1070	0.0022	0.1172	0.1172	-0.0035	0.1171	0.1171
	LAD_{ipw}	0.00033	0.1094	0.1094	0.0043	0.1275	0.1275	-0.0046	0.1284	0.1285
	LS_{ipw}	0.00081	0.0879	0.0879	0.0056	0.1016	0.1017	-0.0047	0.1060	0.1060
	Proposed	-0.00008	0.1291	0.1290	-0.0067	0.1884	0.1885	-0.0040	0.2173	0.2173
2	LS_{cc}	1.0014	0.3969	1.0771	-0.0057	0.4400	0.4398	-0.0121	0.4445	0.4445
	LMS_{cc}	-0.0085	0.8507	0.8503	-0.1223	1.1803	1.1861	-0.0173	1.1953	1.1948
	MM_{cc}	-0.0042	0.2487	0.2486	-0.00028	0.2907	0.2906	0.0056	0.2842	0.2841
	S_{cc}	-0.0038	0.1570	0.1570	0.0024	0.1832	0.1832	-0.00056	0.1821	0.1821
	$Huber_{cc}$	0.1316	0.1105	0.1719	0.00078	0.1196	0.1196	-0.0043	0.1205	0.1205
	LAD_{cc}	0.0968	0.1228	0.1564	0.00093	0.1379	0.1378	-0.0019	0.1315	0.1315
	LAD_{ipw}	0.0957	0.1255	0.1577	0.00096	0.1486	0.1485	-0.0035	0.1400	0.1399
	LS_{ipw}	1.0045	0.4054	1.0831	-0.0064	0.4730	0.4728	-0.0212	0.4731	0.4734
	Proposed	-0.0036	0.1254	0.1254	-0.0057	0.2002	0.2002	0.0017	0.2011	0.2010
3	LS_{cc}	0.6840	0.3869	0.7858	-206396	0.4106	2.6713	-2.6230	0.4103	2.6549
	LMS_{cc}	0.0112	0.9035	0.9031	-0.0838	1.2398	1.2420	-0.1307	1.2901	1.2961
	MM_{cc}	0.00096	0.2226	0.2225	-0.0051	0.2825	0.2824	0.0035	0.2801	0.2800
	S_{cc}	0.0085	0.1563	0.1564	-0.0118	0.2302	0.2304	-0.0090	0.2566	0.2567
	$Huber_{cc}$	0.3966	0.3410	0.5229	-1.6061	0.7869	1.7883	-1.5902	0.7948	1.7776
	LAD_{cc}	0.3487	0.3624	0.5027	-1.4316	0.8572	1.6683	-1.4142	0.8516	1.6506
	LAD_{ipw}	0.1175	0.1998	0.2317	-0.6912	0.3050	0.7554	-0.6833	0.3157	0.7527
	LS_{ipw}	0.4824	0.4322	0.6476	-2.5635	0.4023	2.5949	-2.5495	0.4032	2.5811
	Proposed	0.0046	0.1457	0.1457	-0.0109	0.2053	0.2054	0.0032	0.2177	0.2177
4	LS_{cc}	0.7373	0.4084	0.8428	-2.8083	0.4006	2.8367	-2.8132	0.4032	2.8420
	LMS_{cc}	-0.0457	0.8542	0.8550	-0.1649	1.2557	1.2658	-0.0990	1.2294	1.2328
	MM_{cc}	0.0090	0.2794	0.2794	-0.0056	0.3186	0.3185	-0.0072	0.3037	0.3036
	S_{cc}	0.0215	0.2241	0.2250	-0.0949	0.5456	0.5535	-0.0820	0.5325	0.5385
	$Huber_{cc}$	0.6869	0.4663	0.8301	-2.7575	0.4746	2.7980	-2.7980	0.4752	2.8380
	LAD_{cc}	0.6825	0.5002	0.8460	-2.7567	0.5112	2.8037	-2.7975	0.5139	2.8443
	LAD_{ipw}	0.5879	0.5008	0.7721	-2.7183	0.5000	2.7638	-2.7579	0.5023	2.8032
	LS_{ipw}	0.6469	0.4059	0.7636	-2.7815	0.3968	2.8096	-2.7857	0.3976	2.8139
	Proposed	0.0022	0.1129	0.1128	-0.0042	0.1719	0.1719	0.0014	0.1763	0.1763

表 **4.3 :** 当 $n = 200$ 时基于 **probit** 回归缺失模型下不同估计方法的模拟结果**.**

		β_1			β_2			β_3		
Case	Method	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
1	LS_{cc}	-0.00045	0.0580	0.0580	0.00004	0.0649	0.0649	-0.00035	0.0653	0.0653
	LMS_{cc}	0.0174	0.6566	0.6576	-0.0007	0.8062	0.8058	-0.0305	0.9284	0.9285
	MM_{cc}	-0.0052	0.0863	0.0864	0.0014	0.1037	0.1036	-0.0024	0.1001	0.1001
	S_{cc}	-0.0084	0.1107	0.1109	-0.00056	0.1327	0.1327	-0.0027	0.1325	0.1324
	$Huber_{cc}$	-0.0012	0.0594	0.0594	0.0005	0.0667	0.0667	-0.0004	0.0672	0.0671
	LAD_{cc}	-0.0042	0.0709	0.0710	-0.00026	0.0817	0.0816	-0.0012	0.0817	0.0817
	LAD_{ipw}	-0.0041	0.0711	0.0711	0.0001	0.0826	0.0825	-0.0012	0.0822	0.0822
	LS_{ipw}	-0.0005	0.0581	0.0580	0.0001	0.0653	0.0653	-0.0006	0.0654	0.0654
	Proposed	-0.0017	0.0798	0.0797	-0.00095	0.1284	0.1284	-0.0021	0.1274	0.1274
2	LS_{cc}	0.9871	0.2596	1.0206	-0.0034	0.2999	0.2997	0.0048	0.2971	0.2970
	LMS_{cc}	-0.0092	0.7954	0.7950	-0.0086	1.1212	1.1206	-0.1085	1.0762	1.0812
	MM_{cc}	0.0021	0.0807	0.0806	0.0019	0.0910	0.0910	-0.0026	0.0922	0.0922
	S_{cc}	-0.0014	0.1028	0.1028	-0.0031	0.1224	0.1224	-0.0005	0.1261	0.1261
	$Huber_{cc}$	0.1310	0.0701	0.1486	0.00056	0.0774	0.0773	-0.0010	0.0812	0.0811
	LAD_{cc}	0.0951	0.0777	0.1228	-0.0019	0.0884	0.0883	-0.00034	0.0930	0.0929
	LAD_{ipw}	0.0946	0.0777	0.1224	-0.0023	0.0882	0.0882	-0.0005	0.0932	0.0932
	LS_{ipw}	0.9869	0.2595	1.0204	-0.0023	0.2988	0.2996	0.0052	0.2991	0.2990
	Proposed	0.0050	0.0831	0.0832	0.0040	0.1370	0.1370	0.0039	0.1323	0.1323
3	LS_{cc}	-0.0206	0.2834	0.2840	-2.5799	0.2943	2.5966	-2.6329	0.2948	2.6493
	LMS_{cc}	-0.0043	0.8859	0.8855	-0.0755	1.2362	1.2379	-0.0658	1.2894	1.2904
	MM_{cc}	-0.0043	0.0813	0.0814	-0.00009	0.0960	0.0959	-0.0001	0.0978	0.0978
	S_{cc}	-0.0047	0.0948	0.0949	0.0011	0.1153	0.1153	0.0009	0.1167	0.1167
	$Huber_{cc}$	-0.0132	0.1831	0.1835	-1.2990	0.5573	1.4133	-1.3182	0.5453	1.4265
	LAD_{cc}	-0.0110	0.1789	0.1792	-1.0644	0.5634	1.2042	-1.0797	0.5425	1.2082
	LAD_{ipw}	-0.0294	0.1340	0.1371	-0.7106	0.2215	0.7443	-0.7260	0.2192	0.7584
	LS_{ipw}	-0.0867	0.2792	0.2922	-2.5451	0.2880	2.5614	-2.5996	0.2901	2.6158
	Proposed	-0.0021	0.0979	0.0979	0.00029	0.1465	0.1464	-0.0018	0.1458	0.1457
4	LS_{cc}	-0.0353	0.2919	0.2939	-2.7731	0.2933	2.7885	-2.8046	0.3009	2.8207
	LMS_{cc}	-0.1051	0.8241	0.8304	-0.0982	1.1934	1.1969	-0.1321	1.2157	1.2223
	MM_{cc}	-0.00046	0.0808	0.0807	0.0045	0.0919	0.0920	-0.0033	0.0896	0.0896
	S_{cc}	-0.0034	0.1316	0.1316	-0.0567	0.4273	0.4308	-0.0669	0.4413	0.4461
	$Huber_{cc}$	-0.0470	0.3324	0.3356	-2.7383	0.3433	2.7597	-2.7682	0.3534	2.7907
	LAD_{cc}	-0.0474	0.3527	0.3557	-2.7428	0.3669	2.7672	-2.7627	0.3762	2.7882
	LAD_{ipw}	-0.0766	0.3532	0.3613	-2.7251	0.3682	2.7498	-2.7451	0.3768	2.7709
	LS_{ipw}	-0.0678	0.2916	0.2992	-2.7617	0.2907	2.7769	-2.7942	0.2976	2.8100
	Proposed	-0.0031	0.0817	0.0817	-0.0006	0.1174	0.1173	-0.0037	0.1200	0.1200

表 4.4 : 当 $n = 400$ 时基于 probit 回归缺失模型下不同估计方法的模拟结果.

