

硕 士 学 位 论 文

加速失效时间模型下基于倾向得分的平均因果 效应分析

**The Analysis of the Average of Causal Effects via the Propensity
Score under the Accelerated Failure Time Model**

作 者 姓 名： 马文婷

学 科、 专 业： 概率论与数理统计

学 号： 21801045

指 导 教 师： 王晓光

完 成 日 期： 2021 年 5 月 5 日

大连理工大学

Dalian University of Technology

摘 要

在观察性研究中,倾向得分被广泛地用于估计因果效应.近些年提出的协变量平衡倾向得分方法,在估计倾向得分时,通过优化协变量平衡得到更为稳健的估计.在医学和流行病学研究中的某些情况下,感兴趣的结果变量是生存时间.在估计处理分配机制对生存时间的影响时,由于生存数据具有删失和截断等特征,使得完整数据下的估计方法不能直接使用.本文考虑的结果变量为带有右删失的生存时间,基于调整后的协变量平衡倾向得分和加权最小二乘方法来估计倾向得分模型和加速失效时间模型(结果模型),然后运用四种估计方法给出平均因果效应的估计.在本文中,四种估计方法分别取为 Horvitz-Thompson 估计、逆倾向得分加权估计、加权最小二乘估计和双稳健估计.在模拟研究中,考虑以下四种情形: (a) 倾向得分模型与结果模型均被正确指定; (b) 结果模型被正确指定而倾向得分模型被错误指定; (c) 倾向得分模型被正确指定而结果模型被错误指定; (d) 倾向得分模型与结果模型均被错误指定.我们在这四种情形下分别对四种估计方法进行比较,并给出模型中参数的估计情况.同时,还考虑了删失率和样本量对参数估计以及平均因果效应估计的影响.最后,结合原发性胆汁性肝硬化数据和乳腺癌数据,将本文的方法应用到实际的例子中.

本文内容安排如下:第一章,阐述论文的研究背景及意义,分析国内外研究现状并提出本文的研究内容;第二章,介绍本文涉及的相关的理论知识;第三章,详细论述模型及其参数估计方法,并给出平均因果效应的四种估计;第四章,通过在不同情形下的数值模拟对四种估计的性能进行比较;第五章,将本文的方法应用到原发性胆汁性肝硬化数据和乳腺癌数据中.最后进行总结,给出结论.

关键词: 因果推断; 协变量平衡倾向得分; 加速失效时间模型; 右删失数据

The Analysis of the Average of Causal Effects via the Propensity Score under the Accelerated Failure Time Model

Abstract

The propensity score is widely used to estimate causal effects in observational studies. The covariate balancing propensity score methodology recently proposed can obtain a more robust estimator by optimizing the covariate balance when estimating the propensity score. In some cases of medical and epidemiological studies, the outcome of interest is the survival time. When estimating the effect of the treatment assignment mechanism on the survival time, many approaches for complete data can't be used directly for the survival data with characters such as censoring, truncation mechanisms, etc. In this paper, the outcome is the right-censored survival time and we adopt a propensity score model and an accelerated failure time model (outcome model) which are based on methods of the adjusted covariate balancing propensity score and the weighted least squares. Then, we estimate the average of causal effects by four estimation methods including Horvitz-Thompson estimation, inverse propensity score weighting estimation, weighted least squares estimation, and doubly robust estimation. In the simulation study, we consider four scenarios as follows: (a) both the propensity score model and the outcome model are correctly specified; (b) the outcome model is correctly specified but the propensity score model is misspecified; (c) the propensity score model is correctly specified but the outcome model is misspecified; (d) both the propensity score model and the outcome model are misspecified. We compare the four estimations respectively under the four scenarios. Meanwhile, we assess the influence of censoring rate and sample size on estimators of the parameter and the average of causal effects. Finally, we apply the method in this paper to the primary biliary cirrhosis data and the breast cancer data.

The contents of this paper are as follows. Section 1 expounds the research background and significance of this paper, analyzes the research status at home and abroad and proposes the main content of this paper. In Section 2, we introduce the related theoretical knowledge involved in this paper. Section 3 discusses the models and its parameter estimation methods in detail and gives the four estimations of the average of causal effects. In Section 4, the performance of the four estimations is compared by the numerical simulation in different situations. In Section 5, we apply the method in this paper to the primary biliary cirrhosis data and the breast cancer data. Finally, we summarize and give the conclusion.

Key Words: Causal Inference; Covariate Balancing Propensity Score; Accelerated Failure Time Model; Right-Censored Data

目 录

摘 要	I
Abstract	II
1 绪论	1
1.1 研究背景及意义	1
1.2 国内外研究现状	2
2 预备知识	6
2.1 因果推断	6
2.1.1 潜在结果模型	6
2.1.2 协变量平衡倾向得分	7
2.2 生存分析	9
3 模型和估计	12
3.1 模型及参数估计	12
3.1.1 倾向得分模型与结果模型	12
3.1.2 倾向得分模型的参数估计	13
3.1.3 结果模型的参数估计	15
3.2 平均因果效应的估计	16
3.2.1 Horvitz-Thompson 估计	16
3.2.2 逆倾向得分加权估计	17
3.2.3 加权最小二乘估计	18
3.2.4 双稳健估计	19
4 模拟研究	21
4.1 参数估计	21
4.1.1 倾向得分模型的参数估计	21
4.1.2 结果模型的参数估计	23
4.2 平均因果效应的估计	28
5 实例分析	30
5.1 原发性胆汁性肝硬化数据	30
5.2 乳腺癌数据	34
结 论	37
参 考 文 献	38

攻读硕士学位期间发表学术论文情况	43
致 谢	44
大连理工大学学位论文版权使用授权书	45

1 绪论

1.1 研究背景及意义

对因果关系的讨论最早可以追溯到 2000 多年前, 亚里士多德指出: 真正的科学知识是关于原因的知识. 从古至今, 几乎所有科学研究的最终目的就是探讨事物之间的因果关系, 例如在哲学、自然科学、社会科学、医学等领域. 因为相关性对可能的因果关系没有足够的约束, 所以相关性并不意味着因果关系. 由于这一点的存在, 使得探索因果关系变得十分困难. 在某些情况下, 即使没有因果关系的两个因素, 仍有一定几率会建立起所谓的相关关系, 这使得大部分统计方法只能够适用于相关性研究而不能适用于因果关系的研究. 同样地, 即使表现为相互独立的两个因素之间仍可能存在着因果关系. 两个变量(如 A 和 B) 之间的高相关性的存在可能是下列条件的结果: (1) A 和 B 都由第三个变量 C 决定. 通过控制 C , A 和 B 之间的高相关性消失, 如果是这样的话, 我们就说这种相关性是假的; (2) A 导致 B . 在这种情况下, 即使控制另一组变量, 还能观察到 A 和 B 之间高度的相关关系; (3) B 导致 A . 在这种情况下, 相关性本身并没有告诉我们因果关系的方向. 关于虚假相关性的例子有很多, 一个国家的巧克力消费与诺贝尔奖获得者的数量之间存在很强的正相关关系, 但是人为地改变巧克力消费, 不会引起诺贝尔获奖者的数量发生变化; 学生的逻辑思维能力与鞋子尺码具有相关性, 但却并不存在因果关系. 此外, 还有很多具有因果关系的因素之间却表现出了虚假的独立性. 例如, 在放射性条件下工作的工人的寿命与人群平均寿命相比可能并不会表现出差距, 但这可能是由于该工作在招聘时就对身体健康有更为严格的要求, 如果这些工人不暴露在放射性条件下的话, 其寿命很可能会高于人群的平均寿命. 因此, 尽管工人寿命没有出现缩短的现象, 但也不能证明暴露在放射性环境下不会对健康造成危害. 这种现象称之为健康工人效应. 因果关系非常重要, 如何研究因果效应一直以来都是科学家们所关注的话题. 在统计学发展的早期阶段, 针对因果推断提出了列联表、路径分析等方法. Neyman^{[1][2]}在 1923 年提出了“潜在结果”这一概念. Fisher^[3]在 1935 年提出随机化试验方法. Rubin^[4]在 1974 年将 Fisher 的思想推广到因果推断的研究中, 给出了在观察性研究中的因果效应模型. 在之后的几十年里, 研究人员越来越关注对因果推断的研究, 到目前为止已有上万篇与之相关的文献.

Simpson^[5]在 1951 年描述了一种与人们常识相悖的现象, 总体上看男生比女生更容易被研究生院录取, 但根据学部划分后发现结果恰恰相反, 这种现象被称之为 Simpson 悖论(也被称为混杂偏倚). 无因果关系的两个因素可能会有相关关系, 而相关关系是由

于其它因素对它们的共同影响所造成的,我们称引起这种混杂偏倚现象的因素为混杂因素.除此之外,混杂因素也可能会使原本有因果关系的两个因素变得互不相关.那么自然而然地,判断是否存在混杂因素以及如何处理混杂因素是研究因果推断的关键问题之一.对于随机对照试验(Randomized Controlled Trial, RCT),由于进行了随机化处理(即将个体随机地分配到处理组或控制组),有效的避免了混杂因素的影响,所以采用 RCT 是研究因果推断最可靠的方法.但在某些问题上,可能存在重要的伦理或实践问题使得 RCT 并非总是可行的,所以在真实世界中,大多采用的是观察性研究(即研究者仅观察对象而不控制任何变量的研究).然而,对于观察性研究,由于处理组与控制组间基线协变量分布的差异而引起的混杂偏倚,除非加上很强的假定,否则很难得到有关因果推断的结论.

在生物医药、流行病学等领域,因果关系也扮演着十分重要的角色.例如,我们想要知道一种新开发的药物是否可以应用于癌症治疗,就可以将治疗效果量化为因果效应(也可以称之为处理效应).在研究过程中,我们经常遇到的数据是生存数据,但是由于生存数据通常是不完整的(即存在删失、截断等情况),这使得将因果推断方法应用到生存分析中时遇到了挑战.

综上所述,分析观察性研究中处理分配机制对生存数据的因果效应具有重要的理论意义和应用价值.

1.2 国内外研究现状

因果推断一直是许多领域中备受关注的研究问题. Rubin^[4]在 1974 年给出了潜在结果模型. Holland^[6]在 1986 年针对因果推断的各种统计模型进行了系统的总结与阐述,强调了统计实验关注因果关系的三个观点: (1) 对因果关系的分析应该从研究原因的结果开始,而不是试图定义某一既定结果的原因是什么; (2) 原因的结果总是相对于其他原因而言(即需要两个原因才能定义一个结果); (3) 不是每件事都能成为原因,特别是个体的属性从来都不是原因. Freedman^[7]在 1991 年介绍了从相关关系研究到因果关系研究的发展过程. Pearl^[8]在 1995 年通过将因果推断的基础理论与观察数据相结合,提出了因果网络图和外部干预的概念,给出了因果网络图模型. Pearl^[9]在 2000 年介绍了因果效应模型与因果网络图模型的等价性.

考虑到 Simpson 悖论的存在,混杂因素一直是学者关注的问题.对混杂因素的研究主要是从两个方向展开,一是研究判断混杂因素的准则,如可压缩性准则(可参见 Kleinbaum、Kupper 和 Morgenstern^[10], Whittemore^[11])、可比较性准则(可参见 Miettinen

和 Cook^[12], Greenland 和 Robins^[13])等; 另一方向是研究控制混杂因素的方法. 根据观测的混杂因素的情况, 现阶段处理混杂因素的方法分为两种. 对于已测量的混杂因素的情况, 有传统的统计分析方法, 如分层分析法、配对分析法、协方差分析法和多因素调整法等; 以及非传统的方法, 如匹配法, 倾向得分法及疾病风险评估(Disease Risk Score)法等. 对于未知或未测量的混杂因素的情况, 有工具变量(Instrumental Variable)法、双重差分模型(Difference-in-Differences)、事件率比较校正(Prior Event Rate Ratio Adjustment)法, 及其衍生方法等^{[14][15]}. 本文着重关注倾向得分法以及其之后的研究发展状况.

倾向得分(P propensity Score)被定义为在给定协变量的情况下接受处理的条件概率, 在因果推断的研究中起着核心作用. Rosenbaum 和 Rubin^[16]在 1983 年的文章中表明在给定协变量的情况下如果处理分配机制是强可忽略的, 那么可以通过调整倾向得分而不是高维的混杂因素来获得平均因果效应的无偏估计. 在接下来的三十年里, 基于倾向得分的几种新方法被开发出来, 它们已经成为跨学科应用研究人员的工具包中重要的组成部分. 大多数基于倾向得分的方法共有两个主要步骤:

第一步: 估计倾向得分模型, 最常用的是通过最大化一些似然准则的方法. 广义线性模型(McCullagh 和 Nelder^[17])已经成为这一目的的主要工具, 但更复杂的替代方案, 也在文献中提出, 如非参数回归(Hirano 等^[18])和提升树(McCaffrey 等^[19]).

第二步: 使用步骤一中估计的倾向得分来校正协变量失衡. 针对这一步骤已经提出了许多方法, 包括: 匹配(如 Rosenbaum 和 Rubin^[20], Rosenbaum^[21]), 亚分类(如 Rosenbaum 和 Rubin^[22], Rosenbaum^[23]), 加权(如 Rosenbaum^[24], Robins 等^[25]), 回归(如 Heckman 等^[26]), 或它们的组合(如 Robins 和 Zhao^[27], Ho 等^[28]). Imbens^[29]及 Imbens 和 Rubin^[30]对此进行了全面的综述.

尽管上述方法在理论上很受欢迎, 但这些方法面临的主要困难是必须估计倾向得分. 事实上, 研究人员已经发现, 倾向得分模型的微小误差会导致因果效应的估计出现严重偏差(参见 Kang 和 Schafer^[31], Smith 和 Todd^[32]), 但第二步中所有的校正方法都假设估计的倾向得分非常接近真实情况, 所以这通常需要一个正确指定的模型或对倾向得分进行非参数回归. 在真实的世界中, 正确的模型指定往往是不现实的, 而且由于维数诅咒问题, 非参数回归只有当样本量很大, 协变量很少时, 才是一个明智的选择. 上述挑战突出了倾向得分的矛盾性质——倾向得分的设计初衷是为了减少协变量的维度, 然而它的估计却需要对高维协变量进行建模. 在实践中, 研究者通过重复更改模型并检查由此产生的协变量平衡这一过程来寻找合适的倾向得分模型. Imai 等^[33]称之为“倾向得分重言式(Propensity Score Tautology)”——如果平衡了协变量, 那么估计的倾向得分就是恰当的. 尽管已经提出了各种方法来完善和改进倾向得分匹配和加权技术(如 Robins 等^[34]),

但 Imai 和 Ratkovic^[35]仍认为开发一种估计倾向得分的稳健方法是非常有必要的,于是在 2014 年提出了协变量平衡倾向得分(Covariate Balance Propensity Score)方法,协变量平衡倾向得分作为协变量平衡得分和处理分配的条件概率,开发了倾向得分的双重特性. Fong 和 Imai^[36]将协变量平衡倾向得分中的二值型处理变量推广到一般情况,即多值和连续值. Wyss 等^[37]使用模拟和实验数据来比较协变量平衡倾向得分与逻辑回归、增强分类和回归树的性能,表明协变量平衡倾向得分提高了倾向得分模型的稳健性. Fan 等^[38]将协变量平衡倾向得分进行了改进,给出了其渐近性质的证明,还提出了一个双重稳健且渐近有效的方法,并解决了 Imai 和 Ratkovic^[35]的文献中关于如何以最佳方式选择协变量平衡倾向得分方法中的协变量平衡函数这一开放性问题,又进一步提出了新的逆倾向得分加权(Inverse Propensity Score Weighting)估计. Imai 和 Ratkovic^[39]将协变量平衡倾向得分推广到纵向数据中. Oyenubi^[40]证明在存在工具变量和未测量混杂因素时,协变量平衡倾向得分的估计效果优于其它方法,原因是协变量平衡倾向得分减少了工具变量对倾向得分的影响,使得基于协变量平衡倾向得分的模型中参数估计的偏差和均方误差更低. Ning 等^[41]将协变量平衡倾向得分推广到高维协变量的情况.

