arXiv:2601.06685v1 [stat.ME] 10 Jan 2026

右删失数据的 R估计

Glen A. Satten 生殖与产科系，埃默里大学医学院，美国佐治亚州亚特兰大。Mo Li 数学系，路易斯安那大学拉法叶分校，美国路易斯安那州拉法叶。Ni Zhao 生物统计系，约翰斯·霍普金斯大学，美国马里兰州巴尔的摩。Robert L. Strawderman *生物统计与计算生物学系，罗切斯特大学，美国纽约州罗切斯特。

摘要

本文探讨在线性模型框架下，如何将 R估计量直接推广至右删失结局。针对 Wilcoxon（即秩的线性函数）得分函数的情形，我们提出 R估计量所用秩及相应估计方程的一种自然推广，并说明它们如何分别精确表示为Ritov [1990] 和 Tsiatis [1990]提出的估计方程类中的成员。随后，我们为一大类有界的秩的非线性函数得分建立了类似结果。渐近理论和方差估计可直接由这些表示结果推导得出。残差分布函数的自洽估计量以及中间累积分布函数（必要时还包括其推广形式），在这些推导中发挥着关键作用。

KEYWORDS: 线性秩检验；中秩；半参数加速失效时间模型；加权对数秩估计方程。

*通讯作者：robert strawderman@urmc.rochester.edu

1 引言

线性模型是统计学中许多领域的基础工具。尽管普通最小二乘法是拟合线性模型最常用的方法，但众所周知，最小二乘估计量对离群观测很敏感。因此，人们提出了多种用于估计线性模型参数的稳健估计量，其中包括但不限于 R估计量。由于 R估计量基于秩，因此对底层误差（即残差）分布的形状具有更强的稳健性。该领域早期的一些重要研究包括Sen [1968], Jurečková [1971] 和 Jaeckel [1972]；例如，参见 Hettmansperger [1984], Chang等人 [1999], Hettmansperger 和 McKean [2011] 关于后续发展的精选研究。

线性模型广受欢迎，但众所周知，Cox比例风险模型是个例外，它长期以来一直是生存分析中的主力模型。加速失效时间（AFT）模型 [例如，Kalbfleisch 和 Prentice，2002年，第 7] 与线性模型直接类似；在最常见的形式下，它将失效时间的自然对数建模为可用协变量的线性函数。当对“误差”分布施加参数假设时，可以轻松地用最大似然法拟合AFT模型 [例如，Lawless，1982年，第6节。5]；在误差分布未作具体设定的半参数设定下，计算和推断则会变得极具挑战性。当对数线性模型中的参数是直接关注的对象，或出于其他原因必须估计这些参数时，AFT模型往往更受青睐 [例如，Satten等人， 2025]。

在线性模型中，若假设误差分布的均值为零，并且（通常还会假设）关于均值对称，则回归函数可解释为（条件）均值。在此背景下，由于对偏度和离群值的敏感性较低，R估计量仍是参数估计的理想选择 [Hettmansperger 和 McKean， 2011]。对于AFT模型，回归函数通常不被解释为均值函数模型。这是因为“误差”分布通常允许具有非零均值，旨在刻画将所有协变量设为零所定义的亚群体的基线生存经历。相应地，该底层分布可能会产生相对于其均值呈非对称分布的“误差”；而R估计量对这类行为的敏感性较低，这一优点使其成为构建估计量的良好基础。

关于右删失数据下AFT回归模型参数估计量的研究已有众多提案。半参数AFT模型早期极具影响力的研究包括 Prentice [1978] 和 Tsiatis [1990]的工作。Prentice [1978] 在AFT模型下检验回归系数假设的背景下，发展了针对右删失数据的线性秩检验理论，并建议将提出的得分检验统计量转化为目标函数，以此估计该系数；相关综述见 Lawless [1982, 第8节。2] 以及 Kalbfleisch 和 Prentice [2002, 第 7] 章。Tsiatis [1990]在此基础上，建立了相应估计量的渐近理论；另见 Mehrotra 等人 [1982], Cuzick [1985], 和Ritov [1990] 的相关研究，它们将这两种方法联系起来。由于半参数模型本身缺乏光滑性，估计过程较为复杂。在这方面，Fygenson和Ritov [1994] 的研究尤为突出：他们提出了一种有效的估计程序，通过最小化凸目标函数构造“Gehan估计量”。此后，研究者投入了大量精力，以改进AFT回归模型参数估计及相关推断和方差估计方法的计算方面；其中两个（众多）例子包括 Jin等人 [2003]以及 Johnson 和 Strawderman [2009]。

Gehan估计量具有一些很好的性质，这是因为它是通过最小化一个凸目标函数推导出来的。然而，若将其视为更广泛的加权对数秩估计量类别中的一员，那么从渐近效率的角度来看，它存在问题。具体而言，

众所周知，在这一类别中，效率最高的权重函数选择仅取决于潜在的失效分布 [例如，Tsiatis，1990]；相比之下，在加权对数秩估计量类别中，能够导出Gehan估计量的权重函数选择必然取决于潜在的删失分布。此外，结合本文的具体研究兴趣，Gehan估计量并非真正的R估计量推广，因为它仅利用能够明确排序的观测值之间的“秩”。

本文旨在针对右删失数据，构建一个尽可能贴近相应完整数据R估计量推导过程的加速失效时间模型估计方程。受删失无偏变换这一概念的启发，我们通过用具有类似性质的插补排序替代实际秩，实现了这一目标。对于使用秩的线性变换的估计函数，所谓的中间累积分布函数（CDF）在这一推导过程中发挥着重要作用；对于非线性变换，则需要并引入中间累积分布函数的推广。本文其余部分安排如下。第2节介绍一些符号说明后，我们在第3.2节推导所提出的针对右删失数据的R估计量推广，重点讨论Wilcoxon得分（即线性变换）的情形。在第3.2.2节，我们说明这一推广与另外两类重要的基于秩的估计量有直接联系。在第3.3节，我们利用这些结果推导参数估计的方差估计量。在第4 节，我们进一步拓展这一方法，使其能够使用秩的非线性变换。我们在第5节通过模拟研究所提出估计量的性能。最后，我们在第6节通过讨论总结全文。

2 符号说明

我们假设结局 Y_i^*的期望值是协变量 X_i组成的 p–向量的线性函数，具体由线性模型指定

$$Y_i^* = X_i^T \beta_0 + \epsilon_i. \tag{1}$$

在此模型中，“误差” ϵ_i 假设服从一个共同的未知累积分布函数 F。更常见的公式（1）可能会将 α 作为截距项，并假设 $E[\epsilon_i|X_i] = 0$；我们有意避免在（1）中引入截距，而是在本文余下部分将其视为干扰参数。因此， β 不包含截距参数，并且有 $E[\epsilon_i|X_i] = \int u dF$（$u$） $=\alpha$，其中 $\alpha \in \mathbb{R}$ 假设为有限值。

当 $Y_i^* =$log(T_i^*) 时，加速失效时间模型（AFT）对应于式（1），其中 $T_i^* > 0$表示失效时间。在这种情况下，条件生存函数为 $P\{T_i^* > u|X_i\} = F$(log $u - X_i^T \beta_0$)，其中 $\epsilon_i \sim F.$ 。失效时间数据通常会发生右删失；在这种情况下，我们还假设存在一个潜在的右删失时间 C_i^*，并假设只能观测到$T_i =$ min(T_i^*, C_i^*)、 $Y_i =$log(T_i) 和 $\Delta_i = I\{C_i^* \geq T_i^*\}$。我们对每个 i [施加标准的可识别条件 $T_i^* \perp C_i^*|X_i$（Kalbfleisch 和 Prentice， 2002]）。目前，我们不对 T^* 和 C^*.的分布施加任何连续性假设。

3 R估计与推断

3.1 简要回顾：完全数据情形

对于未删失数据，回归系数 β_0 在（1）中的取值可以通过所谓的 R-估计量 β来估计，该估计量定义为目标函数的任意一个极小值点

$$\sum_{i=1}^{n} a(R(\tilde{E}_{i,\beta}))\tilde{E}_{i,\beta}$$

其中 $a:\ [0,1] \to \mathbb{R}$ 是一个单调递增评分函数，而 $nR(E_{i,\ \beta})$ 表示该i^{th} “残差”的秩

$$\tilde{E}_{i,\beta} = Y_i - X_i^T\beta, \tag{2}$$

在所有 n β–相关残差 [例如，Jaeckel， 1972]。等价地，可以求解

$$\sum_{i=1}^{n} a(R(\tilde{E}_{i,\beta}))\left(X_i - \bar{X}\right) = 0 \tag{3}$$

其中 $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$；例如，参见 Hettmansperger [1984], Hettmansperger 和 McKean[2011]。重要的是，式（3）的解在任意位置平移下保持不变；因此，在模型（ ϵ_i1）下，对于任意 i 不一定具有均值为零这一事实，并不影响其有效性。

Jurečkovˊa [1971] 证明了式（3）在 β 中的渐近线性，从而为推断奠定了基础，适用于一般得分函数 $a(\cdot)$；Jaeckel [1972] 则证明了相应的 β渐近分布；这些结果的总结见 Hettmansperger [1984, 定理 5.2。3] 同时还对得分 $a(u) = u$ （即“Wilcoxon”得分函数，其位置和尺度已知，定义至该位置和尺度）进行了详细分析。在这种情况下， F 的一个简单而自然的估计量是估计残差E_i、$\hat{\beta}$、 $i = 1$、.....、 n. 的经验累积分布函数。相应地，当还假设 F具有对称性时，可以用 F 的中位数作为未知截距项 α的简单估计量（若需要）；事实上，即使不假设 F 对称， F 的中位数仍可作为位置估计量 [（Hettmansperger，1984年，第 5.2]节）。

3.2 将R估计扩展到右删失数据

3.2.1 估计 β_0：Wilcoxon得分的情形

正如引言中所述，人们已经提出了许多适用于右删失数据下加速失效时间模型的基于秩的估计量；该领域的重要论文包括Tsiatis [1990] 和Fygenson和Ritov [1994] ，相关的重要工作还包括但不限于 Prentice [1978], Cuzick[1985], Ritov [1990]和Jin等人 [2003]。对于右删失数据，基于秩的方法所面临的核心问题是，无法完全观测到删失观测的秩。通常，现有方法的区别在于各自如何处理这一问题。

Sen [1968] 通过考虑对 Kendall 的修正，提出了 β 在（1）条件下的估计量。该统计量为 τ。Fygenson 和 Ritov [1994]将这一方法直接推广到加速失效时间模型的估计中，由此得到一个单调估计函数，该函数仅涉及那些残差顺序可以明确判定（即可以确定秩）的残差对的贡献。若定义

$$\mathcal{D}_{i,j}(\beta) = \Delta_i I\{\tilde{E}_{i,\beta} < \tilde{E}_{j,\beta}\}, \tag{4}$$

其中$E_{i,\beta}$的定义与（2）中的定义相同，则可以证明，Fygenson 和 Ritov 估计量等价于求解

$$\sum_{i=1}^{n} \left(\mathcal{D}_{i,+}(\beta) - \mathcal{D}_{+,i}(\beta)\right)\left(X_i - \bar{X}\right) = 0, \tag{5}$$

其中，‘+’下标表示对该下标求和。在不存在删失且残差之间没有并列值时，估计函数（5）涉及考察每个残差的秩与反秩之间的 X–加权差异，并且可以证明，对于 Wilcoxon 得分，该函数可化简为（3） $a(u) = u$（即相差中心化和尺度常数）。然而，存在右删失时，该估计函数只涉及未删失残差之间的所有成对排序，且没有明显的对应概念可用于表示右删失样本中的“秩”。

这里，我们采用另一种方法，尝试通过为每个观测赋予一个秩，直接推广（3）；该秩与其在未删失试验中本应具有的秩直接相关。本节将重点讨论 Wilcoxon 得分，即 $a(u) = u$；更一般的秩变换将在第4节中讨论。我们的动机在于，对于未删失连续数据，“残差” i^{th} ϵ_i 的秩等于 $n\widehat{F}(\epsilon_i)$，其中 $\widehat{F}$ 表示 ϵ_i 的经验分布。对于右删失样本，可以类似地为未删失观测定义秩，只需将经验分布 F 替换为能够恰当处理删失数据的估计量，例如 Kaplan-Meier 估计量或自洽估计量 [（例如，见 Strawderman 和 Baer， 2023]）。然而，这一定义不能用于删失观测。相反，考虑到 $nF(\epsilon_i)$ 的总体对应量是 $nF(\epsilon_i)$，一个显而易见的选择是，对于任何在时刻 u 右删失的观测，使用 $E[nF(\epsilon_i)|\epsilon_i > u]$ ，这相当于对所有大于 u 的“残差”所对应的未观测秩取平均。

出于稍后会明了的原因，在我们的推导中使用中秩会很有帮助。具体而言，对于残差的累积分布函数 F，令相应的中间累积分布函数为

$$F^{\star}(y) = \frac{1}{2}\left[F(y) + F(y-)\right]; \tag{6}$$

类似地，给定一个合适的估计量 F，令 $F^{\star}$ 为根据（6）得到的相应代入估计量。在没有删失、但可能存在并列值的情况下，中秩可定义为 $nF^{\star}(\epsilon_i)$ [Ruymgaart， 1980]。按照上一段针对删失数据提出的方法，对于给定的 β，我们将右删失数据的插补秩（或插补中秩）定义为 $n\mathcal{R}(E_{i,\beta})$，其中

$$\widehat{\mathcal{R}}(\tilde{E}_{i,\beta}) \;=\; \Delta_i \widehat{F}^{\star}(\tilde{E}_{i,\beta}) + (1-\Delta_i)\int_{\tilde{E}_{i,\beta}}^{\infty} \widehat{F}^{\star}(y)\frac{d\widehat{F}(y)}{\widehat{S}(\tilde{E}_{i,\beta})}\;, \tag{7}$$

而 $S = 1 - F$ 是相应的估计生存函数。

在本文的其余部分，我们假设 S 是自洽估计量，例如 Strawderman 和 Baer， 2023]。对于 $t < T_{(n)}$ ，该自洽估计量等价于 Kaplan-Meier 估计量；对于 $t \geq T_{(n)}$ [，其值为零，例如 Strawderman 和 Baer，2023年，命题 2]；等价地说，它是在修改后的数据集中计算得到的 Kaplan-Meier 估计量，在该数据集中，所有满足 $T_i = T_{(n)}$ 的时间 $\Delta_i = 1$ （也就是说，不论其实际失效状态如何）。接下来，我们将假定所有经验表达式均采用这一修改后的数据集；这样做可确保与使用 F 及由其导出的量相兼容。通常，估计量 S，因而 F 和 $F^{\star}$，也取决于 β；为简化符号说明，在本小节余下部分中，我们省略这一依赖关系。

将 F 替换为 $F^\star$ 在（7）中，一个直接且有用的推论是

$$\int_{\tilde{E}_{i,\beta}}^{\infty} \widehat{F}^\star(y)\frac{d\widehat{F}(y)}{\widehat{S}(\tilde{E}_{i,\beta})} = \frac{1}{2}\frac{\widehat{F}^2(\infty)-\widehat{F}^2(\tilde{E}_{i,\beta})}{\widehat{S}(\tilde{E}_{i,\beta})} = \frac{1}{2}\left[1+\widehat{F}(\tilde{E}_{i,\beta})\right], \tag{8}$$

其中第二个等号成立，是因为 $F(\infty) = 1$ 对于自洽估计量成立；更多细节请参见附录。因此，(7) 可化简为易于计算的量

$$\widehat{\mathcal{R}}(\tilde{E}_{i,\beta}) = \Delta_i \widehat{F}^\star(\tilde{E}_{i,\beta}) + (1-\Delta_i)\left[1-\frac{1}{2}\widehat{S}(\tilde{E}_{i,\beta})\right]. \tag{9}$$