		β_1			β_2			β_3		
Case	Method	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
1	LS_{cc}	0.0016	0.0605	0.0605	0.0017	0.0688	0.0688	-0.0025	0.0691	0.0691
	LMS_{cc}	-0.0058	0.6504	0.6501	-0.0028	0.8778	0.8773	0.0207	0.8359	0.8357
	MM_{cc}	0.0006	0.0919	0.0919	-0.00025	0.1071	0.1070	0.0016	0.1064	0.1064
	S_{cc}	0.0027	0.1148	0.1148	0.0017	0.1421	0.1420	-0.00097	0.1408	0.1408
	$Huber_{cc}$	0.0011	0.0619	0.0619	0.0017	0.0705	0.0705	-0.0017	0.0712	0.0712
	LAD_{cc}	0.0021	0.0734	0.0734	0.0009	0.0858	0.0857	-0.0017	0.0853	0.0853
	LAD_{ipw}	0.0019	0.0739	0.0739	0.0020	0.0927	0.0927	-0.0011	0.0904	0.0904
	LS_{ipw}	0.0010	0.0614	0.0614	0.0025	0.0739	0.0739	-0.0014	0.0729	0.0729
	Proposed	0.0011	0.0819	0.0819	0.0038	0.1281	0.1281	-0.0083	0.1309	0.1311
2	LS_{cc}	0.9927	0.2722	1.0293	0.0035	0.3104	0.3103	0.0142	0.3103	0.3104
	LMS_{cc}	-0.0246	0.8159	0.8159	-0.1022	1.1808	1.1846	-0.0682	1.1202	1.1217
	MM_{cc}	0.0019	0.0864	0.0864	-0.0017	0.0959	0.0959	0.0020	0.0976	0.0975
	S_{cc}	0.0023	0.1079	0.1079	-0.0006	0.1216	0.1215	0.0026	0.1249	0.1248
	$Huber_{cc}$	0.1285	0.0765	0.1496	0.0015	0.0806	0.0806	0.0035	0.0876	0.0876
	LAD_{cc}	0.0954	0.0830	0.1264	0.00003	0.0923	0.0922	0.0049	0.0940	0.0941
	LAD_{ipw}	0.0937	0.0843	0.1260	0.0018	0.0980	0.0979	0.0064	0.1010	0.1011
	LS_{ipw}	0.9908	0.2756	1.0284	0.0072	0.3296	0.3295	0.0159	0.3272	0.3274
	Proposed	0.0060	0.0858	0.0860	0.0003	0.1353	0.1353	0.0010	0.1411	0.1410
3	LS_{cc}	0.6894	0.2691	0.7400	-2.6479	0.2760	2.6623	-2.6289	0.2820	2.6440
	LMS_{cc}	-0.0879	0.9288	0.9325	-0.0754	1.2570	1.2586	-0.0610	1.3054	1.3062
	MM_{cc}	0.0039	0.0897	0.0897	-0.0027	0.0964	0.0964	-0.0009	0.1026	0.1026
	S_{cc}	0.0084	0.1246	0.1248	-0.0164	0.2112	0.2117	-0.0101	0.2192	0.2193
	$Huber_{cc}$	0.3943	0.2421	0.4626	-1.5608	0.6255	1.6814	-1.5500	0.6392	1.6766
	LAD_{cc}	0.3322	0.2462	0.4134	-1.3309	0.6715	1.4905	-1.3261	0.6807	1.4904
	LAD_{ipw}	0.1281	0.1243	0.1784	-0.6531	0.2067	0.6850	-0.6486	0.2096	0.6816
	LS_{ipw}	0.5209	0.2700	0.5867	-2.5686	0.2670	2.5824	-2.5503	0.2741	2.5649
	Proposed	0.0030	0.0985	0.0985	-0.00001	0.1497	0.1496	-0.0017	0.1517	0.1516
4	LS_{cc}	0.7283	0.2848	0.7819	-2.8212	0.2766	2.8347	-2.7940	0.2816	2.8082
	LMS_{cc}	-0.0267	0.8116	0.8116	-0.1564	1.1559	1.1658	-0.0729	1.2174	1.2190
	MM_{cc}	-0.0027	0.0865	0.0865	-0.0021	0.0946	0.0946	0.0022	0.0995	0.0995
	S_{cc}	0.0182	0.2029	0.2036	-0.0882	0.4946	0.5021	-0.0822	0.5073	0.5137
	$Huber_{cc}$	0.6883	0.3311	0.7637	-2.7920	0.3212	2.8104	-2.7693	0.3249	2.7883
	LAD_{cc}	0.6853	0.3579	0.7731	-2.7867	0.3430	2.8077	-2.7749	0.3447	2.7962
	LAD_{ipw}	0.6050	0.3525	0.7001	-2.7444	0.3417	2.7655	-2.7260	0.3438	2.7475
	LS_{ipw}	0.6474	0.2826	0.7064	-2.7930	0.2733	2.8063	-2.7663	0.2788	2.7803
	Proposed	0.0009	0.0851	0.0850	-0.0018	0.1176	0.1175	0.0016	0.1216	0.1216

表 **4.1 (续):**当 $n = 200$ 时基于 **logistic** 回归缺失模型下不同估计方法的模拟结果.

		β_1			β_2			β_3		
Case	Method	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
5	LS_{cc}	0.4462	0.4612	0.6415	-2.6025	0.4964	2.6494	-2.6261	0.4914	2.6717
	LMS_{cc}	-0.0332	0.7952	0.7955	-0.0617	1.1751	1.1762	-0.0225	1.1440	1.1437
	MM_{cc}	-0.0108	0.2122	0.2124	0.0043	0.2633	0.2632	-0.0054	0.2593	0.2592
	S_{cc}	-0.0044	0.1514	0.1514	-0.0127	0.2627	0.2629	-0.0118	0.2418	0.2420
	$Huber_{cc}$	0.0924	0.2795	0.2943	-1.3389	0.7532	1.5360	-1.3565	0.7784	1.5638
	LAD_{cc}	0.0744	0.2816	0.2911	-1.1530	0.8018	1.4042	-1.1723	0.8139	1.4269
	LAD_{ipw}	0.0189	0.1810	0.1819	-0.6872	0.3163	0.7565	-0.6892	0.3159	0.7581
	LS_{ipw}	0.3572	0.4600	0.5822	-2.5683	0.4895	2.6145	-2.5972	0.4851	2.6421
	Proposed	-0.0022	0.1159	0.1159	-0.0031	0.1884	0.1883	-0.0029	0.1820	0.1819

表 **4.2 (续):**当 $n = 400$ 时基于 **logistic** 回归缺失模型下不同估计方法的模拟结果.

		β_1			β_2			β_3		
Case	Method	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
5	LS_{cc}	1.2134	0.4806	1.3050	-2.6183	0.4723	2.6605	-2.6386	0.4774	2.6814
	LMS_{cc}	-0.0248	0.8406	0.8405	-0.0172	1.1869	1.1864	-0.1232	1.2397	1.2452
	MM_{cc}	0.00041	0.2425	0.2423	0.0052	0.3249	0.3247	0.0056	0.2906	0.2905
	S_{cc}	0.0019	0.1684	0.1684	-0.00092	0.2535	0.2534	-0.0170	0.2560	0.2564
	$Huber_{cc}$	0.5186	0.3899	0.6487	-1.5020	0.8049	1.7039	-1.5289	0.8336	1.7412
	LAD_{cc}	0.4581	0.4040	0.6107	-1.3293	0.8632	1.5847	-1.3622	0.8979	1.6313
	LAD_{ipw}	0.1709	0.1816	0.2493	-0.6110	0.2978	0.6797	-0.6216	0.3067	1.6930
	LS_{ipw}	1.003	0.5044	1.1201	-2.5385	0.4664	2.5810	-2.5650	0.4665	2.6070
	Proposed	0.00097	0.1263	0.1262	-0.00067	0.1856	0.1855	-0.0086	0.1849	0.1850

表 **4.3 (续):**当 $n = 200$ 时基于 **probit** 回归缺失模型下不同估计方法的模拟结果.

		β_1			β_2			β_3		
Case	Method	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
5	LS_{cc}	0.4519	0.3351	0.5625	-2.6016	0.3244	2.6217	-2.6147	0.3294	2.6354
	LMS_{cc}	-0.0545	0.7371	0.7388	-0.0601	1.1161	1.1171	-0.1228	1.0800	1.0864
	MM_{cc}	0.0006	0.0820	0.0820	0.0010	0.0928	0.0928	-0.0001	0.0939	0.0938
	S_{cc}	0.0011	0.1071	0.1070	-0.0122	0.2125	0.2127	-0.0078	0.1769	0.1770
	$Huber_{cc}$	0.0897	0.1685	0.1908	-1.1789	0.5320	1.2933	-1.1905	0.5364	1.3057
	LAD_{cc}	0.0742	0.1639	0.1798	-0.9374	0.5210	1.0724	-0.9505	0.5305	1.0884
	LAD_{ipw}	0.0418	0.1149	0.1222	-0.6305	0.1835	0.6566	-0.6374	0.1847	0.6635
	LS_{ipw}	0.3906	0.3308	0.5117	-2.5685	0.3204	2.5884	-2.5827	0.3241	2.6030
	Proposed	0.00038	0.0816	0.0815	0.0003	0.1297	0.1297	-0.0041	0.1262	0.1262

表 4.4 (续):当 n = 400 时基于 probit 回归缺失模型下不同估计方法的模拟结果.