对于生存数据,由于其具有删失、截断等特征,使得上述的方法不能被直接使用.当潜在结果是生存数据时,因果效应也可以量化为由处理分配机制导致的平均或中位生存时间的改变,或在指定的随访时间内发生事件的概率的绝对减少等. Xie 和 Liu^[42]提出了一个校正后的 Kaplan-Meier 生存函数估计,并推导出了一个校正后的对数秩检验用于生存比较. Sugihara^[43]将该方法扩展到多值处理分配的情况. Xu 等^[44]进一步将该方法扩展到具有稳定逆倾向得分加权的时变治疗. Cole 和 Hernan^[45]建议使用稳定的逆倾向得分加权且拟合 Cox 比例风险模型,以获得无混杂生存曲线. Austin^[46]提倡使用生存概率或风险变化来估计治疗效果. Austin 和 Schuster^[47]比较了不同倾向得分方法在估计某一处理分配机制对生存结果的绝对影响方面的表现,如平均/中位生存时间或生存概率的差异. Austin 和 Stuart^{[48][49]}以及 Gayat 等^[50]研究了使用倾向得分估计风险比的方法. Li 等^[51]探索了基于倾向得分的方法来估计删失医疗费用的因果效应,并扩展了协变量调整分层、加权和双重稳健方法来处理删失医疗费用. Cho, Hu 和 Ghosh^[52]对非独立删失的半竞争风险数据提出了一种在加速失效时间模型下对平均因果效应进行加权估计的策略. Henderson 等^[53]利用非参数加速失效时间模型分析当结果是生存时间时的异质治疗效果.

与以上方法不同,我们根据协变量平衡策略来估计倾向得分模型. 本文将协变量平衡倾向得分方法推广到生存结果中,利用倾向得分的双重特性得到最优的协变量平衡并

对处理分配机制进行建模.

本文研究的主要内容是受到 Li 等^[51], Ning 等^[41]以及 Henderson 等^[53]这三篇文献的启发, 令潜在结果为具有右删失的生存时间, 基于调整后的协变量平衡倾向得分和加权最小二乘方法来估计倾向得分模型和加速失效时间模型(结果模型), 然后运用四种估计方法给出平均因果效应的估计. 这四种估计方法分别取为 Horvitz-Thompson 估计、逆倾向得分加权估计、加权最小二乘估计和双稳健(Doubly Robust)估计. 在模拟研究中, 将四种估计分别在倾向得分模型和结果模型均被正确指定、倾向得分模型或结果模型只有一个被正确指定以及两个模型均被错误指定这四种情形下进行对比研究, 并给出模型参数的估计情况. 通过分别更改删失率和样本量的大小, 得出删失率和样本量的改变对四种估计的影响. 最后, 将本文的方法应用到原发性胆汁性肝硬化数据和乳腺癌数据中.

2 预备知识

本章主要回顾有关因果推断和生存分析的基本知识.

2.1 因果推断

2.1.1 潜在结果模型

由于因果推断基础理论的持续创新与发展以及其重要的应用价值与实际意义, 目前因果推断的相关问题已经成为多个领域的研究热点.

在因果推断的理论模型中, 潜在结果模型是其中重要的模型之一. 潜在结果模型的主要思想是比较同一个体在接受处理和不接受处理时的结果差异, 潜在结果模型的提出为之后的理论发展奠定了重要基础. 该模型由统计学家 Rubin 提出, 所以潜在结果模型又被称之为鲁宾因果模型(Rubin Causal Model). 但是通常情况下, 对于同一个体我们不能同时观察到两种处理的结果, 所以, 对于接受处理的个体, 不接受处理是一种“反事实”的状态, 反之亦然. 故该模型也被称之为反事实框架(Counterfactual Framework). 但 Rubin^[54]在 2005 年发表的一篇文献中提出了这样一个观点: 因为这些结果是在处理分配后才会是反事实的, 而将所有潜在结果称为“反事实”肯定会混淆那些永远无法观察到的量. 因此, Rubin 更喜欢 Neyman 提出的“潜在结果”这个概念. Fisher^[3]在 1935 年提出“随机试验”的概念. 这两个概念的提出对因果推断的发展产生了非常重要的影响. 接下来, Rubin^[4]对这两个概念进一步深入整合, 提出了潜在结果模型, 并将其应用于随机试验和观察性研究中.

许多因素可能会导致因果效应估计产生偏差, 所以潜在结果模型在实际应用之前必须做出一些基本假设. 当不满足假设时, 有些学者也进行了研究(如 Morgan 和 Winship^[55], Vander^[56]).

假设总体的个数为 n , 对于个体 $i (i=1, \dots, n)$, 令 A_i 是一个二元处理变量表示个体 i 处理分配的情况, 分配到控制组取值为 0, 分配到处理组取值为 1; $Y_i(0)$, $Y_i(1)$ 表示个体 i 分配到控制组或处理组时的潜在结果; $Y_i = A_i Y_i(1) + (1 - A_i) Y_i(0)$ 表示个体 i 观察到的结果变量; \mathbf{X}_i 为 d 维的协变量.

假设 2.1 稳定单位处理值假设(Stable Unit Treatment Value Assumption, SUTVA): 假设个体 i 的处理变量 A_i 对应一个唯一的值 $Y_i(A_i)$.

这个假设保证了: (1) 无干扰性, 即对一个个体的处理不会影响另一个个体的结果; (2) 一致性, 即对于每个个体, 每个处理方式都没有可以导致不同潜在结果的不同形式或版本; (3) 排他性, 即接受特定处理方式的个体不能接受该处理的其他方式. SUTVA 在决定哪些问题被充分表述为具有因果关系时起着关键作用. 只有在 SUTVA 下, 潜在结果模型对结果的表示才是适当的.

假设 2.2 强可忽略假设(Strongly Ignorable Treatment Assignment): (1) 在给定协变量 \mathbf{X}_i 的条件下, A_i 与 $(Y_i(1), Y_i(0))$ 条件独立, 即 $(Y_i(1), Y_i(0)) \perp A_i | \mathbf{X}_i$; (2) $0 < P(A_i = 1 | \mathbf{X}_i) < 1$.

该假设中(1)也可称为条件独立假设(Conditional Independence Assumption)^[57]、无混杂假设(Unconfoundedness Assumption)^[16]、依可测变量选择假设(Selection on Observables Assumption)或外生性假设(Exogeneity Assumption)^[29]等, 表明不存在不可观测的混杂因素. (2)也可称之为公共支撑假设(Overlap Assumption)、概率分配假设(Probabilistic Assignment Assumption)或正值假设(Positivity Assumption)等, 表明在协变量给定的情况下接受处理的概率严格限制于 0 到 1 之间.

基于以上这些假设, 下面给出平均因果效应(Average of Causal Effects, ACE)^[4]的概念. 对于一个感兴趣的总体, 平均因果效应为:

$$ACE = E\{Y_i(1) - Y_i(0)\}.$$

2.1.2 协变量平衡倾向得分

研究人员越来越多地使用观察性研究来估计处理分配机制对结果的影响. 在观察性研究中, 由于缺乏随机处理分配, 接受处理的受试者往往不同于未接受处理的受试者. 因此, 在估计因果效应时, 使用统计方法来消除或最小化由观察或测量的基线协变量在处理组和控制组之间分布的差异所造成的混杂影响是非常重要的. 倾向得分这个词最早出现在 1983 年, Rosenbaum 和 Rubin^[16]提出了从观察数据中对因果效应的估计方法. 倾向得分法越来越多地用于减少或最小化在观察性研究中经常发生的处理分配机制对结果的混杂影响.

Rosenbaum 和 Rubin^[16]在 1983 年给出平衡得分(Balancing Score)以及倾向得分的概念. 平衡得分 $b(\mathbf{X}_i)$ 是关于协变量 \mathbf{X}_i 的一个函数, 满足在给定 $b(\mathbf{X}_i)$ 的条件下 \mathbf{X}_i 的条件分布分别在控制组和处理组中是相同的, 即

$$\mathbf{X}_i \perp A_i | b(\mathbf{X}_i).$$

倾向得分表示为给定协变量 \mathbf{X}_i 的条件下, 接受处理的条件概率, 即

$$\pi(\mathbf{X}_i) = P(A_i = 1 | \mathbf{X}_i).$$

该文献还证明了 $\pi(\mathbf{X}_i)$ 是最粗糙(Coarsest)的平衡得分.

常用的倾向得分法有匹配法、加权法、分层法和回归调整法或它们的结合等. 尽管这些方法很受欢迎, 在理论上也很有吸引力, 但它们在实际应用中面临的主要困难是必须估计倾向得分. 事实上, 研究人员发现, 倾向得分模型的轻微误差可能导致因果效应估计的重大偏差. Imai 和 Ratkovic^[35]在 2014 年开发了一种估计倾向得分的稳健方法——协变量平衡倾向得分.

协变量平衡倾向得分旨在用广义矩估计(Generalized Method of Moments, GMM)代替最大似然估计(Maximum Likelihood Estimation, MLE), 同时优化处理分配预测和协变量平衡, 它假设倾向得分遵循逻辑回归模型, 即

$$\pi_{\beta}(\mathbf{X}_i) = \frac{\exp(\mathbf{X}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i^T \boldsymbol{\beta})},$$

其中 $\boldsymbol{\beta} \in \Theta$ 是未知的参数向量, Θ 表示参数空间.

一般来说, 人们将通过最大化对数似然函数来估计 $\boldsymbol{\beta}$. 参数 $\boldsymbol{\beta}$ 可由以下得分方程求解:

$$\frac{1}{n} \sum_{i=1}^n S_{\beta}(A_i, \mathbf{X}_i) = 0,$$

$$\text{其中 } S_{\beta}(A_i, \mathbf{X}_i) = \left\{ \frac{A_i}{\pi_{\beta}(\mathbf{X}_i)} - \frac{(1-A_i)}{1-\pi_{\beta}(\mathbf{X}_i)} \right\} \pi'_{\beta}(\mathbf{X}_i), \text{ 且 } \pi'_{\beta}(\mathbf{X}_i) = \frac{\partial \pi_{\beta}(\mathbf{X}_i)}{\partial \boldsymbol{\beta}}.$$

在协变量平衡倾向得分中, Imai 和 Ratkovic^[35]提出用指定的关于 \mathbf{X}_i 的 m 维可测向量值函数 $\mathbf{f}(\mathbf{X}_i)$ 来替代 $\pi'_{\beta}(\mathbf{X}_i)$. 得分方程是 $\mathbf{f}(\mathbf{X}_i)$ 的一个特殊的形式, 即 $\mathbf{f}(\mathbf{X}_i) = \pi'_{\beta}(\mathbf{X}_i)$. 协变量平衡倾向得分方法通过求解以下 m 维的估计方程来估计 $\boldsymbol{\beta}$:

$$\bar{g}_{\beta}(A, \mathbf{X}) = \frac{1}{n} \sum_{i=1}^n g_{\beta}(A_i, \mathbf{X}_i) = 0. \quad (2.1)$$

$$\text{其中 } g_{\beta}(A_i, \mathbf{X}_i) = \left\{ \frac{A_i}{\pi_{\beta}(\mathbf{X}_i)} - \frac{(1-A_i)}{1-\pi_{\beta}(\mathbf{X}_i)} \right\} \mathbf{f}(\mathbf{X}_i). \text{ } \mathbf{f}(\mathbf{X}_i) \text{ 称为协变量平衡函数, 因为等式(2.1)}$$

也可以表示为协变量平衡方程:

$$\sum_{i=1}^n \frac{A_i}{\pi_{\beta}(\mathbf{X}_i)} \mathbf{f}(\mathbf{X}_i) = \sum_{i=1}^n \frac{1-A_i}{1-\pi_{\beta}(\mathbf{X}_i)} \mathbf{f}(\mathbf{X}_i).$$

协变量平衡方程使平均因果效应的估计具有稳健性和有效性. 如果我们选择 $\mathbf{f}(\mathbf{X}_i)$ 为 $\{\mathbf{X}_i^T, \dots, (\mathbf{X}_i^m)^T\}$ (其中 \mathbf{X}_i^m 表示为向量 \mathbf{X}_i 中每一元素的 m 次方), 如上述所言, 估计方程就成为协变量的矩条件. 如果参数的数量等于方程的数量, 它将是恰好识别 (Just-identified) 的估计. 当协变量平衡方程与得分函数条件相结合时, 可导出过度识别的 (Over-identified) 连续更新 GMM 估计:

$$\hat{\beta}_{GMM} = \arg \min_{\beta \in \Theta} \bar{g}_{\beta}(A, \mathbf{X})^T \hat{\mathbf{W}} \bar{g}_{\beta}(A, \mathbf{X}),$$

其中 $\hat{\mathbf{W}}$ 为 $m \times m$ 维协方差矩阵. 一旦参数 β 被估计, 就可以通过逆倾向得分加权方法估计平均因果效应. 除此之外, 也可以用 (2.1) 式中给出的协变量平衡方程作为约束条件, 通过经验似然框架进行估计.

上述的协变量平衡倾向得分方法还有一个没有解决的问题就是 $\mathbf{f}(\mathbf{X}_i)$ 的最优选择是什么? Fan 等^[38]探讨了 $\mathbf{f}(\mathbf{X}_i)$ 的最优选择和基于协变量平衡倾向得分的逆倾向得分加权估计的理论特性. 如果存在一个 α , 使 $\alpha^T \mathbf{f}(\mathbf{X}_i)$ 等于两个潜在结果的条件均值函数的加权平均, 那么当倾向得分模型被错误指定时, 基于协变量平衡倾向得分的逆倾向得分加权估计仍然是相合的. 一旦协变量平衡函数以这种方式构建, 当倾向得分模型也被正确指定时, 基于协变量平衡倾向得分的逆倾向得分加权估计将是半参数有效的.

2.2 生存分析

生存数据经常出现的一个特殊特征是“删失”, 广义地说, 当已知某些生命周期只在特定时间间隔内发生时, 就会发生删失. 删失有多种类型, 如右删失、左删失和区间删失. 本文主要关注右删失, 其有如下三种类型.

第一种右删失是 I 型删失, 即仅当事件发生在某个预先指定的时间之前时才被观察到, 删失时间因人而异. 一种特殊的动物研究或临床试验从固定数量的动物或病人开始, 并对其进行治疗. 由于时间或成本的考虑, 研究者将终止研究或在所有受试者意识到他们的事件之前报告结果. 在这种情况下, 如果没有意外损失或受试者退出, 所有经过删失的观察的时间等于研究期间的长度.

假设总体的个数为 n , 对于个体 $i (i=1, \dots, n)$, 我们假设 T_i 表示存活时间, C_i 表示删失时间. 假设所有的 T_i 是独立同分布的, 且其概率密度函数为 $f(t)$, 生存函数为 $S(t)$.

当且仅当个体 i 的生存时间 T_i 小于或等于 C_i 时, 个体的确切寿命 T_i 才会被观测. 如果 T_i 大于 C_i , 则该个体是存活者, 他的事件发生时间在 C_i 处被删失. 实验数据可以用成对的随机变量 (U_i, δ_i) 表示, 其中 δ_i 表示生存时间 T_i 是否删失 ($\delta_i = 1$ 为未删失, $\delta_i = 0$ 为删失). 若生存时间被观察到, 则 $U_i = T_i$; 若发生删失, 则 $U_i = C_i$, 即 $U_i = \min(T_i, C_i)$.

当存在不同的固定删失时间时, 这种 I 型删失形式称为递增型 I 型删失. 这种删失形式的一个优点是删失的个体提供了关于非致命疾病的自然史的信息.

另一个 I 型删失的例子是, 个体在不同的时间进入研究, 研究的终点是研究者预先确定的, 当个体进入研究时, 删失时间是已知的. 在这类研究中, 个体有自己特定的、固定的删失时间. 这种形式的删失被称为广义 I 型删失.

第二种右删失是 II 型删失, 研究一直进行到前 r 个个体失效为止, 其中 r 是一个预定的整数 ($r < n$). II 型删失经常用于设备寿命测试. 所有项目都同时进行测试, 当 n 项中有 r 项失效时, 测试终止.

