删失数据下分位数回归曲线的比较

Lorenzo Tedesco * 和 Ingrid Van Keilegom *

ORSTAT，鲁汶大学，比利时

arXiv:2302.03440v1 [stat.ME] 7 Feb 2023

摘要

本文提出了一种新的检验方法，用于比较条件分位数曲线，其中所关注的结果变量通常为持续时间，并且受到右删失的影响。该检验既适用于两个独立样本的情形，也适用于配对数据；既可用于比较固定分位数水平上的分位数，也可用于比较有限个分位数水平或一系列分位数水平上的分位数。在原假设和局部备择假设下，本文分别推导了所提出检验统计量的渐近分布。我们描述了一种用于近似临界值的自助法程序，并给出了一项模拟研究的结果。在该研究中，我们考察并比较了检验方法在小样本量和中等样本量下的表现，以及备择检验的表现。最后，我们将所提出的检验应用于一组有关糖尿病视网膜病变的数据。

关键词：自助法、删失数据、分位数曲线比较、分位数回归、生存分析。

*研究经费由欧洲研究委员会提供（2016—2022年，地平线2020/欧洲研究委员会资助协议第694409号）。

1 引言

检验分布或分位数曲线是否相等，是实证研究中的一项重要任务。例如，著名的配对 t 检验、双样本 t 检验或 Wilxocon-Mann Whitney 检验，都是基础统计学教科书中的内容。本文旨在研究两个独立样本的比较以及配对数据的比较，并考虑上述基础检验设定中的若干复杂情况。首先，在我们的设定中存在协变量，因此我们将基于这些协变量对样本进行条件比较。其次，我们将通过线性分位数回归模型（其误差分布未知）所得到的条件分位数来比较样本。第三，该分位数模型中的响应变量允许存在右删失。

在关于两个独立样本或配对数据比较的文献中，已有若干检验部分满足我们设定中的约束，但据我们所知，迄今尚无论文研究我们感兴趣的设定。例如，Koul 和 Schick（1997）、Sun（2006）、Dette 等人（2011）、Dette 等人（2013）、Delgado 和 Escanciano（2013）等论文都比较了条件分位数函数。然而，这些论文假设响应变量完全可观测，而在涉及持续时间变量的实验中，情况往往并非如此。另一方面，Li 等人（1996）介绍了删失数据情形下分位数曲线的比较，但其研究设定不包含协变量。在分位数曲线比较中纳入协变量，可能会推动这一领域取得重大进展。最后，Vilar-Fernández 和 González-Manteiga（2004）提出了相关误差下曲线的非参数比较方法，但其程序不适用于分位数曲线或删失数据；此外，他们所指的相关性存在于各样本内部的观测值之间，而非样本之间。

该领域的另一项重要贡献是 Pardo-Fernández 和 Van Keilegom（2006）的论文，他们研究了删失结局情形下回归曲线的比较。不过，他们回归模型中的位置泛函属于某一类 L-泛函，其中包括某些截尾均值作为特例，但不包括任何分位数。此外，他们假设模型中的误差与协变量相互独立，而我们不希望作此假设。因此，他们的研究框架仍与我们想要研究的问题不同。Lee（2009）的论文提出了一种比较删失响应情形下条件分布的方法，这同样是一个相关但不同于我们想要研究的问题。该论文中的协变量在两个样本中的分布不一定相同，我们的设定也是如此；但比较分布函数不同于比较单个或有限个分位数水平上的分位数。此外还需注意，比较条件中位数等量，可以让我们具体了解两个样本的相对表现。最后一篇与此有些相关的论文是 Sant'Anna（2014）的研究，该文提出了一种在删失响应情形下比较条件分布和单条分位数曲线的方法。然而，该方法要求两个样本中的协变量具有相同的分布，这在某些实际情形中可能会造成问题。此外，该论文中的回归模型完全采用非参数形式，而我们更倾向于使用线性回归模型，因为这类模型更易于解释，也不会遭遇维数灾难问题。

本文　　组织结构如下。第2节介绍分位数回归模型，并提出

我们希望检验的原假设。第3节介绍用于比较的检验统计量，分别适用于配对数据和独立样本的情形，并在关于条件分位数估计量的高层条件下，给出所提出检验在原假设和局部备择假设下的渐近分布。第4节针对若干特定估计量验证这些高层条件。第5节介绍一种自助法程序，用以近似计算检验的临界值。第6节通过模拟研究所提出检验的表现；第7节则运用所提出的方法分析糖尿病视网膜病变数据。最后，第8节总结全文并探讨未来研究方向；附录包含主要结果的证明及补充模拟结果。

2 模型

我们考虑一个用于比较两条分位数回归曲线的通用框架，这两条曲线可以基于配对数据或独立样本进行估计。该框架允许比较一个分位数水平、有限个分位数水平或一个分位数水平区间上的分位数曲线。我们重点讨论两个样本中的响应变量均受到随机右删失的情形，但也可以类似地将该方法推广到其他类型的不完整数据。此外，所提出的方法基于两个样本中分位数曲线的通用估计量。这些估计量需要满足某些高层条件，而特定估计量可以满足这些条件。

我们需要引入以下记号。对于 $j = 1,2$，生存时间 $T_j > 0$ 满足

$$\log T_j = \beta_{j0}(\tau) + \beta_{j1}(\tau)Z_{j,1} + \ldots + \beta_{jp}(\tau)Z_{j,p} + \epsilon_j(\tau). \tag{2.1}$$

这里，$Z_j = (1, Z_{j,1}, \ldots, Z_{j,p})^T$ 是一个协变量向量，误差 $\epsilon_j(\tau)$ 的分布未作具体规定，唯一的限制是 $P(\epsilon_j(\tau) \leq 0 | Z_j) = \tau$。因此，给定 Z_j，误差 $\epsilon_j(\tau)$ 的 τ 分位数等于 0。对于给定的 $\tau \in (0,1)$，令 $\beta_j(\tau) = (\beta_{j0}(\tau) \ \beta_{j1}(\tau) \ldots, \beta_{jp}(\tau))^T$ 为系数向量，因此我们也可以写成 $\log T_j = Z_j^T \beta_j(\tau) + \epsilon_j(\tau)$。时间变量的对数变换并非严格必需，但通常会将（正值）响应变量 T_j 转换为定义在整个实数轴上的变量，以便在实际应用中获得更好的拟合。

我们感兴趣的是检验

$$H_0 : \beta_1(\tau) = \beta_2(\tau) \quad \text{for all } \tau \in A, \tag{2.2}$$

与一般备择假设相对，在该假设下，至少存在一个 τ 中的值，使得 $\beta_1(\tau)$ 不等于 $\beta_2(\tau)$。这里，A 可以是单元素集、有限点集，或 $(0, 1)$ 中的一个区间。请注意，可以考虑对原假设 H_0 进行调整。例如，可以检验 $R\beta_1(\tau) = R\beta_2(\tau)$，其中 R 是一个矩阵；也可以检验 $\gamma(\beta_1(\tau)) = \gamma(\beta_2(\tau))$，其中 γ 是一个可能非线性但可微的函数。特别是，这意味着可以针对固定的协变量向量 z 检验 $\beta_1(\tau)^T z = \beta_2(\tau)^T z$；也就是说，我们可以检验在特定协变量取值下两个分位数之间是否存在显著差异。我们不再对此进一步展开，但了解这一点是可行的很有用。

由于随机右删失，我们观察到的不是 T_j for $j = 1,2$，而是 $X_j = \min(T_j, C_j)$ 和 $\Delta_j = I(T_j \leq C_j)$，其中假设删失时间 C_j 在给定 Z_j 的条件下与 T_j 独立。数据

由两个独立同分布样本 $(X_{ji}, \Delta_{ji}, Z_{ji})_{i=1}^{n_j}$ 组成，其分布与 (X_j, Δ_j, Z_j) 相同，其中 $j = 1, 2$；对于独立样本，n_1 和 n_2 是样本量，而对于配对数据，样本量为 $n_1 = n_2 = n$。后一种情况下，数据来自相同的受试者（例如，来自双眼或两个不同时间点），因此对于每个固定的 i，$(X_{1i}, \Delta_{1i}, Z_{1i})$ 和 $(X_{2i}, \Delta_{2i}, Z_{2i})$ 相依。

基于这些数据，我们将估计回归系数向量 $\beta_1(\tau)$ 和 $\beta_2(\tau)$。关于含删失数据的线性分位数回归，已有大量研究，例如 Portnoy（2003）、Peng 和 Huang（2008）、Wang 和 Wang（2009）、Portnoy and Lin (2010)、Yang 等人（2018）、De Backer 等（2019）和 De Backer 等人（2020）。有关这些估计量的更多详情，请参阅 Peng（2021）近期的综述。本文将考察 Portnoy（2003）、Peng 和 Huang（2008）、Wang 和 Wang（2009）以及 De Backer 等（2019）提出的估计量。后两种估计量可用于单个 τ 值，而 Portnoy（2003）和 Peng 和 Huang（2008）的估计量则在一定范围的 τ 值上满足某些渐近结果。