		β_1			β_2			β_3		
Case	Method	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
5	LS_{cc}	1.1955	0.3311	1.2405	-2.6241	0.3250	2.6442	-2.6426	0.3320	2.6633
	LMS_{cc}	-0.0156	0.8831	0.8828	-0.0851	1.1811	1.1835	-0.1327	1.2161	1.2227
	MM_{cc}	0.0034	0.0882	0.0882	0.0022	0.0967	0.0966	-0.0018	0.0996	0.0996
	S_{cc}	0.0031	0.1288	0.1288	-0.0071	0.1968	0.1968	-0.0210	0.2433	0.2441
	$Huber_{cc}$	0.4815	0.2691	0.5515	-1.4172	0.6288	1.5503	-1.4234	0.6413	1.5611
	LAD_{cc}	0.4013	0.2754	0.4867	-1.1820	0.6678	1.3574	-1.1917	0.6858	1.3748
	LAD_{ipw}	0.1736	0.1135	0.2074	-0.5667	0.1732	0.5925	-0.5787	0.1869	0.6082
	LS_{ipw}	1.0284	0.3304	1.0801	-2.5468	0.3187	2.5667	-2.5634	0.3245	2.5838
	Proposed	0.0056	0.0850	0.0851	0.0031	0.1277	0.1277	0.0006	0.1273	0.1273

表 4.5 :不同估计方法在不同情形下的 MMAPE.

		方法								
		logistic 缺失机制								
	Case	LS_{cc}	LMS_{cc}	MM_{cc}	S_{cc}	$Huber_{cc}$	LAD_{cc}	LAD_{ipw}	LS_{ipw}	Proposed
	1	0.6651	0.9898	0.6743	0.6652	0.6621	0.6589	0.6585	0.6651	0.6771
	2	1.2198	1.2958	0.7595	1.2193	0.7566	0.7486	0.7487	1.2211	0.7651
n = 200	3	3.2917	1.4406	0.8139	3.2892	1.9784	1.7790	1.3078	3.2779	0.8357
	4	3.3673	1.3215	0.7645	3.3646	3.2953	3.3059	3.3118	3.3666	0.7661
	5	3.3384	1.3054	0.7583	3.3358	1.8644	1.6739	1.1569	3.3208	0.7686
	1	0.6715	1.0098	0.6627	0.6716	0.6703	0.6681	0.6681	0.6716	0.6779
	2	1.1836	1.2935	0.7470	1.1831	0.7605	0.7549	0.7548	1.1835	0.7626
n = 400	3	3.3088	1.4560	0.8059	3.3062	1.8825	1.6358	1.2635	3.2880	0.8226
	4	3.4081	1.3785	0.7560	3.4055	3.3708	3.3743	3.3699	3.4038	0.7686
	5	3.3481	1.2910	0.7493	3.3454	1.6982	1.4416	1.1107	3.3188	0.7620
		probit 缺失机制								
	1	0.6689	0.9956	0.6939	0.6688	0.6666	0.6608	0.6619	0.6701	0.6821
	2	1.2320	1.3000	0.7664	1.2314	0.7573	0.7501	0.7512	1.2424	0.7680
n = 200	3	3.1428	1.4385	0.8270	3.1403	2.0617	1.8922	1.1926	3.1020	0.8353
	4	3.2073	1.3681	0.7792	3.2048	3.1448	3.1520	3.1491	3.2028	0.7675
	5	3.1939	1.3192	0.7667	3.1915	1.9524	1.7851	1.0604	3.1393	0.7616
	1	0.6722	1.0011	0.6626	0.6722	0.6705	0.6675	0.6678	0.6723	0.6765
	2	1.1932	1.3073	0.7469	1.1926	0.7593	0.7552	0.7556	1.1959	0.7615
n = 400	3	3.1490	1.4555	0.8041	3.1467	2.0458	1.8210	1.1632	3.1043	0.8217
	4	3.2226	1.3356	0.7534	3.2202	3.1931	3.1935	3.1853	3.2193	0.7645
	5	3.1970	1.3543	0.7478	3.1945	1.8772	1.6407	1.0287	3.1310	0.7614

这表明新方法不仅有准确的参数估计值, 还有较好的预测值.

4.6 实例分析

在这一节中, 选用 Boston Housing Study 数据集来检验我们所提出的稳健估计方法, 该数据可在 http://lib.stat.cmu.edu/datasets/boston 上找到. 该数据包含 506 个样本和 14 个变量. 和 Chang et al. (2018) 一样, 我们取房价中位数的对数作为响应变量 (y), 另外 13 个变量分别记为 (CRIM, x_1), (ZN, x_2), (INDUS, x_3), (CHAS, x_4), (NOX, x_5), (RM, x_6), (AGE, x_7), (DIS, x_8), (RAD, x_9), (TAX, x_{10}), (OTRATIO, x_{11}), (B, x_{12}), (LSTAT, x_{13}). 各个协变量的具体介绍请参看 Chang et al. (2018) 和 Jiang et al. (2019). 为了消除量纲的影响, 我们将 13 个协变量进行标准化处理.

根据 Harrison and Rubinfeld (1978), 考虑使用线性模型

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\theta} + \epsilon_i, \quad i = 1, \ldots, 506$$

拟合上面介绍的 Boston Housing 数据集, 其中 $\boldsymbol{x}_i = (1, x_{i,1}, \ldots, x_{i,13})^\top$, ϵ_i 是误差项, $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_{13})^\top$. 为了说明协变量中含有异常值对估计的影响, 根据 Chang et al.(2019) 的想法, 对三个重要的显著协变量(RM, AGE, DIS) 上加上 10% 的异常值, 具体地, $x_6 = x_6 + 10$, $x_7 = x_7 + 10$, $x_8 = x_8 + 10$. 同时, 人工产生响应变量缺失, 缺失机制模型为$\delta_i = bernoulli(\pi_i)$,$\pi_i = \omega(\boldsymbol{x}_i, \boldsymbol{\eta}) = \{1 + \exp(-\eta_1 - \boldsymbol{\eta}_2^\top \boldsymbol{x}_i)\}^{-1}$,其中$\boldsymbol{\eta} = (\eta_1, \boldsymbol{\eta}_2^\top)^\top = (1.5, 0.2 * \mathbf{1}_{13})^\top$, 其中 $\mathbf{1}_{13}$ 代表一个元素全为 1 的 13 维向量. 这样, 就产生了一个既含有缺失数据也含有异常值的实例数据集合 $\{(\delta_i, y_i, \boldsymbol{x}_i), i = 1, \ldots, n\}$.

下面, 我们使用 LS_{cc}, LMS_{cc}, MM_{cc}, $Huber_{cc}$, S_{cc}, LAD_{cc}, LS_{ipw}, LAD_{ipw} 以及所提出的方法分析上述数据. 相对应的参数估计值将在表 4.6 中给出, 且基于 200 次 Bootstap 的参数估计的标准差将在表 4.7 给出. 对每一次 bootstrap, 我们从 506 个样本中随机抽取 400 个样本作参数估计. 此外, 针对不同的估计方法, 我们还计算完全可观测数据的预测值与真实值之间的绝对中位数 (MAPE). 从表 4.6 和 表 4.7, 我们可以看到: (i) 与所提出的方法相比, LS_{cc}, LMS_{cc}, MM_{cc}, $Huber_{cc}$, LS_{ipw}, LAD_{ipw} 有较大的参数估计值; (ii) 与其他方法相比, 所提出的估计

表 4.6 : 基于 Boston Housing Study 数据的参数估计.

变量	LS_{cc}	LMS_{cc}	MM_{cc}	$Huber_{cc}$	S_{cc}	LAD_{cc}	LS_{ipw}	LAD_{ipw}	Proposed
常数项	4.4372	2.5672	2.9801	3.0153	3.0515	3.0430	3.0421	4.4369	2.9524
x_1	-0.1043	0.1671	-0.2574	-0.0770	-0.0716	-0.0702	-0.0719	-0.1101	-0.4095
x_2	-0.0057	0.4160	0.0293	0.0235	0.0258	0.0261	0.0227	0.0054	0.0087
x_3	1.2122	0.2773	0.0107	0.0138	0.0371	0.0328	0.0276	1.2548	0.0044
x_4	0.0112	-0.0770	0.0172	0.0163	0.0215	0.0271	0.0224	0.0001	0.0158
x_5	-0.6753	0.4528	-0.0614	-0.0838	-0.0885	-0.0869	-0.0888	-0.6839	0.0033
x_6	0.1351	-0.4710	0.1257	0.1553	0.0824	0.1017	0.1139	0.1321	0.1452
x_7	0.5109	0.0458	-0.0616	-0.0696	0.0046	-0.0146	-0.0248	0.5212	-0.0552
x_8	0.1752	-0.1030	-0.1003	-0.1056	-0.0820	-0.0899	-0.0939	0.2211	-0.0777
x_9	-0.6567	2.8944	0.1246	0.0844	0.0848	0.0788	0.0685	-0.5508	0.1727
x_{10}	0.3943	-3.1698	-0.0740	-0.0614	-0.0825	-0.0732	-0.0586	0.2804	-0.0992
x_{11}	-0.4401	0.1098	-0.0661	-0.0668	-0.0851	-0.0799	-0.0847	-0.4429	-0.0539
x_{12}	0.2465	-0.2244	0.0560	0.0784	0.0453	0.0459	0.0482	0.2574	0.0434
x_{13}	-0.0341	-1.3521	-0.0982	-0.0702	-0.1754	-0.1471	-0.1279	-0.0088	-0.1000
MAPE	0.1098	0.2427	0.0972	0.0819	0.1026	0.0985	0.0946	0.1023	0.0939

表 4.7 : 基于 Boston Housing Study 数据的参数估计的标准差.