我们强调，$n\mathcal{R}(E_{i,\beta})$ 同时涉及自洽估计量 $S(\cdot)$ 以及使用相应中间累积分布函数计算得到的对应量 $F^\star(\cdot)$，其中 $F = 1 - S$. 重要的是，若将 $\widehat{}$ 替换为 $F^\star(\cdot)$, $F(\cdot)$，则这些结果不再如所述成立，这凸显了中间累积分布函数在这两个恒等式中的重要作用。下面的结果为 (7)，等价地，(9) 的总体层面行为提供了依据和进一步的见解。

定理 1 定义 F 为 ϵ的累积分布函数，并令

$$\mathcal{R}_{\beta_0} = \Delta F^\star(\tilde{E}_{\beta_0}) + (1-\Delta)\frac{1}{S(\tilde{E}_{\beta_0})}\int_{\tilde{E}_{\beta_0}}^{\infty} F^\star(u)dF(u),$$

其中 $\tilde{E}_{\beta_0} = Y - X^T\beta_0$ 且 Δ 分别是真实的观测残差和失效状态。于是

$$\mathcal{R}_{\beta_0} = \Delta(1-S^\star(\tilde{E}_{\beta_0})) + (1-\Delta)\left[1-\frac{1}{2}S(\tilde{E}_{\beta_0})\right], \tag{10}$$

以及 $E(\mathcal{R}_{\beta_0} \mid X) = E(\mathcal{R}_{\beta_0}) = 1/2$. 当 F 是连续变量时，还可以证明 th at

$$Var(\mathcal{R}_{\beta_0}|X) = \frac{1}{12}\left(1-E(S^3(E_{\beta_0})|X)\right),$$

其中 $E_{\beta_0} = \log C^* - X^T\beta_0$，此时 $Var(\mathcal{R}_{\beta_0}) = \frac{1}{12}(1-E(S^3(E_{\beta_0}))$。

对于一组未删失残差，且 F连续， $F(E_{\beta_0,\,i})$ 在 (0，1)上均匀分布。因此，定理 1 的一个推论是，$\mathcal{R}_{\beta_0,\,i}$ 的前两个矩与均匀随机变量的前两个矩非常接近，只要删失程度不是太高。接下来我们将证明，当使用经验中间累积分布函数 $F^\star$ 来定义 (9) 时，插补秩之和是一个与 β无关的常数。

T定理 2 当使用 F 的自洽估计量且 $\|\beta\| <$ ∞,

$$\sum_{i=1}^{n}\widehat{\mathcal{R}}(\tilde{E}_{i,\beta}) = \frac{n}{2}. \tag{11}$$

定理 2 表明，中秩之和，即 $\sum_{i=1}^n n\widehat{\mathcal{R}}(\tilde{E}_{i,\,\beta})$，等于 $n^2/2$，与 β. 无关。该结果的证明是定理5的一个特例，后文将在非线性 $a(\cdot)$的情形下给出，因此暂且略去。同样，若在（ $F^\star(\cdot)$ 中用 $F(\cdot)$ 替换（9），则该结果不再如上述表述那样成立。

结合（3）及上述推导，我们可以通过“求解”方程来推广（3）$\Psi_n(\beta)=0$，其中

$$\Psi_n(\beta)=\sum_{i=1}^{n}\widehat{\mathcal{R}}(\tilde{E}_{i,\beta})(X_i-\bar{X}); \tag{12}$$

等价地，根据定理 2，我们可以使用

$$\Psi_n^{(c)}(\beta)=\sum_{i=1}^{n}\Big(\widehat{\mathcal{R}}(\tilde{E}_{i,\beta})-\frac{1}{2}\Big)X_i$$

代替 $\Psi_n(\beta)$。由于 $\Psi_n(\beta)$ 的不连续性，可能不存在精确解；在这种情况下，可以将解 β 定义为 $\Psi_n(\beta)$ 的任意零点穿越。由此得到的估计量直接将求解 Wilcoxon得分 $a(u)=u$ （或等价地，均值为零的分数 $a(u)=u-1/2$）的方程 (3) 的 R估计量推广到右删失数据情形，并为求解方程 (5) 的 Fygenson 和 Ritov 估计量提供了一种有趣的替代方案。关于如何在实践中估计 β 的进一步讨论见第3.3节。

3.2.2 （(12) 与 Ritov 和 Tsiatis 类之间的联系

设 $\widehat{S}_\beta(\cdot)$ 为由 $(\Delta_i, \tilde{E}_{i,\beta})$ $i=1,\ldots\ldots n$. 得到的 S 的自洽估计量。在模型 (1) 下，若 Y 可能被右删失，Ritov（1990年）提出通过（近似地）求解 $\mathcal{S}_{n,R}(\beta,\gamma)=0$, 来估计 β，其中

$$\mathcal{S}_{n,R}(\beta,\gamma)=\sum_{i=1}^{n}(X_i-\bar{X})\left[\Delta_i\gamma(\tilde{E}_{i,\beta})+(1-\Delta_i)\frac{1}{\widehat{S}_\beta(\tilde{E}_{i,\beta})}\int_{\tilde{E}_{i,\beta}}^{\infty}\gamma(u)d\widehat{F}_\beta(u)\right] \tag{13}$$

其中，给定函数 $\gamma(\cdot)$ 满足某些正则条件。选择 $\gamma(u)=u$即可得到 Buckley-James 估计量；当 ϵ_i 具有良好光滑的分布函数时，在最大似然框架下可得到该函数的最优选择，其形式为$\gamma_{opt}(u)=\frac{d}{du}\log \underline{f}(u)$，其中 f是与 F相对应的概率密度函数（PDF）。当 F 为逻辑斯蒂分布时，取$\gamma(u)=F(u)$ 即为最优选择。

Ritov [1990]证明，类（13）中的任何估计函数都等价于 Tsiatis（1990年）提出的加权 logrank（或线性秩）估计方程类中的一个成员；相关结果见Cuzick [1985], ，而对于线性秩检验的情形，相关结果由Mehrotra 等人 [1982]给出。Tsiatis 估计方程类的形式为

$$\mathcal{S}_{n,T}(\beta;\gamma)=\sum_{i=1}^{n}\int_{-\infty}^{\infty}w(u;\beta)(X_i-\bar{X}(u;\beta))dN_i(u;\beta), \tag{14}$$

其中 $N_i(u;\beta)=\Delta_i I\{\tilde{E}_{i,\beta}\le u\}$, 且对于 $Y_i(u;\beta)=I\{\tilde{E}_{i,\beta}\ge u\}$,

$$\bar{X}(u;\beta)=\frac{\sum_{j=1}^{n}X_jY_j(u;\beta)}{\sum_{j=1}^{n}Y_j(u;\beta)}.$$

Ritov [1990](另见 Andersen 等人 [1993], 第 582–583 页）建立了（13）与（14）之间的直接联系；特别是，当.....时，两者在数学上等价

$$w(u;\beta)=\gamma(u)-\frac{1}{\widehat{S}_\beta(u)}\int_u^{\infty}\gamma(u)\,d\widehat{F}_\beta(u). \tag{15}$$

^比较（12）和（13），由（7）可立即看出，在（13）中令 $\gamma(u) = F^{\star}_{\beta}(u)$ ^即可得到（12）；也就是说，$\mathcal{S}_{n,R}(\beta, F^{\star}_{\beta}(\cdot)) = \Psi_n(\beta)$。将 $\gamma(u) = F^{\star}_{\beta}(u)$ 代入（15），并使用附录中的（42），也很容易证明

$$w(u;\beta) = \widehat{F}^{\star}_{\beta}(u) - \frac{1}{\widehat{S}_{\beta}(u)} \int_{u}^{\infty} \widehat{F}^{\star}_{\beta}(r)\, d\widehat{F}_{\beta}(r) = -\frac{1}{2}\widehat{S}_{\beta}(u-).$$

综合上述结果，可得以下一组等价关系： $\Psi_n(\beta) = \mathcal{S}_{n,R}(\beta; \widehat{F}^{\star}_{\beta}(\cdot)) = \mathcal{S}_{n,T}(\beta; -\frac{1}{2}\widehat{S}^{\star}_{\beta}(\cdot-))$；也就是说，

$$\Psi_n(\beta) \equiv -\frac{1}{2}\sum_{i=1}^{n} \int_{-\infty}^{\infty} \widehat{S}_{\beta}(u-)(X_i - \bar{X}(u;\beta))dN_i(u;\beta), \tag{16}$$

^在此情形下，为了满足数学等式，必须使用负号；当 $\gamma(u) = F^{\star}_{\beta}(u)$，但除此之外，负号并不重要。特别地，将 R估计量推广至右删失数据的（12）是 Ritov 和 Tsiatis 所研究的一类估计方程的具体示例。这些结果，尤其是数学等价性，很大程度上依赖于^使用中间累积分布函数 $F^{\star}_{\beta}(\cdot)$。

^选择 $\gamma(u) = F^{\star}_{\beta}(u)$ 不满足 Ritov 在 $\gamma(\cdot)$ 中提出的所有假设。因此，为求完整起见，我们在下文将（12）与（16）之间的等价关系陈述为一个定理，并在附录的A.2 节给出独立的代数证明。

定理3 估计函数 $\Psi_n(\beta)$ 在 (12) 中与 (16)中给出的一致。

3.2.3 渐近理论：简要总结

(12) 与 (16) 等价，因此可以相对直接地推导渐近结果，为所提出的 R估计量提供推断依据，其中包括若干种估计参数渐近方差的估计方法。假设 $\widehat{\beta}$满足 $n^{-1}\Psi_n(\widehat{\beta}) = o_P(n^{-1/2})$[，例如 Huang、 2013], ，并且 Jin 和 Ying [2004] 中给出的正则条件成立；另见附录 A.3。这些条件包括一项假设，该假设意味着 F 绝对连续，且具有可微的风险函数 $\lambda_F(\cdot)$。为避免过于技术性的推导，我们还假设 (T_i, Δ_i, X_i) $i = 1$, ..., n 独立同分布。定义

$$\Sigma = \frac{1}{4}\int_{-\infty}^{\infty} S^2(u-)\mathcal{H}(u)\Theta_0(u)d\Lambda_F(u), \tag{17}$$

and

$$\Xi = \frac{1}{2}\int_{-\infty}^{\infty} S(u-)\mathcal{H}(u)\Theta_0(u)\frac{\lambda'_F(u)}{\lambda_F(u)}d\Lambda_F(u)\ , \tag{18}$$

其中， $\Lambda_F(t) = \int_0^t \lambda_F(u)du$ 是 $F(\cdot)$ 的累积风险函数，

$$\mathcal{H}(u) = \frac{\Theta_2(u)}{\Theta_0(u)} - \left(\frac{\Theta_1(u)}{\Theta_0(u)}\right)^{\otimes 2}$$

and

$$\Theta_j(u) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} E[X_i^{\otimes j} Y_i(u,\beta_0)]\,.$$

然后，在所示条件以及一个额外条件下（当$w(u;\beta) = -\widehat{S}_\beta(u-)/2$(例如，参见附录 A.3），我们有 $n^{-1/2}\Psi_n(\beta_0) \xrightarrow{d} N(0,\Sigma)$ 当 $n \to \infty$；因此，Σ 是 $n^{-\frac{1}{2}}\Psi_n(\beta_0)$ 的渐近方差-协方差矩阵。假设 Ξ 为正定矩阵，我们还可得

$$\sqrt{n}(\widehat{\beta} - \beta_0) \xrightarrow{d} N(0, \Xi^{-1}\Sigma\Xi^{-1}).$$

尽管假设的 $F(\cdot)$ 连续性意味着 $S(u-)$ 可以替换为 $S(u)$，且 $d\Lambda_F(u)$ 可以替换为 $\lambda_F(u)du$，但为便于后续使用，我们将采用这些更一般的表达式。

3.2.4 Σ 的分解

矩阵 Σ 在（17）中具有熟悉的形式，它是直接利用鞅论证由（16）推导得到的。然而，对于将（ Σ 视为 R 估计方程时，其与（12）的渐近方差之间的关系则不太明确；人们可能会预期，该方差更直接地取决于秩的方差与协变量方差 X的乘积。下面的结果给出了 Σ的一个有信息量的分解，从而进一步阐明了这种联系。

定理 4 假设上一节的正则条件成立，并定义 μ_X 和 Σ_X 分别为 X 的边际均值和协方差矩阵。则 $\Sigma = \Sigma_1 - \Sigma_2$，其中半正定矩阵 Σ_i 分别为

$$\Sigma_1 = \frac{1}{4}\int_{-\infty}^{\infty} E\big[I\{E_{\beta_0} \geq u\}(X - \mu_X)^{\otimes 2}\big]S^2(u)dF(u)$$

and

$$\Sigma_2 = \frac{1}{4}\int_{-\infty}^{\infty} (\mu_{X|E_{\beta_0}}(u) - \mu_X)^{\otimes 2}P\{E_{\beta_0} \geq u\}S^2(u)dF(u),$$

其中 $\mu_{X|E_{\beta_0}}(u) = \Theta_1(u)/\Theta_0(u) = E[X|E_{\beta_0} \geq u]$. 此外，

$$\Sigma_1 = \frac{\Sigma_X}{12} - \frac{1}{12}E\big[(X - \mu_X)^{\otimes 2}S^3(E_{\beta_0})\big]. \tag{19}$$

分解式 $\Sigma = \Sigma_1 - \Sigma_2$ 的推导，正如定理中前两个展示公式所给出的那样，是直接的。另一种表达式（19）关于 Σ_1 的推导则不那么直观，并运用了分部积分。

有关更多细节和评论，请参阅附录的第A.4 节；其中还证明了，只要将 $\widehat{\mathcal{R}}(\tilde{E}_{i,\beta_0})$ 替换为 $R_{\beta_0,i}$，Σ_1 就等于 $n^{-1/2}\Psi_n(\beta_0)$ 的渐近方差。正如该处所示，对 Σ_1 的这种解释直接源于以下观察：对于每个 $i.$ ， X_i和 $R_{\beta_0,i}$ 不相关。综合而言，分解式$\Sigma = \Sigma_1 - \Sigma_2$ 中出现 Σ_2 ，是因为在定义 $n^{-1/2}\Psi_n(\beta_0)$ 时，需要使用 $\widehat{F}_{\beta_0}(\cdot)$ 来替代 $F(\cdot)$。有趣的是，由于 Σ_2 通常是正定的，这种替换会降低渐近方差。

最后，在既没有删失也没有并列观测的情况下，注意到 Σ_1 化简为$\Sigma_X/12$，这就是在3式于 $\beta = \beta_0$ 且 $a(u) = u$ [时的渐近方差（例如，Hettmansperger，1984年，问题5.5）。3]。

3.3 R估计与推断的实践应用

对 β_0 的推断通常基于第3.2.3节总结的渐近结果。具体而言，只要有适当的 β_0 估计值，就可以按常规方式构造其 Wald型置信区间和检验

；给定 β_0 和 Ω 的适当估计值即可；相反，对 β_0 进行准得分检验只需要对 Σ.作出适当估计

除了准得分检验之外，在使用此类结果进行^估计和推断时，还存在一些众所周知的挑战，主要源于：(i) 需要找到一个解 β ，该解位于 $n^{-1/2}$–的邻域内 β_0 ，这是由于 $\Psi_n(\beta$缺乏单调性；以及 (ii) 在 (λ_F'/λ_F 18) 中存在，而通常很难对其进行良好估计。我们将在接下来的章节讨论这两个问题；在此之前，先考虑 Σ的估计和准得分检验。