我们以一条关于配对数据情形的说明结束本节。原假设意味着分位数之差等于零，但对于配对数据，也可以检验一对配对数据点之差的分位数是否等于零。与均值函数不同，分位数函数不是线性的（也就是说，差值的分位数不等于分位数之差），因此这两个量并不相等。我们认为，在当前语境下，分位数之差更有意义，因为它能够用于一般性的分布比较。

3 检验与渐近理论

本节将提出用于检验 H_0 的检验统计量，并分别针对两种设定（配对样本或独立样本）推导这些检验统计量的渐近分布。检验统计量将基于两个样本的某些估计量 $\beta_1(\cdot)$ 和 $\beta_2(\cdot)$。这些估计量取决于实际采用的估计方法。因此，本节将使用满足某些正则条件的一般估计量。下一节将验证特定估计量满足这些正则条件。

在独立样本的情况下，$n = n_1 + n_2$ 定义为；而在配对样本的情况下，n 表示配对数据的数量。所提出的检验统计量取决于权重函数 $w(\tau)$，并针对集合 A 为区间的情况给出。如果 A 是有限点集（或只包含一个分位数水平），则下文的积分需要替换为对 A中各元素的求和。我们考虑三种检验统计量：

- 平均 L_2-范数：

$$T_{L_2} = n^{1/2} \int_A \|\hat{\beta}_1(\tau) - \hat{\beta}_2(\tau)\|_2 w(\tau)\, d\tau, \tag{3.1}$$

  其中，$\|\cdot\|_2$ 是 $\mathbb{R}^{p+1}$中的欧几里得范数。

- 平均 L_∞-范数：

$$T_{L_\infty} = n^{1/2} \int_A \|\hat{\beta}_1(\tau) - \hat{\beta}_2(\tau)\|_\infty w(\tau)\, d\tau, \tag{3.2}$$

  其中，$\|\cdot\|_\infty$ 是 $\mathbb{R}^{p+1}$中的上确界范数。

- 经 Bonferroni 校正的平均绝对值：

$$T_{B,k} = n^{1/2} \int_A |\hat{\beta}_{1k}(\tau) - \hat{\beta}_{2k}(\tau)| w(\tau)\, d\tau \tag{3.3}$$

对于 $k = 0, \ldots, p$，其做法是：若与检验统计量 $T_{B,0}, \ldots, T_{B,p}$ 对应的 p 值中至少有一个小于 $\alpha/(p+1)$，则拒绝 H_0，其中 α 为选定的显著性水平。

需要指出的是，也可以考虑其他检验统计量（例如上确界范数$n^{1/2}\sup_\tau \|\hat{\beta}_1(\tau) - \hat{\beta}_2(\tau)\|_\infty$），但由于上述检验统计量的模拟结果更好，我们选择了它们。还需指出的是，可以很自然地调整检验统计量，以检验 β_1 的某个分量是否等于 β_2中对应的分量。如果只关注某个特定协变量的效应，这种做法可能会有用。

现在，我们分别针对配对样本和独立样本讨论这些检验统计量的渐近理论。对于这两种情形，我们假设估计量 $\beta_1(\cdot)$ 和 $\beta_2(\cdot)$ 均可写为独立同分布项之和，再加上一个渐近阶可忽略的余项。下一节将验证特定估计量是否满足这一条件。

$\hat{}$ (A) 估计量 $\beta_j(\cdot)$ $(j = 1, 2)$ 对所有 $\tau \in A$ 和 $k = 0, \ldots, p$ 都满足以下独立同分布表示：

$$\hat{\beta}_{jk}(\tau) - \beta_{jk}(\tau) = n_j^{-1} \sum_{i=1}^{n_j} g_{jk}(\tau, X_{ji}, \Delta_{ji}, Z_{ji}) + R_{jk}(\tau),$$

$\hat{}$ 对某个函数 g_{jk}，其满足 $E[g_{jk}(\tau, X_j, \Delta_j, Z_j)|Z_j] = 0$，且依赖于估计量 $\beta_j(\tau)$，其中 $\sup_{\tau \in A} |R_{jk}(\tau)| = o_P(n_j^{-1/2})$。此外，类 $\mathcal{F}_{jk} = \{(x, \delta, z) \to g_{jk}(\tau, x, \delta, z) : \tau \in A\}$ 是 P_j-Donsker 类，其中 P_j 是 (X_j, Δ_j, Z_j) 的联合概率分布。

条件（A）中的表示通常不会对所有 $0 < \tau < 1$ 都成立。由于右删失机制，对于大于某个阈值的 τ，无法估计 $\beta_j(\tau)$。更多细节请参见第4节。因此，集合 A 的选择取决于这个阈值。

设 $g_j = (g_{j0}, \ldots, g_{jp})$ 满足 $j = 1, 2$。正如我们将在下一节中看到的，对于某个单一的 τ值，第2节中提到的所有估计量都满足这一条件；对于 Peng 和 Huang（2008）以及 Portnoy（2003）的估计量，某个 τ–值区间内也满足这一条件。

3.1 独立样本

下文中，我们将采用以下关于样本量 n_1 和n_2的等比例增长（EG）假设：

(EG) 当 n 趋于无穷时， $n_1/n \to p_1$ 和 $n_2/n \to p_2$，其中 $0 < p_1, p_2 < 1$。

于是我们得到以下结果。

定理 3.1。 假设 *(A)* 和 *(EG)* 成立。则在 H_0 下，以 $\tau \in A$ 为索引的 $p+1)$ 维过程 $n^{1/2}(\hat{\beta}_1(\tau) - \hat{\beta}_2(\tau))$弱收敛于均值为零的高斯过程 W_I，其协方差函数为

$$\begin{aligned} Cov(W_I(\tau_1), W_I(\tau_2)) = & p_1^{-1} E\Big[g_1(\tau_1, X_1, \Delta_1, Z_1) g_1(\tau_2, X_1, \Delta_1, Z_1)^T\Big], \\ & + p_2^{-1} E\Big[g_2(\tau_1, X_2, \Delta_2, Z_2) g_2(\tau_2, X_2, \Delta_2, Z_2)^T\Big], \end{aligned}$$

对于 $\tau_1, \tau_2 \in A$。

证明见附录。

请注意，当 A 是有限点集而非区间时，高斯过程会化为多元正态随机向量。根据这一结果和连续映射定理，我们得到以下关于检验统计量渐近分布的推论。

推论 3.1. 假设 *(A)* 和 *(EG)* 成立。那么，在 H_0 下，

$$T_{L_2} \xrightarrow{d} \int_A \|W_I(\tau)\|_2 w(\tau)\, d\tau, \qquad T_{L_\infty} \xrightarrow{d} \int_A \|W_I(\tau)\|_\infty w(\tau)\, d\tau,$$
$$T_{B,k} \xrightarrow{d} \int_A |W_{I,k}(\tau)| w(\tau)\, d\tau,$$

对于 $k = 0$、.....、 p。

与前文一样，上述结果针对 A 为区间的情形推导而得。如果 A 是有限点集，只需将积分替换为求和即可。

Co现在考虑检验统计量在以下局部备择假设下的极限行为 :

$$H_1 : \beta_1(\tau) - \beta_2(\tau) = n^{-1/2} b(\tau) \quad \text{for all } \tau \in A,$$

其中 b 是某个连续且可积的函数。那么，我们有以下结果:

推论 3.2。 假设 *(A)* 和 *(EG)* 成立。那么，在 H_1 下，

$$T_{L_2} \xrightarrow{d} \int_A \|W_I(\tau) + b(\tau)\|_2 w(\tau)\, d\tau, \qquad T_{L_\infty} \xrightarrow{d} \int_A \|W_I(\tau) + b(\tau)\|_\infty w(\tau)\, d\tau,$$
$$T_{B,k} \xrightarrow{d} \int_A |W_{I,k}(\tau) + b_k(\tau)| w(\tau)\, d\tau,$$

对于 $k = 0, \ldots, p$。

证明见附录。

3.2 配对样本

配对样本情形下的结果可用与独立样本情形类似的方法得到。主要区别在于极限过程协方差函数的推导，下一条结果将对此进行说明。

定理 3.2。 假设 *(A)* 成立。那么，在 H_0 下，由 $\tau \in A$ 索引的 $(p+1)$ 维过程 $n^{1/2}(\hat{}\beta_1(\tau) - \hat{}\beta_2(\tau))$ 弱收敛于一个均值为零的高斯过程 W_P，其协方差函数为

$$Cov(W_P(\tau_1), W_P(\tau_2)) = E\Big[\big\{g_1(\tau_1, X_1, \Delta_1, Z_1) - g_2(\tau_1, X_2, \Delta_2, Z_2)\big\} \times \big\{g_1(\tau_2, X_1, \Delta_1, Z_1) - g_2(\tau_2, X_2, \Delta_2, Z_2)\big\}^T\Big],$$