变量	LS_{cc}	LMS_{cc}	MM_{cc}	$Huber_{cc}$	S_{cc}	LAD_{cc}	LS_{ipw}	LAD_{ipw}	Proposed
常数项	0.1229	0.2120	0.3069	0.1969	0.0062	0.0080	0.1224	0.0085	0.0087
x_1	0.1263	0.4912	0.1753	0.0747	0.0110	0.0128	0.1304	0.0145	0.0328
x_2	0.1494	0.0991	0.0154	0.0115	0.0079	0.0123	0.1478	0.0108	0.0097
x_3	0.2894	0.1648	0.0271	0.0155	0.0106	0.0122	0.2818	0.0124	0.0094
x_4	0.1249	0.5894	1.0899	0.7124	0.0049	0.0086	0.1197	0.0080	0.0044
x_5	0.2152	0.1301	0.0323	0.0201	0.0088	0.0114	0.2153	0.0123	0.0154
x_6	0.1709	0.1233	0.0315	0.0178	0.0108	0.0144	0.1589	0.0134	0.0124
x_7	0.2143	0.1148	0.0216	0.0174	0.0120	0.0117	0.2075	0.0099	0.0072
x_8	0.1946	0.1011	0.0216	0.0165	0.0088	0.0100	0.1923	0.0104	0.0169
x_9	0.4785	0.2275	0.0399	0.0217	0.0151	0.0214	0.4798	0.0207	0.0230
x_{10}	0.5380	0.2160	0.0327	0.0228	0.0158	0.0207	0.5365	0.0216	0.0224
x_{11}	0.1793	0.0657	0.0186	0.0103	0.0070	0.0119	0.1776	0.0117	0.0059
x_{12}	0.1109	0.1690	0.0620	0.0172	0.0070	0.0090	0.1170	0.0089	0.0139
x_{13}	0.2465	0.1724	0.0454	0.0279	0.0155	0.0224	0.2315	0.0198	0.0110

拥有较小的标准差, 这说明新提出的估计要更有效; (iii) S_{cc} 和所提出的估计方法拥有较小的 MAPE 值. 以上事实说明: 当响应变量存在随机缺失且数据集中含有异常值时, 本章提出的估计方法不仅能获得无偏估计, 还能有效地提高参数估计的有效性.

4.7 定理证明

在这一节, 我们将考虑定理 4.1 和定理 4.2 的证明. 本文主要考虑非光滑矩条件的大样本性质, 但是其理论结果可以推广到光滑矩条件上. 首先给出一些正则条件.

条件 A (一些有关缺失数据模型的正则条件).

(A1) 对所有的 $\boldsymbol{x}_i$, 在 $\boldsymbol{\eta}_0$ 的局部领域内的 $\boldsymbol{\eta}$, $\omega(\boldsymbol{x}_i, \boldsymbol{\eta})$ 关于 $\boldsymbol{\eta}$ 的三阶导数 $\frac{\partial^3 \omega(\boldsymbol{x}_i, \boldsymbol{\eta})}{\partial \eta_{l_1} \eta_{l_2} \eta_{l_3}}$ 存在且 $\frac{\partial^3 \omega(\boldsymbol{x}_i, \boldsymbol{\eta})}{\partial \eta_{l_1} \eta_{l_2} \eta_{l_3}} \leq K(\boldsymbol{x}_i)$, $\mathbb{E}[K(\boldsymbol{x}_i)] < \infty$, 其中 $1 \leq \eta_{l_1}, \eta_{l_2}, \eta_{l_3} \leq m$;

(A2) 对所有的 $\boldsymbol{x}_i$, 在 $\boldsymbol{\eta}_0$ 的局部领域内的 $\boldsymbol{\theta}$, $\omega(\boldsymbol{x}_i, \boldsymbol{\eta}) \in (0, 1)$;

(A3) 矩阵 S_{11} 是正定矩阵, 且 $\mathbb{E}[\|\omega_1(\boldsymbol{x}_i, \boldsymbol{\eta}_0)\|] < \infty$, $\mathbb{E}[\omega(\boldsymbol{x}_i, \boldsymbol{\eta}_0)\|\omega_1(\boldsymbol{x}_i, \boldsymbol{\eta}_0)\|] < \infty$;

(A4) 对所有的 $\boldsymbol{x}_i$, $|h(\boldsymbol{x}_i)| < \infty$ 且 $\max_{1 \leq i \leq n} w(h(\boldsymbol{x}_i)) < \infty$.

条件 B (一些有关矩条件的正则条件).

(B1) $\mathbb{E}\{\boldsymbol{g}_1(y_i, \boldsymbol{x}_i, \delta_i, \boldsymbol{\beta}_0, \sigma_0, \boldsymbol{\eta}_0)\} = 0$;

(B2) $\boldsymbol{\theta} = (\boldsymbol{\theta}^\top, \sigma)^\top \in \Theta$ 且 $\boldsymbol{\theta}_0$ 是 Θ 的一个内点;

(B3) $\mathbb{E}\{\boldsymbol{g}_1(y_i, \boldsymbol{x}_i, \delta_i, \boldsymbol{\beta}_0, \sigma_0, \boldsymbol{\eta}_0)\}$ 在 $\boldsymbol{\theta}_0$ 出关于 $\boldsymbol{\theta}$ 的导数 $\mathbf{B}_{\boldsymbol{\theta}}$ 存在且使得 $\mathbf{B}_{\boldsymbol{\theta}}^\top W \mathbf{B}_{\boldsymbol{\theta}}$ 非奇异;

(B4) $\sqrt{n}\bar{\boldsymbol{g}}_1(\boldsymbol{\theta}_0, \hat{\boldsymbol{\eta}}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_2)$, 其中 $\Sigma_2 = Var(\boldsymbol{g}_1(z_i, \boldsymbol{\theta}_0, \boldsymbol{\eta}_0) + \mathbf{B}_{\boldsymbol{\eta}} \psi(\boldsymbol{x}_i, \boldsymbol{\eta}_0))$;

(B5) 对任意 $\zeta_n \to 0$, $\sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 \leq \zeta_n} \sqrt{n}\|\bar{\boldsymbol{g}}_1(\boldsymbol{\theta}, \boldsymbol{\eta}_0) - \bar{\boldsymbol{g}}_1(\boldsymbol{\theta}_0, \boldsymbol{\eta}_0) - \mathbb{E}\{\boldsymbol{g}_1(z_i, \boldsymbol{\theta}_0, \boldsymbol{\eta}_0)\}\|_2 / [1 + \sqrt{n}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2] \xrightarrow{p} 0$.

条件 (A1)-(A3)经常用于缺失机制模型的刻画, 譬如 Qin et al.(2009). 条件 (A4) 是用来限制加权极大似然函数 (Maronna et al., 2006). 条件 (B1)-(B5) 是非光滑矩条件研究中常见的条件, 参看 Newey and Mcfadden (1994).

定理 4.1 的证明. 首先证明相合解的存在性, 再证明渐近正态性.

1. 相合解的存在性. 对一些 $\varrho > 0$, 令 $S_\varrho = \{\boldsymbol{\eta} : \|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| \le \varrho\}$ 是一个在 $\boldsymbol{\eta}_0$ 的局部领域内的紧集. 根据 Ferguson (1996) 的定理 17, 当 $\boldsymbol{\eta}$ 的参数空间 $\boldsymbol{\Theta} = S_\varrho$ 时, 存在一个强相合的的序列 $\hat{\boldsymbol{\eta}}_r$ 是 $\widetilde{S}_n(\boldsymbol{\eta}) = 0$ 的解.

下面来验证 Ferguson (1996) 的定理 17 (pp. 114) 的条件. 显然, Ferguson (1996) 定理 17 的条件 (1)、(2) 和 (5) 自动满足. 另外, 因为 $\omega(\boldsymbol{x}_i, \boldsymbol{\eta})$ 关于 $\boldsymbol{\eta}$ 是连续函数, 所以定理 17 的条件 (4) 也成立. 最后验证定理 17 的条件 (3). 为此, 令

$$U(\delta, \boldsymbol{x}, \boldsymbol{\eta}) = w\{h(\boldsymbol{x})\}\Big[\delta \log \omega(\boldsymbol{x}, \boldsymbol{\eta}) + (1-\delta)\log\{1-\omega(\boldsymbol{x}, \boldsymbol{\eta})\} - w\{h(\boldsymbol{x})\}\Big[\delta \log \omega(\boldsymbol{x}, \boldsymbol{\eta}_0) + (1-\delta)\log\{1-\omega(\boldsymbol{x}, \boldsymbol{\eta}_0)\},$$

$$\Psi(\delta, \boldsymbol{x}, \boldsymbol{\eta}) = w\{h(\boldsymbol{x}_i)\}\frac{\omega_1(\boldsymbol{x}_i, \boldsymbol{\eta})\omega_1(\boldsymbol{x}_i, \boldsymbol{\eta})^\top}{\omega(\boldsymbol{x}_i, \boldsymbol{\eta})(1-\omega(\boldsymbol{x}_i, \boldsymbol{\eta}))}$$

和

$$\dot{\Psi}(\delta, \boldsymbol{x}, \boldsymbol{\eta}) = \partial\Psi(\delta, \boldsymbol{x}, \boldsymbol{\eta})/\partial\boldsymbol{\eta}^\top.$$

将 $U(\delta, \boldsymbol{x}, \boldsymbol{\eta})$ 在 $\boldsymbol{\eta}_0$ 出展开有

$$\begin{aligned} U(\delta, \boldsymbol{x}, \boldsymbol{\eta}) &= U(\delta, \boldsymbol{x}, \boldsymbol{\eta}_0) + \Psi(\delta, \boldsymbol{x}, \boldsymbol{\eta})^\top(\boldsymbol{\eta} - \boldsymbol{\eta}_0) \\ &\quad + (\boldsymbol{\eta} - \boldsymbol{\eta}_0)^\top\Big[\int_0^1\int_0^1 \lambda\dot{\Psi}(\delta, \boldsymbol{x}, \lambda\mu(\boldsymbol{\eta} - \boldsymbol{\eta}_0))d\lambda d\mu\Big](\boldsymbol{\eta} - \boldsymbol{\eta}_0). \end{aligned}$$

因为 $U(\delta, \boldsymbol{x}, \boldsymbol{\eta}_0) = 0$, $\Psi(\delta, \boldsymbol{x}, \boldsymbol{\eta}_0) = 0$ 以及条件 (A3) 成立, 根据可积函数的性质有 $U(\delta, \boldsymbol{x}, \boldsymbol{\eta})$ 在 S_ϱ 上一致有界. 故, 完成相合解和存在性问题的证明.