II 型删失的推广与 I 型删失的推广类似, 但删失时间不同, 称为递增型 II 型删失. 这里观察并记录了 n 个设备(或动物)样本中第一个 r_1 (研究开始前选择的整数)个失效设备(或动物). 然后将剩余的 $n - r_1$ 个未失效的设备(或动物)中的 $n_1 - r_1$ 个从实验中移除(或删失), 留下 $n - n_1$ 个设备(或动物)在研究中. 当下一个 r_2 (研究开始前选择的另一个整数)个设备(或动物)失效时, 未失效的 $n_2 - r_2$ 个设备被移除(或删失). 这个过程持续进行, 直到完成预定的一系列重复. 同样, r_i 和 n_i ($i = 1, 2$)是固定整数, 且两个删失时间 T_{r_1} 和 $T_{n_1+r_2}$ 是随机的.

第三种右删失是竞争风险删失, 称之为 III 型删失. 竞争风险删失的一个特例是随机删失. 当我们对某些事件的边际分布估计感兴趣, 但一些被研究的个体可能经历了一些竞争事件, 导致他们被排除在研究之外时, 就会出现这种删失.

在生存数据分析中, 加速失效时间模型(Wei^[58])可以作为 Cox 比例风险模型的替代模型. 加速失效时间模型以对数失效时间作为响应变量, 提供了协变量和失效时间之间相互关系的直接解释. 此外, 该模型可以直接根据两种不同处理方式的潜在失效时间来定义因果效应. 加速失效时间模型的这些特点, 使它们在个性化医疗中以及在比较不同处理方式的效果的背景下具有非常大吸引力.

令个体失效时间为 T_i , $i = 1, \dots, n$, 删失时间为 C_i , $\delta_i = I(T_i \leq C_i)$ 为删失指标, $U_i = \min(T_i, C_i)$ 为观察时间. 假设在给定 d 维的协变量 \mathbf{X}_i 的条件下, T_i 和 C_i 独立. 加速失效时间模型为:

$$\log(T_i) = \mathbf{X}_i^T \boldsymbol{\alpha} + \varepsilon_i,$$

其中 $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)^T$ 为参数向量, ε_i 为残差项.

3 模型和估计

本章为全文的主要内容, 主要论述了模型及其参数估计方法, 并给出平均因果效应的四种估计.

3.1 模型及参数估计

3.1.1 倾向得分模型与结果模型

假设我们从一个感兴趣的总体中观察到一个大小为 n 的简单随机样本, 对每个个体 i , 我们观察到的数据为 $\{(U_i, \delta_i, A_i, \mathbf{X}_i), i=1, \dots, n\}$. 令 T_i 为失效时间, C_i 为删失时间, $U_i = \min(T_i, C_i)$ 为观察到的时间, $\delta_i = I(T_i \leq C_i)$ 为删失指标, A_i 为二元处理变量表示个体 i 处理分配的情况(取值为 1 时个体 i 分配到处理组, 取值为 0 时分配到控制组), \mathbf{X}_i 为 d 维的协变量, $\log(T_i(0))$ 和 $\log(T_i(1))$ 表示个体 i 分配到控制组或处理组时的潜在结果, $\log(T_i) = A_i \log(T_i(1)) + (1 - A_i) \log(T_i(0))$ 为只在 $\delta_i = 1$ 时观察到的结果, $K_0(u)$, $K_1(u)$ 为处理变量 A_i 取 0 或 1 时的删失时间 C_i 的生存函数, 即

$$K_j(u) = P(C_i \geq u | A_i = j), \quad j = 0, 1.$$

假设删失是无信息的, 即给定 (A_i, \mathbf{X}_i) 下 C_i 与 T_i 相互独立.

由 2.2.2 节可知, 加速失效时间模型为:

$$\log(T_i) = \mathbf{X}_i^T \boldsymbol{\alpha} + \varepsilon_i,$$

其中参数 $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)^T$ 为 d 维的参数向量, 假设残差项 ε_i 为独立同分布的, 且满足 $E(\varepsilon_i) = 0$.

在 2.1.1 节中的假设下, 平均因果效应可以表示为^[53]:

$$\mu^* = E\{\log T_i(1)\} - E\{\log T_i(0)\},$$

令 $\mu_1^* = E\{\log T_i(1)\}$, $\mu_0^* = E\{\log T_i(0)\}$. 这里我们关注 μ_1^* 的估计, 因为 μ_0^* 可以由相似的方法估计. 我们为处理分配机制施加了一个有效的参数模型 $\pi(\mathbf{X}_i^T \boldsymbol{\beta})$, 这被称为倾向得分, 即 $P(A_i = 1 | \mathbf{X}_i)$, 其中 $\pi(\cdot)$ 是一个已知的函数, 参数 $\boldsymbol{\beta}$ 是一个未知的 d 维向量. 当倾向得分模型被正确指定时, 我们有:

$$\pi_i = P(A_i = 1 | \mathbf{X}_i) = \pi(\mathbf{X}_i^T \boldsymbol{\beta}^*),$$

其中 $\boldsymbol{\beta}^* \in \mathbb{R}^d$ 为 $\boldsymbol{\beta}$ 的真值. 对于结果变量, 我们采用加速失效时间模型. 当结果模型被正确指定时, 我们有:

$$E\{\log T_i(1) | \mathbf{X}_i\} = \mathbf{X}_i^T \boldsymbol{\alpha}^*, \quad (3.1)$$

其中 $\boldsymbol{\alpha}^* \in \mathbb{R}^d$ 为 $\boldsymbol{\alpha}$ 的真值.

定义 3.1 协变量平衡性质^[41]: 令 $\hat{\pi}_i = \pi(\mathbf{X}_i^T \hat{\boldsymbol{\beta}})$ 为倾向得分的一个估计, $\hat{\boldsymbol{\beta}}$ 是 $\boldsymbol{\beta}^*$ 的估计值, $\boldsymbol{\beta}^*$ 是 $\boldsymbol{\beta}$ 的真值.

(a) 若下式成立,

$$\sum_{i=1}^n \left(\frac{A_i}{\hat{\pi}_i} - 1 \right) \mathbf{X}_i = 0, \quad (3.2)$$

我们称 $\hat{\pi}_i$ 满足强协变量平衡性质.

(b) 若下式成立,

$$\sum_{i=1}^n \left(\frac{A_i}{\hat{\pi}_i} - 1 \right) \mathbf{X}_i^T \boldsymbol{\alpha}^* = 0,$$

其中 $\boldsymbol{\alpha}^*$ 为(3.1)式中的参数向量, 我们称 $\hat{\pi}_i$ 满足弱协变量平衡性质.

虽然强协变量平衡性质蕴含弱协变量平衡性质, 但反之不一定成立. 现有的协变量平衡倾向得分旨在实现强协变量平衡性质, 从而平衡 \mathbf{X}_i 的每个分量的均值(例如, Imai 和 Ratkovic^[35], Fan 等^[38]). 下面我们分别对倾向得分模型和结果模型的参数进行估计.

3.1.2 倾向得分模型的参数估计

本文将 Imai 和 Ratkovic^[35]提出的协变量平衡倾向得分进行了调整, 将倾向得分模型从常见的逻辑回归模型推广至广义线性模型.

定义一个广义似然函数:

$$Q_n(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \int_0^{\mathbf{X}_i^T \boldsymbol{\beta}} \left\{ \frac{A_i}{\pi(u)} - 1 \right\} \omega_1(u) du, \quad (3.3)$$

其中 $\omega_1(u)$ 是一个正的加权函数. 假设 $\pi(\cdot)$ 关于 $\boldsymbol{\beta}$ 二阶连续可微, 则有得分条件:

$$\frac{1}{n} \sum_{i=1}^n s_{\beta}(A_i, \mathbf{X}_i) = 0,$$

其中 $s_{\beta}(A_i, \mathbf{X}_i) = \left\{ \frac{A_i}{\pi(\mathbf{X}_i^T \boldsymbol{\beta})} - 1 \right\} \omega_1(\mathbf{X}_i^T \boldsymbol{\beta}) \mathbf{X}_i$. 由协变量平衡性质, 协变量平衡矩条件为:

$$\frac{1}{n} \sum_{i=1}^n \omega_{\beta}(A_i, \mathbf{X}_i) \mathbf{X}_i = 0,$$

其中 $\omega_{\beta}(A_i, \mathbf{X}_i) = \left\{ \frac{A_i}{\pi(\mathbf{X}_i^T \boldsymbol{\beta})} - 1 \right\}$. 利用 GMM 估计参数 $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^d} \left\{ \bar{g}_{\beta}(A, \mathbf{X})^T \Sigma_{\beta}(A, \mathbf{X})^{-1} \bar{g}_{\beta}(A, \mathbf{X}) \right\},$$

其中 $\bar{g}_{\beta}(A, \mathbf{X}) = \frac{1}{n} \sum_{i=1}^n g_{\beta}(A_i, \mathbf{X}_i)$ 是矩条件的样本均值, $\Sigma_{\beta}(A, \mathbf{X})$ 为 $g_{\beta}(A_i, \mathbf{X}_i)$ 的协方差矩

阵, $g_{\beta}(A_i, \mathbf{X}_i) = \begin{pmatrix} s_{\beta}(A_i, \mathbf{X}_i) \\ \omega_{\beta}(A_i, \mathbf{X}_i) \mathbf{X}_i \end{pmatrix}$. 我们假设 $\pi(\cdot)$ 满足 GMM 估计的标准正则条件(Newey 和 McFadden^[59]). 由下式给出 $g_{\beta}(A_i, \mathbf{X}_i)$ 的协方差矩阵:

$$\Sigma_{\beta}(A, \mathbf{X}) = \frac{1}{n} \sum_{i=1}^n E \left\{ g_{\beta}(A_i, \mathbf{X}_i) g_{\beta}(A_i, \mathbf{X}_i)^T \middle| \mathbf{X}_i \right\}.$$

在倾向得分模型为逻辑回归模型的情况下, 即 $\pi_i = \frac{\exp(\mathbf{X}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i^T \boldsymbol{\beta})}$, 则 $\Sigma_{\beta}(A, \mathbf{X})$ 的表达式为:

$$\Sigma_{\beta}(A, \mathbf{X}) = \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} (1/\pi_i - 1) \omega_1^2(\mathbf{X}_i^T \boldsymbol{\beta}) \mathbf{X}_i \mathbf{X}_i^T & (1/\pi_i - 1) \omega_1(\mathbf{X}_i^T \boldsymbol{\beta}) \mathbf{X}_i \mathbf{X}_i^T \\ (1/\pi_i - 1) \omega_1(\mathbf{X}_i^T \boldsymbol{\beta}) \mathbf{X}_i \mathbf{X}_i^T & (1/\pi_i - 1) \mathbf{X}_i \mathbf{X}_i^T \end{pmatrix}.$$

于是我们得到倾向得分的估计 $\hat{\pi}_i = \pi(\mathbf{X}_i^T \hat{\boldsymbol{\beta}})$.

我们把(3.3)式中的函数 $Q_n(\boldsymbol{\beta})$ 称为广义拟似然函数, 因为它的构造类似于广义线性模型的拟似然函数(Wedderburn^[60]). 为了理解广义拟似然函数是如何产生的, 我们计算相应的拟得分函数:

$$\frac{\partial Q_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{1}{n} \sum_{i=1}^n s_{\boldsymbol{\beta}}(A_i, \mathbf{X}_i) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{A_i}{\pi(\mathbf{X}_i^T \boldsymbol{\beta})} - 1 \right\} \omega_1(\mathbf{X}_i^T \boldsymbol{\beta}) \mathbf{X}_i. \quad (3.4)$$

由于(3.4)是参数 $\boldsymbol{\beta}$ 的无偏估计函数, 所以 $Q_n(\boldsymbol{\beta})$ 是一个结合了拟得分函数(3.4)的合理的拟似然函数. 拟似然函数 $Q_n(\boldsymbol{\beta})$ 依赖于权重 $\omega_1(u)$ 的选择. 特别地, 我们考虑以下两个例子:

(a) 如果 $\omega_1(u) = \pi(u)$, 则(3.4)与逻辑回归的得分函数相同, 因此 $Q_n(\boldsymbol{\beta})$ 为处理分配机制的标准似然函数.

(b) 如果 $\omega_1(u) = 1$, 则拟得分函数(3.4)就是强协变量平衡方程(3.2). 因此, 我们称 $\omega_1(u) = 1$ 时的 $Q_n(\boldsymbol{\beta})$ 为协变量平衡损失函数.

因此, 对应于不同的 $\omega_1(u)$, 我们使用拟最大似然估计和其它的惩罚 M-估计作为一个矩条件. 通过计算 $Q_n(\boldsymbol{\beta})$ 的 Hessian 矩阵, 我们发现(3.3)是一个取决于 $\omega_1(u)$ 选择的非凸优化问题. 非凸性可能会给计算带来挑战. 例如, 梯度下降算法可能会被困在一个远离全局最大化点的局部解. 为了避免计算问题, 我们主要关注凹拟似然函数 $Q_n(\boldsymbol{\beta})$. 在(a)情况下, 当 $\omega_1(u) = \pi(u)$ 时, $Q_n(\boldsymbol{\beta})$ 是凹的; 在(b)情况下, $\omega_1(u) = 1$ 时, $Q_n(\boldsymbol{\beta})$ 也是凹的.

3.1.3 结果模型的参数估计

根据结果模型, 定义一个处理组中加权最小二乘损失函数:

$$L_n(\boldsymbol{\alpha}) = \frac{1}{n} \sum_{i=1}^n A_{(i)} \omega_{n(i)} \omega_2(\mathbf{X}_{(i)}^T \boldsymbol{\beta}) (\log(U_{(i)}) - \mathbf{X}_{(i)}^T \boldsymbol{\alpha})^2,$$

其中 $U_{(1)} \leq \dots \leq U_{(n)}$ 是 U_i 的次序统计量, $\mathbf{X}_{(i)}$ 是与 $U_{(i)}$ 对应的协变量, $A_{(i)}$ 为与 $U_{(i)}$ 对应的处理变量, $\omega_2(\mathbf{X}_{(i)}^T \boldsymbol{\beta})$ 是一个正的加权函数, $\omega_{n(i)}$ 是 Kaplan-Meier 权, 表达式为:

$$\omega_{n(i)} = \frac{\delta_{(i)}}{K_1(U_{(i)} -)},$$

其中 $\delta_{(i)}$ 是与 $U_{(i)}$ 对应的删失指标, $K_1(\cdot)$ 为处理组中删失时间的生存函数的左连续形式.

我们将损失函数最小化, 得到参数 $\boldsymbol{\alpha}$ 的估计:

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha} \in \mathbb{R}^d} L_n(\boldsymbol{\alpha}) = \left(\sum_{i=1}^n A_{(i)} \hat{\omega}_{n(i)} \omega_2(\mathbf{X}_{(i)}^T \hat{\boldsymbol{\beta}}) \mathbf{X}_{(i)} \mathbf{X}_{(i)}^T \right)^{-1} \sum_{i=1}^n A_{(i)} \hat{\omega}_{n(i)} \omega_2(\mathbf{X}_{(i)}^T \hat{\boldsymbol{\beta}}) \log(U_{(i)}) \mathbf{X}_{(i)},$$

其中 $\hat{\boldsymbol{\beta}}$ 为 3.1.2 节倾向得分模型的参数估计值, $\hat{\omega}_{n(i)} = \frac{\delta_{(i)}}{\hat{K}_1(U_{(i)}^-)}$, $\hat{K}_1(\cdot)$ 是处理组中删失时间的生存函数的 Kaplan-Meier 估计, 可表示为:

$$\hat{K}_1(u) = \prod_{i=1}^n \left[\frac{n-i}{n-i+1} \right]^{I(A_{(i)}=1, U_{(i)} \leq u, \delta_{(i)}=0)}. \quad (3.5)$$

同理可表示出 $\hat{K}_0(\cdot)$.

在本节中, 我们用加权最小二乘估计拟合结果模型. 允许权重 $\omega_2(\mathbf{X}_i^T \boldsymbol{\beta})$ 依赖于 \mathbf{X}_i . 我们考虑以下例子:

(a) 如果 $\omega_2(u)=1$, 则 $L_n(\boldsymbol{\alpha})$ 为处理组的经典 Kaplan-Meier 权最小二乘损失函数. 令 $\boldsymbol{\alpha}$ 的估计为 $\hat{\boldsymbol{\alpha}}_{WLS1}$.