3.3.1 估计 Σ和准得分检验

准得分检验基于 $\Psi_n(\beta)$，如式 (16) 所示（等价地，式 (12)），等价于一种删失数据线性秩检验，该检验使用 Wilcoxon得分；由于其对估计的要求较少，因此可能优于 Wald 检验。对于 H_0 : $\beta_0 = 0$，准得分检验统计量满足$n^{-1}[\Psi_n(0)]^T\Sigma^{-1}\Psi_n(0) \to \chi^2_p$ 在 H_0 下；此外，当 ϵ 服从逻辑斯蒂分布时，该检验在局部上也具有最大检验效能；更多讨论见 Prentice [1978]。在此检验中， Σ 可以用^替代 Σ(0)，它通过以下针对 Σ的代入估计量定义：

$$\widehat{\Sigma}(\beta) = \frac{1}{4}\int_{-\infty}^{\infty} \widehat{S}^2_\beta(u-)\widehat{\mathcal{H}}(u;\beta)\widehat{\Theta}_0(u;\beta)d\widehat{\Lambda}_\beta(u) \tag{20}$$

其中

$$\widehat{\mathcal{H}}(u;\beta) = \frac{\widehat{\Theta}_2(u;\beta)}{\widehat{\Theta}_0(u;\beta)} - \left(\frac{\widehat{\Theta}_1(u;\beta)}{\widehat{\Theta}_0(u;\beta)}\right)^{\otimes 2},$$

$$\widehat{\Theta}_j(u;\beta) = \frac{1}{n}\sum_{i=1}^{n} X_i^{\otimes j} Y_i(u,\beta),$$

and

$$\widehat{\Lambda}_\beta(u) = \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{u} \frac{dN_i(s,\beta)}{\widehat{\Theta}_0(s;\beta)}. \tag{21}$$

公式（20）可大幅简化：

$$\widehat{\Sigma}(\beta) = \frac{1}{4n}\sum_{i=1}^{} \Delta_i[\widehat{S}_\beta(\tilde{E}_{i,\beta}-)]^2\widehat{\mathcal{H}}(\tilde{E}_{i,\beta};\beta). \tag{22}$$

^ ^ ^ ^可以使用 $\Sigma = \Sigma(\beta)$ 代替 Σ(0)，来检验以下全局零假设：$\beta_0 = 0$，但计算负担会有所增加。更一般地说，可以使用 β_0 的假设约束估计量构造复合假设的得分型检验。

3.3.2 估计 β_0

估计 β_0 实际上涉及两个相关问题。一个问题是在所示的 $n^{-1/2}$–邻域中找到 β_0的一个解；文献中通常利用一个众所周知的事实来处理这一问题：Gehan加权对数秩估计函数对应于凸目标函数的（次）梯度，并且（在某些正则条件下）保证^ ^能够得到一个 $\sqrt{n}$–一致的初始估计量 $\beta_{G^\cdot}$。找到后， β_G 可用作起点，以确定一个 $\widehat{\beta}$，使得 $n^{-1}\Psi_n(\widehat{\beta}) = o_P(n^{-1/2})$。另一个问题是实际求出这样的^一个 β，也就是说，找到一个使（12）“等于”零的解。这个求根问题具有挑战性，主要是因为

（12）作为 β的函数并不光滑。事实上，即使给定 β_G，标准求根软件包也可能无法很好地执行，因为得到的解通常只是（12）的一个“零点穿越”，而非精确的零点。

针对本文后续开展的模拟，我们设计并采用了一种新的序贯单变量方程求解器，该求解器反复使用简单的二分算法；该算法能够轻松识别任何一维求根问题中的零点穿越。在大量模拟研究中，使用 $\beta = 0$ 作为初始值估计参数时，没有遇到任何困难。这并不是说一般情况下不必或不宜使用 β_G ，只是我们最终发现，所考虑的问题并不要求我们先确定，从而增加总体 β_G计算负担。在需要或有益于使用 β_G 的情况下，目前已有许多R 软件包可以计算或近似求得 β_G 。如果需要，相关的“诱导平滑”方法也可以显著减轻在较大问题中确定 β_G 所带来的计算负担；例如参见Brown 和 Wang [2005] 以及Johnson 和 Strawderman [2009]。Huang [2013] 还提供了另一种方法。

3.3.3 估计Ω、Wald 检验和置信区间

估计 $\Omega = \Xi^{-1}\Sigma\Xi^{-1}$（ $\sqrt{n}(\widehat{\beta} - \beta_0)$ 的渐近协方差矩阵）比估计 Σ 更具挑战性，因为“斜率”矩阵 Ξ 无法轻易通过对非光滑估计函数 $\Psi_n(\beta)$ 求一阶导数得到。Zeng 和 Lin [2008] 提出了一种较新的蒙特卡罗最小二乘法，用于估计 Ω 并规避这一问题；另见 Strawderman [2005], ，他较早提出了一个变体，将其应用于加速失效时间模型中的复发事件。这两种方法的一个重要特点是：只需要 $\Sigma = \Sigma(\beta)$，并且一旦计算出来，便可通过结合蒙特卡洛抽样与线性回归，利用 $\Psi_n(\beta)$ 的渐近线性，进而得到估计量 Ξ 。具体而言，可以按如下方式估计 Ξ，从而估计 Ω:

1. 生成 $\beta_b = \beta + Z_b/\sqrt{n}$ ，其中 Z_b 是一个均值为零、协方差矩阵由用户指定的 $p-$向量 $\mathcal{D}_Z$， $b = 1, \ldots, B$。2. 计算 $n^{-1/2}\Psi_n(\tilde{\beta}_b)$ $b = 1, \ldots, B$。3. 分别用各自的样本均值对 $n^{-1/2}\Psi_n(\tilde{\beta}_b)$和 $\tilde{\beta}_b - \widehat{\beta}$ 进行中心化，通过将前者对后者进行回归来计算 $\tilde{\Xi}$ ，并用 $\widehat{\Xi} = 0.5(\tilde{\Xi} + \tilde{\Xi}^T)$估计 Ξ 。4. 通过 $\widehat{\Xi}^{-1}\widehat{\Sigma}\widehat{\Xi}^{-1}$.估计Ω，即 $\sqrt{n}(\widehat{\beta} - \beta_0)$的协方差矩阵，

–

这种方法的主要优点是，无需反复求解一个非光滑、可能是非单调的估计方程的零点。其主要缺点是，可能会引入一定的不稳定性，因为它增加了一个额外的变异性来源（即蒙特卡洛近似误差），而该误差在一定程度上取决于用户指定的方差矩阵 $\mathcal{D}_Z$. Zeng 和 Lin[2008] 没有具体说明如何最佳地选择 $\mathcal{D}_Z$. 在实践中，重要的是选择 $\mathcal{D}_Z$ 使得 β_b, $b = 1, \ldots, B$ 通常落在 β附近的区域内，其中 $\Psi_n(\beta)$ 近似为线性。尽管渐近地可以保证这一点，但第 3 步中得到的估计量 Ξ 未必总能表现良好，尤其是在样本量较小或 $\mathcal{D}_Z$ 选择不当时。

一种估计 Ω 的替代方法可以通过采用 Huang [2002]提出的“逆数值微分”方法推导得出。与上述方法类似，Huang 的方法只需要能够获得 Σ，并利用渐近线性来证明其合理性。不过，该方法不依赖蒙特卡洛抽样，而是需要 p 额外的求根步骤。下面，

我们对 Huang 的原始方法提出一种小幅变体。设 $n^{-1}\widehat{\Sigma}$ 的对称平方根为 $\widehat{\mathcal{C}}$。对于 $\widehat{\mathcal{C}}$ 的每一列，我们首先求出一个满足下式的“正”解 $\widehat{\beta}_u^{(k)}$

$$n^{-1}\Psi_n(\beta_u^{(k)}) = \widehat{\mathcal{C}}_k \tag{23}$$

然后确定相应的“负”解 $\widehat{\beta}_l^{(k)}$，使其满足

$$n^{-1}\Psi_n(\beta_l^{(k)}) = -\widehat{\mathcal{C}}_k. \tag{24}$$

由于 $\widehat{\mathcal{C}}_k$ 是 $O_p(n^{-\frac{1}{2}})$，而 $n^{-1}\Psi_n(\widehat{\beta})$ 是 $o_p(n^{-1/2})$，因此解 $\widehat{\beta}_u^{(k)}$ 和 $\widehat{\beta}_l^{(k)}$ 都应接近 $\widehat{\beta}$，并且可以用相对高效的数值方法确定。接下来，我们构造矩阵 $\widehat{\mathcal{B}}$，其第 k 列定义为 $(\widehat{\beta}_u^{(k)} - \widehat{\beta}_l^{(k)})/2$，其中 $k = 1, \ldots\ldots, K.$。Huang 第3.2节中的结果 [2002] 表明，$\widehat{\mathcal{B}}$ 能够一致估计 $n^{-1/2}\Xi^{-1}\Sigma^{1/2}$；因此，可以通过以下方式估计 Ω：

$$\widehat{\Omega} = n\widehat{\mathcal{B}}\widehat{\mathcal{B}}^T. \tag{25}$$

如果（23）无解，但（24）可解，我们就将 $\widehat{\beta} - \widehat{\beta}_l^{(k)}$ 用于 k列，即 $\widehat{\boldsymbol{B}}$的第 列；类似地，如果（23）可解，但（24）无解，我们就使用 。

后在文将会看到，这在我们的模拟中表现良好。部分原因是，$\widehat{\mathcal{C}}_k$ 为 $n^{-1}\Psi_n(\beta)$ 定义了恰当的偏差尺度；求解 $n^{-1}\Psi_n(\beta^{(k)}) = \pm\widehat{\mathcal{C}}_k$ 关于 $\beta^{(k)}$ 的方程，便可隐式确定恰当缩放的偏差 $\beta^{(k)} - \widehat{\beta}$. 不过，重要的是，计算 Ω（以及 β）会随着参数数量的增加而变得越来越繁琐（也就是说，该方法需要反复使用求根程序）。对于参数较多的数据，只要能够估计 β_0 并选择恰当的 $\mathcal{D}_Z.$，上文概述的蒙特卡洛方法在计算上或许更有优势。此前提到的 Strawderman [2005],提出的这种蒙特卡洛方法变体，通过在 $\widehat{\beta}_G$ 的邻域内进行抽样，而非在由任意矩阵 β 所定义的 $\mathcal{D}_Z$.邻域内抽样，为解决这一问题提供了一种可能的方案；该方法使用的是 的估计协方差矩阵。

结定可以按常规方法构造具有渐近正态分布或 χ^2 分布的 Wald 检验。例如，H_0：可以使用 $n(\widehat{\beta} - \beta_0)^T\widehat{\Omega}^{-1}(\widehat{\beta} - \beta_0)$检验 $\beta = \beta_0$，其具有渐近 χ_p^2 分布。

4 将 R估计量推广到一般得分 s

前几节的结果聚焦于线性函数 $a(u) = u$ 其中 $u \in [0, 1]$. 许多作者此前已经考虑过使用式（3）估计 β 的问题，其中 $a(\cdot)$ 取非线性形式；例如，对于未删失数据，可参见 Hettmansperger [1984], Hettmansperger 和 McKean [2011] 。在线性秩检验的背景下，Prentice [1978] 考虑了 $a(\cdot)$ 的若干种可能选择，包括 Wilcoxon 得分，以及由极值分布导出的所谓“logrank”得分；更一般地，对于某个假设的概率密度函数 $f_0(\cdot)$（即，对于 ϵ）及其对应的累积分布函数 $F_0(\cdot)$ [例如 Lai 和 Ying, 1992], ，可以定义 $a(u) = -f_0'(F_0^{-1}(u))/f_0(F_0^{-1}(u))$；或者也可以通过 $S_0(\cdot) = 1 - F_0(\cdot)$ [例如 Prentice, 1978, Cuzick, 1985]进行等价参数化。当 F_0 且 f_0 对应于ϵ. 的真实分布时，这一特定的得分函数在渐近意义下是最优的。下面，我们将说明如何把前几节的结果扩展到右删失结局的情形。我们首先考虑有界得分函数的情形；然后讨论得分函数可能无界的情形。

4.1 有界得分函数

设 $a: [0,1] \to \mathbb{R}$ 为给定的连续可微单调递增（可能为非线性）得分函数，并定义 $A(u) = \int_0^u a(s)ds$ 对于 $u \in (0,1]$. 注意， $a(u) = u$ 意味着 $A(u) = u^2/2$， $\widehat{\ }$ $\widetilde{\ }$这对应于通过式（9）定义的中秩 $n\mathcal{R}(E_{i,\beta})$。更一般地，如果定义$a(u)$，则使用

$$a(u) = -\frac{f_0'(F_0^{-1}(u))}{f_0(F_0^{-1}(u))} \tag{26}$$

对于用户指定的 $f_0(\cdot)$，很容易证明

$$A(u) = f_0(F_0^{-1}(0)) - f_0(F_0^{-1}(u)) \tag{27}$$

因此，只要 $F_0^{-1}(\cdot)$ 可直接获得， $A(u)$ 就很容易计算。在本节中，假设 $a(u)$ 对于 $u \in [0,1]$ 有界；无界情形的讨论留待下一小节。

(9) 的简洁形式直接依赖于中秩分布的使用，后者对于简化（7）中出现的积分项以及定理 2中的秩和保持结果尤为关键。因此，很自然会问，$\widehat{\ }$中间累积分布函数 $F^\star(\cdot)$ 是否在将（7）推广到非线性函数$a(\cdot)$ 的情形时发挥类似的核心作用；也就是说，是否

$$\Delta_i a(\widehat{F}_\beta^\star(\tilde{E}_{i,\beta})) + (1-\Delta_i)\int_{\tilde{E}_{i,\beta}}^\infty a(\widehat{F}_\beta^\star(y))\frac{d\widehat{F}_\beta(y)}{\widehat{S}_\beta(\widetilde{E}_{i,\beta})},$$

确实是（7）最恰当的推广。这里的答案是否定的，至少就保持精确对应关系（而非渐近对应关系）而言如此；尽管该表达式定义良好，但由于

$$\int_{\tilde{E}_{i,\beta}}^\infty a(\widehat{F}_\beta^\star(y))\frac{d\widehat{F}_\beta(y)}{\widehat{S}_\beta(\widetilde{E}_{i,\beta})} \neq \frac{A(1) - A(\widehat{F}_\beta(\widetilde{E}_{i,\beta}))}{\widehat{S}_\beta(\widetilde{E}_{i,\beta})}.$$

$\widehat{\ }$一般而言，这个积分等式并不成立，因为 $F_\beta(\cdot)$ 只有广义逆。

因此可能会有人问，是否存在某个函数，例如 $\widehat{F}_{a,\beta}^\star(t)$，使得

$$\widehat{\mathcal{R}}^{(a)}(\tilde{E}_{i,\beta}) = \Delta_i a\big(\widehat{F}_{a,\beta}^\star(\tilde{E}_{i,\beta})\big) + (1-\Delta_i)\frac{1}{\widehat{S}_\beta(\tilde{E}_{i,\beta})}\int_{\tilde{E}_{i,\beta}}^\infty a\big(\widehat{F}_{a,\beta}^\star(s)\big)\,d\widehat{F}_\beta(s) \tag{28}$$

$$= \Delta_i a\big(\widehat{F}_{a,\beta}^\star(\tilde{E}_{i,\beta})\big) + (1-\Delta_i)\left[\frac{A(1) - A(\widehat{F}_\beta(\tilde{E}_{i,\beta}))}{\widehat{S}_\beta(\tilde{E}_{i,\beta})}\right]; \tag{29}$$

注意，此处等号成立的条件是

$$\int_t^\infty a\big(\widehat{F}_{a,\beta}^\star(s)\big)\,d\widehat{F}_\beta(s) = A(1) - A(\widehat{F}_\beta(t)). \tag{30}$$