2. 渐近正态性. 注意到 $\widetilde{S}_n(\boldsymbol{\eta}) = \frac{1}{n}\sum_{i=1}^n w\{h(\boldsymbol{x}_i)\}\frac{\delta_i - \omega(\boldsymbol{x}_i, \boldsymbol{\eta})}{\omega(\boldsymbol{x}_i, \boldsymbol{\eta})(1-\omega(\boldsymbol{x}_i, \boldsymbol{\eta}))}\frac{\partial\omega(\boldsymbol{x}_i, \boldsymbol{\eta})}{\partial\boldsymbol{\eta}}$. 将$\widetilde{S}_n(\boldsymbol{\eta})$ 泰勒展开有

$$\widetilde{S}_n(\boldsymbol{\eta}) = \widetilde{S}_n(\boldsymbol{\eta}_0) + \Big[n^{-1}\sum_{i=1}^n \dot{\Psi}(\delta, \boldsymbol{x}, \boldsymbol{\eta}_0)\Big](\boldsymbol{\eta} - \boldsymbol{\eta}_0) + \boldsymbol{c}$$

其中 $\boldsymbol{c} = (c_1, \ldots, c_m)^\top$, $c_l = n^{-1}\sum_{i=1}^n\sum_{j=1}^m\sum_{k=1}^m \frac{\partial\widetilde{S}_{n,l}(\boldsymbol{\eta}^*)}{\partial\eta_j\partial\eta_k}(\eta_j - \eta_{0j})(\eta_k - \eta_{0k})(l = 1, \ldots, m)$, $\widetilde{S}_{n,l}(\boldsymbol{\eta})$ 是 $\widetilde{S}_n(\boldsymbol{\eta})$ 的第 l 个分量, $\|\boldsymbol{\eta}^* - \boldsymbol{\eta}\| \le \|\boldsymbol{\eta} - \boldsymbol{\eta}_0\|$.

令 $\boldsymbol{\eta} = \hat{\boldsymbol{\eta}}_r$, 其中 $\hat{\boldsymbol{\eta}}_r$ 是 $\boldsymbol{\eta}_0$ 的任意强相合估计量且满足 $\widetilde{S}_n(\hat{\boldsymbol{\eta}}_r) = 0$. 这样, $\|\hat{\boldsymbol{\eta}}_r - \boldsymbol{\eta}_0\| = o_p(1)$. 于是根据条件 (A1), 很容易有 $\boldsymbol{c} = o_p(n^{-1/2})$.

综上, 有

$$\boldsymbol{\eta} - \boldsymbol{\eta}_0 = \Big[- n^{-1} \sum_{i=1}^{n} \dot{\Psi}(\delta, \boldsymbol{x}, \boldsymbol{\eta}_0) \Big]^{-1} \widetilde{S}_n(\boldsymbol{\eta}_0) + o_p(n^{-1/2}).$$

根据大数定律有 $-n^{-1} \sum_{i=1}^{n} \dot{\Psi}(\delta, \boldsymbol{x}, \boldsymbol{\eta}_0) = S_{11}(\boldsymbol{\eta}_0) + o_p(1)$. 注意到 $\mathbb{E}\{\Psi(\delta, \boldsymbol{x}, \boldsymbol{\eta}_0)\} = 0$ 和 $\mathrm{Var}(\Psi(\delta, \boldsymbol{x}, \boldsymbol{\eta}_0)) = S_{11}(\boldsymbol{\eta}_0)$.

根据中心极限定理, 得到

$$\sqrt{n}\widetilde{S}_n(\boldsymbol{\eta}_0) \xrightarrow{\mathcal{L}} N(\mathbf{0}, S_{11}(\boldsymbol{\eta}_0)).$$

注意到 $\widetilde{S}_n(\boldsymbol{\eta}_0) = O_p(n^{-1/2})$, 则

$$\boldsymbol{\eta} - \boldsymbol{\eta}_0 = S_{11}(\boldsymbol{\eta}_0)^{-1} \widetilde{S}_n(\boldsymbol{\eta}_0) + o_p(n^{-1/2}).$$

将上式两边同时乘以 $\sqrt{n}$, 有 $\sqrt{n}(\boldsymbol{\eta} - \boldsymbol{\eta}_0) = S_{11}(\boldsymbol{\eta}_0)^{-1} \sqrt{n}\widetilde{S}_n(\boldsymbol{\eta}_0) + o_p(1)$. 最后, 由 S_{11} 是正定矩阵和 Slutsky's 定理有

$$\sqrt{n}(\boldsymbol{\eta} - \boldsymbol{\eta}_0) \xrightarrow{\mathcal{L}} N(\mathbf{0}, S_{11}^{-1}(\boldsymbol{\eta}_0)).$$

定理 4.2 的证明. 定理 4.2 的证明与 Newey & Mcfadden (1994) 的定理 7.1 的证明非常类似. 在这里, 令 $\hat{\boldsymbol{g}}_1(\boldsymbol{\theta}) = \bar{\boldsymbol{g}}_1(\boldsymbol{\theta})$ 和 $\boldsymbol{g}_0(\boldsymbol{\theta}) = \mathbb{E}[\boldsymbol{g}_1(z_i, \boldsymbol{\theta}, \boldsymbol{\eta})]$. 那么本定理的证明就简化成验证 Newey & Mcfadden (1994) 的定理 7.1 的条件.

记 $Q(\boldsymbol{\theta}) = -\boldsymbol{g}_0(\boldsymbol{\theta})^{\top} W \boldsymbol{g}_0(\boldsymbol{\theta})/2$, $\hat{Q}(\boldsymbol{\theta}) = -\hat{\boldsymbol{g}}_1(\boldsymbol{\theta})^{\top} \hat{W} \hat{\boldsymbol{g}}_1(\boldsymbol{\theta}) + \hat{\Delta}(\boldsymbol{\theta})$, 其中 $\hat{\Delta}(\boldsymbol{\theta})$ 是如下给出的特定函数. 根据条件 (B1) 和 (B2), 有

$$\begin{aligned} Q(\boldsymbol{\theta}) &= -[\mathbf{B}_{\boldsymbol{\theta}}^{\top}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|)]^{\top} W [\mathbf{B}_{\boldsymbol{\theta}}^{\top}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|)]/2 \\ &= Q(\boldsymbol{\theta}_0) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^{\top} H (\boldsymbol{\theta} - \boldsymbol{\theta}_0\|)/2 + o(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2), \end{aligned}$$

其中 $H = -\mathbf{B}_{\boldsymbol{\theta}}^{\top} W \mathbf{B}_{\boldsymbol{\theta}}$ 和 $Q(\boldsymbol{\theta}_0) = 0$. 于是, $Q(\boldsymbol{\theta})$ 在 $\boldsymbol{\theta}_0$ 处二阶连续可导. 同时, 根据 W 正定和 $\mathbf{B}_{\boldsymbol{\theta}}^{\top} W \mathbf{B}_{\boldsymbol{\theta}}$ 非奇异以及 H 非负定可以得到, 在 $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ 的局部领域内存在参

数 $\boldsymbol{\theta}$ 使得 $Q(\boldsymbol{\theta})$ 有唯一最小值(最小值为零). 这样, Newey & Mcfadden (1994) 的定理 7.1 的假设 (i)-(ii) 成立.

根据 Slusky Theorem 和条件 (B4),

$$\hat{D} = -\mathbf{B}_{\boldsymbol{\theta}}^{\top}\hat{W}\sqrt{n}\hat{\boldsymbol{g}}_1(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} N(0,\Omega),$$

其中 $\Omega = \mathbf{B}_{\boldsymbol{\theta}}^{\top}W\Sigma_2\hat{W}\mathbf{B}_{\boldsymbol{\theta}}$. 故, Newey & Mcfadden (1994) 的定理 7.1 的假设 (v) 也成立.

剩下的问题就变成了验证 Newey & Mcfadden (1994) 的定理7.1 的初始假设和条件 (v). 令$\hat{\varepsilon}(\boldsymbol{\theta}) = [\hat{\boldsymbol{g}}_1(\boldsymbol{\theta}) - \hat{\boldsymbol{g}}_1(\boldsymbol{\theta}_0) - \boldsymbol{g}_0(\boldsymbol{\theta})]/[1 + \sqrt{n}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|]$. 我们可以得到

$$\begin{aligned}\hat{\boldsymbol{g}}_1(\boldsymbol{\theta})^{\top}\hat{W}\hat{\boldsymbol{g}}_1(\boldsymbol{\theta}) &= (1 + 2\sqrt{n}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| + \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2)\hat{\varepsilon}(\boldsymbol{\theta})^{\top}\hat{W}\hat{\varepsilon}(\boldsymbol{\theta}) + \boldsymbol{g}_0(\boldsymbol{\theta})^{\top}\hat{W}\boldsymbol{g}_0(\boldsymbol{\theta})\\ &\quad +\hat{\boldsymbol{g}}_1(\boldsymbol{\theta}_0)^{\top}\hat{W}\hat{\boldsymbol{g}}_1(\boldsymbol{\theta}_0) + 2\boldsymbol{g}_0(\boldsymbol{\theta})^{\top}\hat{W}\hat{\boldsymbol{g}}_1(\boldsymbol{\theta})\\ &\quad +2[\boldsymbol{g}_0(\boldsymbol{\theta}) + \hat{\boldsymbol{g}}_1(\boldsymbol{\theta})]^{\top}\hat{W}\hat{\varepsilon}(\boldsymbol{\theta})[1 + \sqrt{n}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|].\end{aligned}$$

令 $\hat{Q}(\boldsymbol{\theta}) = -\hat{\boldsymbol{g}}_1(\boldsymbol{\theta})^{\top}\hat{W}\hat{\boldsymbol{g}}_1(\boldsymbol{\theta})/2 + \hat{\varepsilon}(\boldsymbol{\theta})^{\top}\hat{W}\hat{\varepsilon}(\boldsymbol{\theta})/2 + \hat{\boldsymbol{g}}_1(\boldsymbol{\theta})^{\top}\hat{W}\varepsilon(\boldsymbol{\theta})$. 对任意 $\zeta_n \to 0$, 根据条件 (B5), 有

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le\zeta_n}\left|\hat{Q}(\boldsymbol{\theta})-\{-\hat{\boldsymbol{g}}_1(\boldsymbol{\theta})^{\top}\hat{W}\hat{\boldsymbol{g}}_1(\boldsymbol{\theta})/2\}\right| \le O_p(1)\cdot\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le\zeta_n}\left\{\|\hat{\varepsilon}(\boldsymbol{\theta})\|\|\hat{\varepsilon}(\boldsymbol{\theta})\|+O_p(n^{-1/2})\right\} = o_p(n^{-1}).$$

于是, $\hat{Q}(\hat{\boldsymbol{\theta}}) \ge \sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le\zeta_n}\hat{Q}(\boldsymbol{\theta}) - o_p(n^{-1})$. 所以, Newey & Mcfadden (1994) 的定理 7.1 的初始假设成立.