(b) 如果 $\omega_2(u)=1/\pi(u)$, 则 $L_n(\boldsymbol{\alpha})$ 为处理组的逆概率 Kaplan-Meier 权最小二乘损失函数. 令 $\boldsymbol{\alpha}$ 的估计为 $\hat{\boldsymbol{\alpha}}_{WLS2}$.

3.2 平均因果效应的估计

Kang 和 Schafer^[61]研究了四种倾向得分加权估计. 他们使用的倾向得分模型是以协变量 \mathbf{X}_i 作为线性预测因子的一个逻辑回归模型. 这是一个错误的模型, 因为真正的倾向得分是以协变量 \mathbf{X}_i^* 为线性预测因子的一个逻辑回归模型(但在观察到的 \mathbf{X}_i 中是非线性的). 他们检验的加权估计是 Horvitz-Thompson 估计, 逆倾向得分加权估计, 加权最小二乘估计以及双稳健估计. 我们将这些估计推广到具有右删失的生存数据中. 下面给出平均因果效应的四种估计.

3.2.1 Horvitz-Thompson 估计

Horvitz 和 Thompson^[62]在有限总体的背景下, 提出了用逆概率加权观测值的方法来估计平均因果效应, Rosenbaum^[24]将其扩展到倾向得分. 对于删失数据, 我们使用 Kaplan-Meier 权 $\delta_i/K_j(U_i)$ 对删失数据进行校正.

由于在强可忽略假设下,

$$\begin{aligned} E \left\{ \frac{A_i \delta_i \log U_i}{\pi_i K_1(U_i)} - \mu_1^* \right\} &= E \left\{ \frac{I(A_i=1) I(C_i \geq T_i(1)) \log T_i(1)}{\pi_i K_1(T_i(1))} - \mu_1^* \right\} \\ &= E \left\{ E \left\{ \frac{I(A_i=1) I(C_i \geq T_i(1)) \log T_i(1)}{\pi_i K_1(T_i(1))} \middle| T_i(1), \mathbf{X}_i \right\} \right\} - \mu_1^* \end{aligned}$$

$$\begin{aligned}
 &= E \left\{ \frac{\pi_i K_1(T_i(1)) \log T_i(1)}{\pi_i K_1(T_i(1))} \right\} - \mu_1^* \\
 &= E \{ \log T_i(1) \} - \mu_1^* = 0.
 \end{aligned}$$

因此, 当倾向得分模型被正确指定时, 我们可以很容易的看出 $n^{-1} \sum_{i=1}^n A_i \delta_i \log U_i \left(\hat{\pi}_i \hat{K}_1(U_i) \right)^{-1}$ 可以单独从处理组中计算出来, 无偏地估计整个处理组潜在结果的平均值. 同理对于控制组有:

$$E \left\{ \frac{(1-A_i) \delta_i \log U_i}{(1-\pi_i) K_0(U_i)} - \mu_0^* \right\} = E \{ \log T_i(0) - \mu_0^* \} = 0.$$

故由上述可知, $\mu_1^* = E \{ \log T_i(1) \}$, $\mu_0^* = E \{ \log T_i(0) \}$ 的 Horvitz-Thompson 估计为:

$$\begin{aligned}
 \hat{\mu}_{HT1}^* &= \frac{1}{n} \sum_{i=1}^n \frac{A_i \delta_i \log U_i}{\hat{\pi}_i \hat{K}_1(U_i)}, \\
 \hat{\mu}_{HT0}^* &= \frac{1}{n} \sum_{i=1}^n \frac{(1-A_i) \delta_i \log U_i}{(1-\hat{\pi}_i) \hat{K}_0(U_i)},
 \end{aligned}$$

其中 $\hat{\pi}_i$ 是由 3.1.2 节得到的倾向得分的估计, $\hat{K}_0(U_i)$ 和 $\hat{K}_1(U_i)$ 是由(3.5)式估计的当 A_i 分别取 0 或 1 时删失时间的生存函数. 因此, 平均因果效应的 Horvitz-Thompson 估计为 $\hat{\mu}_{HT}^* = \hat{\mu}_{HT1}^* - \hat{\mu}_{HT0}^*$, 且为 μ^* 的无偏估计.

3.2.2 逆倾向得分加权估计

在 Horvitz-Thompson 估计中, 如果我们将分母 n 替换为 $\sum_{i=1}^n A_i \delta_i \left(\hat{\pi}_i \hat{K}_1(U_i) \right)^{-1}$, 那么精确度通常会增强. 我们以这种方式对权重进行归一化, 用由 3.1.2 节得到的倾向得分的估计代替未知的倾向得分, 从而得到 μ_1^* 的逆倾向得分加权估计:

$$\hat{\mu}_{IPW1}^* = \frac{\sum_{i=1}^n \frac{A_i \delta_i \log U_i}{\hat{\pi}_i \hat{K}_1(U_i)}}{\sum_{i=1}^n \frac{A_i \delta_i}{\hat{\pi}_i \hat{K}_1(U_i)}}.$$

同理可得到 $\hat{\mu}_{IPW0}^*$, 则平均因果效应的逆倾向得分加权估计为 $\hat{\mu}_{IPW}^* = \hat{\mu}_{IPW1}^* - \hat{\mu}_{IPW0}^*$.

由于在强可忽略条件下,

$$E \left\{ \frac{A_i \delta_i (\log U_i - \mu_1^*)}{\pi_i K_1(U_i)} \right\} = E \left\{ \frac{I(A_i = 1) I(C_i \geq T_i(1)) (\log T_i(1) - \mu_1^*)}{\pi_i K_1(T_i(1))} \right\}$$

$$\begin{aligned}
 &= E \left\{ E \left\{ \frac{I(A_i = 1)I(C_i \geq T_i(1))(\log T_i(1) - \mu_1^*)}{\pi_i K_1(T_i(1))} \middle| T_i(1), \mathbf{X}_i \right\} \right\} \\
 &= E \left\{ \frac{\pi_i K_1(T_i(1))(\log T_i(1) - \mu_1^*)}{\pi_i K_1(T_i(1))} \right\} \\
 &= E \{ \log T_i(1) - \mu_1^* \} = 0.
 \end{aligned}$$

同理,

$$E \left\{ \frac{(1 - A_i)\delta_i(\log U_i - \mu_0^*)}{(1 - \pi_i)K_0(U_i)} \right\} = E \{ \log T_i(0) - \mu_0^* \} = 0.$$

故当倾向得分模型被正确指定时, $\hat{\mu}_{IPW}^*$ 为 μ^* 的无偏估计.

3.2.3 加权最小二乘估计

Horvitz-Thompson 估计和逆倾向得分加权估计不考虑协变量与潜在结果的关系, 我们通过构建加速失效时间模型, 将 $\log(T_i)$ 与协变量 \mathbf{X}_i 联系起来并可以对缺失值进行预测. 因为生存数据具有右删失, 所以我们用 Kaplan-Meier 权 $\delta_i/K_j(U_i)$ 对删失进行校正.

当结果模型被正确指定时, 可知,

$$E \{ \log(T_i) | \mathbf{X}_i \} = \mathbf{X}_i^T \boldsymbol{\alpha}.$$

则,

$$\mu_1^* = E \{ \log T_i(1) \} = E \{ E \{ \log T_i(1) | \mathbf{X}_i \} \} = E \{ \mathbf{X}_i^T \boldsymbol{\alpha} \}.$$

接下来, 我们利用结果模型对参数 $\boldsymbol{\alpha}$ 进行加权最小二乘估计, 即 3.1.3 小节中的 $\hat{\boldsymbol{\alpha}}_{WLS2}$. 由 Kang 和 Schafer^[61] 的文献中对于参数 $\boldsymbol{\alpha}$ 估计的讨论可知, 与 $\boldsymbol{\alpha}$ 的普通最小二乘估计相比, 如果倾向得分模型是正确的, $\boldsymbol{\alpha}$ 的加权最小二乘估计有效地纠正了来自错误的结果模型的偏差, 但如果倾向得分模型是错的, 则估计表现不佳. 当两模型均被错误指定时, $\boldsymbol{\alpha}$ 的加权最小二乘估计比普通最小二乘估计表现的更好. 由此, 我们选用参数 $\boldsymbol{\alpha}$ 的估计 $\hat{\boldsymbol{\alpha}}_{WLS2}$ 而不是 $\hat{\boldsymbol{\alpha}}_{WLS1}$, 故可以得到 μ_1^* 的加权最小二乘估计:

$$\hat{\mu}_{WLS1} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i^T \hat{\mathbf{a}}_{WLS2}.$$

同理可求出 $\hat{\mu}_{WLS0}^*$, 则平均因果效应的加权最小二乘估计为 $\hat{\mu}_{WLS}^* = \hat{\mu}_{WLS1}^* - \hat{\mu}_{WLS0}^*$.

3.2.4 双稳健估计

Robins 等^[34]提出双稳健估计, 并证明了双稳健估计在加权估计中具有最小的大样本方差, 且局部半参数有效. 双稳健估计要求我们指定结果模型和倾向得分模型, 它对其中一个(而不是两个)模型的错误具有稳健性. 在本文中, 我们用 Kaplan-Meier 权 $\delta_i/K_j(U_i)$ 对删失数据进行校正.

下面给出 μ_1^* 的双稳健估计:

$$\hat{\mu}_{DR1} = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{A_i \delta_i (\log U_i)}{\hat{\pi}_i \hat{K}_1(U_i)} - \frac{(A_i - \hat{\pi}_i) \delta_i \mathbf{X}_i^T \hat{\mathbf{a}}_{WLS1}}{\hat{\pi}_i \hat{K}_1(U_i)} \right\}.$$

同理可求出 $\hat{\mu}_{DR0}^*$, 因此平均因果效应的双稳健估计为 $\hat{\mu}_{DR}^* = \hat{\mu}_{DR1}^* - \hat{\mu}_{DR0}^*$.

由于在强可忽略条件下,

$$\begin{aligned} & E \left\{ \frac{A_i \delta_i \log U_i}{\pi_i K_1(U_i)} - \frac{\delta_i (A_i - \pi_i) \mathbf{X}_i^T \mathbf{a}}{\pi_i K_1(U)} \right\} \\ &= E \left\{ \frac{A_i \delta_i \log T_i(1)}{\pi_i K_1(T_i(1))} - \frac{\delta_i (A_i - \pi_i) \mathbf{X}_i^T \mathbf{a}}{\pi_i K_1(T_i(1))} \right\} \\ &= E \left\{ \frac{\delta_i \log T_i(1)}{K_1(T_i(1))} \right\} + E \left\{ \left(\frac{A_i}{\pi_i} - 1 \right) \left(\frac{\delta_i}{K_1(T_i(1))} \log T_i(1) - \mathbf{X}_i^T \mathbf{a} \right) \right\} \\ &= E \left\{ E \left\{ \frac{I(C_i \geq T_i(1)) \log T_i(1)}{K_1(T_i(1))} \middle| T_i(1), \mathbf{X}_i \right\} \right\} + E \left\{ \left(\frac{A_i}{\pi_i} - 1 \right) \left(\frac{\delta_i}{K_1(T_i(1))} \log T_i(1) - \mathbf{X}_i^T \mathbf{a} \right) \right\} \\ &= E \{ \log T_i(1) \} + E \left\{ \left(\frac{A_i}{\pi_i} - 1 \right) \left(\frac{\delta_i}{K_1(T_i(1))} \log T_i(1) - \mathbf{X}_i^T \mathbf{a} \right) \right\} \\ &= \mu_1^* + E \left\{ \left(\frac{A_i}{\pi_i} - 1 \right) \left(\frac{\delta_i}{K_1(T_i(1))} \log T_i(1) - \mathbf{X}_i^T \mathbf{a} \right) \right\}. \end{aligned}$$

因此只有当上式第二项 $S = E \left\{ \left(\frac{A_i}{\pi_i} - 1 \right) \left(\frac{\delta_i}{K_1(T_i(1))} \log T_i(1) - \mathbf{X}_i^T \mathbf{a} \right) \right\}$ 为 0 时, $\hat{\mu}_{DR1}$ 是无偏估计.

当倾向得分模型被正确指定时,

$$E(A_i|T_i(1), \mathbf{X}_i) = E(I(A_i = 1)|T_i(1), \mathbf{X}_i) = \pi_i,$$

则

$$\begin{aligned} S &= E \left\{ E \left\{ \left(\frac{A_i}{\pi_i} - 1 \right) \left(\frac{\delta_i}{K_1(T_i(1))} \log T_i(1) - \mathbf{X}_i^T \boldsymbol{\alpha} \right) \middle| T_i(1), \mathbf{X}_i \right\} \right\} \\ &= E \left\{ \left(\frac{E(A_i|T_i(1), \mathbf{X}_i)}{\pi_i} - 1 \right) \left(\frac{\delta_i}{K_1(T_i(1))} \log T_i(1) - \mathbf{X}_i^T \boldsymbol{\alpha} \right) \right\} = 0. \end{aligned}$$

当结果模型被正确指定时,

$$E(\log T_i(1)|\mathbf{X}_i) = \mathbf{X}_i^T \boldsymbol{\alpha},$$

则

$$\begin{aligned} S &= E \left\{ E \left\{ \left(\frac{A_i}{\pi_i} - 1 \right) \left(\frac{\delta_i}{K_1(T_i(1))} \log T_i(1) - \mathbf{X}_i^T \boldsymbol{\alpha} \right) \middle| \mathbf{X}_i \right\} \right\} \\ &= E \left\{ \left(\frac{A_i}{\pi_i} - 1 \right) \left(E \left(\frac{\delta_i}{K_1(T_i(1))} \log T_i(1) \middle| \mathbf{X}_i \right) - \mathbf{X}_i^T \boldsymbol{\alpha} \right) \right\} \\ &= E \left\{ \left(\frac{A_i}{\pi_i} - 1 \right) \left(E(\log T_i(1)|\mathbf{X}_i) - \mathbf{X}_i^T \boldsymbol{\alpha} \right) \right\} = 0. \end{aligned}$$

因此, 如果正确指定了倾向得分模型或结果模型, 则双稳健估计 $\hat{\mu}_{DR}^*$ 是无偏的.

4 模拟研究

在本章中, 我们通过数值模拟在倾向得分模型和结果模型均被正确指定、倾向得分模型或结果模型只有一个被正确指定以及两个模型均被错误指定这四种情形下, 分别对平均因果效应的四种估计进行对比研究, 并给出倾向得分模型和结果模型中参数的估计情况.

4.1 参数估计

令样本数量为 $n(500, 1000)$, 我们先产生 4 维的协变量 $\mathbf{X}_i \sim N(0, I)$, 其中 I 为单位阵, $i = 1, \dots, n$. 接下来利用形如 $\pi_i = 1 - 1 / \{1 + \exp(-X_{i1} + X_{i2}/2 - X_{i3}/4 - X_{i4}/10)\}$ 的逻辑回归模型产生二元处理变量 A_i . 对于潜在结果, 我们使用加速失效时间模型产生生存数据 $\{(U_i, \delta_i)\}$, $U_i = \min(T_i, C_i)$ 为观察到的时间, 删失时间 $C_i \sim U(0, c)$ (c 由不同删失率确定), $\delta_i = I(T_i \leq C_i)$ 为删失指标. 在每种设置下重复模拟 500 次.

考虑潜在结果模型为如下的加速失效时间模型:

$$\log(T_i(1)) = 2 - 0.291X_{i1} - 0.137(X_{i2} + X_{i3} + X_{i4}) + \varepsilon_{1i},$$

$$\log(T_i(0)) = 2 - 0.137(X_{i2} + X_{i3} + X_{i4}) + \varepsilon_{0i},$$

其中 ε_{1i} 和 ε_{0i} 为独立的标准正态随机变量. 在这种设置下, 我们考虑以下四种情形:

- (a) 倾向得分模型和结果模型均被正确指定.
- (b) 结果模型被正确指定, 但倾向得分模型被错误指定. 我们使用变换变量:

$$\mathbf{X}_{i_miss} = \left\{ \exp(X_{i1}/3), \exp(X_{i2}/3), \exp(X_{i3}/2), (X_{i4})^2 \right\},$$

来产生处理变量, 但用原始协变量 \mathbf{X}_i 产生结果变量.