$\widehat{\ }$若将 $F_\beta(t)$ 替换为关于 t的连续函数，则很容易证明等式 (30) 成立[例如，Falkner 和 Teschl, 2012]。更一般地，我们将在下文证明，阶跃函数 $F_\beta(\cdot)$ 也$\widehat{\ }$ $\widehat{\ }$存在肯定的答案。特别是，对每个 t，令 $F_{a,\beta}^\star(t)$ 由下列等式定义

$$a(\widehat{F}_{a,\beta}^\star(t)) = \frac{A(\widehat{F}_\beta(t)) - A(\widehat{F}_\beta(t-))}{\widehat{F}_\beta(t) - \widehat{F}_\beta(t-)}; \tag{31}$$

由中值定理可知，$\widehat{F}^{\star}_{a,\beta}(t)$ 保证存在唯一解，因为 $a(\cdot)$ 是连续且单调的。现在很容易看出，(30) 可简化为：

$$\int_t^{\infty} a\big(\widehat{F}^{\star}_{a,\beta}(s)\big)\, d\widehat{F}_{\beta}(s) = \sum_{s>t} a\big(\widehat{F}^{\star}_{a,\beta}(s)\big) \times \big(\widehat{F}_{\beta}(s) - \widehat{F}_{\beta}(s-)\big) = \sum_{s>t} \big(A(\widehat{F}_{\beta}(s)) - A(\widehat{F}_{\beta}(s-))\big),$$

最后一个求和项通过望远镜求和化简为 $A(1) - A(\widehat{F}_{\beta}(t))$，因为自洽估计量满足 $\widehat{F}_{\beta}(\infty) = 1$. 因此，要保持（30）成立，就需要使用 $\widehat{F}^{\star}_{\beta}(t)$ 的一种特定推广，而这种推广取决于 $a(\cdot)$ 的选择。幸运的是，实际应用（29）时，无须显式计算 $\widehat{F}^{\star}_{a,\beta}(t)$，因为可以直接利用（31）中给出的关系。

以上推导的结果表明，现在可以通过求解 $\Psi_n(\beta; a) = 0$，来估计 β，其中

$$\Psi_n(\beta; a) = \sum_{i=1}^{n} \widehat{\mathcal{R}}^{(a)}(\tilde{E}_{i,\beta})(X_i - \bar{X}) \tag{32}$$

以及（29）和（31）结合可得

$$\widehat{\mathcal{R}}^{(a)}(\tilde{E}_{i,\beta}) = \Delta_i \left(\frac{A(\widehat{F}_{\beta}(\tilde{E}_{i,\beta})) - A(\widehat{F}_{\beta}(\tilde{E}_{i,\beta}-))}{\widehat{F}_{\beta}(\tilde{E}_{i,\beta}) - \widehat{F}_{\beta}(\tilde{E}_{i,\beta}-)} \right) + (1 - \Delta_i) \left[\frac{A(1) - A(\widehat{F}_{\beta}(\tilde{E}_{i,\beta}))}{\widehat{S}_{\beta}(\tilde{E}_{i,\beta})} \right].$$

容易证明，$\widehat{\mathcal{R}}^{(a)}(\tilde{E}_{i,\beta})$ 在 $a(u) = u$ 且 $A(u) = u^2/2$ 时可化简为（7），因此（32）可化简为（12）。

上述推导依赖于自洽估计量 $F_{\beta}(\cdot$的两个特征：(i) 它是一个阶梯函数；(ii) 对每个有限的 β，它都会导出一个适当的离散累积分布函数。现在我们证明，这些论证的关键要素可推广至任意正规的累积分布函数 $H(\cdot)$。这些推导本身颇有意义，同时也为推广定理 1 和 2奠定了有益基础。令 $U \sim \text{Unif}(0,1)$，并对任意 t

$$\gamma_a(t; H) = E\left[a(UH(t) + (1-U)H(t-))\right]$$

其中，期望是对 U 取的。函数 $\gamma_a(t; H)$ 初看可能不太直观。不过，以下观察显而易见：

- 当 $H(\cdot)$ 连续且 $a(\cdot)$ 非线性时，我们有 $\gamma_a(t; H) = a(H(t))$;
- 对于一般的 $H(\cdot)$ 和 $a(u) = u$，我们有 $\gamma_a(t; H) = H^{\star}(t)$，其中 $H^{\star}(\cdot)$ 是相应的中间累积分布函数；如果 $H(\cdot)$ 也连续，则 $\gamma_a(t; H) = H(t)$;

因此，在若干重要应用场景中，$\gamma_a(t; H)$ 会简化为 $a(H^{\star}(t))$，但一般而言，它与 $a(H^{\star}(t))$ 并不相同。值得注意的是，$\gamma_a(t; H)$ 也有简单的闭式表达：对于每个固定的$H(t)$ 和 t,

$$\gamma_a(t; H) = \int_0^1 a(sH(t) + (1-s)H(t-))ds = \frac{A(H(t)) - A(H(t-))}{H(t) - H(t-)}, \tag{33}$$

其中，右侧应解释为等于 $a(H(t))$，当 $H(\cdot)$ 在 t处连续时。对于特定的选择 $H(t) = F_{\beta}(t)$，我们有 $\gamma_a(t; F_{\beta}) = a(\widehat{F}^{\star}_{a,\beta}(t))$，如式（31）所示，这也

^化为中间累积分布函数 $F^{\star}_{\beta}(t)$，当 $a(u) = u$. 因此，式（33）将式（31）推广至任意正规累积分布函数 H。关键的是，若随后定义

$$\Gamma_a(t; H) = \frac{1}{1 - H(t)} \int_t^{\infty} \gamma_a(s; H)\, dH(s), \tag{34}$$

M´arquez Alb´es 和 Slav´ık [2024, 定理 7.2] 中关于斯蒂尔切斯积分的链式法则和代换法则的推广进一步意味着 $\int_t^{\infty} \gamma_a(s; H) dH(s) = A(1) - A(H(t))$，这是对 (30) 的推广。因此，对于满足 $H(t) < 1$ 的 t，

$$\Gamma_a(t; H) = \frac{A(1) - A(H(t))}{1 - H(t)}; \tag{35}$$

因为 $a(\cdot)$ 有界且连续可微，可以证明 (35) 收敛至 $a(1)$，当 $H(t) \to 1$.

定理 2 给出了关于 (9) 之和的一个简洁结果，等价地也适用于 (7)。下面的结果利用 (33) 以及 (34) 与 (35) 的等价性，将定理1和2 推广到非线性 $a(\cdot)$的情形。

定理 5 定义 F为 ϵ的累积分布函数，并令 $\mathcal{R}^{(a)}_{\beta_0} = \Delta\gamma_a(\tilde{E}_{\beta_0}; F) + (1 - \Delta)\Gamma_a(\tilde{E}_{\beta_0}; F)$ 其中 $\tilde{E}_{\beta_0} = Y - X^T\beta_0$ 和 Δ 分别是真实的观测残差和失效状态， F 是 ϵ =的对数 累积分布函数。于是， $E(\mathcal{R}^{(a)}_{\beta_0} \mid X) = E(\mathcal{R}^{(a)}_{\beta_0}) = A(1) - A(0)$。此外，当使用 F 的自洽估计量且 $\|\beta\| < \infty$ 时，

$$\sum_{i=1}^{n} \widehat{\mathcal{R}}^{(a)}(\tilde{E}_{i,\beta}) = n(A(1) - A(0)), \tag{36}$$

其中 $\widehat{\mathcal{R}}^{(a)}(\tilde{E}_{i,\beta})$ 的定义同(29)。

定理5 在附录 A.5中证明。要看出定理 5包含定理1 和2作为特例，请回想 $a(u) = u$ 意味着 $A(u) = u^2/2$；因此， $A(1) - A(0) = 1/2$.

沿着此前的发展脉络，我们猜想，(32) 在 Tsiatis 加权对数秩统计量类中有一个精确的对应形式。回想一下，Ritov 和 Tsiatis 类^估计方程之间的等价性依赖于选择形式为 (15) 的权重。在 (15) 中选择 $\gamma(\cdot) = \gamma_a(t; F_{\beta})$，并利用 (34) 与 (35) 之间的等价性，估计方程 (32) 应等于权重函数为 $w(u; \beta) = w_{\beta,\ a}(u)$，其中

$$w_{\beta,a}(u) = \frac{A(\widehat{F}_{\beta}(u)) - A(\widehat{F}_{\beta}(u-))}{\widehat{F}_{\beta}(u) - \widehat{F}_{\beta}(u-)} - \frac{A(1) - A(\widehat{F}_{\beta}(u))}{\widehat{S}_{\beta}(u)}; \tag{37}$$

正如紧接在（35）下方所述，（37）右侧的第二项对所有满足 $F_{\beta}(u) = 1.$ 的 u 而言，都化简为 $a(1)$。在关于 $a(\cdot)$ 的一些温和附加条件下，第 3.2.3 节和第 3.3 节分别介绍的大样本理论以及推断和方差估计方法，可以扩展到更一般的得分情形。例如，在第3.3.3节的方差估计程序中，可以使用一个对（32）的方差进行估计的估计量，将（22）推广到非线性 $a(\cdot)$；也就是说，可以用 $\Sigma_a(\beta)$ 替换 $\Sigma = \Sigma(\beta)$，其中

$$\widehat{\Sigma}_a(\beta) = \frac{1}{n} \sum_{i=1} \Delta_i [w_{\beta,a}(\tilde{E}_{i,\beta})]^2 \widehat{\mathcal{H}}(\tilde{E}_{i,\beta}; \beta). \tag{38}$$

进一步推测，使用任何相合估计量来估计 F 都不会改变其渐近行为 $\sqrt{n}(\beta-\beta)$ 或其方差估计量，前提是满足与第3.2.3 节中所用条件类似的条件。例如，使用

$$\widetilde{w}_{\beta,a}(u)=a(\widehat{F}_{\beta}^{\star}(u))-\frac{A(1)-A(\widehat{F}_{\beta}(u))}{\widehat{S}_{\beta}(u)}$$

代替（37）不应改变渐近行为，至少当底层分布F 是连续的情况下如此。此外，我们预计，第3.2.4节的结果可以扩展，使得 Σ_1 表示 $n^{-1/2}\Psi_n(\beta_0;\ a)$ 的渐近方差，其中 $\widehat{F}_{\beta}(u)$ 被$F(u)$ 所替代于（32）中。

给定方差估计量（38），一般权重函数 $a(u)$ 对应的伪得分统计量 $n^{-1}[\Psi_n(0,\ a)]^T\widehat{\Sigma}_a^{-1}(0)\Psi_n(0,\ a)$ 在 $H_0:\ \beta_0=0$下将具有渐近 χ^2 分布；当 $a(u)$ 是与 ϵ的分布相对应的最优得分时，它是局部最强检验。此外，可以使用3.3.3中所述的任一方法来获得Ω_a，

– 即适用于使用权重函数 $a(\cdot)$ 时的 $\sqrt{n}(\beta-\beta)$ 的估计方差-协方差矩阵，并可按通常方式构造Wald置信区间和Wald 检验。根据我们截至目前的经验，Huang的方法表现不逊于或优于所述的用于估计Ω_a.的蒙特卡洛方法

注1 定义

$$w_a(u)=a(u)-\frac{A(1)-A(u)}{S(u)};$$

那么，*Cuzick [1985,* 第 *3]* 节表明，式（*14*）中取 $w(u;\beta)=w_a(F_\beta(u))$ 时，所得秩检验统计量渐近地等价于 *Prentice[1978]* 最初提出的线性秩统计量类。这种特定的权重选择及相应的等价性结果，实质上预示了 *Ritov [1990]* 提出的一个关键观察。具体而言，在 *Ritov* 的估计函数中选择 $\gamma(u)=a(F_\beta(t))$，会得到一个加权 *logrank* 统计量（*14*），其中 $w(u;\beta)=W_Fa(F_\beta(u))$，其中

$$\begin{aligned}
W_Fa(\widehat{F}_{\beta}(u)) &= a(\widehat{F}_{\beta}(u))-\frac{\int_u^{\infty}a(\widehat{F}_{\beta}(s))d\widehat{F}_{\beta}(s)}{1-\widehat{F}_{\beta}(u)}\\
&\approx a(\widehat{F}_{\beta}(u))-\frac{\int_{\widehat{F}_{\beta}(t)}^{1}a(s)ds}{1-\widehat{F}_{\beta}(u)}\\
&= a(\widehat{F}_{\beta}(u))-\frac{A(1)-A(\widehat{F}_{\beta}(u))}{\widehat{S}_{\beta}(u)}\\
&= w_a(\widehat{F}_{\beta}(u)).
\end{aligned}$$

阶梯函数 $F_\beta(u)$ 只有广义逆，这就是第一行与第二行之间的转换涉及近似的原因。

注*2* 使用自洽估计量意味着，对于超过最大观测残差的 u 值，$F_\beta(u)$ 和 $F_\beta(u-)$ 均等于 *1*。在 $a'(u)$ 保持有界的条件下，对于 $u\in(0,1],)$，可以使用分部积分证明 $\gamma_a(u;\ F_\beta)\to a(1)$ 以及 $\Gamma_a(u;\ F_\beta)\to a(1)$，从而 $w_\beta(u)\to 0$，当 $F_\beta(u)\to 1$ 时。

注3 假设 $a(\cdot)$ 二阶连续可微，且其三阶导数有界，则可以在 $H(t) = H^\star(t)$; 处对 $A(H(t))$ 进行泰勒级数展开；类似地，可以在 $H(t-) = H^\star(t)$处对 $A(H(t-))$ 进行泰勒级数展开。通过进行这两次展开并将其相减，可以证明

$$\gamma_a(t; H) = a(H^\star(t)) + \frac{1}{24}a''(H^\star(t))(H(t) - H(t-))^2 + O((H(t) - H(t-))^4).$$

因此， $\gamma_a(t; H)$ 可以用 $a(H^\star(t))$ 很好地近似，误差项取决于$(H(t) - H(t-))^2$ 的大小，以及 $a''(H^\star(t))$ 的符号和大小。

4.2 无界得分函数

上一节的推导假设 $a(\cdot)$ 有界，同时满足一些较弱的连续性和可微性条件。在线性秩检验中，无界得分函数的两个常见例子包括 $a(u) = \Phi^{-1}(u)$（即正态得分）和$a(u) = -1 - \log(1-u)$（即极小值或 logrank 得分）。若要考虑使用无界得分函数，一种实用的解决方案是将其替换为 $a([nu + 0.5]/[n + 1])$（用于正态得分）或 $a([nu]/[n + 1])$（用于极值分数），从而将 $a(u)$ 的取值限制在远离 $u = 0$ 和 $u = 1$ 的范围内，其中 n 为样本量。此前已在对误差分布尾部施加略严格条件的情况下，为无界得分下的线性秩检验（及其在回归中的扩展）建立了渐近结果（即，相较于有界得分情形）；例如，Jurečková [1971] 针对完整数据的情形，以及 Lai 和 Ying [1992] 针对右删失（可能伴有左截断）数据的情形。然而，正如 Hettmansperger 和 McKean [2011, 第 3.6 节], 所指出的，使用无界得分函数 $a(\cdot)$，在未删失数据下会得到针对 β 的 R估计量，其影响函数无界，因而在存在较极端残差时稳健性会降低。因此，建议使用有界得分以增强稳健性。

另一种从无界得分推导有界得分的简单方法，是对后者进行“温莎化”。例如，定义 $z_{1-\alpha}$ 使得 $\Phi(z_{1-\alpha}) = 1 - \alpha$ 对于 $\alpha \in [0, 1/2]$, 则可将正态得分 $a(u) = \Phi^{-1}(u)$ 温莎化，分别令其在 $u < \alpha$ 时等于 $-z_{1-\alpha}$ ，或在 $u > 1 - \alpha$. 时等于 $z_{1-\alpha}$。另外，更一般地，McKean 和 Sievers [1989] 提出使用广义 F–分布 [（例如，Kalbfleisch 和 Prentice，2002年，第 2.2 节）7] ，并选择有限的自由度参数，以生成适用于未删失数据回归参数估计的有界得分，这适用于 F 可能不对称的情况。在其原始参数化下，广义$F_{m_1、\ m_2}$ 分布指失效时间T^*的分布，并对应于式 (1) 的位置-尺度模型，其中残差 ϵ 服从一个$F_{2m_1、\ 2m_2}$ 随机变量的自然对数（即，差一个尺度参数 $\sigma > 0$）。广义 F–分布具有很强的灵活性。在 ϵ的尺度上，它包含逻辑斯蒂分布这一特例；此外，它还能生成许多其他分布，包括极小值或极大值的极值分布，以及作为重要边界情形的（退化）正态分布，这些分布的得分函数均为无界。通过重新参数化，直接将 m_1 和 m_2 与新参数 p 和 q关联起来，可以达到同样的通用性，同时还能使标准正态分布在 m_1 和 m_2 均趋于 ∞ [时成为极限情形（Kalbfleisch 和 Prentice， 2002]）。