对于 $\tau_1, \tau_2 \in A$。

证明见附录。

与独立样本的情形类似，我们现在得到以下推论，其中结合了 H_0 和 H_1 下的结果（只需在 H_0下取 $b \equiv 0$ ）：

推论 3.3. 假设 *(A)* 成立。那么，在 H_1 下，

$$T_{L_2} \xrightarrow{d} \int_A \|W_P(\tau) + b(\tau)\|_2 w(\tau)\, d\tau, \qquad T_{L_\infty} \xrightarrow{d} \int_A \|W_P(\tau) + b(\tau)\|_\infty w(\tau)\, d\tau,$$
$$T_{B,k} \xrightarrow{d} \int_A |W_{P,k}(\tau) + b_k(\tau)| w(\tau)\, d\tau,$$

对于 $k = 0, \ldots, p$。

4 分位数回归估计量的选择

本节将介绍文献中针对响应变量被删失情形提出的一些分位数估计量。我们将说明这些估计量的定义，以及对于集合 A 和函数 g_j ($j = 1$，2) 的哪些取值，这些估计量满足假设 (A)。由此，我们可以得出结论：上一节给出的渐近结果适用于这些估计量。

4.1 Peng 和 Huang（2008）的估计量

该估计量利用随机删失数据所蕴含的鞅结构，为模型 (2.1) 构造估计方程。具体而言，令 $\Lambda_T(t|Z) = -\log(1-P(T \le t|Z))$ 为给定 Z 时 T 的累积风险函数，其中 $N(t) = I(X \le t,\ \Delta = 1)$ 且 $M(t) = N(t) - \Lambda_T(t \wedge X|Z)$。令 $N_{ji}(t)$ 和 $M_{ji}(t)$ 分别为样本 $j = 1$，2 中观测 i 的 $N(t)$ 和 $M(t)$ 的对应形式。那么，$M_{ji}(t)$ 是与计数过程 $N_{ji}(t)$ 相关联的鞅过程。因此，对于 $t > 0$，$E[M_{ji}(t)|Z_{ji}] = 0$

$$E\Big\{ \sum_{i=1}^{n_j} Z_{ji} \Big[N_{ji}(\exp\big(Z_{ji}^T \beta_j(\tau)\big)) - \Lambda_T\big(\exp(Z_{ji}^T \beta_j(\tau)) \wedge X_{ji} | Z_{ji}\big) \Big] \Big\} = 0$$

对每个 τ。现在，$Z_{ji}^T \beta_j(\tau)$ 关于 τ 单调，因此很容易看出 $\Lambda_T(\exp(Z_{ji}^T \beta_j(\tau)) \wedge X_{ji}|Z_{ji}) = \int_0^\tau I(X_{ji} \ge \exp(Z_{ji}^T \beta_j(u)))dH(u)$，其中 $H(x) = -\log(1-x)$，前提是模型 (2.1) 对所有$0 \le u \le \tau$都成立。这些结果促使我们求解估计方程

$$S_{jn_j}(\beta_j, \tau) = 0, \tag{4.1}$$

其中

$$S_{jn_j}(\beta_j, \tau) = \frac{1}{n_j} \sum_{i=1}^{n_j} Z_{ji} \Big[N_{ji}(\exp\big(Z_{ji}^T \beta_j(\tau)\big)) - \int_0^\tau I\big(X_{ji} \ge \exp(Z_{ji}^T \beta_j(u))\big) dH(u) \Big]. \tag{4.2}$$

ˆ因此，通过近似求解 (4.1) 的随机解，可以得到 $\beta_j(\tau)$ 的估计量 $\beta_j(\tau)$。在温和条件下（参见 Peng 和 Huang （ 2008 ）中的第 3 节），该估计量一致相合，并弱收敛到一个高斯过程。

尚待证明的是，假设 (A) 对 $A = [\tau_L, \tau_R]$ 成立，其中 $0 < \tau_L < \tau_R$ 和τ_R 必须小于 C_j 的支撑集上界处 T_j 的累积分布，且$j = 1$，2。正如 Peng 和 Huang（2008）定理 2 的证明所示，该类

$$\mathcal{C} = \Big\{ Z_{ji} \Big[N_{ji}(\exp\big(Z_{ji}^T \beta_j(\tau)\big)) - \int_0^\tau I\big(X_{ji} \ge \exp(Z_{ji}^T \beta_j(u))\big) dH(u) \Big] : \tau \in [\tau_L, \tau_R] \Big\} \tag{4.3}$$

当 $\beta_j(\tau)$ 等于真实参数时，它属于 Donsker 类；此外，

$$E\Big\{ Z_{ji} \Big[N_{ji}\big(\exp(Z_{ji}^T \beta_j(\tau))\big) - \int_0^\tau I\big(X_{ji} \ge \exp(Z_{ji}^T \beta_j(u))\big) dH(u) \Big] \Big| Z_{ji} \Big\} = 0 \tag{4.4}$$

根据鞅性质。因此，量 S_{jn_j} 可以看作是在 Donsker 类 (4.3) 上求值的经验过程。现在，Peng 和 Huang（2008）证明，差值 $n_j^{1/2}(\hat{}\beta_j(\tau) - \beta_j(\tau))$ 可以写成

$$n_j^{1/2}(\hat{\beta}_j(\tau) - \beta_j(\tau)) = n_j^{1/2}\tilde{\phi}(-S_{jn_j}(\beta_j, \tau)) + o_{[\tau_L,\tau_R]}(1), \tag{4.5}$$

其中，$o_I(1)$ 是一个关于 $\tau \in I$ 依概率一致收敛于零的量，而 ϕ 是一个线性映射，因此它保留了 Donsker 性质（$\tilde{\phi}(\cdot)(\tau)$ 是定义在第 648 页上的映射 $B(\beta(\tau))^{-1}\phi(\cdot)(\tau)$，见 Peng 和 Huang（2008））。注意，(4.3)-(4.5) 等价于说 β_j 满足条件（A），其中 g_{jk} 对应于 $g_j(\tau,\ X_j,\ \Delta_j,\ Z_j) = Z_j\left[N_j(\exp(Z_j^T\beta_j(\tau))) - \int_0^\tau I(X_j \geq \exp(Z_j^T\beta_j(u)))dH(u)\right]$ 的第 k 个分量。

4.2 其他估计量

本节简要讨论文献中提出的另外三种估计量，即 Wang 和 Wang（2009）、Portnoy（2003）以及 De Backer 等（2019）提出的估计量，并解释为什么这些估计量满足条件（A）。

4.2.1 Wang 和 Wang（2009）估计量

Wang 和 Wang（2009）提出的估计量基于对标准分位数损失函数的修改，并融入了自洽Kaplan-Meier估计量的思想。这意味着，该估计量通过局部加权方案，将每个删失数据点所对应的概率质量重新分配到右侧。现在，通过证明 $n_j^{1/2}(\hat{}\beta_j(\tau) - \beta_j(\tau))$ 是均值为零的独立同分布随机向量之和，再加上一个阶为$o_P(1)$的误差项，并应用中心极限定理，可以得到其渐近分布。因此，对于单个 τ 值，$n_j^{1/2}(\hat{}\beta_j(\tau) - \beta_j(\tau))$ 可视为由单个函数组成的函数族所对应的经验过程，因此它是 Donsker 的。总之，当集合 A 为单元素集时，Wang 和 Wang（2009）估计量也满足条件（A）。该估计量不要求模型（2.1）对所有小于 τ 的分位数水平都成立。

4.2.2 Portnoy（2003）的估计量

通过调整自洽Kaplan-Meier估计量中重新分配删失概率的思路，Portnoy（2003）的估计量采用迭代自洽算法来估计全局线性分位数回归模型。Neocleous等（2006）将原论文提出的算法简化为基于网格的序贯估计程序，其相应的渐近研究见 Portnoy and Lin (2010)。后者证明，对于固定的 τ，$n_j^{1/2}(\hat{}\beta_j(\tau) - \beta_j(\tau))$ 收敛于一个正态随机变量。Peng (2012) 基于随机积分方程提出了自洽方法的另一种表述，并证明该估计量（记为 $\beta_{P,\ j}(\tau)$）$\hat{}$ 与 Peng and Huang（2008）估计量（记为 $\beta_{PH,\ j}(\tau)$）$\hat{}$ 渐近等价，即 $\sup_{\tau\in[\tau_L,\tau_R]} \|n_j^{1/2}(\hat{}\beta_{P,\ j}(\tau) - \hat{}\beta_{PH,\ j}(\tau))\|_2 \to_p 0$。总之，为证明 Portnoy（2003）估计量满足条件（A），当集合 A 为单元素集时，可以采用 Wang 和 Wang（2009）类似的论证。对于 τ 的多个取值，尚不清楚如何直接证明条件（A），但可以利用其与 Peng and Huang（2008）估计量的等价性，因为我们已经证明后者满足条件（A）。