- (c) 倾向得分模型被正确指定, 但结果模型被错误指定. 我们使用同样的变换变量 \mathbf{X}_{i_miss} 产生结果变量, 但用原始的协变量 \mathbf{X}_i 产生处理变量.

- (d) 倾向得分模型和结果模型均被错误指定. 我们使用上述变换变量 \mathbf{X}_{i_miss} 产生处理变量和结果变量.

4.1.1 倾向得分模型的参数估计

在本小节中我们分别在上述四种情形下, 对倾向得分模型中的参数进行估计. 我们也讨论了不同样本量 ($n = 500$, $n = 1000$) 和不同删失率 (20%, 40%) 对参数估计的有限样本性能的影响.

表 4.1 和表 4.2 展示了在四种情形以及不同样本量(Size)和不同删失率(Censoring rate)的设置下倾向得分模型中参数估计的偏差(Bias)、标准误差(Std. err)、标准差(Sd)以及 95% 置信区间的覆盖率(Cp). 通过表 4.1 和表 4.2, 我们可以发现, 在四种情形下, 删失率的改变对倾向得分模型中参数估计量的有限样本性能影响较小. 在表 4.1(a) 和表 4.2(c) 的设置下(即倾向得分模型被正确指定时), 参数估计的偏差、标准误差、标准差都较小, 置信区间覆盖率都在 95%左右, 且随着样本量的增大, 这四项指标几乎都在减小. 由此可见, 我们对参数估计的方法在倾向得分模型正确的情况下表现良好. 在表 4.1(b) 和表 4.2(d) 的设置下(即倾向得分模型被错误指定时), 参数估计量的偏差较大且置信区间覆盖率远小于 95%, 估计的效果不佳, 这也符合设置的情况.

表 4.1 两模型均被正确指定时、结果模型被正确指定而倾向得分模型被错误指定时倾向得分模型中的参数估计

Tab. 4.1 Estimators of the parameter in the propensity score model when both models are correct and when the outcome model is correctly specified but the propensity score model is misspecified

Censoring rate	Size	Parameter	(a)				(b)			
			Bias	Std. err	Sd	Cp	Bias	Std. err	Sd	Cp
40%	n=500	β_0	-0.010	0.107	0.105	0.956	0.945	0.105	0.104	0.000
		β_1	0.050	0.177	0.188	0.930	-0.673	0.104	0.106	0.000
		β_2	-0.010	0.114	0.115	0.952	0.337	0.103	0.100	0.104
		β_3	-0.003	0.109	0.108	0.958	0.096	0.105	0.105	0.836
		β_4	0.014	0.107	0.108	0.954	-0.182	0.098	0.100	0.528
	n=1000	β_0	0.002	0.073	0.074	0.944	0.940	0.073	0.073	0.000
		β_1	0.026	0.119	0.127	0.926	-0.676	0.071	0.068	0.000
		β_2	-0.006	0.078	0.076	0.952	0.332	0.072	0.071	0.004
		β_3	-0.003	0.075	0.071	0.956	0.099	0.073	0.074	0.718
		β_4	0.006	0.074	0.071	0.964	-0.181	0.068	0.064	0.228
20%	n=500	β_0	0.003	0.105	0.105	0.954	0.940	0.105	0.099	0.000
		β_1	0.042	0.181	0.182	0.956	-0.672	0.104	0.099	0.000
		β_2	-0.014	0.114	0.115	0.944	0.334	0.104	0.103	0.108
		β_3	-0.009	0.108	0.105	0.968	0.108	0.105	0.103	0.824
		β_4	0.003	0.107	0.108	0.948	-0.181	0.097	0.101	0.524
	n=1000	β_0	0.000	0.073	0.072	0.950	0.942	0.072	0.075	0.000
		β_1	0.018	0.120	0.122	0.950	-0.672	0.072	0.069	0.000
		β_2	-0.006	0.078	0.076	0.940	0.330	0.072	0.074	0.002
		β_3	-0.004	0.074	0.072	0.968	0.105	0.073	0.069	0.678
		β_4	0.000	0.073	0.080	0.922	-0.178	0.067	0.067	0.222

表 4.2 倾向得分被正确指定而结果模型被错误指定时、两模型均被错误指定时倾向得分模型中的参数估计

Tab. 4.2 Estimators of the parameter in the propensity score model when the propensity score model is correctly specified but the outcome model is misspecified and when both models are misspecified

Censoring rate	Size	Parameter	(c)				(d)			
			Bias	Std. err	Sd	Cp	Bias	Std. err	Sd	Cp
40%	n=500	β_0	-0.003	0.104	0.105	0.938	0.938	0.105	0.098	0.000
		β_1	0.046	0.177	0.186	0.924	-0.676	0.104	0.099	0.000
		β_2	-0.019	0.114	0.113	0.956	0.338	0.104	0.107	0.130
		β_3	0.000	0.109	0.104	0.938	0.105	0.105	0.105	0.820
		β_4	-0.003	0.108	0.104	0.964	-0.172	0.097	0.095	0.556
	n=1000	β_0	0.004	0.074	0.071	0.960	0.938	0.073	0.071	0.000
		β_1	0.036	0.121	0.125	0.932	-0.673	0.072	0.071	0.000
		β_2	-0.012	0.079	0.075	0.958	0.330	0.072	0.072	0.006
		β_3	0.001	0.075	0.072	0.962	0.097	0.073	0.070	0.740
		β_4	0.000	0.074	0.071	0.956	-0.176	0.068	0.068	0.270
20%	n=500	β_0	-0.001	0.106	0.098	0.970	0.948	0.105	0.100	0.000
		β_1	0.053	0.180	0.170	0.952	-0.672	0.105	0.102	0.002
		β_2	-0.020	0.114	0.107	0.960	0.329	0.104	0.098	0.096
		β_3	-0.011	0.109	0.105	0.960	0.099	0.105	0.103	0.838
		β_4	0.006	0.107	0.108	0.962	-0.170	0.098	0.092	0.608
	n=1000	β_0	-0.003	0.072	0.069	0.956	0.937	0.072	0.072	0.000
		β_1	0.017	0.118	0.126	0.924	-0.676	0.071	0.071	0.000
		β_2	-0.014	0.078	0.074	0.944	0.336	0.072	0.071	0.006
		β_3	-0.002	0.075	0.076	0.948	0.103	0.073	0.070	0.698
		β_4	0.000	0.073	0.072	0.952	-0.179	0.068	0.068	0.254

4.1.2 结果模型的参数估计

在本小节中我们分别在上述四种情形下,对结果模型中的参数进行估计.我们也讨论了不同样本量($n=500$, $n=1000$)和不同删失率(20%, 40%)以及 3.1.3 节中不同权函数 $\omega_2(u)$ 的选择(WLS1, WLS2)对参数估计量的有限样本性能的影响.

在四种情形以及不同样本量、不同删失率和不同权函数(WLS1, WLS2)的设置下,表 4.3-4.6 展示了结果模型中参数估计的偏差、标准误差、标准差以及 95%置信区间的覆盖率.总的来说,删失率越小,估计的效果越好.随着样本量的增大,偏差、标准误差和标准差都在减小.在表 4.3 和表 4.4 的情形下(即结果模型被正确指定时),除了截距项外,参数估计的偏差、标准误差、标准差都较小,置信区间覆盖率都在 95%左右.由此可见,我们对参数估计的方法在结果模型正确的情况下表现良好.在表 4.5 和表 4.6 的情形下(即结果模型被错误指定时),参数估计量的偏差较大,但标准误差较小,覆盖率远小于

95%，估计的效果不好，这也符合设置的情况。在表 4.3-4.6 中，我们发现权函数 WLS1 与 WLS2 的估计效果相差不多。

表 4.3 两模型均被正确指定时结果模型中的参数估计

Tab. 4.3 Estimators of the parameter in the outcome model when both models are correct

Censoring rate	Size	Parameter	WLS1				WLS2			
			Bias	Std. err	Sd	Cp	Bias	Std. err	Sd	Cp
40%	n=500	α_{00}	-0.370	0.076	0.083	0.006	-0.372	0.079	0.089	0.006
		α_{01}	0.002	0.075	0.078	0.940	0.009	0.086	0.099	0.906
		α_{02}	0.050	0.070	0.070	0.890	0.051	0.081	0.082	0.898
		α_{03}	0.043	0.069	0.073	0.872	0.045	0.078	0.085	0.866
		α_{04}	0.051	0.070	0.072	0.872	0.047	0.079	0.082	0.870
		α_{10}	-0.364	0.077	0.085	0.016	-0.366	0.080	0.088	0.018
		α_{11}	0.102	0.077	0.076	0.756	0.092	0.091	0.100	0.788
		α_{12}	0.050	0.072	0.076	0.880	0.049	0.082	0.093	0.880
		α_{13}	0.043	0.071	0.075	0.892	0.036	0.081	0.086	0.906
		α_{14}	0.051	0.072	0.076	0.880	0.047	0.082	0.090	0.888
	n=1000	α_{00}	-0.361	0.054	0.055	0.000	-0.362	0.056	0.058	0.000
		α_{01}	-0.002	0.053	0.055	0.926	-0.002	0.064	0.070	0.928
		α_{02}	0.050	0.051	0.052	0.834	0.049	0.060	0.068	0.826
		α_{03}	0.047	0.049	0.049	0.842	0.047	0.056	0.059	0.844
		α_{04}	0.047	0.048	0.054	0.790	0.050	0.055	0.063	0.820
		α_{10}	-0.353	0.055	0.058	0.000	-0.354	0.057	0.062	0.000
		α_{11}	0.106	0.055	0.060	0.492	0.098	0.067	0.080	0.650
		α_{12}	0.046	0.052	0.050	0.836	0.043	0.061	0.066	0.860
		α_{13}	0.052	0.051	0.053	0.816	0.050	0.059	0.065	0.834
		α_{14}	0.047	0.050	0.053	0.820	0.045	0.058	0.065	0.854
20%	n=500	α_{00}	-0.179	0.071	0.079	0.292	-0.180	0.074	0.084	0.330
		α_{01}	-0.006	0.071	0.076	0.920	-0.007	0.085	0.096	0.906
		α_{02}	0.033	0.067	0.068	0.896	0.028	0.078	0.086	0.894
		α_{03}	0.025	0.065	0.064	0.940	0.026	0.076	0.080	0.922
		α_{04}	0.031	0.065	0.070	0.910	0.032	0.075	0.085	0.898
		α_{10}	-0.173	0.073	0.078	0.346	-0.175	0.076	0.081	0.370
		α_{11}	0.068	0.073	0.070	0.968	0.062	0.087	0.094	0.948
		α_{12}	0.028	0.068	0.070	0.912	0.028	0.079	0.081	0.916
		α_{13}	0.039	0.067	0.068	0.986	0.036	0.077	0.080	0.906
		α_{14}	0.027	0.067	0.068	0.920	0.021	0.076	0.080	0.922
	n=1000	α_{00}	-0.187	0.051	0.056	0.054	-0.187	0.053	0.059	0.086
		α_{01}	0.001	0.050	0.053	0.926	0.001	0.062	0.071	0.910
		α_{02}	0.029	0.047	0.047	0.912	0.027	0.056	0.058	0.926
		α_{03}	0.030	0.046	0.045	0.902	0.028	0.054	0.056	0.902
		α_{04}	0.029	0.046	0.047	0.892	0.029	0.053	0.054	0.902
		α_{10}	-0.180	0.052	0.054	0.088	-0.182	0.054	0.057	0.108
		α_{11}	0.066	0.052	0.053	0.750	0.059	0.064	0.067	0.818
		α_{12}	0.034	0.048	0.047	0.990	0.032	0.056	0.059	0.988
		α_{13}	0.030	0.047	0.048	0.900	0.026	0.055	0.058	0.908
		α_{14}	0.032	0.047	0.048	0.864	0.029	0.054	0.059	0.868

表 4.4 结果模型被正确指定而倾向得分模型被错误指定时结果模型中的参数估计

Tab. 4.4 Estimators of the parameter in the outcome model when the outcome model is correctly specified but the propensity score model is misspecified

Censoring rate	Size	Parameter	WLS1				WLS2			
			Bias	Std. err	Sd	Cp	Bias	Std. err	Sd	Cp
40%	n=500	α_{00}	-0.299	0.059	0.069	0.018	-0.300	0.058	0.069	0.018
		α_{01}	0.001	0.060	0.060	0.950	0.002	0.060	0.061	0.958
		α_{02}	0.047	0.060	0.063	0.856	0.046	0.060	0.064	0.854
		α_{03}	0.044	0.059	0.061	0.860	0.045	0.059	0.061	0.848
		α_{04}	0.045	0.058	0.059	0.862	0.045	0.058	0.060	0.866
		α_{10}	-0.443	0.113	0.117	0.032	-0.443	0.113	0.118	0.030
		α_{11}	0.113	0.119	0.112	0.830	0.108	0.128	0.128	0.826
		α_{12}	0.051	0.110	0.111	0.908	0.050	0.117	0.121	0.896
		α_{13}	0.054	0.113	0.116	0.906	0.055	0.120	0.124	0.888
		α_{14}	0.056	0.118	0.117	0.912	0.054	0.125	0.126	0.912
	n=1000	α_{00}	-0.298	0.042	0.048	0.002	-0.298	0.042	0.048	0.004
		α_{01}	0.002	0.043	0.044	0.958	0.003	0.043	0.044	0.954
		α_{02}	0.044	0.043	0.046	0.812	0.044	0.043	0.046	0.824
		α_{03}	0.045	0.043	0.043	0.834	0.045	0.043	0.043	0.822
		α_{04}	0.048	0.042	0.045	0.768	0.048	0.042	0.045	0.776
		α_{10}	-0.450	0.077	0.078	0.000	-0.451	0.078	0.078	0.000
		α_{11}	0.110	0.081	0.079	0.716	0.104	0.088	0.091	0.760
		α_{12}	0.050	0.075	0.080	0.852	0.049	0.080	0.085	0.874
		α_{13}	0.057	0.078	0.084	0.866	0.056	0.083	0.089	0.864
		α_{14}	0.053	0.082	0.080	0.896	0.053	0.087	0.086	0.898
20%	n=500	α_{00}	-0.287	0.085	0.094	0.088	-0.286	0.085	0.094	0.092
		α_{01}	0.004	0.086	0.087	0.930	0.003	0.089	0.093	0.920
		α_{02}	0.032	0.082	0.083	0.916	0.031	0.084	0.087	0.906
		α_{03}	0.034	0.079	0.080	0.912	0.033	0.082	0.083	0.918
		α_{04}	0.038	0.092	0.090	0.932	0.041	0.095	0.095	0.934
		α_{10}	-0.364	0.054	0.059	0.000	-0.364	0.054	0.059	0.000
		α_{11}	0.038	0.050	0.048	0.892	0.037	0.051	0.049	0.898
		α_{12}	0.077	0.051	0.052	0.966	0.077	0.052	0.053	0.986
		α_{13}	0.078	0.052	0.052	0.966	0.077	0.052	0.053	0.976
		α_{14}	0.073	0.049	0.049	0.962	0.073	0.049	0.050	0.964
	n=1000	α_{00}	-0.138	0.038	0.042	0.072	-0.138	0.038	0.042	0.074
		α_{01}	0.000	0.039	0.044	0.906	0.000	0.040	0.044	0.906
		α_{02}	0.027	0.040	0.042	0.908	0.027	0.041	0.043	0.904
		α_{03}	0.025	0.040	0.042	0.894	0.025	0.040	0.042	0.886
		α_{04}	0.023	0.039	0.042	0.900	0.023	0.039	0.042	0.898
		α_{10}	-0.235	0.074	0.076	0.110	-0.235	0.074	0.076	0.114
		α_{11}	0.077	0.076	0.074	0.830	0.072	0.083	0.084	0.824
		α_{12}	0.032	0.071	0.072	0.922	0.028	0.076	0.079	0.916
		α_{13}	0.037	0.074	0.073	0.916	0.036	0.078	0.080	0.898
		α_{14}	0.032	0.077	0.078	0.918	0.031	0.081	0.084	0.928

表 4.5 倾向得分模型被正确指定而结果模型被错误指定时结果模型中的参数估计

Tab. 4.5 Estimators of the parameter in the outcome model when the propensity score model is correctly specified but the outcome model is misspecified