令$\mathcal{F}_{m_1,\ m_2}(\cdot)$ 表示对数$F_{2m_1,\ 2m_2}$ 随机变量的累积分布函数，并令 $f_0(\cdot)$ 表示其对应的概率密度函数 [Kalbfleisch 和 Prentice，2002年，公式 (2.5)]。然后，使用（26）并按指定方式取 $f_0(\cdot)$ 后，得分函数可简化为 [McKean 和 Sievers， 1989]

$$a(u) = \frac{m_1 m_2(\exp\{\mathcal{F}^{-1}_{m_1,m_2}(u)\} - 1)}{m_2 + m_1 \exp\{\mathcal{F}^{-1}_{m_1,m_2}(u)\}}. \tag{39}$$

可以证明 $-m_1 \le a(u) \le m_2$；因此，前一节的结果适用于 $a(\cdot)$ 的这一取法，其中 m_1 和 m_2. 取有限值。如果定义 $F^{-1}_{2m_1,2m_2}(u)$ 为具有 $2m_1$ 和 $2m_2$ 自由度的标准 F 分布的分位数函数，则上述表达式可以简化；特别地，我们只需将 $\exp\{\mathcal{F}^{-1}_{m_1,m_2}(u)\}$ 替换为 $F^{-1}_{2m_1,2m_2}(u)$。由 (27) 可得，与 (39) 对应的函数 $A(u)$ 为

$$A(u) = -\frac{m_1^{m_1} m_2^{m_2}}{B(m_1, m_2)} \frac{[F^{-1}_{2m_1,2m_2}(u)]^{m_1}}{[m_2 + m_1 F^{-1}_{2m_1,2m_2}(u)]^{m_1+m_2}} \tag{40}$$

其中， $B(m_1, m_2)$ 是贝塔函数。R 和许多其他语言都支持对 $F^{-1}_{m_1, m_2}(\cdot)$、计算 $a(u)$ 和 $A(u)$ 所需的快速数值计算。

5 模拟研究

我们开展模拟研究的目标如下。首先，我们希望确认参数估计及其方差-协方差估计，以及准得分检验和 Wald 检验在有限样本下具有良好表现。其次，我们希望考察变换 $a(r)$ 的选择如何影响估计量和检验的性能。最后，我们希望将所得结果与第3.2节所述的 Fygenson和Ritov 方法得到的结果进行比较。为实现这些目标，我们根据模型（1）生成数据，其中 ϵ 服从均值为零、尺度参数为 $k = \frac{1}{2}$ 的中心极值（Gumbel）分布；也就是说，所得风险函数为 $\lambda_F(u) = k \exp(ku - \xi)$，其中 $\xi = 0.57721\cdots$ 是欧拉常数。这一选择对应于形状参数为 $k = \frac{1}{2}$、尺度参数为 $\exp\{2\xi + X_i^T\beta_0\}$ 的服从 Weibull 分布的失效时间；等价地，条件分布 $T^*|X$ 由风险函数 $\lambda_{T^*}(t;\, X) = k \exp\{-(\xi + kX^\top\beta_0)\}t^{k-1}$. 刻画。对于模拟，我们考虑了包含 $n = 200$ 个观测值和两个协变量 X_1 和 X_2. 的数据。我们生成了 $X_2 \sim$ Bernoulli(0.5)，并且为了在 X_1 和 X_2 之间产生一定相关性，独立于 X_2. 生成了 $X_1 \sim Z + \frac{1}{2}X_2$，其服从 $Z \sim N(0, 1)$ 分布。删失时间服从 $\log(C_i) \sim N(\mu_c = 1.5, \sigma_c = 2)$ 分布，所得平均删失率约为34%。为简便起见，我们通过关系式 $\beta_0 = b \times (1, -1)^\top$ 定义了一系列 β_0 值，其中 b 在区间 $[-1, 1]$. 内变化。

我们在分析中考虑了 5 个候选函数 $a(u)$： $a(u) = u$ 对应于采用 Wilcoxon得分，或等价地，采用第3.2节中的“无权重”R估计量（记为 raft.NoW）；$a(u) = -1 - \log(1 - \frac{n}{n+1}u)$，其中 n 为样本量，对应于 Gumbel 数据的（平移后）最优选择（记为 raft.WW）；以及式 (39) 中给出的 $a(u)$，分别对应于 (m_1, m_2) =(1, 10)（记为 raft.FW1）、(m_1, m_2) =(10, 1)（记为 raft.FW2）和 (m_1, m_2) = (3, 3)（记为 raft.FW3）。在广义 F 族中，Gumbel 分布对应于 $m_1 = 1, m_2 \to \infty$；因此，我们可以预期 (m_1, m_2) =(1, 10) 表现良好，而 (m_1, m_2) =(10, 1) 表现相对较差（即在效率方面）。作为比较，我们还纳入了使用第3.2节所述 Fygenson-Ritov 估计量（记为 fraft）获得的结果。

表1中汇总的结果表明，我们提出的 R估计量偏差较小，且对于所有五种 $a(u)$ 的选择，式（22）和（38）中的估计方差，以及式（25）给出的参数估计，表现良好。重要的是， Σ的尺度取决于秩变换 $a(u)$ 的选择，而 Ω 的尺度则不受影响。因此，可以使用 Ω 的相应经验估计来比较不同 $a(\cdot)$ 选择下的相对效率。例如，与最优权重函数（即当 26）中给出，且 F_0 是 ϵ的累积分布函数时）相比，使用 Wilcoxon得分时，参数 $1.14 = 1.506/1.322$ 的相对效率为 β_1，当 $\beta_0 = (0,0)$。请注意，Fygenson 和 Ritov 估计量的相对效率为 $1.28 = 1.689/1.322$

		Emp. var of Ψ ($\times 10^{-4}$)		*Mean est var of* Ψ ($\times 10^{-4}$)		Bias ($\times 10^{-2}$)	*Emp. var of* $\widehat{\beta}$ ($\times 10^{-1}$)		*Mean est var of* $\widehat{\beta}$ ($\times 10^{-1}$)	
β	方法	Σ_{11}	Σ_{22}	$\widehat{\Sigma}_{11}$	$\widehat{\Sigma}_{22}$	$\overline{\beta}-\beta$	Ω_{11}	Ω_{22}	$\widehat{\Omega}_{11}$	$\widehat{\Omega}_{22}$
(0, 0)	raft.NoW	0.923	3.916	0.894	3.893	(1.419,-0.794)	1.506	0.381	1.571	0.377
(0, 0)	raft.WW	7.472	31.686	7.331	30.678	(1.268,-0.459)	1.322	0.318	1.368	0.329
(0, 0)	raft.WF1	7.350	31.170	7.205	30.197	(1.258,-0.490)	1.324	0.319	1.371	0.330
(0, 0)	raft.WF2	10.011	42.538	9.898	45.811	(2.700,-1.109)	2.290	0.610	2.322	0.556
(0, 0)	raft.WF3	14.043	59.617	13.765	60.051	(1.730,-0.793)	1.527	0.386	1.570	0.376
(0, 0)	fraft	3.115	13.226	3.018	13.436	(1.535,-0.959)	1.689	0.437	1.756	0.423
(1, −1)	raft.NoW	0.927	3.755	0.914	3.962	(2.027,-1.479)	1.559	0.426	1.630	0.413
(1, −1)	raft.WW	7.384	28.941	7.409	29.785	(1.777,-1.094)	1.378	0.361	1.413	0.359
(1, −1)	raft.WF1	7.274	28.596	7.288	29.501	(1.795,-1.133)	1.376	0.362	1.420	0.360
(1, −1)	raft.WF2	10.038	41.745	9.757	43.881	(3.358,-1.818)	2.360	0.661	2.414	0.607
(1, −1)	raft.WF3	14.071	57.148	13.882	60.344	(2.340,-1.473)	1.577	0.431	1.626	0.412
(1, −1)	fraft	3.174	12.968	3.135	13.985	(2.228,-1.709)	1.759	0.488	1.835	0.465
(−1, 1)	raft.NoW	0.927	3.768	0.913	3.846	(1.227,-0.572)	1.608	0.432	1.623	0.411
(−1, 1)	raft.WW	7.385	29.106	7.379	29.144	(1.277,-0.387)	1.425	0.365	1.406	0.360
(−1, 1)	raft.WF1	7.276	28.754	7.259	28.805	(1.258,-0.405)	1.428	0.367	1.410	0.360
(−1, 1)	raft.WF2	10.039	41.823	10.040	45.101	(2.473,-0.730)	2.395	0.675	2.395	0.601
(−1, 1)	raft.WF3	14.074	57.339	13.980	59.144	(1.484,-0.577)	1.627	0.441	1.618	0.410
(−1, 1)	fraft	3.175	13.011	3.104	13.546	(1.309,-0.620)	1.795	0.493	1.822	0.460

表1：极值误差下的模拟。样本量为 200。结果基于 1,000 次重复模拟。

当 β_1 时，与最优选择相比表现较差，因此不如 Wilcoxon得分。在此例中，广义 F估计量在 (m_1, m_2) =(1,10) 下的相对效率约为 1.0（即相对于最优权重选择）；这与以下观点一致：$m_2 = 10$ 已经足够大，能够近似相关的极限结果（同时仍保持得分函数有界）。

为更好地说明本研究中各估计量的相对效率，图1还考察了每种方法下 Wald 检验和准得分检验拒绝原假设 H_0 ： β_0 = (0, 0) 的检验效能。所有 $a(\cdot)$ 的选择以及 Fygenson 和 Ritov 估计量，在 H_0. 下都能产生检验水平适当的检验。图1还展示了所示双侧检验的估计检验效能，其中 $\beta_0 = (b, -b)^\top$ 沿着由 $b \in [-1, 1]$. 定义的方向偏离 H_0。最优权重选择和参数为 (m_1, m_2) = (1, 10) 的广义 F 得分具有最高的检验效能，这也在预料之中，两者的检验效能非常相近。使用 Wilcoxon 得分 $a(u) = u$ 所达到的检验效能（即，与参数为 (m_1, m_2) =(1, 1) 的广义 F 得分在位置和尺度上等价），以及参数为 (m_1, m_2) =(3, 3) 的广义 F 得分，其次；这里，两种得分都对应于对称的 ϵ 分布。最后，在此处考察的所有方法中，Fygenson 和 Ritov 方法的检验效能倒数第二低，其后是选择不当的、参数为 (m_1, m_2) =(10, 1) 的广义 F 得分。

为确认 $\widehat{\beta}$ 的渐近正态性，$\beta_0 = (\beta_1, \beta_2)^\top$ 并进一步说明估计方差 $\widehat{\Omega}$ 的准确性，我们还比较了 β_j、$j = 1$、2 的（边际）置信区间的经验覆盖率与名义覆盖率。图 2 展示了 $a(u)$ 取最优值时的结果，图 3 则展示了表现最差的（m_1、m_2）=(10、1) 的结果。经验覆盖率与名义覆盖率高度吻合，这表明对于线性模型的删失数据分析，在 $a(\cdot)$ 的各种取值下，我们的 *R*估计量表现良好。

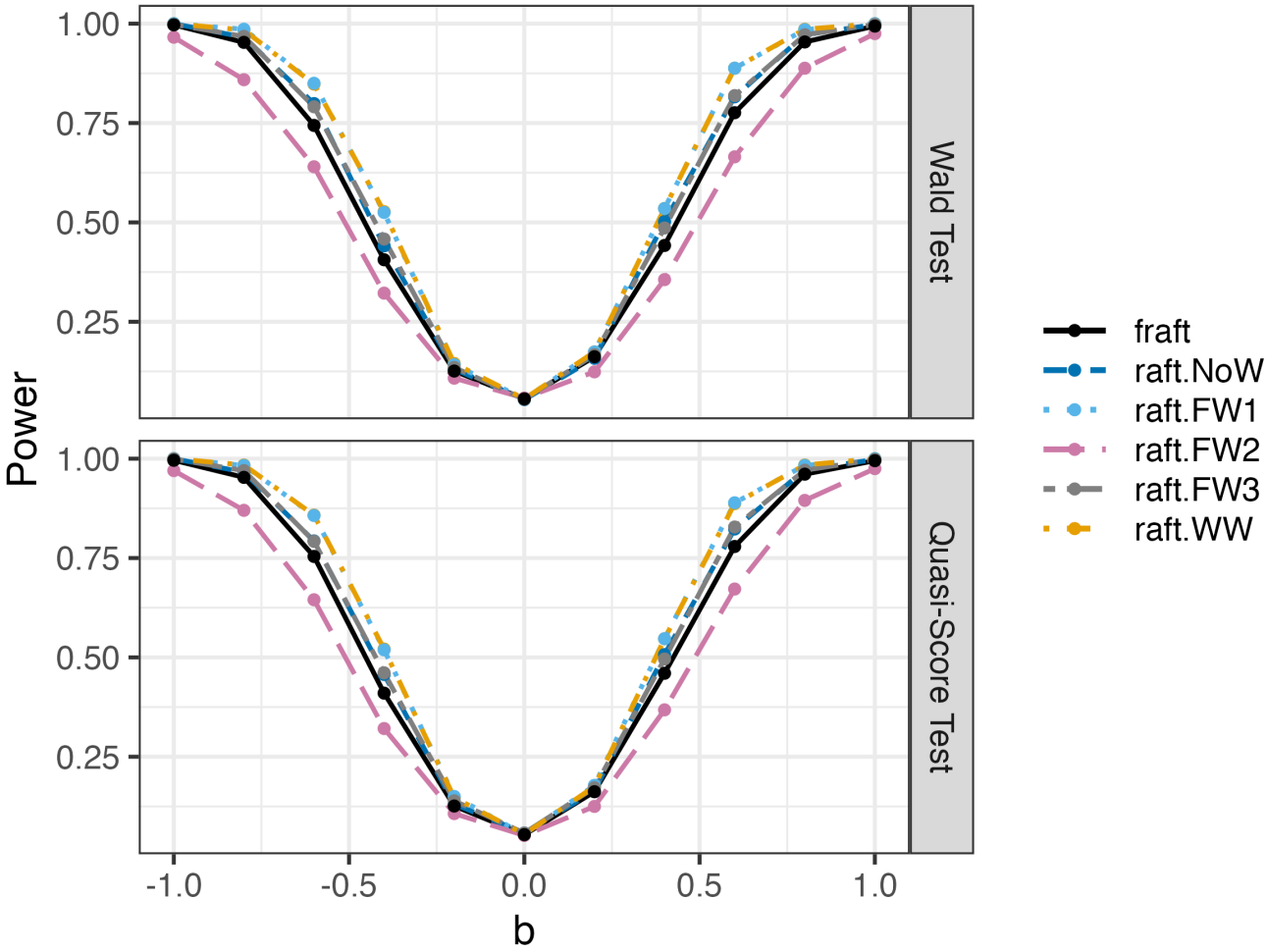

图 1：在权重函数 $a(u)$ 各异时，Wald 检验和准得分检验拒绝假设 $\beta_0 = (0,0)^\top$ 的检验效能，以及 Fygenson 和 Ritov 估计量（fraft）针对形式为$\beta_0 = (b, -b)^\top$, $b \in [-1, 1]$.的备择假设的检验效能