4.2.3 De Backer 等人（2019）的估计量

De Backer 等人（2019）提出了另一种可能的估计量。该估计量基于经过调整的标准分位数损失函数，可适用于随机删失数据。更准确地说，定义$G_{C|Z}(c|Z) = \mathbb{P}(C \geq c|Z)$，则 $\phi_\tau(t, Y, G_{C|Z}(\cdot|Z)) = (Y - t)\{\tau - I(Y \leq t)\} - (1-\tau)\int_0^t \{1 - G_{C|Z}(s|Z)\}ds$ 对 t 的导数等于 $-\{I(Y > t) - G_{C|Z}(t|Z)(1-\tau)\}$。因此，在给定 Z 的条件下，当模型（2.1）下 $t = \beta(\tau)^\top Z$ 时，该导数的期望为零。在上述论文中，作者利用这一性质为删失分位数回归构造了一个调整后的损失函数，即 $\sum_{i=1}^n \phi_\tau(Z_i^\top \beta, Y_i, \hat{G}_{C|Z}(\cdot|Z_i))$，其中 $\hat{G}_{C|Z}(\cdot|Z_i)$ 是给定 Z 时 C 的生存函数的一致估计量。根据 C 和 T 之间的依赖关系，$G_{C|Z}$ 可以有不同的取法；更多细节请参见原论文及 Peng（2021）。随后，通过最小化非凸的调整后损失函数，可以得到 $\beta(\tau)$ 的一致且渐近正态估计量 $\beta(\tau)$。De Backer 等人（2019）还为此类最小化问题提供了一种数值稳定的MM算法。当集合 A 为单元素集时，条件（A）的有效性由 De Backer 等人（2019）的定理3.1和定理3.2保证。

5 自助法程序

现在，我们将提出一种通用的自助法，可用于近似第2节所述检验统计量的临界值，从而避免估计极限过程表达式中出现的未知密度函数，或其他难以估计的函数。文献中已探讨过在曲线比较背景下使用自助法技术来近似检验统计量的分布，相关例子可参见 Vilar-Fernández 等人（2007）。我们将结合文献中已有的分位数估计量来讨论这一通用方法。

假设我们有一个满足以下条件的自助法估计量 $\beta_j^*(\cdot)$：

(A*) 自助法估计量 $\beta_j^*(\cdot)$ $(j = 1, 2)$ 对所有 $\tau \in A$ 和$k = 0, \ldots, p$ 都满足以下独立同分布表示：

$$\hat{\beta}_{jk}^*(\tau) - \hat{\beta}_{jk}(\tau) = n_j^{-1} \sum_{i=1}^{n_j} \eta_{ji} g_{jk}(\tau, X_{ji}, \Delta_{ji}, Z_{ji}) + R_{jk}^*(\tau),$$

对于独立样本的情形，取某些均值为零且方差等于 1 的独立同分布变量 η_{ji}，$i = 1, \ldots, n_j$，$j = 1, 2$；对于配对样本的情形，取某些均值为零且方差等于 1 的独立同分布变量 $\eta_{1i} = \eta_{2i}$，$i = 1, \ldots, n = n_1 = n_2$。这里，对于数据的几乎所有实现（$X_{ji}$，$\Delta_{ji}$，$Z_{ji}$），都有 $\sup_{\tau \in A} |R_{jk}^*(\tau)| = o_{P_j^*}(n_j^{-1/2})_{i=1}^{n_j}$，函数 g_{jk} 与条件（A）中的相同，而P_j^* 表示 η_{ji}-变量的概率测度。

基于此，我们将证明，可以使用 $\beta_1^*(\tau) - \beta_2^*(\tau) - \beta_1(\tau) + \beta_2(\tau)$ 的经验分位数来近似 $\beta_1(\tau) - \beta_2(\tau) - \beta_1(\tau) + \beta_2(\tau)$ 的经验分位数。因此，可以通过将自助法过程与 $\beta_1(\tau) - \beta_2(\tau)$ 进行比较来检验 H_0。以下定理同时适用于独立样本和配对样本的情形（在后一种情形下 $n_1 = n_2 = n$）。

定理 5.1。 假设 *(A*)* 成立，并且在独立样本情况下满足定理 *3.1* 的假设、在配对样本情况下满足定理 *3.2* 的假设。则， $n^{1/2}(\hat{\beta}_1 - \hat{\beta}_2 - \beta_1 + \beta_2)(\tau)$ 收敛于一个

以 $\tau \in A$ 为索引的高斯过程 $W(\tau)$，并且对于 $\tau \in A$，条件于数据，几乎处处地，$n^{1/2}(\hat{}\beta_1^* - \hat{}\beta_2^* - \hat{}\beta_1 + \hat{}\beta_2)(\tau)$ 收敛于同一个高斯过程 $W(\tau)$。

证明见附录。

5.1 Peng 和 Huang（2008）的估计量

为了证明 Peng 和 Huang（2008）估计量满足条件 (A*)，我们定义

$$S_{jn_j}^*(\beta, \tau) = \frac{1}{n_j} \sum_{i=1}^{n_j} \tilde{\eta}_{ji} Z_{ji} \Big[N_{ji}\big(\exp(Z_{ji}^T \beta(\tau))\big) - \int_0^\tau I\big(X_{ji} \geq \exp(Z_{ji}^T \beta(u))\big) dH(u) \Big], \tag{5.1}$$

其中，在独立样本的情形下，$\tilde{\eta}_{ji} (i = 1, \ldots, n_j,\ j = 1, 2)$ 独立同分布且服从单位指数分布；在配对样本的情形下，$\tilde{\eta}_{1i} = \tilde{\eta}_{2i} (i = 1, \ldots, n)$ 独立同分布且服从单位指数分布。Peng 和 Huang（2008）将自助法估计量 $\beta_j^*(\tau)$ 定义为方程 $S_{jn_j}^*(\beta, \tau) = 0 \text{in } \beta$ 的近似解。他们在附录 C 中证明，$\beta_j^*(\tau)$ 满足

$$n_j^{1/2} S_{jn_j}^*(\hat{\beta}_j^*, \tau) = o_{[\tau_L, \tau_R]}(1)\ a.s. \tag{5.2}$$

$$n_j^{1/2} \{\hat{\beta}_j^*(\tau) - \hat{\beta}_j(\tau)\} = \tilde{\phi}(-n_j^{1/2} S_{jn_j}^*(\hat{\beta}_j, \tau)) + o_{[\tau_L, \tau_R]}(1), \tag{5.3}$$

~其中，$o_{[\tau_L, \tau_R]}(1)$ 再次表示一个依概率一致收敛于零的量，且该收敛关于 $[\tau_L, \tau_R]$ 一致；ϕ 是与 (4.5) 中相同的线性映射。

我们在第 4.1 节证明了，k 阶分量 $\tilde{\phi}(-n_j^{1/2} S_{jn_j}(\hat{}\beta_j,\ \tau))$ 等价于条件（A）右侧出现的和~，并且还给出了函数 g_{jk}。由于 ϕ 是线性的，很容易看出，量 $\tilde{\phi}(-n_j^{1/2} S_{jn_j}^*(\hat{}\beta_j,\ \tau))$ 的 k 阶分量对应于 $\tilde{\phi}(-n_j^{1/2} S_{jn_j}(\hat{}\beta_j,\ \tau))$ 的 k 阶分量，其中 g_{jk} 被 $\eta_{ji} g_{jk}$ 和 $\eta_{ji} = \tilde{\eta}_{ji} - 1$ 替换。因此，条件（A*）成立。该重新加权自助法程序的实现详见 Peng 和 Huang（2008）的第 4.1 节。

5.2 其他估计量

在 Wang 和 Wang（2009）、Portnoy（2003）以及 De Backer 等（2019）的论文中，尚未证明自助法程序的渐近有效性。如这些论文所建议的，我们将采用一种朴素自助法：对于独立样本，分别对三元组 $\{(X_{ji},\ \Delta_{ji},\ Z_{ji})\}_{i=1}^{n_j}$ 进行有放回重抽样，分别对应 $j = 1$ 和 $j = 2$；而对于配对样本，则从联合样本 $\{(X_{1i},\ \Delta_{1i},\ Z_{1i},\ X_{2i},\ \Delta_{2i},\ Z_{2i})\}_{i=1}^{n}$ 中有放回地抽取重抽样样本。请注意，该自助法程序可由经典自助法理论证明其合理性。更多信息请参阅 Wang 和 Wang（2009）的第 3.3 节、Portnoy（2003）的第 6.1 节以及 De Backer 等（2019）的第 3 节。