最后, 验证 Newey & Mcfadden (1994) 的定理 7.1 的条件 (v). 注意到 $\hat{\varepsilon}(\boldsymbol{\theta}_0) = 0$. 记 $\hat{R}(\boldsymbol{\theta}) = \sqrt{n}[\hat{Q}(\hat{\boldsymbol{\theta}}) - \hat{Q}(\hat{\boldsymbol{\theta}}_0) - \{\partial\hat{Q}(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \{Q(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}_0)\}/\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|$, 则

$$\begin{aligned}&\sqrt{n}\left|\hat{R}(\boldsymbol{\theta})/[1 + \sqrt{n}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|]\right| \le \textstyle\sum_{j=1}^{r}\hat{r}_j(\boldsymbol{\theta}),\\ &\hat{r}_1(\boldsymbol{\theta}) = \sqrt{n}(2\sqrt{n}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| + \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2)/[\|\boldsymbol{\theta} - \boldsymbol{\theta}_0(1 + \sqrt{n}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|)],\\ &\hat{r}_2(\boldsymbol{\theta}) = \sqrt{n}|[\boldsymbol{g}_0(\boldsymbol{\theta}) + \mathbf{B}_{\boldsymbol{\theta}}^{\top}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)]^{\top}\hat{W}\boldsymbol{g}_0(\boldsymbol{\theta})|/[\|\boldsymbol{\theta} - \boldsymbol{\theta}_0(1 + \sqrt{n}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|)],\\ &\hat{r}_3(\boldsymbol{\theta}) = n|[\boldsymbol{g}_0(\boldsymbol{\theta}) + \hat{\boldsymbol{g}}_1(\boldsymbol{\theta}_0)]^{\top}\hat{W}\hat{\varepsilon}(\boldsymbol{\theta})|[1 + \sqrt{n}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|],\\ &\hat{r}_4(\boldsymbol{\theta}) = \sqrt{n}|\boldsymbol{g}_0(\boldsymbol{\theta})^{\top}\hat{W}\hat{\varepsilon}(\boldsymbol{\theta})|/\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|,\\ &\hat{r}_5(\boldsymbol{\theta}) = \sqrt{n}|\boldsymbol{g}_0(\boldsymbol{\theta})^{\top}[\hat{W} - W]\boldsymbol{g}_0(\boldsymbol{\theta})|/[\|\boldsymbol{\theta} - \boldsymbol{\theta}_0(1 + \sqrt{n}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|)].\end{aligned}$$

然后, 对 $\zeta_n \to 0$ 和 $U = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \le \zeta_n\}$, 有

$\sup_U \hat{r}_1(\boldsymbol{\theta}) \le Cn \cdot \sup_U \|\hat{\varepsilon}(\boldsymbol{\theta})\|^2 \|\hat{W}\| = o_p(1)$,

$\sup_U \hat{r}_2(\boldsymbol{\theta}) \le \sqrt{n} \sup_U \{o_p(\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|)\|\hat{W}\|\|\boldsymbol{g}_0(\boldsymbol{\theta})\|\} = \sup_U \{o_p(\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|)O_p(1) = o_p(1)$,

$\sup_U \hat{r}_3(\boldsymbol{\theta}) \le \left\{ \sup_U \sqrt{n}\|\boldsymbol{g}_0(\boldsymbol{\theta})\|/(\sqrt{n}\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|) + \sqrt{n}\|\hat{\boldsymbol{g}}_1(\boldsymbol{\theta}_0)\| \right\} \|\hat{W}\| \sup_U \sqrt{n}\|\hat{\varepsilon}(\boldsymbol{\theta})\| =$
$\{\sup_U O(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|) + O_p(1)\}o_p(1) = o_p(1)$,

$\sup_U \hat{r}_4(\boldsymbol{\theta}) \le \sup_U (\|\boldsymbol{g}_0(\boldsymbol{\theta})\|/\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|)\|\hat{W}\| \sup_U \sqrt{n}\|\hat{\varepsilon}(\boldsymbol{\theta})\| = o_p(1)$,

$\sup_U \hat{r}_5(\boldsymbol{\theta}) \le \sup_U (\|\boldsymbol{g}_0(\boldsymbol{\theta})\|^2/\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2)\|\hat{W} - W\| = o_p(1)$.

这样, 就完成 Newey & Mcfadden (1994) 的定理 7.1 的条件的验证.

4.8 本章小结

在本章中考虑响应变量随机缺失下回归模型的稳健估计问题. 在给定恰当的缺失机制模型条件下, 先建立稳健的似然函数, 利用极大似然方法实现缺失机制模型的稳健参数估计的目的. 然后, 利用估计方程的思想构建包含回归参数的稳健估计方程, 继而使用广义矩方法去估计未知参数. 在一定的正则条件下, 证明了参数估计的相合性和渐近正态性. 利用数值模拟和实例分析来说明所提出的方法的表现, 结果表明所提方法能很好地处理缺失数据和异常值共存的问题.

第五章 总结及展望

本文在响应变量缺失数据的背景下, 研究了超高维数据下的特征筛选, 广义部分线性单指标模型的参数估计以及数据存在异常点时的稳健估计.

首先针对超高维响应变量随机缺失数据, 提出了一种基于带有非参数插补的调整 Spearman 秩相关关系 (ASRC) 的新的特征筛选方法. 所提的特征筛选方法有以下几个优点. 它不需要假定任何模型; 它对异常值、重尾数据、相关协变量、模型的错误指定和缺失数据机制的错误指定都有较强的稳健性; 它对于响应变量和协变量的单调变换也是不变的. 在一些正则化的条件下, 建立所提筛选方法的确定筛选和秩相合的性质. 再次针对具有不可忽略的响应变量的广义部分线性单指标模型, 提出了一种参数估计和未知连接函数的 PS-WEE 估计方法. 虽然已有文献研究了 GPLSM, 但所有的研究都只关注完全观测数据. 本文中提出的方法推广了现有的方法, 允许考虑不可忽略缺失响应变量的信息. 假设响应概率模型为半参数逻辑回归模型, 其中非参数部分采用非参数回归方法估计, 参数部分采用 GMM 方法估计. 而且系统地研究了所得估计量的理论性质. 模拟研究表明了该方法的优越性, 最后考虑响应变量随机缺失下回归模型的稳健估计问题. 在给定恰当的缺失机制模型条件下, 先建立稳健的似然函数, 利用极大似然方法实现缺失机制模型的稳健参数估计的目的. 然后, 利用估计方程的思想构建包含回归参数的稳健估计方程, 继而使用广义矩方法去估计未知参数. 在一定的正则条件下, 获得了参数估计的相合性和渐进正态性. 利用数值模拟和实例分析来说明所提出的方法的表现, 结果表明此方法能很好地处理缺失数据和异常值共存的问题.

基于本文的研究结果, 后续研究方向有如下几个方面:

(1) 假设条件 MCI 是为了分别估计 X_k 和 Y 的分布函数 $F_k(x)$, $F(y)$. 当条件 MCI 不满足时, 很难从理论上获得想要的确定筛选和秩相合的性质, 这是未来需要研究的问题.

(2) 近来, 采用迭代筛选和模型拟合的方法, 研究提高以模型为基础的特征

筛选方法. 但是对于所提的特征筛选方法, 本文还没有采用类似的迭代过程,这是一个值得进一步研究的有趣课题. 这篇文章仅仅考虑 MAR 缺失机制和没有删失数据, 将所提的特征筛选方法推广到非随机缺失的响应变量和具有截尾和缺失数据的生存数据中是有一个有待研究的课题. 而且, 所提的特征筛选方法是在假设候选预测变量是连续的且数据不作为一组的前提下提出的. 当候选预测变量和响应变量是离散的或数据作为一组时, 需要提出一种新的特征筛选方法, 这也是一个值得进一步研究的有趣课题.

(3) 更复杂的统计模型的稳健估计是一个值得研究的课题. 同时, 当参数维数是高维情形时, 处理缺失数据和异常值共存将更加复杂, 其原因是一个异常值点可能会导致模型选择的结果不相合. 故, 高维条件下缺失数据和异常值共存时的参数估计和模型选择也是今后需要研究且有实际意义的问题.

参考文献

[1] 唐年胜, 李会琼. 应用回归分析[M]. 科学出版社, 2014.

[2] Candes, E., Tao, T. The Dantzig selector: statistical estimation when p is much larger than n (with discussion)[J]. Annals of Statistics, 2007, 35, 2313-2404.

[3] Carroll, R., Fan, J., Gijbels, I., and Wand, M. Generalized Partially Linear Single-Index Models[J]. Journal of the American Statistical Association, 1997, 92, 477-489.

[4] Chang, J., Tang, C.Y., Wu, Y. Marginal empirical likelihood and sure independence feature screening[J]. Annals of Statistics, 2013, 41, 2123-2148.