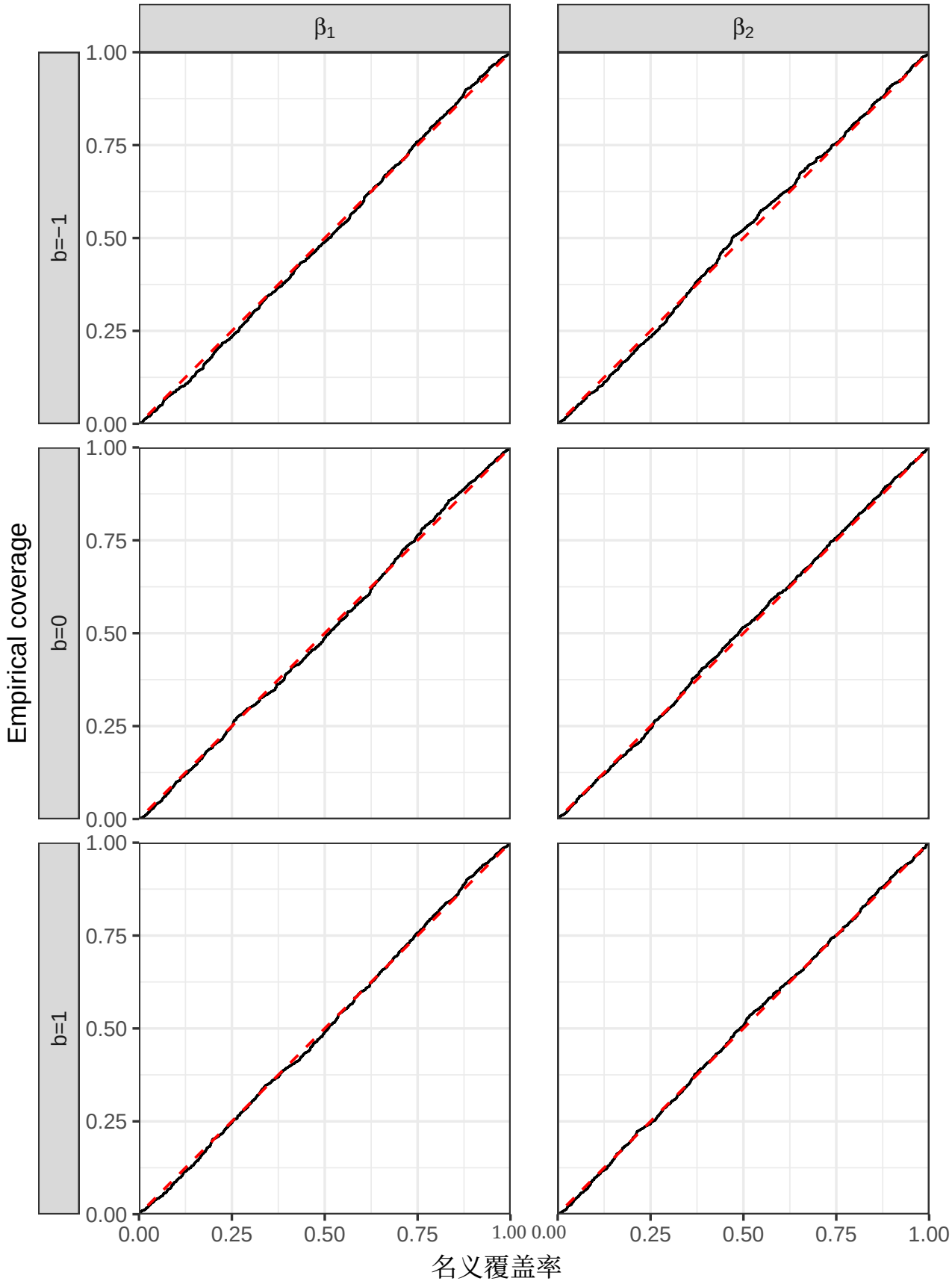

图2：当 $\beta_0 = (b, -b)^\top$, $b \in [-1, 1]$, 时，使用最优权重函数 $a(u) = -1 - \log(1 - \frac{200}{201}u)$所得 β_1（第1列）和 β_2（第2列）的（边际）Wald型置信区间的经验覆盖率与名义覆盖率对比。红色虚线表示名义覆盖率与经验覆盖率相等。[20]

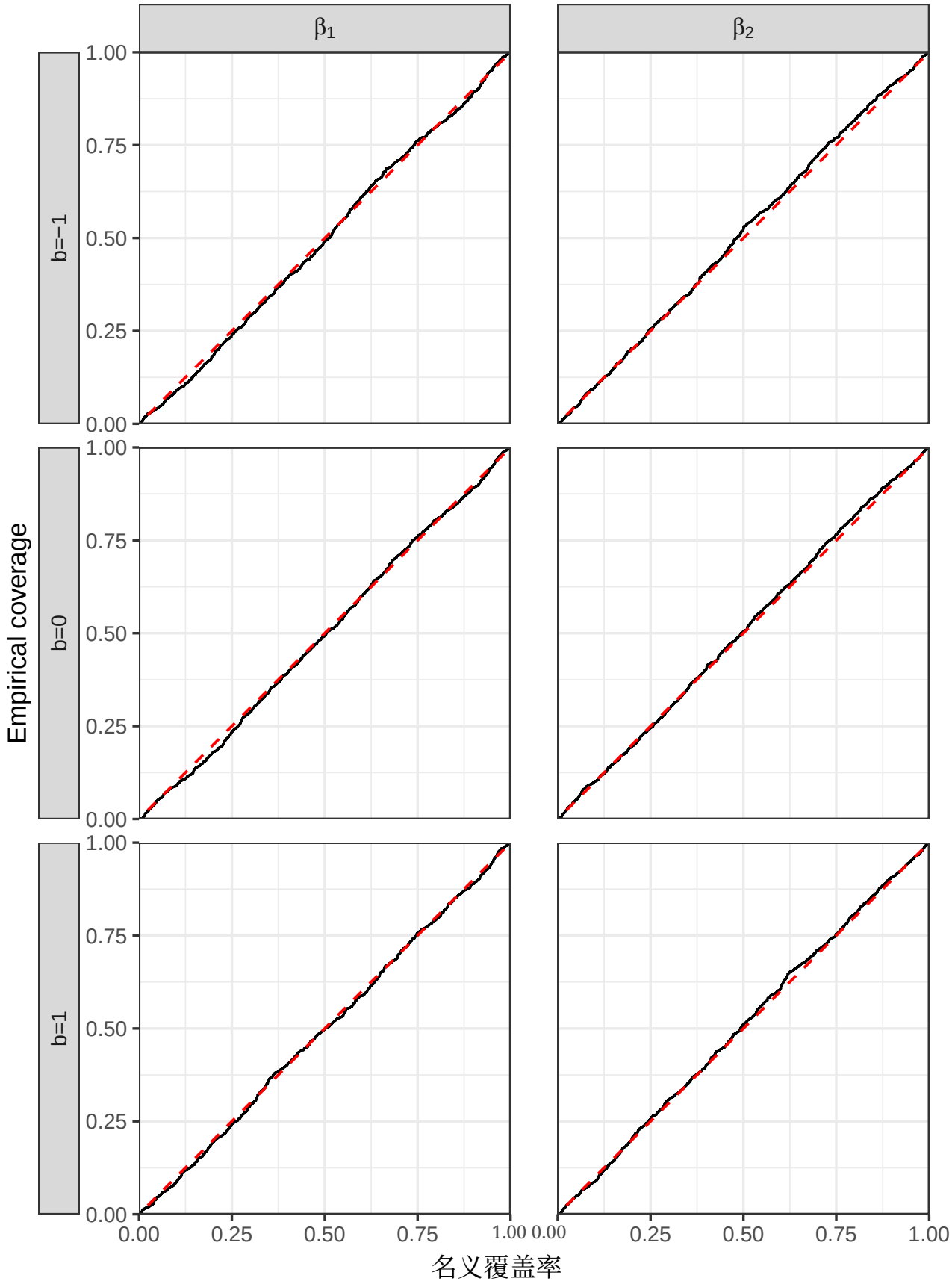

图3：当 $\beta_0 = (b, -b)^\top$， $b \in [-1, 1]$，使用广义 F权重函数，且 $(m_1, m_2) = (10, 1)$ 时，β_1（第1列）和 β_2（第2列）的（边际）Wald型置信区间的经验覆盖率与名义覆盖率。红色虚线表示名义覆盖率与经验覆盖率相等。

6 讨论

长期以来，线性回归参数的 R 估计量一直被视为普通最小二乘估计量的一种颇具吸引力的替代方法。本文提出了秩概念的一种自然推广，并相应地自然扩展了 R 估计方程（3），以便在线性和非线性得分函数的情形下处理右删失数据 $a(\cdot)$。模拟结果表明，对于一系列得分函数，所提出的估计量在失效时间服从 Weibull 分布时表现良好 $a(\cdot)$，但同时也表明，$a(\cdot)$ 的选择会影响所得估计量的检验效能和效率。正如预期，选择最优权重函数可获得最佳表现；使用仅近似最优权重的权重函数，表现几乎相同。我们的结果还表明，与 Fygenson和Ritov 的估计量相比，使用 Wilcoxon 得分（其对应的加权 logrank 统计量采用 Kaplan–Meier（即自洽）残差分布估计量作为权重）可带来更高的检验效能和效率 Fygenson和Ritov [1994], 该估计量对应于 Gehan加权对数秩统计量，其权重为$\sum_{j=1}^{n} Y_j(u;\beta)$。

本文首先聚焦于秩的线性函数得分函数这一情形，即 $a(u) = u$. 我们通过（7）对秩所作定义的一个有用推论，来自定理 2：该定理表明，无论 β如何，秩的总和都是常数，这与完全数据情形相同。我们还表明，相应的秩估计函数（12）与Ritov [1990]和 Tsiatis [1990]分别针对右删失结局提出的估计量类别之间存在精确关系。这些结果都直接依赖于 S 的自洽估计量（因此 F）以及定义（7）中相应的中间累积分布函数；这些非渐近结果适用于一般的 F. 当然，从渐近角度看， S (F还可以有许多其他估计量；值得探究其他有效选择是否会实质性地影响有限样本表现。

据作者所知，第3.2.2节中针对非线性得分函数的研究是新的，尤其是（28）与（29）之间的精确对应关系（即通过（30）建立），以及相关的估计方程（32）。这些结果适用于一般的 F，并以一种颇有意思的方式直接推广了线性得分函数的相应结果。值得注意的是，这些推导并不特别依赖于使用 $\widehat{}$中间累积分布函数 $F^{\star}_{\beta}(t)$，而是依赖于（31）中规定的关系；这是一种特定的推广，它（i）依赖于 $a(\cdot)$，且（ii）旨在确保（30）成立。我们进一步展示了如何将这些结果推广至任意正规累积分布函数 $H(\cdot)$；具体做法是引入（33），对每个t，将其定义为 $H(t)$ 与 $H(t-)$ 的均匀加权平均经过非线性函数变换后的期望。$\widehat{}$这一 $\gamma_a(t;H)$ 的定义在 $H(t) = F_{\beta}(t)$ 时还原为（31）；更一般地，它的使用保证了（34）与（35）在数学上等价，从而推广了（30）。

本文的一项新颖贡献，是展示了右删失数据 R估计量推广与加权 logrank 统计量类别之间确切的内在联系；具体而言，分别是（12）与（16），以及（32）与采用权重函数（37）的加权 logrank 统计量之间的联系。当然，作为这些内在联系的结果，由此得到的一般回归参数估计量类别，以及第 3.3 节所概述的推断方法，严格来说并非全新，尤其是在 $a(u) = u$. 这些等价性结果的一个特点是，它们可用于说明采用某些特定权重函数的理由，而这些权重函数在构造加权对数秩统计量时并不常用。例如，文献通常关注对紧密符合 Cox比例风险模型的数据具有高效率和／或高检验效能的权重。本文考察的权重函数类别，则对应于误差分布给定时，针对服从加速失效时间（AFT）模型（等价地，对数失效时间的线性模型）的数据，R估计量的最优选择，并将其推广以考虑右删失。我们所提出的删失数据 R估计量形式相对简单，这也使得将这些方法推广到更复杂的数据结构成为可能，

例如具有右删失结局的聚类失效时间数据；参见 Bilgic 和 Susmann[2013]及其中的参考文献。

在本研究中，与未删失数据R估计量的相关文献一致，我们聚焦于右删失结局，并假定 $a(\cdot)$ 单调。在未删失数据的情形下，式 (3) 中 $a(\cdot)$ 的单调性使得该估计量同时也是 β 中某个凸目标函数的极小值，即 Jaeckel离散度 [Jaeckel, 1972]。然而，对于右删失数据，即使 $a(u) = u$，也不能期望式 (32) 关于 β 单调；因此，我们提出的估计量已无法成为某个凸目标函数的极小值。此外，最优得分的选择 (26)（即 $F_0(\cdot)$ 为 ϵ的累积分布函数时）本身也不保证单调。例如，可以证明，当 ϵ 服从柯西分布时，(26) 意味着 $a(u) = \sin(2\pi u)$。这些观察结果，以及在不存在删失的情况下，类似第 3.3 节和第 4 节所总结的结果仍然适用，共同拓展了未删失数据问题中可用的新型秩估计量类别，使其涵盖得分非单调的加权R估计量。使用柯西分布及其最优权重函数进行的初步模拟（结果未展示）表明，这种方法对于删失数据和未删失数据都可能有用。

致谢

本研究部分获得美国国立卫生研究院资助，资助编号为 R01ES034021（RLS）和 R01GM147162（NZ、GS）。

参考文献

P. K. Andersen、Ø. Borgan、R. D. Gill 和 N. Keiding。《基于计数过程的统计模型》。Springer, 纽约，1993年。Y. K. Bilgic 和 H. Susmann。“rlme：用于随机效应嵌套模型中基于秩的估计与预测的 R软件包。”《R期刊》，5(2)，2013年。ISSN ISSN 2073-4859。B. M. Brown 和 Y.-G. Wang。“平滑秩估计量的标准误与协方差矩阵。”《生物统计学》，92(1)：149–158，03 2005年。W. H. Chang、J. W. Mckean、J. D. Naranjo 和 S. J. Sheather。“高崩溃点秩回归。”《美国统计学会会刊》，94(445)：205–219，1999年。J. Cuzick。“删失线性秩检验的渐近性质。”《统计年鉴》，13(1)：133–141，1985年。N. Falkner 和 G. Teschl。“关于勒贝格–斯蒂尔切斯积分的代换法则。”《数学展望》，30(4)：412–418，2012年。ISSN 0723-0869。M. Fygenson 和 Y. Ritov。“删失数据的单调估计方程。”《统计年鉴》，22(2)：732–746，1994年。T. P. Hettmansperger。《基于秩的统计推断》。Wiley，纽约，1984年。T. P. Hettmansperger 和 J. W. McKean。《稳健非参数统计方法》，第二版。CRC Press，佛罗里达州博卡拉顿，2011年。