Censoring rate	Size	Parameter	WLS1				WLS2			
			Bias	Std. err	Sd	Cp	Bias	Std. err	Sd	Cp
40%	n=500	α_{00}	-0.718	0.080	0.083	0.000	-0.723	0.081	0.086	0.000
		α_{01}	0.001	0.079	0.080	0.938	0.006	0.089	0.100	0.898
		α_{02}	0.110	0.074	0.076	0.674	0.105	0.081	0.087	0.740
		α_{03}	0.089	0.073	0.075	0.750	0.092	0.081	0.088	0.766
		α_{04}	0.123	0.072	0.076	0.578	0.130	0.079	0.088	0.596
		α_{10}	-0.942	0.076	0.084	0.000	-0.946	0.079	0.090	0.000
		α_{11}	0.226	0.076	0.076	0.160	0.218	0.089	0.101	0.310
		α_{12}	0.104	0.071	0.071	0.700	0.104	0.080	0.087	0.724
		α_{13}	0.081	0.070	0.068	0.792	0.077	0.078	0.083	0.814
		α_{14}	0.145	0.070	0.066	0.460	0.136	0.079	0.077	0.590
	n=1000	α_{00}	-0.713	0.058	0.061	0.000	-0.715	0.060	0.065	0.000
		α_{01}	0.000	0.057	0.057	0.962	0.003	0.068	0.078	0.902
		α_{02}	0.110	0.053	0.053	0.432	0.110	0.062	0.067	0.532
		α_{03}	0.086	0.052	0.058	0.612	0.088	0.060	0.069	0.658
		α_{04}	0.128	0.052	0.051	0.312	0.135	0.059	0.061	0.354
		α_{10}	-0.938	0.054	0.058	0.000	-0.943	0.056	0.060	0.000
		α_{11}	0.230	0.053	0.054	0.014	0.218	0.064	0.068	0.108
		α_{12}	0.110	0.050	0.048	0.400	0.112	0.057	0.061	0.480
		α_{13}	0.092	0.049	0.050	0.550	0.088	0.057	0.062	0.652
		α_{14}	0.150	0.049	0.050	0.154	0.140	0.056	0.059	0.332
20%	n=500	α_{00}	-0.496	0.075	0.086	0.004	-0.499	0.076	0.090	0.002
		α_{01}	0.001	0.075	0.077	0.938	0.006	0.086	0.096	0.902
		α_{02}	0.102	0.069	0.071	0.676	0.103	0.079	0.085	0.694
		α_{03}	0.076	0.068	0.068	0.802	0.081	0.077	0.077	0.806
		α_{04}	0.124	0.070	0.076	0.566	0.134	0.079	0.087	0.614
		α_{10}	-0.763	0.073	0.076	0.000	-0.769	0.076	0.082	0.000
		α_{11}	0.219	0.073	0.073	0.152	0.209	0.085	0.099	0.336
		α_{12}	0.109	0.067	0.066	0.634	0.107	0.077	0.079	0.714
		α_{13}	0.075	0.066	0.063	0.804	0.071	0.075	0.075	0.846
		α_{14}	0.150	0.068	0.072	0.402	0.139	0.076	0.082	0.550
	n=1000	α_{00}	-0.501	0.053	0.056	0.000	-0.503	0.056	0.059	0.000
		α_{01}	-0.001	0.053	0.058	0.932	0.002	0.064	0.072	0.922
		α_{02}	0.105	0.050	0.050	0.424	0.104	0.057	0.061	0.532
		α_{03}	0.077	0.048	0.051	0.622	0.079	0.056	0.059	0.688
		α_{04}	0.133	0.049	0.051	0.234	0.142	0.057	0.061	0.312
		α_{10}	-0.764	0.052	0.053	0.000	-0.768	0.054	0.056	0.000
		α_{11}	0.217	0.051	0.051	0.018	0.206	0.063	0.066	0.116
		α_{12}	0.102	0.048	0.047	0.418	0.099	0.057	0.055	0.582
		α_{13}	0.079	0.046	0.045	0.610	0.076	0.054	0.057	0.692
		α_{14}	0.149	0.048	0.050	0.150	0.142	0.055	0.059	0.276

表 4.6 两模型均被错误指定时结果模型中参数估计

Tab. 4.6 Estimators of the parameter in the outcome model when both models are misspecified

Censoring rate	Size	Parameter	WLS1				WLS2			
			Bias	Std. err	Sd	Cp	Bias	Std. err	Sd	Cp
40%	n=500	α_{00}	-0.609	0.060	0.068	0.002	-0.609	0.060	0.067	0.002
		α_{01}	0.001	0.062	0.077	0.892	0.002	0.063	0.077	0.896
		α_{02}	0.100	0.062	0.067	0.642	0.100	0.063	0.067	0.648
		α_{03}	0.085	0.061	0.066	0.708	0.086	0.061	0.066	0.700
		α_{04}	0.138	0.060	0.068	0.382	0.138	0.060	0.067	0.382
		α_{10}	-0.957	0.111	0.117	0.000	-0.609	0.060	0.067	0.002
		α_{11}	0.232	0.113	0.107	0.454	-0.958	0.112	0.117	0.000
		α_{12}	0.109	0.110	0.108	0.800	0.229	0.121	0.117	0.512
		α_{13}	0.092	0.111	0.111	0.852	0.107	0.116	0.115	0.830
		α_{14}	0.136	0.118	0.118	0.768	0.087	0.116	0.115	0.864
	n=1000	α_{00}	-0.605	0.044	0.056	0.000	-0.605	0.044	0.056	0.000
		α_{01}	0.003	0.045	0.050	0.922	0.004	0.046	0.050	0.916
		α_{02}	0.107	0.046	0.054	0.338	0.107	0.046	0.055	0.340
		α_{03}	0.083	0.046	0.052	0.538	0.084	0.046	0.053	0.524
		α_{04}	0.138	0.044	0.047	0.136	0.138	0.044	0.047	0.142
		α_{10}	-0.952	0.078	0.081	0.000	-0.953	0.078	0.081	0.000
		α_{11}	0.233	0.078	0.079	0.166	0.229	0.084	0.089	0.236
		α_{12}	0.105	0.074	0.075	0.710	0.103	0.078	0.081	0.748
		α_{13}	0.091	0.076	0.079	0.754	0.087	0.081	0.085	0.788
		α_{14}	0.133	0.081	0.080	0.624	0.132	0.085	0.085	0.646
20%	n=500	α_{00}	-0.434	0.055	0.065	0.000	-0.435	0.055	0.065	0.000
		α_{01}	0.003	0.058	0.064	0.934	0.003	0.059	0.066	0.932
		α_{02}	0.100	0.058	0.065	0.566	0.100	0.059	0.065	0.578
		α_{03}	0.075	0.057	0.061	0.734	0.075	0.058	0.062	0.734
		α_{04}	0.135	0.058	0.062	0.380	0.134	0.058	0.062	0.372
		α_{10}	-0.772	0.108	0.109	0.000	-0.774	0.109	0.109	0.000
		α_{11}	0.214	0.109	0.104	0.480	0.209	0.118	0.116	0.530
		α_{12}	0.104	0.103	0.104	0.818	0.103	0.110	0.115	0.824
		α_{13}	0.081	0.107	0.103	0.874	0.079	0.114	0.115	0.882
		α_{14}	0.138	0.115	0.108	0.778	0.137	0.120	0.113	0.792
	n=1000	α_{00}	-0.438	0.040	0.046	0.000	-0.438	0.040	0.046	0.000
		α_{01}	0.004	0.042	0.044	0.938	0.004	0.043	0.044	0.940
		α_{02}	0.103	0.043	0.046	0.326	0.103	0.043	0.046	0.322
		α_{03}	0.076	0.041	0.044	0.514	0.077	0.041	0.044	0.498
		α_{04}	0.137	0.042	0.047	0.106	0.137	0.041	0.047	0.098
		α_{10}	-0.772	0.075	0.082	0.000	-0.774	0.075	0.083	0.000
		α_{11}	0.220	0.075	0.074	0.176	0.214	0.081	0.084	0.272
		α_{12}	0.104	0.071	0.073	0.684	0.105	0.076	0.080	0.696
		α_{13}	0.083	0.072	0.077	0.758	0.079	0.077	0.082	0.798
		α_{14}	0.132	0.079	0.078	0.616	0.130	0.082	0.080	0.642

4.2 平均因果效应的估计

在本节中，我们利用倾向得分模型和结果模型中参数的估计构建平均因果效应(ACE)的四种估计，分别为 Horvitz-Thompson 估计(HT)，逆倾向得分加权估计(IPW)，加权最小二乘估计(WLS)以及双稳健估计(DR). 将这四种估计分别在 4.1 节所述的四种情形下进行对比研究.

表 4.7 两模型均被正确指定时、结果模型被正确指定而倾向得分模型被错误指定时平均因果效应的四种估计

Tab. 4.7 Four estimators of the average of casual effects when both models are correct and when the outcome model is correctly specified but the propensity score model is misspecified

Censoring rate	Size	Method	(a)				(b)			
			Bias	Std. err	Sd	Cp	Bias	Std. err	Sd	Cp
40%	n=500	HT	-0.046	0.185	0.160	0.994	-0.668	0.070	0.068	0.000
		IPW	-0.050	0.124	0.129	0.992	-0.201	0.105	0.118	0.516
		WLS	-0.031	0.115	0.123	0.968	-0.138	0.106	0.120	0.788
		DR	-0.028	0.099	0.100	0.982	-0.139	0.106	0.120	0.992
	n=1000	HT	-0.045	0.111	0.115	0.980	0.109	0.109	0.161	0.880
		IPW	-0.046	0.089	0.091	0.976	0.176	0.087	0.108	0.480
		WLS	-0.027	0.082	0.084	0.982	0.189	0.082	0.097	0.312
		DR	-0.021	0.069	0.068	0.978	-0.016	0.092	0.131	0.902
20%	n=500	HT	-0.071	0.136	0.165	0.994	-0.485	0.093	0.086	0.000
		IPW	-0.058	0.119	0.120	0.986	-0.164	0.101	0.115	0.702
		WLS	-0.047	0.109	0.114	0.978	-0.119	0.100	0.116	0.888
		DR	-0.037	0.872	0.112	0.980	-0.119	0.100	0.115	0.982
	n=1000	HT	-0.065	0.112	0.107	0.978	0.156	0.090	0.106	0.666
		IPW	-0.059	0.083	0.083	0.960	0.109	0.077	0.088	0.824
		WLS	-0.048	0.079	0.079	0.964	0.114	0.074	0.084	0.760
		DR	-0.038	0.078	0.077	0.968	0.067	0.081	0.093	0.922

表 4.7 和表 4.8 展示了平均因果效应的四种估计方法在不同删失率和样本量的设置下的偏差、标准误差、标准差和 95%置信区间的覆盖率. 在表 4.7(a) 情形下，当两个模型均被正确指定时 Horvitz-Thompson 估计、逆倾向得分加权估计、加权最小二乘估计和双稳健估计均有相对较小的偏差、标准误差和标准差且覆盖率接近 95%. 表 4.7(b) 显示了当结果模型被正确指定但是倾向得分模型被错误指定时各种估计的性能. 正如预期的那样，我们发现仅依赖于倾向得分的 Horvitz-Thompson 估计和逆倾向得分加权估计在使用标准逻辑回归模型时具有较大的偏差. 相比之下，双稳健估计的偏差在大样本下较小. 在表 4.8(c) 的情形下，当倾向得分模型被正确指定但结果模型被错误指定时，正如预期的那样，四种估计的效果与表 4.1(a) 相似. 结合表 4.7(b) 和表 4.8(c) 可以证实双稳健估计的双重稳健性，即只要正确指定倾向得分模型或结果模型，双稳健估计表现良好. 表

4.8(d) 表明了 Kang 和 Schafer(2007)提出的最重要的一点, 即当倾向得分模型和结果模型均被错误指定时, 四种估计的性能均下降. 其中有些覆盖率特别小, 甚至为 0 的主要原因为偏差大, 但标准误差及标准差较小.

表 4.8 倾向得分模型被正确指定而结果模型被错误指定时、两模型均被错误指定时平均因果效应的四种估计

Tab. 4.8 Four estimators of the average of casual effects when the propensity score model is correctly specified but the outcome model is misspecified and when when both models are misspecified

Censoring rate	Size	Method	(c)				(d)			
			Bias	Std. err	Sd	Cp	Bias	Std. err	Sd	Cp
40%	n=500	HT	0.054	0.147	0.135	0.948	-0.238	0.133	0.142	0.634
		IPW	-0.065	0.125	0.121	0.994	-0.196	0.130	0.132	0.802
		WLS	-0.059	0.116	0.114	0.984	-0.186	0.128	0.129	0.770
		DR	-0.022	0.098	0.094	0.986	-0.316	0.108	0.116	0.060
	n=1000	HT	0.059	0.099	0.101	0.924	-0.242	0.096	0.110	0.196
		IPW	-0.069	0.090	0.093	0.970	-0.199	0.094	0.099	0.364
		WLS	-0.064	0.084	0.090	0.840	-0.191	0.091	0.095	0.382
		DR	-0.027	0.069	0.071	0.980	-0.319	0.078	0.091	0.008
20%	n=500	HT	-0.022	0.166	0.160	0.994	-0.219	0.132	0.137	0.686
		IPW	-0.072	0.117	0.124	0.978	-0.152	0.125	0.126	0.930
		WLS	-0.067	0.110	0.121	0.968	-0.146	0.124	0.125	0.940
		DR	-0.063	0.112	0.116	0.980	-0.275	0.116	0.120	0.256
	n=1000	HT	-0.003	0.103	0.102	1.000	-0.221	0.093	0.105	0.248
		IPW	-0.070	0.084	0.081	0.962	-0.152	0.088	0.097	0.712
		WLS	-0.067	0.080	0.077	0.952	-0.149	0.087	0.096	0.732
		DR	-0.061	0.077	0.074	0.960	-0.280	0.082	0.091	0.004

5 实例分析

本章主要将四种平均因果效应的估计方法运用到原发性胆汁性肝硬化数据和乳腺癌数据中进行分析.

5.1 原发性胆汁性肝硬化数据

原发性胆汁性肝硬化(Primary Biliary Cirrhosis, PBC)梅奥临床试验的时间是从 1974 年 1 月到 1984 年 5 月, 目的是评价 D-青霉胺治疗 PBC 的临床疗效, 数据列于 Fleming 和 Harrington^[63]的附录 D 中. 在这项临床研究中, 已经研究了共计 424 例符合试验标准的 PBC 患者. 然而, 在 424 例 PBC 患者中, 只有数据集中的前 312 例患者完全参与了随机试验. 因此, 完整的观察数据均来自这 312 例患者. 我们重点分析了 276 例无缺失值的观测数据.

关于测量单位和变量表示的描述如下. 连续变量: age(年龄, 年)、time(生存时间, 天)、bili(血清胆红素, mg/dl)、chol(血清胆固醇, mg/dl)、albumin(血清白蛋白, g/dl)、copper(尿铜, u/day)、alk.phos(碱性磷酸酶, u/liter)、ast(天冬氨酸转氨酶, u/ml)、trig(甘油三酸酯, mg/dl)、platelet(血小板, 1000ml/每立方)和 protime(标准凝血时间, 秒); 分类变量: status(状态, 0 表示删失、1 表示移植、2 表示死亡)、trt(处理, 1 表示 D-青霉胺、2 表示安慰剂)、stage(疾病的组织学分期)、sex(性别, m 为男性、f 为女性)、ascites(出现腹水, 0 表示没有腹水、1 表示存在腹水)、hepato(肝脏肿大, 0 表示没有肝肿大、1 表示存在肝肿大)、spiders(皮肤血管畸形, 0 表示没有畸形、1 表示有畸形)、edema(水肿, 0 表示没有水肿、0.5 表示未治疗或成功治疗的水肿、1 表示未成功治疗的水肿). 表 5.1 对 PBC 数据根据 trt 变量分组进行简单的描述性统计分析, 可以看到删失率约为 60%, 各协变量分布较为平衡

接下来我们将两个分类变量 stage 和 edema 转化为哑变量; trt 变量中用 0 表示控制组(安慰剂)、1 表示处理组(D-青霉胺); sex 变量中用 0 表示男性、1 表示女性; status 变量中的 0 和 1 一同记为 0 表示删失, 2 记为 1 表示死亡. 图 5.1 中的左图比较了不同 trt 的存活率, 图中生存曲线交叉显示出控制组和处理组之间的非比例危险. 同样, 图 5.1 中的右图比较了男性和女性在不同 trt 下的存活率. 曲线显示, 与服用 D-青霉胺治疗 PBC 的男性相比, 女性似乎更有生存优势.