6 模拟

本节报告模拟结果，以评估所提出的用于检验 H_0 的检验统计量在有限样本中的性能。对于独立样本和配对样本两种情形，我们检验单条分位数曲线（具体而言是中位数）的相等性，以及一系列分位数水平上的相等性。我们研究第3节所述检验统计量以及第4节所讨论估计量的性能。

检验中位数是否相等时，集合 $A = \{0.5\}$ 是单元素集，因此可以使用第4节所述的四种估计量中的任意一种。相比之下，检验一系列分位数水平上的相等性（或检验相应分布在一系列时间点上的相等性）时，只能使用 Peng和Huang（2008）的估计量和 Portnoy（2003）的估计量，因为只有针对这两种估计量，才证明了关于一组 τ 值的渐近结果的有效性。不过，我们也决定纳入 Wang 和 Wang（2009）以及 De Backer 等（2019）的估计量结果，以便比较检验性能。

所提出的检验基于第5节介绍的自助法，决定接受还是拒绝原假设。具体而言，对于第3节定义的任一检验统计量，令 t_a^* 为自助检验统计量的 $0 < a < 1$ 水平分位数。于是，若给定样本的检验统计量落在区间 $[t^*_{\alpha/2}, t^*_{1-\alpha/2}]$之外，则拒绝原假设。不过，对过程 $n^{1/2}(\hat{}\beta_1 - \hat{}\beta_2)$ 进行标准化，并对相应的自助法过程进行标准化，可以改善检验统计量的表现。更确切地说，我们将使用 $n^{1/2}\hat{}\Sigma(\tau)^{-1/2}(\hat{}\beta_1 - \hat{}\beta_2)(\tau)$，其中 $\hat{}\Sigma(\tau)$ 是协方差矩阵的自助法估计量。相应的自助法过程为 $n^{1/2}\hat{}\Sigma(\tau)^{-1/2}[(\hat{}\beta_1^* - \hat{}\beta_1)(\tau) - (\hat{}\beta_2^* - \hat{}\beta_2)(\tau)]$。所提出的标准化不会改变检验的渐近性质，但事实证明，它特别有助于提高检验功效。这可以这样解释：当 $\beta_1 - \beta_2$ 的各分量尺度不同时，考虑未标准化过程的范数时，较小分量中的差异无法观察到。

为优化模拟研究的计算时间，我们采用 Giacomini 等（2013）提出的方法来近似拒绝概率。对于一项包含（例如） N 次模拟运行的研究，该方法如下所示。对于每个 N样本，我们计算过程 $\beta_1(\tau) - \beta_2(\tau)\hat{}\hat{}\hat{}\hat{}$，并针对从给定样本中抽取的一个自助样本，计算相应的自助法版本 $\beta_1^*(\tau) - \beta_2^*(\tau) - \beta_1(\tau) + \beta_2(\tau)$。所有 N 次模拟均用于拒绝概率的估计。更多详情请参阅上述论文。

6.1 独立样本

我们首先考虑独立样本的情形，并考察表1中列出的三个不同模型。误差ϵ_j服从均值为零、方差为0.25的正态分布（$N(0, 0.25)$），且与协变量向量 $Z_j = (Z_j,1, Z_j,2)$、 $j = 1$、2相互独立。注意，模型1也可以写成方程（2.1）的形式，其中 $\beta_{j0}(\tau) = \beta_{j0} + \Phi^{-1}(\tau)0.5$， $\beta_{j1}(\tau) = \beta_{j1}$， $\beta_{j2}(\tau) = \beta_{j2}$，且$\epsilon_j(\tau) = \epsilon_j - \Phi^{-1}(\tau)0.5$，其中$\Phi^{-1}(\tau)$是标准正态分布的分位数函数。这样，误差$\epsilon_j(\tau)$满足$P(\epsilon_j(\tau) \leq 0|Z_j) = \tau$，正如（2.1）中所要求的。类似地，表1中的模型3也可以写成（2.1）的形式。最后，对于模型2，我们有 $\beta_{j0}(\tau) = \beta_{j0} + \beta_{j2}\Phi^{-1}(\tau)0.5$， $\beta_{j1}(\tau) = \beta_{j1}$， $\beta_{j2}(\tau) = \Phi^{-1}(\tau)0.5$，且$\epsilon_j(\tau) = (\beta_{j2} + Z_j,2)(\epsilon_j - \Phi^{-1}(\tau)0.5)$，只要 $\beta_{j2} + Z_j,\ _2 > 0$，便满足 $P(\epsilon_j(\tau) \leq 0|Z_j) = \tau$，因为$\epsilon_j$ 与Z_j、2相互独立。

模型1	日志$T_j = \beta_{j0} + \beta_{j1}Z_{j,1} + \beta_{j2}Z_{j,2} + \epsilon_j$
模型2	log$T_j = \beta_{j0} + \beta_{j1}Z_{j,\ 1} + (\beta_{j2} + Z_{j,\ 2})\epsilon_j$
模型3	对数 $T_j = \beta_{j0} + \beta_{j1}Z_{j,1} + \beta_{j2}Z_{j,2} + \frac{1}{2}\epsilon_j$

表1：模拟中所用模型的设定（$j = 1,\ 2$）。

对于每个模型，系数 $\beta_{10} = \beta_{20} = 0$， $\beta_{11} = -0.5$， $\beta_{12} = \beta_{22} = 0.5$ 均固定，而 $\beta_{10} = \beta_{20} = 0$ 的取值

为 β_{21} 的取值会变化，具体为 $\beta_{21}=-0.5$，-0.3 或 -0.1。因此，当 $\beta_{21}=-0.5$ 时，原假设 H_0：$\beta_1(\tau)=\beta_2(\tau)$ 成立，而在另外两种情况下不成立。此外，我们将第一个样本的样本量固定为 $n_1=200$，并通过取 $n_2=100$，200，400 来改变第二个样本的样本量。删失时间 C_j 在区间 $[0,c_j]$ 上服从均匀分布，且与 $(T_j,\ Z_j)$ 相互独立；选择 c_j，以使每个样本中删失数据的平均比例分别为 20% 和 40% $j=1$，2。

正如前文所述，我们的方法不要求两个样本中的协变量分布相同。为了能够与竞争方法进行比较，我们将针对协变量分布相同和不同两种情况开展模拟。

在第一种设定中，两个样本中的协变量具有相同的分布：$Z_{j,\ 1}$ 服从均匀分布 $U[0,1]$，而 $Z_{j,\ 2}$ 服从成功参数为 0.5 的伯努利分布 $B(0.5)$，对每个 $j=1,\ 2$ 均如此。协变量 $Z_{1,\ 1}$，$Z_{1,\ 2}$，$Z_{2,\ 1}$ 和 $Z_{2,\ 2}$ 彼此独立。在此设定下，我们进行分布比较，将集合 A 选为区间 $[\tau_L,\tau_R]=[0.1,0.6]$，步长设为 0.01。我们将本方法的表现与 Sant'Anna (2014) 提出的方法进行比较。该文提出了一种非参数方法，用于比较两个独立样本对应的条件分布，且这两个样本中的协变量分布相同。该方法依赖于倾向得分 $p(\cdot)$；我们使用级数 Logit 估计量对其进行估计，其中针对 $n_2=100$使用 1 阶幂级数，针对 $n_2=200$使用 3 阶幂级数，针对 $n_2=400$使用 5 阶幂级数，这一做法是 Sant'Anna (2014) 第5节所建议的。由于 Sant'Anna (2014) 的方法并不假设分位数与协变量之间呈线性关系，因此，与我们的方法相比，检验功效可能会有所下降。

在第二种设定中，两个样本的协变量具有不同的分布：$Z_{1,\ 1}\sim U[0,1]$，$Z_{2,\ 1}\sim U[0,1.2]$，$Z_{1,\ 2}\sim B(0.5)$ 和 $Z_{2,\ 2}\sim B(0.7)$。在这种情况下，我们根据上述三种模型，比较两个样本在$\tau=0.5$处的分位数，该分位数对应中位数。据我们所知，在协变量分布不相同的独立样本情形下，目前没有可供比较的方法。两种模拟设定的概要见表 2。

设定	$Z_{1,1}$	$Z_{1,2}$	$Z_{2,1}$	$Z_{2,2}$	所进行的比较
1	$U[0,1]$	$B(0.5)$	$U[0,1]$	$B(0.5)$	分布（$\tau\in[0.1,0.6]$）
2	$U[0,1]$	$B(0.5)$	$U[0,1.2]$	$B(0.7)$	分位数函数（$\tau=0.5$）