[5] Chang, L., Roberts, S., and Welsh, A. Robust Lasso Regression Using Tukey’s Biweight Criterion[J]. Technometrics, 2018, 60, 36–47.

[6] Chen, S. X., Wang, D. Empirical likelihood for estimating equations with missing values[J]. Annals of Statistics, 2009, 37, 490-517.

[7] Cheng, P.E. Nonparametric estimation of mean functionals with data missing at random[J]. Journal of the American Statistical Association, 1994, 89, 81-87.

[8] Cheng, P.E., Chu, C.K. Kernel estimation on distribution functions and quantiles with missing data[J]. Statistical Sinica, 1996, 6, 63-78.

[9] Cui, X., Härdle, W.K., and Zhu, L.X. The EFM Approach for Single-Index Models[J]. Annals of Statistics, 2011, 39, 1658-1688.

[10] Dong, C.H., Gao, J.t., Dag Tjøstheim. Estimations for single-index and partially linear single-index integrated models[J]. The Annals of Statistics, 2016, 44(1): 425–453.

[11] Fan, J., Feng, Y., Song, R. Nonparametric independence screening in sparse ultrahigh-dimensional additive models[J]. Journal of the American Statistical Association, 2011, 106, 544-557.

[12] Fan, J., Li, R. Variable selection via non-concave penalized likelihood and its oracle properties[J]. Journal of the American Statistical Association, 2001, 96, 1348-1360.

[13] Fan, J., Lv, J. Sure independence screening for ultrahigh dimensional feature space[J]. Journal of the Royal Statistical Society Series B, 2008, 70, 849-911.

[14] Fan, J., Song, R. Sure independence screening in generalized linear models with NP-dimensionality[J]. The Annals of Statistics, 2010, 38, 3567-3604.

[15] Fan, A.X., Tang, N.S. Sensitivity analysis of partially linear models with response missing at random[J]. Communications in Statistics–Simulation and Computation, 2017, 46, 5323-5339.

[16] Fang, F., Shao, J. Model selection with nonignorable nonresponse[J]. Biometrika, 2016, 103, 861-874.

[17] Gervini, D., and Yohai, V. J. A Class of Robust and Fully Efficient Regression Estimators[J]. The Annals of Statistics, 2002, 30, 583 – 616.

[18] Graciela Boente and Daniela Rodriguez. Robust estimates in generalized partially linear single-index models[J]. Test, 2012, 21: 386 – 411.

[19] Garcia, R.I., Ibrahim, J.G., Zhu, H.Variable selection for regression models with missing data[J]. Statist. Sinica, 2010, 20, 149-165.

[20] Hampel, F. Contributions to the Theory of Robust Estimation[D]. Ph.D. dissertation, University of California Berkeley, CA, 1968.

[21] Hampel, F. A General Qualitative Definition of Robustness[J]. The Annals of Mathematical Statistics, 1971, 42, 1887 – 1896.

[22] Han, P., and Wang, L. Estimation With Missing Data: Beyond Double Robustness[J]. Biometrika, 2013, 100, 417 – 430.

[23] Han, P. Multiply Robust Estimation in Regression Analysis With Missing Data[J]. Journal of the American Statistical Association, 2014, 109:507, 1159-1173.

[24] Han, P. A General Method for Quantile Estimation with Missing Data[J]. Journal of the Royal Statistical Society, Series B, in press.

[25] He, X., Wang, L., Hong, H.G. Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data[J]. The Annals of Statistics, 2013, 41, 342-369.

[26] Horvitz, D,G, Thompson, D,J,. A generalization of sampling without replacement from a finite universe[J]. Journal of the American Statistical Association, 1952, 47, 663-685.

[27] Huber, P.Robust Regression: Asymptotics, Conjectures and Monte Carlo[J]. The Annals of Statistics, 1973, 1, 799 – 821.

[28] Ibrahim, J.G., Zhu, H.T., Tang, N.S. Model selection criteria for missing-data problems using the EM algorithm[J]. Journal of the American Statistical Association, 2008, 103, 1648-1658.

[29] Ibrahim, J.G., Lipsitz, S.R., Chen, M. Missing covariates in generalized linear models when the missing data mechanism is nonignorable[J]. Journal of the Royal Statistical Society Series B, 1999, 61, 173-190.

[30] Jiang, D.P., Zhao, P.Y., Tang, N.S., Azen, S.P. A propensity score adjustment method for regression models with nonignorable missing covariates[J]. Computational Statistics and Data Analysis, 2015, 94, 98-119.

[31] Jiang, Y., Wang, Y., Fu, L., Wang, X. Robust Estimation Using Modified Huber’s Functions With New Tails[J]. Technometrics, 2019, 61(1), 111-122.

[32] Shao, J. Semiparametric inverse propensity weighting for nonignorable missing data[J]. Biometrika, 2016, 103, 1:175-187.

[33] Kim, Y., Choi, H., Oh, H.S. Smoothly clipped absolute deviation on high dimensions[J]. Journal of the American Statistical Association, 2008, 103, 1665-1673.

[34] Kim, J.K., Yu, C.L. A semiparametric estimation of mean functionals with nonignorable missing data[J]. Journal of the American Statistical Association, 2011, 106, 157-165.

[35] Koenker, R. Quantile Regression[M]. Cambridge: Cambridge University, Press, 2005.

[36] Koenker, R., and Bassett, G. J. Regression Quantiles[J]. Econometrica, 1978, 46, 33 – 50.

[37] Lai, P., Liu, Y., Liu, Z., Wan, Y. Model free feature screening for ultrahigh dimensional data with responses missing at random[J]. Computational Statistics and Data Analysis, 2017, 105, 201-216.

[38] Lee, E. R., Noh, H., Park, B.U. Model selection via Bayesian information criterion for quantile regression models[J]. Journal of the American Statistical Association, 2014, 109, 216-229.

[39] Lee, S.Y., Tang, N.S. Bayesian analysis of nonlinear structural equation models with nonignorable missing data[J]. Psychometrika, 2006, 71, 541-564.

[40] Li, G., Peng, H., Zhang, J., Zhu, L. Robust rank correlation based screening[J]. The Annals of Statistics, 2012. 40, 1846-1877.

[41] Li, R., Zhong, W., Zhu, L. Feature screening via distance correlation learning[J]. Journal of the American Statistical Association, 2012, 107, 1129-1139.

[42] Liang, H., Liu, X., Li, R., and Tsai, C. Estimation and Testing for Partially Linear Single-Index Models[J]. The Annals of Statistics, 2010, 38, 3811-3836.

[43] Xue, L. Estimation and empirical likelihood for single-index models with missing data in the covariates[J]. Computational Statistics and Data Analysis. 2013, 68, 82-97.

[44] Little, R.J.A., Rubin, D.B. Statistical Analysis With Missing Data[M]. second ed. Wiley, New York, 2002.

[45] Liu, J., Li, R., Wu, R. Feature selection for varying coefficient models with ultrahigh -dimensional covariates[J]. Journal of the American Statistical Association, 2014, 109, 266-274.

[46] Long, Q., Johnson, B.A. Variable selection in the presence of missing data: resampling and imputation[J]. Biostatistics, 2015, 16, 596 – 610.

[47] Ma, S.J., Li, R., Tsai, C.L. Variable screening via quantile partial correlation[J]. Journal of the American Statistical Association, 2017, 112, 650-663.

[48] Mai, Q., Zou, H. The fused Kolmogorov filter: a nonparametric model-free screening method[J]. The Annals of Statistics, 2015, 43, 1471-1497.

[49] Maronna, R.A., Yohai, V. J. Asymptotic Behavior of General M-Estimates for Regression and Scale With Random Carriers[J]. Probability Theory and Related Fields, 1981, 58, 7 – 20.

[50] Newey, W.K., McFadden, D. Large sample estimation and hypothesis testing[J]. In Handbook of Econometrics, 1994, 4, 2111-2245.

[51] Lai, P., Tian, Y., Lian, H. Estimation and variable selection for generalised partially linear single-index models[J]. Journal of Nonparametric Statistics, 2014, 26(1): 171-185.

[52] Xu, P.R., Zhang, J., Huang, X.F., Wang, T. Efficient estimation for marginal generalized partially linear single-index models with longitudinal data[J]. Test, 2016, 25: 413 – 431.

[53] Wang, Q.H., Zhong, T., Wolfgang Karl HÄrdle. An Extended Single-index Model with Missing Response at Random[J]. Scandinavian Journal of Statistics, 2016, 43(4): 1140-1152.

[54] Zou, Q.M., Zhu, Z.Y., Wang, J.L. Local influence analysis for penalized Gaussian likelihood estimation in partially linear single-index models[J]. Ann Inst Stat Math, 2009, 61: 905-918

[55] Qin, J., Shao, J., Zhang, B. Efficient and doubly robust imputation for covariate dependent missing responses[J]. Journal of the American Statistical Association, 2008, 103, 797-810.

[56] Qin, J, Leung, D., Shao, J. Estimation with survey data under nonignorable nonresponse or informative sampling[J]. Journal of the American Statistical Association, 2002, 97, 193-200.

[57] Qin, J., Zhang, B., and Leung, D. H. Y. Empirical Likelihood in Missing Data Problems[J]. Journal of the American Statistical Association, 2009, 104, 1492 – 1503.

[58] Riddles, M.K., Kim, J.K., Im, J. A propensity-score-adjustment method for nonignorable nonresponse[J]. Journal of Survey Statistics and Methodology, 2016, 4, 215-245.

[59] Robins, J., Rotnitzky, A., Zhao, P. Estimation of regression coefficients when some regressors are not always observed[J], Journal of the American Statistical Association, 1994, 89, 846-866.

[60] Robins, J. M., and Rotnitzky, A. Semiparametric Efficiency in Multivariate Regresion Models With Missing Data[J]. Journal of the American Statistical Association, 1995, 90, 122 – 129.