根据 Kaplan-Meier 生存曲线, 似乎 D-青霉胺对治疗 PBC 并没有什么效果, 但由表 5.2 展示的平均因果效应的估计, 我们发现, D-青霉胺具有正的平均因果效应且具有统计

学意义. 表 5.2 括号中的数值表示估计的标准误差, 这里使用非参数 **bootstrap** 方法给出. 由此可以看出, 逆倾向得分加权估计的标准误差最小, 加权最小二乘估计和双稳健估计较为接近, Horvitz -Thompson 估计的标准误差最大. 表 5.3 为结果模型的参数在 3.1.3 节中权函数 $\omega_2(u)$ 的不同选择(WLS1, WLS2)下的估计.

表 5.1 PBC 数据的描述性统计分析
Tab. 5.1 Descriptive statistical analysis of the PBC data

	1 (N=136)	2 (N=140)	Overall (N=276)
status			
0	70 (51.5%)	77 (55.0%)	147 (53.3%)
1	9 (6.6%)	9 (6.4%)	18 (6.5%)
2	57 (41.9%)	54 (38.6%)	111 (40.2%)
time			
Mean (SD)	1960 (1090)	2000 (1140)	1980 (1110)
Median [Min, Max]	1800 [41.0, 4560]	1790 [51.0, 4520]	1790 [41.0, 4560]
age			
Mean (SD)	51.2 (11.0)	48.5 (9.85)	49.8 (10.5)
Median [Min, Max]	51.7 [26.3, 78.4]	47.9 [30.6, 74.5]	49.7 [26.3, 78.4]
sex			
m	20 (14.7%)	14 (10.0%)	34 (12.3%)
f	116 (85.3%)	126 (90.0%)	242 (87.7%)
albumin			
Mean (SD)	3.49 (0.439)	3.54 (0.369)	3.52 (0.405)
Median [Min, Max]	3.55 [2.10, 4.40]	3.55 [1.96, 4.38]	3.55 [1.96, 4.40]
ascites			
0	125 (91.9%)	132 (94.3%)	257 (93.1%)
1	11 (8.1%)	8 (5.7%)	19 (6.9%)
alk.phos			
Mean (SD)	2020 (2130)	1980 (2110)	2000 (2120)
Median [Min, Max]	1220 [369, 11600]	1310 [289, 13900]	1280 [289, 13900]
hepato			
0	72 (52.9%)	62 (44.3%)	134 (48.6%)
1	64 (47.1%)	78 (55.7%)	142 (51.4%)
spiders			
0	96 (70.6%)	100 (71.4%)	196 (71.0%)
1	40 (29.4%)	40 (28.6%)	80 (29.0%)
edema			
0	112 (82.4%)	122 (87.1%)	234 (84.8%)
0.5	15 (11.0%)	10 (7.1%)	25 (9.1%)

	1 (N=136)	2 (N=140)	Overall (N=276)
1	9 (6.6%)	8 (5.7%)	17 (6.2%)
bili			
Mean (SD)	2.96 (3.66)	3.70 (5.35)	3.33 (4.60)
Median [Min, Max]	1.45 [0.300, 20.0]	1.30 [0.300, 28.0]	1.40 [0.300, 28.0]
chol			
Mean (SD)	366 (212)	376 (256)	371 (235)
Median [Min, Max]	316 [127, 1710]	307 [120, 1780]	310 [120, 1780]
copper			
Mean (SD)	103 (94.7)	98.3 (81.8)	101 (88.3)
Median [Min, Max]	74.5 [9.00, 588]	73.0 [4.00, 558]	74.0 [4.00, 588]
stage			
1	9 (6.6%)	3 (2.1%)	12 (4.3%)
2	31 (22.8%)	28 (20.0%)	59 (21.4%)
3	51 (37.5%)	60 (42.9%)	111 (40.2%)
4	45 (33.1%)	49 (35.0%)	94 (34.1%)
trig			
Mean (SD)	124 (71.6)	126 (58.8)	125 (65.3)
Median [Min, Max]	106 [33.0, 598]	114 [44.0, 432]	108 [33.0, 598]
platelet			
Mean (SD)	258 (97.8)	265 (88.5)	262 (93.1)
Median [Min, Max]	253 [62.0, 563]	263 [71.0, 487]	257 [62.0, 563]
protime			
Mean (SD)	10.7 (0.850)	10.8 (1.14)	10.7 (1.01)
Median [Min, Max]	10.6 [9.00, 14.1]	10.6 [9.50, 17.1]	10.6 [9.00, 17.1]

表 5.2 平均因果效应的估计

Tab. 5.2 Estimators of the average of causal effects

	ACE
HT	0.460(0.868)
IPW	0.102(0.183)
WLS	0.100(0.250)
DR	0.112(0.248)

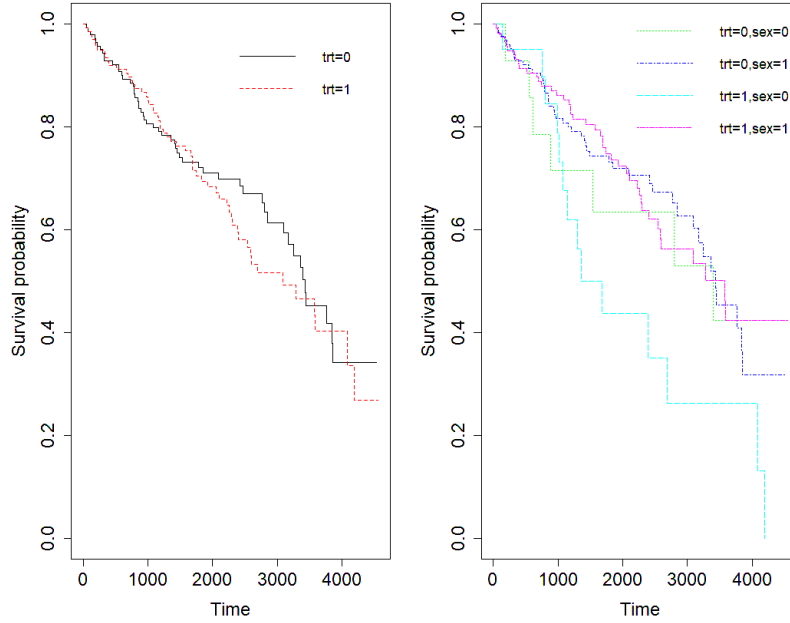


图 5.1 生存概率的 Kaplan-Meier 曲线

Fig. 5.1 Kaplan-Meier curve of survival probability

表 5.3 结果模型中的参数估计

Tab. 5.3 Estimators of the parameter in the outcome model

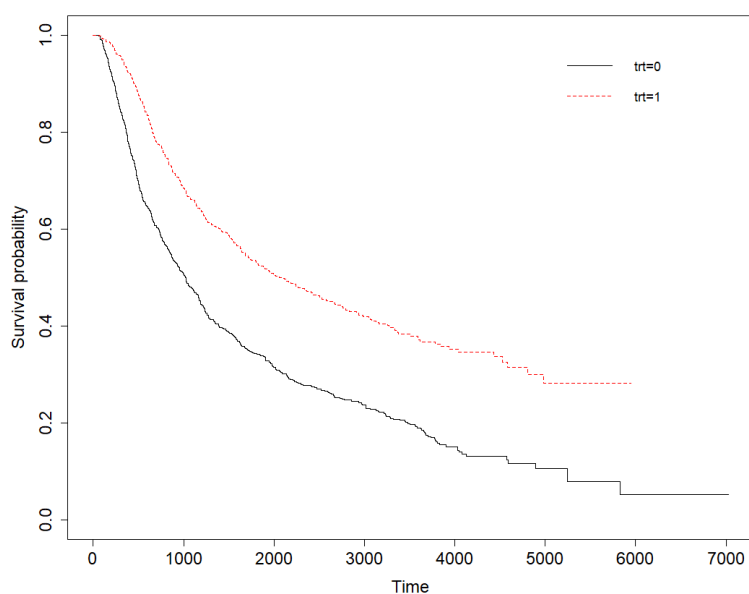
	β	WLS1		WLS2	
		α_0	α_1	α_0	α_1
(Intercept)	3.15e+00	4.61e+00	8.10e+00	4.96e+00	7.84e+00
age	2.82e-02	-1.20e-03	-1.08e-02	-3.06e-03	-1.54e-02
sex	-2.73e-01	-4.62e-02	1.336e-01	-1.92e-02	1.43e-01
ascites	7.31e-01	-3.38e-01	1.091e-01	-4.07e-01	1.21e-01
hepato	-3.98e-01	3.40e-02	1.96e-01	2.48e-02	1.46e-01
spiders	3.40e-01	-3.63e-01	-7.84e-02	-2.96e-01	-1.05 e-01
bili	-8.31e-02	-4.78e-03	2.75e-02	-2.43e-04	6.31e-04
chol	8.02e-04	-9.17e-05	-1.97e-04	-2.02e-04	- 2.23e-04
albumin	-2.22e-01	2.68e-01	-1.02e-01	3.01e-01	-5.861e-02
copper	1.91e-03	-1.45e-03	4.62e-04	-2.09e-03	1.64e-04
alk.phos	2.96e-05	9.19e-05	7.05e-05	7.97e-05	6.64e-05
ast	3.53e-04	3.35e-03	-8.56e-03	3.05e-03	-7.46e-03
trig	-1.89e-04	-3.42e-03	9.85e-04	-3.03e-03	1.30e-03
platelet	-1.13e-03	1.05e-03	-1.19e-03	1.53e-03	-8.61e-04
protime	-3.27e-01	2.805e-02	-8.52e-02	1.67e-02	-4.32e-02
edema1	-1.73e-01	9.65e-01	1.56e+00	6.74e-01	1.43e+00
edema2	6.58e-01	-2.85e-01	8.11e-01	-5.09e-01	8.24e-01
stage1	1.57e+00	1.00e-03	1.57e+00	1.00e-03	1.46e+00
stage2	2.10e-01	4.82e-01	8.50e-01	4.17e-01	7.01e-01
stage3	-4.65e-02	4.92e-01	4.25e-01	4.61e-01	3.37e-01

5.2 乳腺癌数据

原始数据集包括 2982 例原发性乳腺癌患者，他们的记录被收录在 Rotterdam 库中，其中 1546 人患有淋巴结阳性疾病。随访时间为 1~231 个月(中位数为 107 个月)。我们感兴趣的结果是无复发生存时间(Recurrence-free Survival Time, RFS)，其被定义为从初次手术到早期疾病复发或因任何原因死亡的时间。RFS 的结果中有 965(即近 90%)个事件(复发或死亡)是在 0 到 84 个月的时间段内观察到的，而整个随访时间为 231 个月。

关于测量单位和变量表示的描述如下。连续变量: age(年龄, 年)、nodes(阳性淋巴结数目)、pgr(孕酮受体, fmol/l)、er(雌激素受体, fmol/l)、rtime(复发或最后一次随访的天数)、dtime(死亡之日或最后一次随访的天数); 分类变量: meno(绝经状态, 0 表示绝经前、1 表示绝经后)、size(肿瘤大小, ≤ 20 、20-50、 > 50)、grade(肿瘤分级)、hormon(激素治疗, 0 表示否、1 表示是)、chemo(化学疗法, 0 表示否、1 表示是)、recur(复发, 0 表示无复发、1 表示复发)、death(死亡, 0 表示活着、1 表示死亡)。

在本文中，我们主要对化学疗法的因果效应感兴趣，所以将处理组设置为只接受化学疗法，控制组设置为均未接受化学疗法和激素治疗，并令 trt 变量为处理变量。因我们令潜在结果为 RFS，故令 status 变量中 0 表示删失(无复发无死亡)、1 表示复发或死亡，time 变量为 RFS，即当出现复发时表示首次复发时间，未复发时表示死亡时间。



5.2 生存概率的 Kaplan-Meier 曲线

Fig. 5.2 Kaplan-Meier curve of survival probability

表 5.4 Rotterdam 数据的描述性统计分析

Tab. 5.4 Descriptive statistical analysis of the Rotterdam data

	0 (N=655)	1 (N=552)	Overall (N=1207)
time			
Mean (SD)	1540 (1370)	2050 (1400)	1770 (1410)
Median [Min, Max]	1020 [64.0, 7030]	1800 [71.0, 5950]	1300 [64.0, 7030]
status			
0	112 (17.1%)	221 (40.0%)	333 (27.6%)
1	543 (82.9%)	331 (60.0%)	874 (72.4%)
age			
Mean (SD)	62.2 (11.6)	44.6 (7.28)	54.1 (13.2)
Median [Min, Max]	63.0 [28.0, 90.0]	45.0 [24.0, 73.0]	52.0 [24.0, 90.0]
meno			
0	117 (17.9%)	470 (85.1%)	587 (48.6%)
1	538 (82.1%)	82 (14.9%)	620 (51.4%)
size			
<=20	169 (25.8%)	228 (41.3%)	397 (32.9%)
20-50	364 (55.6%)	247 (44.7%)	611 (50.6%)
>50	122 (18.6%)	77 (13.9%)	199 (16.5%)
grade			
2	165 (25.2%)	145 (26.3%)	310 (25.7%)
3	490 (74.8%)	407 (73.7%)	897 (74.3%)
nodes			
Mean (SD)	5.82 (5.07)	4.23 (4.70)	5.09 (4.97)
Median [Min, Max]	4.00 [1.00, 34.0]	2.00 [1.00, 34.0]	3.00 [1.00, 34.0]
pgr			
Mean (SD)	164 (320)	177 (321)	170 (321)
Median [Min, Max]	40.0 [0, 3000]	55.5 [0, 5000]	47.0 [0, 5000]
er			
Mean (SD)	214 (306)	97.2 (190)	161 (266)
Median [Min, Max]	102 [0, 3010]	39.0 [0, 1930]	62.0 [0, 3010]

表 5.5 平均因果效应的估计

Tab. 5.5 Estimators of the average of causal effects

	ACE
HT	-1.028(1.108)
IPW	0.144(0.207)
WLS	0.252(0.173)
DR	0.266(0.187)

表 5.6 结果模型中参数的估计

Tab. 5.6 Estimators of the parameter in the outcome model

	β	WLS1		WLS2	
		α_0	α_1	α_0	α_1
(Intercept)	7.03e+00	6.88e+00	6.20e+00	6.35e+00	6.55e+00
age	-1.30e-01	6.95e-03	-7.90e-03	1.16e-02	1.36e-02
meno	-1.11e+00	-1.65e-01	-6.28e-03	-2.32e-01	-1.67e-01
grade	1.67e-01	-2.43e-01	3.06e-01	-1.75e-01	-1.40e-01
nodes	-8.16e-02	-2.56e-02	-2.56e-02	-2.84e-02	-2.92e-02
pgr	3.06e-04	1.58e-04	4.10e-04	1.37e-04	3.65e-04
er	-6.54e-04	4.93e-04	-3.91e-04	5.02e-04	-1.39e-04
size20	4.37e-01	4.37e-01	4.23e-01	4.44e-01	2.33e-01
size20_50	-2.10e-01	-2.50e-02	2.10e-01	2.02e-01	1.97e-01

根据是否接受化学疗法进行分组对数据进行描述性统计分析, 见表 5.4. 由表可知, 该数据的数据量为 1207, 删失率约为 28%, 协变量分布较不平衡. 我们将分类变量 size 进行哑变量处理.

图 5.2 绘制的是在不同处理下的生存概率的 Kaplan-Meier 曲线, 我们发现采用化学疗法的处理组的生存概率明显优于控制组.