表 2：设定1和设定2中协变量 $Z_{i,\ j}$ 对于 $i=1,\ 2,\ j=1,\ 2$ 的分布设定，以及所进行的比较。这里，$B(p)$ 和 $U[a,b]$ 分别表示参数为 p 的伯努利分布和区间 $[a,b]$ 上的均匀分布。

对于这两种设定，我们使用第4节讨论的估计量，即 De Backer 等（2019）、Peng 和 Huang（2008）、Portnoy（2003）以及 Wang 和 Wang（2009）的估计量，以及式（3.1）、（3.2）和（3.3）中给出的检验统计量 T_{L_2}、T_{L_∞} 和 $T_{B,\ k}$。每个模型重复模拟 500 次，并根据显著性水平 $\alpha=0.05$ 计算拒绝概率（采用 Bonferroni校正时使用 $\alpha=0.05/3$）。对于 Peng 和 Huang（2008）的方法，检验的临界值根据第5.1节所述的加权自助法计算；对于其他估计量，则使用有放回抽样的朴素自助法（见第5.2节）。Sant'Anna（2014）方法基于一种乘子型自助法程序，与第5.1节所述的自助法有些相似。所有自助法的迭代次数均为 $N=500$。

设定1的模拟结果列于表 3 和附录中的表 9，分别对应删失数据占 20% 和 40% 的情形。对于以下设定，所提出的方法能够检验分布是否相等。当 H_0 假设成立时，Peng 和 Huang（2008）的估计量的第一类错误率接近 5% 的理论值。Portnoy（2003）和 Wang 和 Wang（2009）的估计量则并非总是如此。有趣的是，De Backer 等人（2019）的估计量所得结果与 Peng 和 Huang（2008）的结果相当。在所提出的检验统计量（Bonferroni、 T_{L_2} 和 T_{L_∞}）中，当备择假设成立时，结果相近。不过，在 H_0下，Bonferroni统计量的拒绝率更高。还需注意，当出现以下任一情况时，检验功效会如预期般提高：样本量 n_2 增大、差异 $\beta_{21}-\beta_{11}$ 增大，或删失数据的比例降低。Sant'Anna（2014）方法在所有情形下所得结果相当或更差。这可以解释为该方法没有利用分位数线性的假设，而我们的方法利用了这一假设。模型3与模型1相同，唯一的区别是误差项的方差更小，因此比较两者的结果颇有意义。我们发现，在模型3下，备择假设成立时（$\beta_{21}-\beta_{11}=0.2$ 或 0.4）的拒绝率高于模型1下的拒绝率，这符合预期。

			邦弗朗尼	T_{L_2}	T_{L_∞}	邦弗朗尼	T_{L_2}	T_{L_∞}	邦弗朗尼	T_{L_2}	T_{L_∞}	邦弗朗尼	T_{L_2}	T_{L_∞}	
Mod.	n_2	Diff.	De Backer 等（2019）			Peng 和 Huang（2008）			Portnoy（2003）			Wang 和 Wang（2009）			SA
1	100	0	0.070	0.040	0.042	0.040	0.022	0.024	0.012	0.004	0.010	0.024	0.010	0.006	0.066
		0.2	0.202	0.150	0.160	0.144	0.108	0.128	0.018	0.032	0.032	0.197	0.165	0.133	0.148
		0.4	0.636	0.608	0.548	0.654	0.608	0.572	0.462	0.520	0.470	0.597	0.647	0.565	0.516
	200	0	0.052	0.024	0.034	0.054	0.032	0.034	0.076	0.034	0.034	0.050	0.034	0.030	0.032
		0.2	0.282	0.290	0.278	0.300	0.280	0.236	0.214	0.228	0.216	0.322	0.314	0.294	0.254
		0.4	0.836	0.868	0.854	0.860	0.902	0.858	0.870	0.918	0.880	0.864	0.898	0.866	0.702
	400	0	0.046	0.040	0.036	0.034	0.026	0.032	0.030	0.018	0.020	0.056	0.038	0.036	0.022
		0.2	0.350	0.382	0.344	0.406	0.380	0.380	0.198	0.230	0.206	0.372	0.406	0.374	0.358
		0.4	0.902	0.936	0.922	0.926	0.960	0.934	0.860	0.890	0.858	0.902	0.958	0.916	0.792
2	100	0	0.032	0.022	0.018	0.040	0.028	0.034	0.018	0.006	0.010	0.016	0.008	0.016	0.032
		0.2	0.370	0.402	0.346	0.288	0.280	0.232	0.204	0.214	0.162	0.253	0.257	0.204	0.106
		0.4	0.938	0.956	0.948	0.918	0.950	0.912	0.820	0.912	0.840	0.948	0.954	0.908	0.556
	200	0	0.052	0.036	0.028	0.052	0.054	0.046	0.030	0.024	0.026	0.046	0.022	0.026	0.050
		0.2	0.474	0.498	0.430	0.546	0.546	0.524	0.570	0.572	0.520	0.630	0.524	0.500	0.222
		0.4	0.998	0.998	0.998	0.990	0.990	0.986	0.998	0.998	0.996	0.998	1.000	1.000	0.752
	400	0	0.032	0.020	0.020	0.056	0.024	0.040	0.026	0.012	0.018	0.072	0.038	0.042	0.050
		0.2	0.660	0.714	0.666	0.678	0.730	0.676	0.506	0.582	0.512	0.654	0.656	0.620	0.350
		0.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000	1.000	0.844
3	100	0	0.048	0.034	0.034	0.030	0.012	0.016	0.016	0.000	0.000	0.028	0.014	0.018	0.048
		0.2	0.606	0.600	0.580	0.574	0.666	0.604	0.358	0.386	0.316	0.677	0.649	0.629	0.162
		0.4	0.996	0.998	0.996	0.998	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	0.752
	200	0	0.034	0.032	0.030	0.030	0.034	0.024	0.058	0.042	0.056	0.046	0.014	0.022	0.046
		0.2	0.860	0.866	0.852	0.830	0.872	0.830	0.796	0.858	0.804	0.844	0.870	0.830	0.306
		0.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.934
	400	0	0.036	0.022	0.026	0.036	0.026	0.026	0.010	0.006	0.008	0.010	0.008	0.006	0.040
		0.2	0.958	0.966	0.942	0.944	0.964	0.942	0.834	0.890	0.808	0.936	0.938	0.924	0.490
		0.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.970

表 3：删失比例为 20% 时，两个独立样本的分位数曲线比较模拟结果（τ 值范围）。第一个样本的大小为 $n_1=200$，显著性水平为 $\alpha=0.05$，且 Diff $=\beta_{21}-\beta_{11}$，因此当 Diff $=0$ 时原假设成立。缩写 SA 表示 Sant'Anna (2014) 的检验统计量，Bonf 、T_{L_2} 和 T_{L_∞} 分别表示式 (3.3)、(3.1) 和 (3.2) 中定义的检验统计量。

设置 2 的结果列于表 4 中，并分别针对 20% 和 40% 的删失数据列于附录的表 10 中。当 H_0 成立时，Peng 和 Huang

（2008）、De Backer 等（2019）以及 Wang 和 Wang（2009）估计量的第一类错误接近 5% 的理论值。相比之下，对于 Portnoy（2003）估计量，我们再次得到了较差的结果。即使在 $n_2 = 400$情况下，第一类错误也不完全等于 0.05，这一事实可以由 n_1 的值保持不变（$n_1 = 200$）来解释。我们还模拟了 n_1, n_2 同时增加的情形，在该情形下，所有检验统计量都达到了理论上的第一类错误水平。同样，随着样本量n_2 增加、差异 $\beta_{21} - \beta_{11}$ 增大，或删失数据的比例降低，检验功效也会正确地提高。最后，我们注意到，不出所料，考虑一系列 τ值时的检验功效高于针对单条分位数曲线的检验功效。通过比较表 3 和表 4 可以观察到这一点（尽管协变量的分布并不完全相同，但 p 值的巨大差异提供了相当有说服力的证据）。