[61] Rousseeuw, P.J. Least Median of Squares Regression[J]. Journal of the American Statistical Association, 1984, 79, 871 – 880.

[62] Rousseeuw, P.J., Leroy, A.M. Robust Regression and Outlier Detection[M]. New York: Wiley Online Library. 1987.

[63] Rousseeuw, P.J., and Yohai, V. Robust Regression by Means of Estimators[J]. Robust and Nonlinear Time Series, 1984, 26, 256 – 272.

[64] Rosenwald, A., Wright, G., Chan, W.C., Connors, J.M., Campo, E., Fisher, R.I., Gascoyne, R.D., Muller-Hermelink, H.K., Smeland, E.B, Giltnane, J.M, Hurt, E.M., Zhao, H., Averett, L., Yang, L., Wilson, W.H., Jaffe, E.S., Simon, R., Klausner, R.D., Powell, J., Duffey, P.L., Longo, D.L., Greiner, T.C., Weisenburger, D.D., Sanger, W.G., Dave, B.J., Lynch, J.C., Vose, J., Armitage, J.O., Montserrat, E., Lopez-Guillermo, A., Grogan, T.M., Miller, T.P., LeBlanc, M., Ott, G., Kvaloy, S., Delabie, J., Holte, H., Krajci, P, Stokke, T., Staudt, L.M. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma[J]. N. Engl. J. Med., 2002, 346, 1937-1947.

[65] Rubin, D.B., Little, R.J. Statistical Analysis with Missing Data (2nd ed)[M]. New York: Wiley, 2002.

[66] Scheetz, T.E., Kim, K.Y.A., Swiderski, R.E., Philp, A.R., Braun, T.A., Knudtson, K.L., Dorrance, A.M., DiBona, G.F., Huang, J., Casavant, T.L., Sheffield, V.C., Stone, E.M. Regulation of gene expression in the mammalian eye and its relevance to eye disease[J]. Proceedings of the National Academy of Sciences, 2006, 103, 14429-14434.

[67] Tan, Z. A Distributional Approach for Causal Inference Using Propensity Scores[J]. Journal of the American Statistical Association, 2006, 101, 1619 – 1637.

[68] Tan, Z. Bounded, Efficient and Doubly Robust Estimation With Inverse Weighting[J]. Biometrika, 2010, 97, 661 – 682.

[69] Tang, N.S., Zhao, P., Zhu, H. Empirical likelihood for estimation equations with nonignorably missing data[J]. Statistica Sinica, 2014, 24, 723-747.

[70] Tang, N.S., Tang, L. Estimation and variable selection in generalized partially nonlinear models with nonignorable missing responses[J]. Statistics and Its Interface, 2018, 11, 1-18.

[71] Tibshirani, R. Shrinkage and selection via LASSO[J]. Journal of the Royal Statistical Society Series B, 1996, 58, 267-288.

[72] Li, T.T., Yang, H. Inverse probability weighted estimators for single index models with missing covariates[J]. Computational Statistics and Data Analysis, 2016, 94:98-119.

[73] Tsiatis, A.A. Semiparametric Theory and Missing Data[J]. New York: Springer, 2006.

[74] van der Laan, M.J., Gruber, S. Collaborative Double Robust Targeted Maximum Likelihood Estimation[J]. The International Journal of Biostatistics, 2010, 6, Article 17.

[75] Wang, L., Wu, Y., Li, R. Quantile regression for analyzing heterogeneity in ultra-high dimension[J]. Journal of the American Statistical Association, 2012, 107, 214-222.

[76] Wang, J., Xue, L., Zhu, L., and Chong, Y. Estimation for a Partial-Linear Single-Index Model[J]. The Annals of Statistics, 2010, 38: 246-274.

[77] Wang, Q., Linton, O., Hardle, W. Semi-parametric regression analysis with missing response at random[J]. Journal of the American Statistical Association, 2004, 99(466): 334-345.

[78] Wang, Q., Rao, J. Empirical likelihood-based inference under imputation for missing response data[J]. The Annals of Statistics, 2002, 30, 896-924.

[79] Wang, Q., Rao, J. Empirical likelihood for linear regression models under imputation for missing responses[J]. Canadian Journal of Statistics, 2001, 29(4): 597-608.

[80] Wang, S., Shao, J., Kim, J. An instrument variable approach for identification and estimation with nonignorable nonresponse[J]. Statist. Sinica, 2014, 24: 1097-1116.

[81] White, H. Maximum Likelihood Estimation of Misspecified Models[J]. Econometrica, 1982, 50, 1 – 25.

[82] Xie, J., Lin, Y., Yan, X., Tang, N. Category-adaptive variable screening for ultrahigh dimensional heterogeneous categorical data[J]. Journal of the American Statistical Association, 2019. (in press, doi.org/10.1080/01621459.2019.1573734).

[83] Guo, X., Niu, C., Yang, Y., Xu, W. Empirical likelihood for single index model with missing covariates at random[J]. Statistics, 2015, 49(3): 588-601.

[84] Xue, L. Empirical likelihood confidence intervals for response mean with data missing at random[J]. Scandinavian Journal of Statistics, 2009b, 36(4): 671-685.

[85] Xue, L. Empirical likelihood for linear models with missing responses[J]. Journal of Multivariate Analysis, 2009a, 100(7): 1353-1366.

[86] Yan, X., Tang, N., Xie, J., Ding, X., Wang, Z. Fused mean-variance filter for feature screening[J]. Computational Statistics and Data Analysis, 2018, 122, 18-32.

[87] Yohai, V. J. High Breakdown-Point and High Efficiency Robust Estimates for Regression[J]. The Annals of Statistics, 1987, 15, 642 – 656.

[88] Zhao, P.Y., Tang, N.S., Qu, A., Jiang, D.P. Semiparametric estimating equations inference with nonignorable nonresponse[J]. Statistica Sinica, 2017, 27, 89-113.

[89] Zhao, P.Y., Tang, N.S., Jiang, D.P. Efficient inverse probability weighted method for quantile regression with nonignorable missing data[J]. Statistics, 2017, 51, 363-386.

[90] Zhao, S.D., Cai, T.T. and Li, H. Propensity score adjustment with several follow-ups[J]. Biometrika, 2015, 2, 439-448.

[91] Zhu, L.P., Li, L., Li, R., Zhu, L.X. Model-free feature screening for ultrahigh-dimensional data[J]. Journal of the American Statistical Association, 2011, 106, 1464-1475.

[92] Zhu, L.X., Xue, L.G. Empirical likelihood confidence regions in a partially linear single-index model[J]. J R Stat Soc Ser B, 2006, 68: 549-570.

[93] Yu, Z.X., He, B., Chen, M. Empirical Likelihood for Generalized Partially Linear Single-index Models[J]. Communications in Statistics — Theory and Methods, 2014, 43: 4156-4163.

[94] Zou, H. The adaptive lasso and its oracle properties[J]. Journal of the American Statistical Association, 2006, 101, 1418-1429.

[95] Zou, H., Hastie, T. Regularization and variable selection via the elastic net[J]. Journal of the Royal Statistical Society Series B, 2005, 67, 301-320.

[96] Zou, H., Li, R. One-step sparse estimates in non-concave penalized likelihood models[J]. Annals of Statistics, 2008, 36, 1509-1533.

[97] Zou, H., Zhang, H.H. On the adaptive elastic-net with a diverging number of parameters[J]. Annals of Statistics, 2009, 37, 1733-1751.

发表文章目录

[1] Yuan-Yuan Ju, Nian-Sheng Tang, Xiao-Xia Li. Bayesian Local Influence Analysis of Skew-normal Spatial Dynamic Panel Data Models, *Journal of Statistical Computation and Simulation* (SCI), 2018, 88, 2342-2364.

[2] Xiao-xia Li, Jin-Han Xie, Nian-Sheng Tang and Xiao-Dong Yan. A nonparametric feature screening method for ultrahigh-dimensional missing response. *Computational Statistics and Data Analysis* (SCI), accepted.

致　　谢

转眼间, 博士四年就要结束了! 在此向给予我关心和帮助的老师、同事、同学、朋友和家人表示衷心的感谢!

感谢我的导师唐年胜教授, 是他给了我这样一个机会, 让我有幸在云南大学攻读博士学位. 博士刚入学后, 由于从基础数学专业, 又工作了7年, 跨度比较大的转到读统计专业博士, 刚开始很迷茫, 困惑, 不知道该如何开始, 是唐老师耐心地给我指明方向, 并一步一步指导我如何查文献, 看文献、推导理论、编写和调试程序以及整理文稿. 由于自己的数学思维已经根深蒂固, 如何让我有统计的思维思考问题, 唐老师不厌其烦的一直给我强调, 倾注了大量的精力和心血, 用实际的例子告诉我该怎么思考. 唐老师除了在学习上严格要求外, 在生活上, 教给我们做人的道理, 告诉我们要正直, 谦虚, 感恩. 要做一个情商和智商并存的人. 做一个学术和人品并重的人. 他的人格魅力深深影响着我, 这将使我受益终生. 同时, 还要感谢陈老师给我们授课和给我们做学术报告的老师们; 使我们很快的了解学科前言, 开阔视野!

感谢云南大学大学数学与统计学院对我的培养. 感谢李会琼老师、潘冬冬老师、周建军老师、陈黎老师、赵慧老师、唐安民老师、赵普映师兄、严晓东师兄、樊爱霞师姐, 他们在我博士学习期间给予了我的帮助和鼓励. 此外, 特别真诚地感谢读博期间我的同学们谢锦瀚、杨志煌、夏林丽、程伟丽、易凤婷、张韵琪、蒋芬、李伟在学习给予我巨大的帮助和生活中的陪伴.

感谢我的同事们对我的帮助和照顾!

感谢我的父母、公婆、丈夫和儿子, 正是有了他们的支持和鼓励, 我才能顺利完成学业.