表 5.5 估计了总体的平均因果效应, 其中括号里是估计的标准误差. 由表可知, 双稳健估计的标准误差最小, 逆倾向得分加权估计、加权最小二乘估计和双稳健估计都表现出了正的平均因果效应, 但 Horvitz-Thompson 估计为负, 由于其标准误差较大, 其估计的有效性遭到质疑. 表 5.6 为结果模型的参数在 3.1.3 节中权函数 $\omega_2(u)$ 的不同选择 (WLS1, WLS2) 下的估计.

结 论

本文考虑的结果变量为带有右删失的生存时间, 基于调整后的协变量平衡倾向得分和加权最小二乘方法来估计倾向得分模型和加速失效时间模型(结果模型), 然后运用四种估计方法给出平均因果效应的估计. 在本文中, 四种估计方法分别取为 Horvitz-Thompson 估计、逆倾向得分加权估计、加权最小二乘估计和双稳健估计. 本文证明了: (1) 在倾向得分模型被正确指定时, Horvitz-Thompson 估计和逆倾向得分加权估计具有无偏性; (2) 当倾向得分模型或结果模型中有一个模型被正确指定时, 双稳健估计具有无偏性.

本文进行了大量的模拟研究和实例分析. 模拟表明: (1) 删失率越小或样本量越大, 平均因果效应的四种估计效果越好. (2) 当倾向得分模型被正确指定时, 逆倾向得分加权估计、加权最小二乘估计和双稳健估计的估计效果良好, 但当倾向得分模型被错误指定时, 逆倾向得分加权估计和加权最小二乘估计有偏低效. (3) 当结果模型被正确指定时, 双稳健估计的表现更为出色, 由此表明双稳健估计具有双重稳健性. 在实例分析中, 我们将提出的方法应用于原发性胆汁性肝硬化数据和乳腺癌数据, 分别给出了平均因果效应的估计, 表明了 D-青霉胺对治疗原发性胆汁性肝硬化以及化学疗法对治疗乳腺癌均有正的平均因果效应.

本文的创新点在于: (1) 将协变量平衡倾向得分方法中的倾向得分模型从常见的逻辑回归模型扩展为广义线性模型, 从而使协变量平衡倾向得分方法应用更加广泛; (2) 利用生存分析中经典的加速失效时间模型对潜在结果进行建模, 并用调整后的协变量平衡倾向得分方法估计倾向得分模型以及用加权最小二乘法估计结果模型; (3) 利用 Kaplan-Meier 权对右删失数据进行校正, 给出四种平均因果效应的估计.

本文仅提出了平均因果效应的四种估计方法, 但并没有对其渐近性质进行更深一层的讨论. 在以后的研究中, 我们可以将结果模型进行推广, 例如 Cox 比例风险模型、可加风险模型等. 除此之外, 我们也可将处理分配机制推广至非二元的情况, 例如多值处理、连续处理和动态处理等. 我们还可将平均因果效应估计推广至高维删失数据分析中. 这些问题还有待进一步的研究.

参 考 文 献

- [1] Neyman J. On the application of probability theory to agricultural experiments: essay on principles[J]. Statistical Science, 1923, 5(4): 465-480.
- [2] Neyman J, Iwaskiewicz K, Kolodziejczyk S. Statistical problems in agricultural experimentation[J]. Supplement to the Journal of the Royal Statistical Society, 1935, 2(2): 107-180.
- [3] Fisher R A. The design of experiments[M]. London: Oliver and Boyd, 1935.
- [4] Rubin D B. Estimating causal effects if treatment in randomized and nonrandomized studies[J]. Journal of Educational Psychology, 1974, 66(5): 688-701.
- [5] Simpson E H. The interpretation of interaction in contingency tables[J]. Journal of the Royal Statistical Society, Series B, 1951, 13(2): 238-241.
- [6] Holland P W. Statistics and causal inference[J]. Journal of the American Statistical Association, 1986, 81(396): 945-960.
- [7] Freedman D, Pisani R, Purves R A. Statistics[M]. 2nd ed. New York: W.W Norton & Company, Inc., 1991.
- [8] Pearl J. Causal diagrams for empirical research[J]. Biometrika, 1995, 82(4): 669-688.
- [9] Pearl J. Causality: models, reasoning, and inference[M]. Los Angeles: Cambridge University Press, 2000.
- [10] Kleinbaum D G, Kupper L L, Morgenstern H. Epidemiologic research: principles and quantitative methods[M]. New York: Van Nostrand Reinhold, 1982.
- [11] Whittemore A S. Collapsibility of multidimensional contingency tables[J]. Journal of the Royal Statistical Society, Series B, 1978, 40(3): 328-340.
- [12] Miettinen O S, Cook E F. Confounding: essence and detection[J]. American Journal of Epidemiology, 1981, 114(4): 593-603.
- [13] Greenland S, Robins J, Pearl. Confounding and collapsibility in causal inference [J]. Statistical Science, 1999, 14(1): 29-46.
- [14] 黄丽红, 魏永越, 陈峰. 如何控制观察性疗效比较研究中的混杂因素: (一) 已测量混杂因素的统计学分析方法[J]. 中华流行病学杂志, 2019, 40(10): 1302-1307.
- [15] 黄丽红, 魏永越, 陈峰. 如何控制观察性疗效比较研究中的混杂因素: (二) 未知或未测量混杂因素的统计学分析方法[J]. 中华流行病学杂志, 2019, 40(11): 1450-1455.
- [16] Rosenbaum P R, Rubin D B. The central role of the propensity score in observational studies for causal effects[J]. Biometrika, 1983, 70(1): 41-55.

- [17] McCullagh P M, Nelder J A S. Generalized linear models[M]. Boca Raton: Chapman and Hall/CRC, 1989.
- [18] Hirano K, Imbens G, Ridder G. Efficient estimation of average treatment effects using the estimated propensity score[J]. *Econometrica*, 2003, 71(4): 1161-1189.
- [19] McCaffrey D F, Ridgeway G, Morral A R. Propensity score estimation with boosted regression for evaluating causal effects in observational studies[J]. *Psychological Methods*, 2005, 9(4): 403-425.
- [20] Rosenbaum P R, Rubin D B. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score[J]. *The American Statistician*, 1985, 39(1): 33-38.
- [21] Rosenbaum P R. Optimal matching for observational studies[J]. *Journal of the American Statistical Association*, 1989, 84(408): 1024-1032.
- [22] Rosenbaum P R, Rubin D B. Reducing bias in observational studies using subclassification on the propensity score[J]. *Journal of the American Statistical Association*, 1984, 79(387): 516-524.
- [23] Rosenbaum P R. A characterization of optimal designs for observational studies[J]. *Journal of the Royal Statistical Society, Series B*, 1991, 53(3): 597-610.
- [24] Rosenbaum P R. Model-based direct adjustment[J]. *Journal of the American Statistical Association*, 1987, 82(398): 387-394.
- [25] Robins J M, Hernan M A, Brumback B. Marginal structural models and causal inference in epidemiology[J]. *Epidemiology*, 2000, 11(5): 550-560.
- [26] Heckman J J, Ichimura H, Todd P E. Matching as an econometric evaluation estimator [J]. *Review of Economic Studies*, 1998, 65(2): 261-294.
- [27] Robins J M, Zhao R L P. Analysis of semiparametric regression models for repeated outcomes in the presence of missing Data[J]. *Journal of the American Statistical Association*, 1995, 90(429): 106-121.
- [28] Ho D, Imai K, King G, et al. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference[J]. *Political Analysis*, 2007, 15(3): 199-236.
- [29] Imbens G W. Nonparametric estimation of average treatment effects under exogeneity: a review[J]. *The Review of Economics and Statistics*, 2004, 86(1): 4-29.
- [30] Imbens G W, Rubin D B. Causal inference in statistics, social, and biomedical sciences[M]. New York: Cambridge University Press, 2015.
- [31] Kang J D Y, Schafer J L. Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data[J]. *Statistical Science*, 2007, 22(4): 574-580.

- [32] Smith J, Todd P. Does matching address LaLonde's critique of nonexperimental estimators[J]. *Journal of Econometrics*, 2005, 125(1-2): 305-353.
- [33] Imai K, King G, Stuart E. Misunderstandings between experimentalists and observation lists about causal inference[J]. *Journal of the Royal Statistical Society, Series A*, 2008, 171(2): 481-502.
- [34] Robins J M, Rotnitzky A, Zhao L P. Estimation of regression coefficients when some regressors are not always observed[J]. *Journal of the American Statistical Association*, 1994, 89(427): 846-866.
- [35] Imai K, Ratkovic M. Covariate balancing propensity score[J]. *Journal of the Royal Statistical Society, Series B*, 2014, 76(1): 243-263.
- [36] Fong C, Imai K. Covariate balancing propensity score for general treatment regimes [J]. *Princeton Manuscript*, 2014, 1-31.
- [37] Wyss R, Ellis A R, Brookhart M A, Girman C J, et al. The role of prediction modeling in propensity score estimation: an evaluation of logistic regression, BCART, and the covariate-balancing propensity score[J]. *American Journal of Epidemiology*, 2014, 180(6): 645-655.
- [38] Fan J, Imai K, Liu H, et al. Improving covariate balancing propensity score: a doubly robust and efficient approach[R]. Technical report, Princeton: Princeton University. 2016.
- [39] Imai K, Ratkovic M. Robust estimation of inverse probability weights for marginal structural models[J]. *Journal of the American Statistical Association*, 2015, 110(511): 1013-1023.
- [40] Oyenubi A. A note on covariate balancing propensity score and instrument-like variables[J]. *Economic Bulletin*, 2020, 40(1): 202-209.
- [41] Ning Y, Sida P, Imai K. Robust estimation of causal effects via a high-dimensional covariate balancing propensity score[J]. *Biometrika*, 2020, 107(3): 533-554.
- [42] Xie J, Liu C. Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data[J]. *Statistics in Medicine*, 2010, 24(20): 3089-3110.
- [43] Sugihara M. Survival analysis using inverse probability of treatment weighted methods based on the generalized propensity score[J]. *Pharmaceutical Statistics*, 2010, 9(1): 21-34.
- [44] Xu S, Shetterly S, Powers D, et al. Extension of Kaplan-Meier methods in observational studies with time-varying treatment[J]. *Value in Health the Journal of the International Society for Pharmacoeconomics & Outcomes Research*, 2012, 15(1): 167-174.

- [45] Cole S R, Miguel A. Hernán. Adjusted survival curves with inverse probability weights[J]. Computer Methods Programs in Biomedicine, 2004, 75(1): 45-49.
- [46] Austin P C. The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments[J]. Statistics in Medicine, 2014, 33(7): 1242-1258.
- [47] Austin P C, Schuster T. The performance of different propensity score methods for estimating absolute effects of treatments on survival outcomes: a simulation study [J]. Statistical Methods in Medical Research, 2016, 25(5): 2214-2237.
- [48] Austin P C, Stuart E A. Optimal full matching for survival outcomes: a method that merits more widespread use[J]. Statistics in Medicine, 2015, 34(30): 3949-3967.
- [49] Austin P C, Stuart E A. The performance of inverse probability of treatment weighting and full matching on the propensity score in the presence of model misspecification when estimating the effect of treatment on survival outcomes[J]. Statistical Methods in Medical Research, 2017, 26(4): 1654-1670.
- [50] Gayat E, Resche-Rigon M, Mary J Y, et al. Propensity score applied to survival data analysis through proportional hazards models: a Monte Carlo study[J]. Pharmaceutical Statistics, 2012, 11(3): 222-229.
- [51] Li J, Handorf E, Bekelman J, et al. Propensity score and doubly robust methods for estimating the effect of treatment on censored cost[J]. Statistics in Medicine, 2015, 35(12): 1985-1999.
- [52] Cho Y, Hu C, Ghosh D. Covariate adjustment using propensity scores for dependent censoring problems in the accelerated failure time model[J]. Statistics in Medicine, 2018, 37(4): 390-404.
- [53] Henderson N C, Louis T A, Rosner G L, et al. Individualized treatment effects with censored data via fully nonparametric Bayesian accelerated failure time models[J]. Biostatistics, 2020, 21(1): 50-68.
- [54] Rubin D B. Causal inference using potential outcomes[J]. Journal of the American Statistical Association, 2005, 100(469): 322-331.
- [55] Morgan S L, Winship C. Counterfactuals and causal inference: Methods and Principles for Social Research[M]. Cambridge: Cambridge University Press, 2015.
- [56] VanderWeele T. Explanation in causal inference: methods for mediation and interaction[M]. New York: Oxford University Press, 2015.
- [57] Lechner M. Earnings and employment effects of continuous off-the-job training in East Germany after unification[J]. Journal of Business and Economic Statistics, 1999, 17(7): 74-90.

- [58] Wei L J. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis[J]. *Statistics in Medicine*, 1992, 11(14-15): 1871-1879.
- [59] Newey W K, Mcfadden D. Large sample estimation and hypothesis testing[J]. *Handbook of Econometrics*, 1994, 4(5): 2111-2245.
- [60] Wedderburn R W M. Quasi-likelihood functions, generalized linear models, and the Gauss—Newton method[J]. *Biometrika*, 1974, 61(3): 439-447.
- [61] Kang J D Y, Schafer J L. Rejoinder: demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data[J]. *Statistical Science*, 2007, 22(4): 574-580.
- [62] Horvitz D, Thompson D. A generalization of sampling without replacement from a finite universe[J]. *Journal of the American Statistical Association*, 1952, 47(260): 663-685.
- [63] Fleming T R, Harrington D P. *Counting processes and survival analysis*[M]. 2nd ed. Hoboken: John Wiley & Sons, Inc., 2005.
- [64] Hammersley J M, Handscomb D C. *Monte Carlo methods*[M]. London: Methuen, 1964.
- [65] Geweke J. Bayesian inference in econometric models using Monte Carlo integration [J]. *Econometrica*, 1989, 57(6): 1317-1339.

攻读硕士学位期间发表学术论文情况

- 1 加速失效时间模型下基于倾向得分的平均因果效应分析. 马文婷, 王晓光. 大连理工大学网络学刊, 2021 年. 主办单位: 大连理工大学研究生院.

致 谢

我曾于万千世界里感受温暖，也曾于孤独自卑中敏感多疑；我曾感叹世间之大心生向往，也曾眷恋家乡之味驻足停留；我曾努力奋斗试图改变命运，也曾迷茫懒散不知未来何样。这一路上走走停停，阳光与风雨同在，成熟与稚嫩相伴，幸而有人，携同向前，纵有万般不舍，但仍心存感激。

学贵得师，亦贵得友。感谢我的导师王晓光老师常启我明思，指我迷津，言传身教，励我勤勉。承蒙导师之厚爱，幸得教诲，感激涕零，愿您一生长安，幸福美满。感谢鲁大伟老师、牛一老师、宋慧老师、沈新美老师和冯敬海老师传道授业，倾囊相授，助我良多，望诸位老师万事顺遂，桃李满园。

良师为灯，良友为伴。感谢师兄师姐不吝赐教，予我鼓励，助我毕业。感谢同学朋友相励相勉，共话未来。三年同门，相伴相知，就此别过，愿汝前程似锦，快乐相随。

结草衔环，难报之万一。感谢父母予我生命，不辞辛劳，无私奉献。我定不负父母之所望，不畏艰难，全力以赴。毕淑敏有言：“父母在，人生尚有来处。父母去，人生只剩归途。”我只盼二老身体健康，平安喜乐，福如东海，寿比南山。我定珍惜时光，常伴左右。

所遇良人，余生皆欢喜。感谢男友李明洋，虽隔万里，关爱备至，体贴入微，策我进步，助我定稿，望卿学业有成，事事顺利。

回首三年寒窗，求学之路，人生之旅，幸得贵师益友及父母良人。论文之结束，亦意味着学业之终结。入学十九载，几经耕耘，几经收获，不惧成长，未来可期。

大连理工大学学位论文独创性声明

作者郑重声明：所呈交的学位论文，是本人在导师的指导下进行研究工作所取得的成果。尽我所知，除文中已经注明引用内容和致谢的地方外，本论文不包含其他个人或集体已经发表的研究成果，也不包含其他已申请学位或其他用途使用过的成果。与我一同工作的同志对本研究所做的贡献均已在论文中做了明确的说明并表示了谢意。

若有不实之处，本人愿意承担相关法律责任。

学位论文题目： 加速失效时间模型下基于倾向得分的平均因果效应分析

作者签名： 马文停 日期： 2021 年 6 月 17 日