			Bonf	T_{L_2}	T_{L_∞}	Bonf	T_{L_2}	T_{L_∞}	Bonf	T_{L_2}	T_{L_∞}	Bonf	T_{L_2}	T_{L_∞}
Mod.	n_2	Diff.	De Backer			PengHuang			Portnoy			WangWang		
1	100	0	0.006	0.006	0.006	0.016	0.006	0.012	0.006	0.002	0.008	0.008	0.014	0.010
		0.2	0.064	0.096	0.072	0.072	0.056	0.056	0.018	0.032	0.016	0.072	0.070	0.082
		0.4	0.304	0.388	0.282	0.424	0.422	0.394	0.128	0.302	0.178	0.330	0.460	0.350
	200	0	0.016	0.012	0.020	0.022	0.010	0.020	0.042	0.042	0.038	0.006	0.006	0.004
		0.2	0.098	0.124	0.086	0.168	0.224	0.186	0.108	0.188	0.122	0.112	0.130	0.108
		0.4	0.596	0.712	0.614	0.576	0.626	0.596	0.738	0.774	0.720	0.686	0.766	0.714
	400	0	0.038	0.030	0.040	0.024	0.012	0.028	0.002	0.006	0.002	0.028	0.016	0.024
		0.2	0.246	0.266	0.242	0.180	0.228	0.182	0.096	0.110	0.092	0.260	0.280	0.258
		0.4	0.794	0.870	0.816	0.830	0.878	0.794	0.718	0.768	0.684	0.684	0.784	0.688
2	100	0	0.028	0.026	0.022	0.010	0.006	0.010	0.014	0.016	0.010	0.012	0.004	0.010
		0.2	0.112	0.150	0.142	0.134	0.198	0.138	0.062	0.098	0.072	0.124	0.156	0.142
		0.4	0.734	0.832	0.740	0.784	0.786	0.738	0.614	0.752	0.588	0.651	0.782	0.685
	200	0	0.028	0.026	0.024	0.032	0.038	0.026	0.038	0.028	0.038	0.020	0.014	0.016
		0.2	0.322	0.374	0.318	0.372	0.376	0.336	0.406	0.448	0.422	0.372	0.438	0.348
		0.4	0.956	0.978	0.946	0.942	0.960	0.942	0.946	0.976	0.948	0.954	0.976	0.944
	400	0	0.020	0.026	0.022	0.016	0.020	0.022	0.006	0.002	0.006	0.006	0.006	0.006
		0.2	0.558	0.606	0.552	0.292	0.322	0.266	0.308	0.414	0.318	0.460	0.560	0.528
		0.4	0.996	0.998	0.998	1.000	1.000	1.000	0.980	0.994	0.980	0.996	0.998	0.996
3	100	0	0.020	0.012	0.022	0.014	0.018	0.016	0.006	0.000	0.006	0.004	0.002	0.008
		0.2	0.282	0.426	0.324	0.330	0.394	0.326	0.266	0.308	0.230	0.334	0.424	0.364
		0.4	0.984	0.992	0.984	0.988	0.994	0.988	0.960	0.982	0.960	0.978	0.990	0.980
	200	0	0.026	0.022	0.026	0.008	0.002	0.008	0.026	0.016	0.022	0.030	0.022	0.024
		0.2	0.676	0.774	0.660	0.598	0.688	0.582	0.636	0.688	0.638	0.652	0.752	0.646
		0.4	0.996	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	400	0	0.024	0.020	0.024	0.020	0.020	0.022	0.004	0.000	0.004	0.042	0.024	0.030
		0.2	0.838	0.842	0.818	0.792	0.878	0.804	0.648	0.752	0.668	0.748	0.772	0.734
		0.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

表 4：在 20% 删失条件下，两个独立样本中位数（$\tau = 0.5$）比较的模拟结果。第一个样本的大小为 $n_1 = 200$，显著性水平为 $\alpha = 0.05$，Diff 为 $= \beta_{21} - \beta_{11}$，因此当 Diff 为 $= 0$ 时，原假设成立。缩写 Bonf、T_{L_2} 和 T_{L_∞} 分别表示公式（3.3）、（3.1）和（3.2）中定义的检验统计量。

6.2 配对样本

这里报告配对样本情况下的模拟结果，其中考虑表1中指定的三种不同模型。与之前一样，对于每个模型，系数 $\beta_{10} = \beta_{20} = 0$、$\beta_{11} = -0.5$、$\beta_{12} = \beta_{22} =$0.5 均为固定值，而 β_{21} 的取值会变化，分别为 $\beta_{21} = -0.5$、-0.3 或 -0.1。各样本中的协变量取值相同，且服从 $Z_{1、1} = Z_{2、1} \sim U[0, 1]$ 和 $Z_{1、2} = Z_{2、2} \sim B$(0.5)。

两个样本的误差 ϵ_1 和 ϵ_2 现在相互依赖，并服从二元正态分布，具体而言

$$\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix} \sim N_2\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.5 & \eta \\ \eta & 0.5 \end{pmatrix}\Big), \tag{6.1}$$

其中 $\eta = 0.2$，0.4 表示 ϵ_1 与 ϵ_2之间的协方差。注意，对于 $j = 1$，2，ϵ_j的边际分布为 N(0，0.5)。因此，采用与上一节相同的论证，我们可以将每个模型表示为式 (2.1) 的形式。

样本量 $n_1 = n_2 = n$ 为 $n = 100$，200，400。两个样本的删失时间 C_j 相同（因此， $C_1 = C_2 = C$），且 C 在区间 $[0, c]$上服从均匀分布，其中选择 c 是为了使第一个样本中删失数据的平均比例达到 20% 和 40%。请注意，所提出的方法并不要求两个样本的删失时间相同，但我们选择这种设置，是因为在实证实验中，配对个体常常会同时发生删失。同样的理由也适用于所选的协变量取值 $Z_{1,\ j} = Z_{2,\ j}$，其中 $j = 1$，2。

与之前一样，我们纳入了分布比较和单条分位数曲线（$\tau = 0.5$）的相关结果。在这种情况下，无法将我们的方法与其他方法进行比较，因为据我们所知，目前还没有用于基于配对删失数据比较分位数曲线的其他方法。我们纳入了第3节中介绍的检验统计量以及第4节中讨论的估计量的结果。对于每个模型，我们重复模拟500次，并基于显著性水平 $\alpha = 0.05$ 计算拒绝概率（进行 Bonferroni校正时则采用 $\alpha = 0.05/3$ ）。对于 Peng 和 Huang（2008）的方法，检验的临界值根据第5.1节所述的配对加权自助法计算；对于其他估计量，则采用有放回抽样的配对朴素自助法（见第5.2节）。对于分布比较，集合 A 仍为区间 $[\tau_L, \tau_R] = [0.1, 0.6]$，步长为0.01。

配对样本的分布比较结果见表5；附录中的表 11 则分别报告了删失数据占比为 20% 和 40% 时的结果。对于 Peng 和 Huang（2008）以及 Portnoy（2003）的估计量，第一类错误率接近 5% 的理论值；当 A 不同于单元素集时，这些估计量也满足条件（ A ），但总体而言，前者的结果更好。另一方面，当 $n = 400$ 时，Wang 和 Wang（2009）以及 De Backer 等（2019）的估计量所得的第一类错误率更差。请注意，随着样本量 n 或差异 $\beta_{21} - \beta_{11}$ 增大，或者删失数据占比降低，检验功效会如预期般提高。误差之间的协方差增大时，也会出现这一现象。总体而言，该检验较为保守：只有在样本量较大时，第一类错误率才接近 5% 的理论值。这表明，在依赖数据的情况下，分布比较颇具难度。

单条分位数曲线（具体而言，即中位数（$\tau = 0.5$））的比较结果见表 6；对于删失数据比例分别为 20% 和 40% 的情形，结果也列于附录中的表 12。所提出的方法能够正确检验分位数曲线是否相等；与之前一样，随着 $\beta_{21} - \beta_{11}$ 和 n 的增大，或删失数据比例的降低，检验功效会提高。误差之间的协方差增大时，情况也相同。不同估计量和检验统计量得出的结果具有可比性。与之前一样，该检验较为保守：在大多数情况下，第一类错误率低于 5% 的理论值。与分布函数的比较相比，这种保守性更加明显。尽管如此，这些检验仍具有良好的检验功效性质。