Y. Huang。“删失寿命医疗费用的校准回归。”《美国统计学会会刊》，97：457：318–327，2002年。doi：DOI:10.1198/016214502753479446。Y. Huang。“快速删失线性回归。”《斯堪的纳维亚统计学杂志》，40(4)：789–806，2013。L. A. Jaeckel。“通过最小化残差的离散度估计回归系数。”《数学统计年刊》，页码1449–1458，1972年。Z. Jin 和 Z. Ying。《固定删失下加速失效时间模型的秩估计渐近理论》，页码107–120。Birkh¨auser 波士顿，波士顿，MA，2004。Z. Jin、D. Y. Lin、L. J. Wei 和 Z. Ying。“加速失效时间模型的基于秩的推断。”《生物统计学》，90(2)：341–464，2003。L. M. Johnson 和 R. L. Strawderman。“半参数加速失效时间模型的诱导平滑：渐近理论及其向聚类数据的扩展。”《生物统计学》，96(3)：577–590，2009年。J. Jureˇckov´a。“回归系数的非参数估计。”《数学统计年刊》，42(4)：1328–1338，1971年。J. D. Kalbfleisch 和 R. L. Prentice。《失效时间数据的统计分析》。Wiley-Interscience，霍博肯，新泽西州，第二版，2002年。T. L. Lai 和 Z. Ying。“含删失或截断数据的回归分析中的线性秩统计量。”《多元分析杂志》，40(1)：13–45，1992年。J. F. Lawless。《寿命数据的统计模型与方法》。威利概率与数理统计丛书，纽约州纽约，1982年。I. M´arquez Alb´es 和 A. Slav´ık。“广义链式法则及其在斯蒂尔切斯微分方程与积分方程中的应用。”《数学进展》，79(2)：论文编号78，43页，2024年。doi：10.1007/s00025-023-02104-0。J. W. McKean 和 G. L. Sievers。“适用于非对称误差分布下线性模型分析的秩得分。”《技术计量学》，31(2)：207–218，1989年。K. G. Mehrotra、J. E. Michalek 和 D. Mihalko。“删失数据两种线性秩方法之间的关系。”《生物统计学》，69(3)：674–676，1982年。R. L. Prentice。“带右删失数据的线性秩检验。”《生物统计学》，65(1)：167–179，04 1978。Y. Ritov。“删失数据线性回归模型中的估计。”《统计年鉴》，18(1)：303–328，1990年。F. H. Ruymgaart。“某些中秩统计量渐近分布理论的统一方法。”载于 J.-P. Raoult 编，《渐近非参数统计学》，页码1–18，柏林、海德堡，1980年。Springer Berlin Heidelberg。ISBN 978-3-540-38318-5。G. A. Satten、M. Li 和 N. Zhao。“使用对数线性模型对含零单元格的微生物组数据进行基于秩的组成分析。”BioArXiv，2025年。P. K. Sen。“基于肯德尔秩相关系数的回归系数估计。”《美国统计学会会刊》，63(324)：1379–1389，1968年。

B. E. Shepherd、C. Li 和 Q. Liu。连续、离散和删失数据的概率尺度残差。《加拿大统计学杂志》44(4):463–479，2016年。R. L. Strawderman。加速间隔时间模型。《生物统计学》92(3):647–666，2005年9月。R. L. Strawderman 和 B. R. Baer。论 Volterra积分方程在生存函数的自洽估计量、乘积极限估计量、逆删失概率加权估计量及向右再分配估计量中的作用。《终生数据分析》30:649–666，2023年。A. A. Tsiatis。使用线性秩检验估计删失数据的回归参数。《统计年鉴》18:354–372，1990年。Z. Ying。删失回归数据秩估计的大样本研究。《统计年鉴》21(1):76–99，1993年。D. Zeng 和 D. Y. Lin。非光滑估计函数的高效重抽样方法。《生物统计学》9(2):355–363，2008年10月。

A 附录

我们先介绍一个关于中间累积分布函数行为的有用结果。对于任意正规累积分布函数 $H(t)$，定义 $H^\star(t) = \frac{1}{2}(H(t) + H(t-))$ 为其对应的中间累积分布函数。于是：

$$\int_a^b H^\star(t)dH(t) = \frac{1}{2}\left[H^2(b) - H^2(a)\right]; \tag{41}$$

例如，参见 Ross 的《初等分析：微积分理论》第 2 版，定理 35.19）。分部积分公式（41）的一个直接推论是

$$\frac{1}{S(t)}\int_t^\infty F^\star(u)dF(u) = 1 - \frac{S(t)}{2}, \tag{42}$$

其中 $S = 1 - F$.

在某些证明中，我们还会用到计数过程记法，以及对于任意固定的有限 $\beta.$, $S_\beta(t)$ 的自洽估计量的性质。对于 $i = 1, \ldots, n$，回忆一下，$N_i(u;\beta) = \Delta_i I\{E_{i,\beta} \le u\}$，且 $Y_i(u;\beta) = I\{E_{i,\beta} \ge u\}$；令 $N_+(u;\beta) = \sum_{i=1}^n N_i(u,\beta)$，并令 $\curlywedge Y_+(u;\beta) = \sum_{i=1}^n Y_i(u,\beta)$，再将 $\Lambda_\beta(t)$ 定义为对应于 $S_\beta(t)$ 的 Nelson-Aalen估计量；参见（21）。由于 $\widehat{F}_\beta(t) = 1 - \widehat{S}_\beta(t)$，我们还可以写作 $d\widehat{F}_\beta(t) = \widehat{S}_\beta(t-)d\widehat{\Lambda}_\beta(t)$，且 $\int_{-\infty}^\infty d\widehat{F}_\beta(u) = 1.$

这里需要指出，除非另有说明，所有此类量均假定使用修改后的数据集，其中所有与 T_i 并列的观测时间 $T_{(n)}$ 均具有失效状态 $\Delta_i = 1$；这样做只是为了与自洽估计量 F_β的使用保持兼容，这一点尤其重要对于定理 5的秩和证明，以及定理 3 ，因为其中出现了 $\Lambda_\beta(\cdot)$。

A.1 定理 1 的证明

Proof: 公式（10）$\mathcal{R}_{\beta_0}$ 可由积分形式（41）和（42）得到。通过大量使用分部积分，可以证明其均值和方差的所述结果；下文则将利用Shepherd等人 [2016],的结果，从而给出更简单、更简短的证明。

特别地，容易证明（10）可以写成

$$\mathcal{R}_{\beta_0} = \frac{1}{2} + \frac{1}{2}r(\tilde{E}_{\beta_0}, F, \Delta)$$

其中 $r(y, F, \Delta) = F(y) - \Delta(1 - F(y-))$ 是 Shepherd等人 [2016] 为任意底层失效累积分布函数 F 引入的右删失数据概率尺度残差。Shepherd 等人 [2016, 第 5] 节给出的结果现在表明，只要 $T^* \perp C^* \mid X$，就有 $E[r(E_{\beta_0}, F, \Delta)|X] = 0$；因此立即可得 $E(\mathcal{R}_{\beta_0}|X) = E(\mathcal{R}_{\beta_0}) = 1/2$. 此外，当底层分布$F$ 连续时，他们的结果还进一步表明

$$\mathrm{Var}(\mathcal{R}_{\beta_0}|X) = \frac{1}{12}\left(1 - E(S^3(E_{\beta_0})|X)\right);$$

对于 F 有跳跃的情形，该公式还给出了相应的版本，其中离散性校正还涉及跳跃幅度的立方。由于 $E(\mathcal{R}_{\beta_0}|X) = 1/2$，我们还有 $\mathrm{Var}(E(\mathcal{R}_{\beta_0}|X)) = 0$，因此

$$\mathrm{Var}(\mathcal{R}_{\beta_0}) = E\left(\mathrm{Var}(\mathcal{R}_{\beta_0}|X)\right) = \frac{1}{12}\left(1 - E(S^3(E_{\beta_0}))\right),$$

这正是要证明的。

A.2 定理3的证明

^定义 $M(u) = 1 - S_\beta(u)/2$；然后，经过一些代数运算，可注意到（12）可以改写为

$$\begin{aligned}\Psi_n(\beta) &= \sum_{i=1}^{n}(X_i - \bar{X})\left[\Delta_i \widehat{F}^{\star}_{\beta}(\tilde{E}_{i,\beta}) + (1-\Delta_i)M(\tilde{E}_{i,\beta})\right] \\ &= \sum_{i=1}^{n}\int_{-\infty}^{\infty}\widehat{F}^{\star}_{\beta}(u)(X_i - \bar{X}(u;\beta))dN_i(u;\beta) + \int_{-\infty}^{\infty}\widehat{F}^{\star}_{\beta}(u)(\bar{X}(u;\beta) - \bar{X})dN_+(u;\beta) \\ &\quad + \sum_{i=1}^{n}(X_i - \bar{X})M(\tilde{E}_{i,\beta}) - \sum_{i=1}^{n}(X_i - \bar{X})\int_{-\infty}^{\infty}M(u)dN_i(u;\beta).\end{aligned}$$

^令 $W(u) = F^{\star}_{\beta}(u) - M(u)$，进一步代数运算可得

$$\begin{aligned}\Psi_n(\beta) &= \sum_{i=1}^{n}\int_{-\infty}^{\infty}W(u)(X_i - \bar{X}(u;\beta))dN_i(u;\beta) \\ &\quad + \int_{-\infty}^{\infty}W(u)(\bar{X}(u;\beta) - \bar{X})dN_+(u;\beta) + \sum_{i=1}^{n}(X_i - \bar{X})M(\tilde{E}_{i,\beta}).\end{aligned} \tag{43}$$

此外，注意到

$$\begin{aligned}\sum_{i=1}^{n}(X_i - \bar{X})M(\tilde{E}_{i,\beta}) &= -\frac{1}{2}\sum_{i=1}^{n}(X_i - \bar{X})\widehat{S}_\beta(\tilde{E}_{i,\beta}) \\ &= -\sum_{i=1}^{n}(X_i - \bar{X})\Big(\frac{\widehat{S}_\beta(\tilde{E}_{i,\beta})}{2} - \frac{1}{2}\Big).\end{aligned}$$

^由于 $S_\beta(u-)$ 是 S 的自洽估计量，因此有

$$\int_{-\infty}^{\tilde{E}_{i,\beta}}\widehat{S}_\beta(u-)\frac{dN_+(u;\beta)}{Y_+(u;\beta)} = \int_{-\infty}^{\tilde{E}_{i,\beta}}d\widehat{F}_\beta(u) = \widehat{F}_\beta(\tilde{E}_{i,\beta}) = 1 - \widehat{S}_\beta(\tilde{E}_{i,\beta}) = -2\left(\frac{\widehat{S}_\beta(\tilde{E}_{i,\beta})}{2} - \frac{1}{2}.\right)$$

因此，

$$\begin{aligned}\sum_{i=1}^{n}(X_i - \bar{X})M(\tilde{E}_{i,\beta}) &= \frac{1}{2}\sum_{i=1}^{n}(X_i - \bar{X})\int_{-\infty}^{\tilde{E}_{i,\beta}}\widehat{S}_\beta(u-)\frac{dN_+(u;\beta)}{Y_+(u;\beta)} \\ &= \frac{1}{2}\left\{\sum_{i=1}^{n}X_i\int_{-\infty}^{\infty}\widehat{S}_\beta(u-)Y_i(u;\beta)\frac{dN_+(u;\beta)}{Y_+(u;\beta)} - \bar{X}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\widehat{S}_\beta(u-)Y_i(u;\beta)\frac{dN_+(u;\beta)}{Y_+(u;\beta)}\right\} \\ &= -\int_{-\infty}^{\infty}-\frac{\widehat{S}_\beta(u-)}{2}\bar{X}(u;\beta)dN_+(u;\beta) + \bar{X}\int_{-\infty}^{\infty}-\frac{\widehat{S}_\beta(u-)}{2}dN_+(u;\beta) \\ &= -\int_{-\infty}^{\infty}W(u)\bar{X}(u;\beta)dN_+(u;\beta) + \bar{X}\int_{-\infty}^{\infty}W(u)dN_+(u;\beta),\end{aligned}$$

最后一步是由以下事实得出的

$$W(u) = \widehat{F}^{\star}_{\beta}(u) - M(u) = -\frac{1}{2}\hat{S}_\beta(u-). \tag{44}$$

因此，将所得表达式代入（43）并利用（44），

$$\begin{aligned}\Psi_n(\beta) &= \sum_{i=1}^{n}\int_{-\infty}^{\infty} W(u)(X_i - \bar{X}(u;\beta))dN_i(u;\beta) + \int_{-\infty}^{\infty} W(u)(\bar{X}(u;\beta) - \bar{X})dN_+(u;\beta) \\ &\quad - \int_{-\infty}^{\infty} W(u)\bar{X}(u;\beta)dN_+(u;\beta) + \bar{X}\int_{-\infty}^{\infty} W(u)dN_+(u;\beta) \\ &= -\frac{1}{2}\sum_{i=1}^{n}\int_{-\infty}^{\infty} \widehat{S}_\beta(u-)(X_i - \bar{X}(u;\beta))dN_i(u;\beta);\end{aligned}$$

这就证明了所需结果，且无需使用 Ritov [1990] 中所用的恒等式。

A.3 渐近结果的正则条件

在 Ying [1993]，中，假设满足以下正则条件：

1. (T_i^*, C_i^*, X_i) $i = 1, \ldots, n$ 相互独立。

2. 对于 $i = 1, \ldots, n$,

   (a) $\|\beta_0\| < \infty$ 和 $P\{\|X_i\| \le M\} = 1$ 对于某个给定的 $M < \infty$ 成立（即，协变量一致有界）；(b) 给定 X_i, $T_i^* \perp C_i^*$（等价地：$\epsilon_i \perp E_{\beta_0,i}$）。(c) $\epsilon_i | X_i \sim F$，其中 F的一阶（f）和二阶（f'）导数有界，并且

$$\int \left(\frac{d}{ds}\log f(s)\right)^2 f(s)ds < \infty.$$

   (d) $\log C_i^*$ 的分布的累积分布函数为 G_i，其中 G_i 具有有界的概率密度函数。

3. 对某个 $c > 0$.，$\sup_i E|\epsilon_i \wedge E_{\beta_0,\,i}|^c < \infty$

在Jin 和 Ying [2004], 中，假设条件 1 和 2(a) - 2(c) 成立；删失方面的限制性条件，尤其是 2(d) 以及条件 3，不再要求，而由以下条件取代：

- （ Σ的经验类比量，定义于（17）中，其特征值有正的下界，对所有 n 充分大的情况均成立。

上述条件足以确立式 (14) 的渐近行为，以及相应的 β，当 $w(u;\beta) = 1$. 对于更一般的权重，还需要一个附加条件来约束 $w(u;\beta)$ 关于 (u,β) 的函数行为。例如，Ying [1993, 第 3 节的条件 5] 规定了一个由三部分组成的条件，以确保其具有足够良好的行为，并指出这一充分条件由 $\widehat{S}_\beta(\cdot)$ 满足。在 Huang [2002] 和 Strawderman [2005], 中，则施加了另一项充分条件；在当前设定下，该条件要求：(i) $w(u;\beta)$ 具有有界变差；(ii) $w(u;\beta)$ 几乎处处收敛到函数 $w(u)$, 该函数关于 $u \in (-\infty, \infty)$ 一致连续，并且 β 在 β_0. 的某个开邻域内

A.4 定理4的证明

A.4.1 Σ 的分解

回顾

$$\Theta_j(u) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E[X_i^{\otimes j} Y_i(u, \beta_0)].$$

在独立同分布和连续性假设下,

$$\Theta_j(u) = S(u) E\big[X^{\otimes j} Y_0(u)\big],$$

其中 $Y_0(u) = I\{E_{\beta_0} \geq u\}$ 且 E_{β_0} 的定义见第A.1节。由此可得

$$\frac{\Theta_j(u)}{\Theta_0(u)} = \frac{E\big[X^{\otimes j} Y_0(u)\big]}{E\big[Y_0(u)\big]}$$

定义

$$\mu_{X|E_{\beta_0}}(u) = \frac{\Theta_1(u)}{\Theta_0(u)} = \frac{E\big[X Y_0(u)\big]}{E\big[Y_0(u)\big]},$$

于是我们可以写成

$$\mathcal{H}(u) = \frac{\Theta_2(u)}{\Theta_0(u)} - \left(\frac{\Theta_1(u)}{\Theta_0(u)}\right)^{\otimes 2} = \frac{E\big[Y_0(u)(X - \mu_{X|E_{\beta_0}}(u))^{\otimes 2}\big]}{E[Y_0(u)]};$$

因此, $\mathcal{H}(u)\Theta_0(u) = V(u)S(u)$, 其中

$$\mathcal{V}(u) = E\big[Y_0(u)(X - \mu_{X|E_{\beta_0}}(u))^{\otimes 2}\big].$$

由此可得, (17) 可改写为

$$\Sigma = \frac{1}{4} \int_{-\infty}^{\infty} S^2(u) \mathcal{V}(u) dF(u). \tag{45}$$

经简单代数运算可得

$$\mathcal{V}(u) = E\big[Y_0(u)(X - \mu_X)^{\otimes 2}\big] - (\mu_{X|E_{\beta_0}}(u) - \mu_X)^{\otimes 2} E\big[Y_0(u)\big];$$

因此, $\Sigma = \Sigma_1 - \Sigma_2$ 其中

$$\Sigma_1 = \frac{1}{4} \int_{-\infty}^{\infty} E\big[Y_0(u)(X - \mu_X)^{\otimes 2}\big] S^2(u) dF(u)$$

and

$$\Sigma_2 = \frac{1}{4} \int_{-\infty}^{\infty} (\mu_{X|E_{\beta_0}}(u) - \mu_X)^{\otimes 2} E\big[Y_0(u)\big] S^2(u) dF(u).$$

由于 Σ_1 和 Σ_2 均为半正定, 因此矩阵 Σ_1 必然会对变异性作出保守评估。在下一小节中, 我们将展示 Σ_1 可以简化, 并且具有一种有趣且实用的解释。

A.4.2 对 Σ_1 的简化

令 $\Sigma_X = E[(X-\mu_X)^{\otimes 2}]$，我们可以写作

$$\begin{aligned}\Sigma_1 &= \frac{\Sigma_X}{4}\int_{-\infty}^{\infty} S^2(u)dF(u) - \frac{1}{4}\int_{-\infty}^{\infty} E\big[(1-Y_0(u))(X-\mu_X)^{\otimes 2}\big]S^2(u)dF(u)\\ &= \frac{\Sigma_X}{12} - \frac{1}{4}\int_{-\infty}^{\infty} E\big[(1-Y_0(u))(X-\mu_X)^{\otimes 2}\big]S^2(u)dF(u),\end{aligned}$$

由以下事实可得第二个结果：$\int_{-\infty}^{\infty} S^2(u)dF(u) = 1/3$. 回顾第 A.1 节可知$G(u|X) = P\{E_{\beta_0} \le \infty \mid X\}$；那么，由于 $3S^2(u)dF(u) = -dS^3(u)$，

$$\begin{aligned}&\frac{1}{4}\int_{-\infty}^{\infty} E\big[I\{E_{\beta_0}<u\}(X-\mu_X)^{\otimes 2}\big]S^2(u)dF(u)\\ &\quad= -\frac{1}{12}E\left[(X-\mu_X)^{\otimes 2}\left\{\int_{-\infty}^{\infty} G(u-|X)d[S^3(u)]\right\}\right]\\ &\quad= -\frac{1}{12}E\left[(X-\mu_X)^{\otimes 2}\left\{G(\infty\mid X)S^3(\infty) - G(-\infty\mid X)S^3(-\infty) - \int_{-\infty}^{\infty} S^3(u)dG(u\mid X)\right\}\right]\\ &\quad= \frac{1}{12}E\left[(X-\mu_X)^{\otimes 2}\int_{-\infty}^{\infty} S^3(u)dG(u\mid X)\right]\\ &\quad= \frac{1}{12}E\big[(X-\mu_X)^{\otimes 2}S^3(E_{\beta_0})\big];\end{aligned}$$

因此，

$$\Sigma_1 = \frac{\Sigma_X}{12} - \frac{1}{12}E\big[(X-\mu_X)^{\otimes 2}S^3(E_{\beta_0})\big].$$

为了更好地理解这个最后表达式的性质，我们现在推导 Σ_1 与 $\mathrm{Var}(n^{-1/2}\Psi_n(\beta_0))$之间的关系，其中 $\Psi_n(\beta)$ 由式 (12) 给出。定义 $\mathcal{R}_{\beta_0,\,i} = \mathcal{R}(\tilde{E}_{i,\,\beta_0})$ 为式 (9) 的总体对应量；也就是说，

$$\mathcal{R}_{\beta_0,i} = \Delta_i F^{\star}(\tilde{E}_{i,\beta_0}) + (1-\Delta_i)\left[1 - \frac{1}{2}S(\tilde{E}_{i,\beta_0})\right].$$

此外，令 $\tilde{\Psi}_n(\beta_0) = \sum_{i=1}^n (X_i - \bar{X})\mathcal{R}_{\beta_0,\,i} = \sum_{i=1}^n (X_i-\bar{X})(\mathcal{R}_{\beta_0,\,i} - 0.5)$。那么，在所施加的正则条件下，

$$\begin{aligned}n^{-1/2}\tilde{\Psi}_n(\beta_0) &= n^{-1/2}\sum_{i=1}^n (X_i-\mu_X)(\mathcal{R}_{\beta_0,i}-0.5) - n^{-1/2}(\bar{X}-\mu_X)\sum_{i=1}^n(\mathcal{R}_{\beta_0,i}-0.5)\\ &= n^{-1/2}\sum_{i=1}^n (X_i-\mu_X)(\mathcal{R}_{\beta_0,i}-0.5) - n^{1/2}(\bar{X}-\mu_X)\times\frac{1}{n}\sum_{i=1}^n(\mathcal{R}_{\beta_0,i}-0.5)\\ &= n^{-1/2}\sum_{i=1}^n (X_i-\mu_X)(\mathcal{R}_{\beta_0,i}-0.5) + o_P(n^{-1/2})\end{aligned}$$

并且，作为独立同分布假设的结果，

$$\mathrm{Var}(n^{-1/2}\tilde{\Psi}_n(\beta_0)) \;\to\; \mathrm{Var}\{(X_1-\mu_X)\{\mathcal{R}_{\beta_0,1}-0.5\}\} = E\left[(X_1-\mu_X)^{\otimes 2}\mathrm{Var}(\mathcal{R}_{\beta_0,1}|X_1)\right],$$

其中，第二个等式由条件方差公式得出，因为

$$\mathrm{Var}(E[(X_1-\mu_X)^{\otimes 2}\{\mathcal{R}_{\beta_0,1}-0.5\}|X_1])=\mathrm{Var}((X_1-\mu_X)^{\otimes 2}E[\mathcal{R}_{\beta_0,1}-0.5|X_1])=0.$$

利用定理1的结果，

$$\mathrm{Var}(\mathcal{R}_{\beta_0,1}|X_1)=\frac{1}{12}\left(1-E(S^3(E_{\beta_0,1})|X_1)\right);$$

便可立即得到

$$\mathrm{Var}(n^{-1/2}\tilde{\Psi}_n(\beta_0))\to E\Big[(X_1-\mu_X)^{\otimes 2}\mathrm{Var}(\mathcal{R}_{\beta_0,1}|X_1)\Big]=\Sigma_1$$

因此，$n\to\infty$. Σ_1 表示$n^{-1/2}\tilde{\Psi}_n(\beta_0)$的极限方差。该方差的乘积形式与以下观察一致：$X_i$ 和 $R_{\beta_0,\ i}$ 不相关；这是因为

$$E[n^{-1}\tilde{\Psi}_n(\beta_0)]=0=E[(X_1-\mu_X)\{\mathcal{R}_{\beta_0,1}-0.5\}].$$

由于 $\Sigma=\Sigma_1-\Sigma_2$，因此，可以将 Σ_2 的额外贡献视为源于$\tilde{}$ $\hat{}$ $\tilde{}$在估计 β 时，需要通过（12）将 $\mathcal{R}(E_{i,\beta_0})$ 替换为 $\mathcal{R}(E_{i,\beta_0})$；与使用（未知的）残差分布函数相比，这实际上降低了变异性。

A.5 定理5的证明

固定任意有限的 β，并回顾随机过程 $N_i(u;\beta)$ $Y_i(u;\beta)$ 的定义，再令 $M_i(u;\beta)$ $=N_i(u;\beta)$ $-\int_{-\infty}^{u}Y_i(s;\beta)d\Lambda_H(s)$，其中 Λ_H 是与某个任意正规累积分布函数 H 相对应的累积风险函数。此外，令 $\gamma_a(t;H)$ 和 $\Gamma_a(t;H)$ 分别按（33）和（34）定义，并对任意适当函数 $\gamma(\cdot)$ 定义泛函

$$W_H\gamma(t)=\gamma(t)-\frac{\int_t^{\infty}\gamma(s)dH(s)}{1-H(t)};\tag{46}$$

那么，对于每个满足 $H(t)<1$ 的 t，令 $d\Lambda_H(s)=dH(s)/(1-H(s-))$,

$$\int_{-\infty}^{t}W_H\gamma(s)d\Lambda_H(s)=\int_{-\infty}^{\infty}\gamma(s)dH(s)-\frac{\int_t^{\infty}\gamma(s)dH(s)}{1-H(t)}.\tag{47}$$

这个等式在Ritov [1990], 的命题 3.1 的证明中得到证明，其中还假设 $\int_{-\infty}^{\infty}\gamma(s)dH(s)=0$. 那么，

$$\begin{aligned}\int_{-\infty}^{\infty}W_H\gamma_a(u;H)dM_i(u;\beta)&=\int_{-\infty}^{\infty}W_H\gamma_a(u;H)dN_i(u;\beta)-\int_{-\infty}^{\infty}W_H\gamma_a(u;H)Y_i(u;\beta)d\Lambda_H(u)\\&=\int_{-\infty}^{\infty}W_H\gamma_a(u;H)dN_i(u;\beta)-\int_{-\infty}^{\tilde{E}_{i,\beta}}W_H\gamma_a(u;H)d\Lambda_H(u)\\&=\int_{-\infty}^{\infty}W_H\gamma_a(u;H)dN_i(u;\beta)-\left(\int_{-\infty}^{\infty}\gamma_a(u;H)dH(u)-\Gamma_a(\tilde{E}_{i,\beta})\right),\end{aligned}\tag{48}$$

最 t 步可由先应用（47），再使用（34）得到。利用（46），

$$\begin{aligned}\int_{-\infty}^{\infty}W_H\gamma_a(u;H)dN_i(u;\beta)&=\int_{-\infty}^{\infty}\left(\gamma_a(u;H)-\Gamma_a(u;H)\right)dN_i(u;\beta)\\&=\Delta_i\left(\gamma_a(\tilde{E}_{i,\beta};H)-\Gamma_a(\tilde{E}_{i,\beta};H)\right);\end{aligned}$$

代入（48）并化简、重排各项后可得

$$\Delta_i\gamma_a(\tilde{E}_{i,\beta};H)+(1-\Delta_i)\Gamma_a(\tilde{E}_{i,\beta};H) = \int_{-\infty}^{\infty} W_H\gamma_a(u;H)dM_i(u;\beta)+\int_{-\infty}^{\infty}\gamma_a(u;H)dH(u). \tag{49}$$

恒等式（49）是证明定理中所述两个结论的关键。

证明 $E(\mathcal{R}_{\beta_0}^{(a)}\mid X)=E(\mathcal{R}_{\beta_0}^{(a)})=A(1)-A(0)$。

回顾 $\mathcal{R}_{\beta_0}^{(a)}=\Delta\gamma_a(\tilde{E}_{\beta_0};F)+(1-\Delta)\Gamma_a(\tilde{E}_{\beta_0};F)$，其中 $\tilde{E}_{\beta_0}=Y-X^T\beta_0$ 是实际的观测残差，Δ 是实际的失效状态，F 是 $\epsilon=\log T^*-X^T\beta_0$ 的真实累积分布函数。省略所有个体指标，定义 $N(u)=N(u;\beta_0)=I\{E_{\beta_0}\le u,\Delta=1\}$、$Y(u)=Y(u;\beta_0)=I\{E_{\beta_0}\ge u\}$，以及 $M(u)=N(u)-\int_{-\infty}^{u}Y(s)d\Lambda_F(s)$，其中 Λ_F 是对应于 F 的累积风险函数。然后，利用（49），

$$\mathcal{R}_{\beta_0}^{(a)}=\int_{-\infty}^{\infty}W_H\gamma_a(u;F)dM(u)+\int_{-\infty}^{\infty}\gamma_a(u;F)dF(u).$$

在本文的条件下，$M(u)$ 是均值为零的鞅；此外，鉴于（35），

$$\int_{-\infty}^{\infty}\gamma_a(u;F)dF(u)=A(1)-A(0)$$

因此，由于 $\gamma_a(\cdot;F)$ 是一个非随机函数，故为可预测的，

$$E\big[\mathcal{R}_{\beta_0}^{(a)}|X\big]=A(1)-A(0)=E\big[\mathcal{R}_{\beta_0}^{(a)}\big],$$

从而证明了所需结果。

式（36）的证明

下面，我们证明 $\sum_{i=1}^{n}\widehat{\mathcal{R}}^{(a)}(\tilde{E}_{i,\beta})=n\int_{-\infty}^{\infty}\gamma_a(t;\widehat{F}_\beta)\,d\widehat{F}_\beta(t)$，其中

$$\widehat{\mathcal{R}}^{(a)}(\tilde{E}_{i,\beta})=\Delta_i\gamma_a(\tilde{E}_{i,\beta};\widehat{F}_\beta)+(1-\Delta_i)\Gamma_a(\tilde{E}_{i,\beta};\widehat{F}_\beta)$$

以及 $\gamma_a(t;F_\beta)$ 以及 $\Gamma_a(t;F_\beta)$ 定义同（33）和（34）中，且 $H=F_\beta$. 所述结果将直接由以下事实得出

$$\int_{-\infty}^{\infty}\gamma_a(t;\widehat{F}_\beta)d\widehat{F}_\beta(t)=\sum_{-\infty<t<\infty}\frac{A(\widehat{F}_\beta(t))-A(\widehat{F}_\beta(t-))}{\widehat{F}_\beta(t)-\widehat{F}_\beta(t-)}\times\Big(\widehat{F}_\beta(t)-\widehat{F}_\beta(t-)\Big)=A(1)-A(0).$$

为证明 $\sum_{i=1}^{n}\widehat{\mathcal{R}}^{(a)}(\tilde{E}_{i,\beta})=n\int_{-\infty}^{\infty}\gamma_a(t;\widehat{F}_\beta)d\widehat{F}_\beta(t)$，考虑（49），其中 $H=\widehat{F}_\beta$；于是，我们可以写成

$$\widehat{\mathcal{R}}^{(a)}(\tilde{E}_{i,\beta})=\int_{-\infty}^{\infty}W_{\widehat{F}_\beta}\gamma_a(u;\widehat{F}_\beta)dM_i(u;\widehat{\beta})+\int_{-\infty}^{\infty}\gamma_a(u;\widehat{F}_\beta)d\widehat{F}_\beta(u).$$

因此，

$$\sum_{i=1}^{n}\widehat{\mathcal{R}}^{(a)}(\tilde{E}_{i,\beta})=\int_{-\infty}^{\infty}W_{\widehat{F}_\beta}\gamma_a(u;\widehat{F}_\beta)dM_+(u;\widehat{\beta})+n\int_{-\infty}^{\infty}\gamma_a(u;\widehat{F}_\beta)d\widehat{F}_\beta(u).$$

然而，鉴于（21），

$$dM_+(u;\widehat{\beta})=dN_+(u;\widehat{\beta})-Y_+(u;\widehat{\beta})d\widehat{\Lambda}_{\widehat{\beta}}(u)=0,$$

所以 $\sum_{i=1}^{n}\widehat{\mathcal{R}}^{(a)}(\tilde{E}_{i,\beta})=n\int_{-\infty}^{\infty}\gamma_a(u;\widehat{F}_\beta)d\widehat{F}_\beta(u)=n(A(1)-A(0))$，如所要求。