			Bonf	T_{L_2}	T_{L_∞}	Bonf	T_{L_2}	T_{L_∞}	Bonf	T_{L_2}	T_{L_∞}	Bonf	T_{L_2}	T_{L_∞}
			De Backer 等（2019）			Peng 和 Huang（2008）			Portnoy（2003）			Wang 和 Wang（2009）		
Mod.	n	Diff.	$\eta = 0.2$											
1	100	0	0.028	0.014	0.020	0.040	0.036	0.030	0.048	0.032	0.034	0.054	0.014	0.023
		0.2	0.060	0.022	0.028	0.076	0.038	0.052	0.044	0.022	0.026	0.100	0.043	0.048
		0.4	0.230	0.178	0.184	0.171	0.114	0.094	0.162	0.218	0.168	0.179	0.173	0.179
	200	0	0.034	0.024	0.016	0.030	0.020	0.024	0.046	0.030	0.030	0.029	0.016	0.016
		0.2	0.116	0.084	0.082	0.134	0.098	0.104	0.152	0.138	0.124	0.092	0.104	0.086
		0.4	0.454	0.476	0.448	0.470	0.512	0.486	0.458	0.540	0.482	0.495	0.503	0.461
	400	0	0.048	0.034	0.034	0.046	0.046	0.040	0.036	0.016	0.032	0.052	0.032	0.026
		0.2	0.296	0.248	0.236	0.266	0.268	0.234	0.266	0.282	0.266	0.242	0.246	0.212
		0.4	0.862	0.864	0.858	0.794	0.838	0.792	0.826	0.896	0.864	0.816	0.872	0.790
2	100	0	0.030	0.024	0.018	0.032	0.024	0.024	0.024	0.020	0.014	0.022	0.014	0.014
		0.2	0.132	0.092	0.090	0.080	0.074	0.078	0.108	0.072	0.084	0.144	0.103	0.106
		0.4	0.508	0.482	0.452	0.461	0.445	0.373	0.295	0.321	0.259	0.278	0.312	0.253
	200	0	0.048	0.022	0.032	0.038	0.014	0.016	0.066	0.038	0.038	0.029	0.017	0.019
		0.2	0.318	0.226	0.218	0.208	0.200	0.184	0.224	0.286	0.212	0.224	0.212	0.202
		0.4	0.834	0.858	0.798	0.850	0.890	0.858	0.856	0.864	0.828	0.844	0.834	0.741
	400	0	0.046	0.020	0.030	0.060	0.044	0.034	0.058	0.038	0.050	0.064	0.032	0.042
		0.2	0.528	0.466	0.488	0.508	0.554	0.514	0.604	0.566	0.536	0.472	0.518	0.474
		0.4	0.988	0.996	0.988	0.996	0.998	0.998	0.992	0.996	0.994	0.996	0.992	0.994
3	100	0	0.036	0.022	0.028	0.032	0.010	0.012	0.030	0.016	0.016	0.036	0.024	0.028
		0.2	0.192	0.110	0.120	0.210	0.176	0.170	0.198	0.140	0.152	0.211	0.188	0.176
		0.4	0.830	0.814	0.792	0.748	0.806	0.756	0.776	0.772	0.742	0.723	0.762	0.739
	200	0	0.048	0.040	0.042	0.036	0.028	0.024	0.058	0.036	0.052	0.040	0.026	0.024
		0.2	0.354	0.460	0.398	0.542	0.504	0.468	0.440	0.510	0.478	0.428	0.392	0.432
		0.4	0.992	0.992	0.992	0.988	0.994	0.988	0.988	0.996	0.986	0.984	0.990	0.982
	400	0	0.050	0.016	0.028	0.060	0.044	0.048	0.040	0.036	0.044	0.036	0.034	0.038
		0.2	0.836	0.860	0.820	0.822	0.874	0.838	0.838	0.856	0.820	0.784	0.854	0.844
		0.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			$\eta = 0.4$											
1	100	0	0.046	0.010	0.014	0.042	0.020	0.026	0.016	0.002	0.010	0.018	0.005	0.009
		0.2	0.082	0.056	0.046	0.072	0.060	0.054	0.056	0.054	0.050	0.073	0.068	0.073
		0.4	0.190	0.206	0.176	0.242	0.228	0.186	0.293	0.315	0.267	0.325	0.247	0.226
	200	0	0.028	0.020	0.020	0.026	0.010	0.014	0.036	0.022	0.022	0.053	0.024	0.032
		0.2	0.212	0.184	0.174	0.186	0.146	0.148	0.170	0.118	0.156	0.162	0.133	0.131
		0.4	0.740	0.670	0.648	0.676	0.728	0.670	0.678	0.694	0.666	0.563	0.569	0.523
	400	0	0.072	0.032	0.040	0.050	0.036	0.038	0.036	0.024	0.018	0.026	0.026	0.018
		0.2	0.428	0.388	0.372	0.348	0.320	0.286	0.374	0.410	0.376	0.342	0.346	0.344
		0.4	0.940	0.946	0.920	0.976	0.984	0.962	0.968	0.976	0.968	0.954	0.976	0.962
2	100	0	0.038	0.004	0.014	0.022	0.012	0.008	0.024	0.010	0.012	0.022	0.007	0.005
		0.2	0.110	0.106	0.068	0.092	0.076	0.068	0.082	0.064	0.062	0.120	0.074	0.066
		0.4	0.518	0.484	0.434	0.628	0.604	0.519	0.490	0.566	0.474	0.473	0.482	0.430
	200	0	0.046	0.022	0.022	0.038	0.026	0.028	0.028	0.012	0.020	0.047	0.025	0.023
		0.2	0.306	0.276	0.218	0.320	0.274	0.276	0.418	0.394	0.352	0.352	0.305	0.293
		0.4	0.940	0.944	0.922	0.958	0.966	0.954	0.948	0.964	0.948	0.909	0.938	0.922
	400	0	0.046	0.030	0.026	0.044	0.028	0.032	0.042	0.026	0.034	0.030	0.016	0.016
		0.2	0.720	0.710	0.690	0.722	0.706	0.684	0.770	0.798	0.756	0.734	0.720	0.692
		0.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	100	0	0.036	0.008	0.016	0.016	0.004	0.014	0.024	0.010	0.010	0.024	0.006	0.008
		0.2	0.354	0.280	0.264	0.318	0.288	0.270	0.234	0.218	0.216	0.310	0.239	0.237
		0.4	0.886	0.912	0.876	0.928	0.912	0.872	0.922	0.942	0.916	0.915	0.943	0.929
	200	0	0.056	0.040	0.042	0.028	0.012	0.014	0.042	0.042	0.034	0.026	0.012	0.018
		0.2	0.586	0.650	0.572	0.730	0.720	0.632	0.592	0.632	0.596	0.554	0.502	0.488
		0.4	0.998	1.000	0.998	1.000	1.000	1.000	0.998	1.000	0.996	1.000	1.000	0.998
	400	0	0.034	0.020	0.034	0.052	0.032	0.034	0.044	0.032	0.032	0.046	0.020	0.022
		0.2	0.952	0.960	0.956	0.904	0.944	0.914	0.960	0.966	0.940	0.954	0.960	0.954
		0.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

表5：删失比例为 20% 时，两个配对样本的分位数曲线比较模拟结果（ τ-值的范围）。样本量为 $n = 100$，200，400，显著性水平为 $\alpha = 0.05$，Diff $= \beta_{21} - \beta_{11}$，因此 Diff $= 0$ 的原假设成立。误差之间的协方差由 η 表示，见（6.1）。缩写 Bonf、T_{L_2} 和 T_{L_∞} 分别表示式（3.3）、（3.1）和（3.2）中定义的检验统计量。

			Bonf	T_{L_2}	T_{L_∞}	Bonf	T_{L_2}	T_{L_∞}	Bonf	T_{L_2}	T_{L_∞}	Bonf	T_{L_2}	T_{L_∞}
			De Backer 等（2019）			Peng 和 Huang（2008）			Portnoy（2003）			Wang 和 Wang（2009）		
Mod.	n	Diff.	$\eta = 0.2$											
1	100	0	0.022	0.024	0.022	0.010	0.004	0.010	0.020	0.016	0.020	0.020	0.012	0.016
		0.2	0.018	0.018	0.014	0.024	0.028	0.028	0.008	0.012	0.016	0.034	0.030	0.020
		0.4	0.136	0.118	0.136	0.082	0.064	0.074	0.112	0.106	0.128	0.088	0.078	0.056
	200	0	0.010	0.012	0.008	0.014	0.012	0.014	0.020	0.028	0.020	0.016	0.016	0.014
		0.2	0.058	0.056	0.062	0.074	0.090	0.082	0.086	0.084	0.086	0.038	0.050	0.040
		0.4	0.276	0.298	0.228	0.200	0.248	0.236	0.302	0.310	0.290	0.260	0.392	0.288
	400	0	0.040	0.034	0.042	0.024	0.036	0.028	0.020	0.018	0.014	0.014	0.014	0.014
		0.2	0.158	0.168	0.124	0.140	0.148	0.122	0.138	0.164	0.152	0.154	0.114	0.158
		0.4	0.620	0.710	0.640	0.562	0.636	0.554	0.658	0.682	0.652	0.508	0.616	0.510
2	100	0	0.016	0.022	0.010	0.020	0.006	0.020	0.014	0.014	0.008	0.024	0.012	0.024
		0.2	0.022	0.028	0.020	0.024	0.018	0.024	0.036	0.032	0.030	0.065	0.062	0.056
		0.4	0.172	0.196	0.152	0.156	0.182	0.188	0.270	0.282	0.276	0.181	0.301	0.161
	200	0	0.032	0.024	0.030	0.022	0.020	0.022	0.018	0.024	0.018	0.012	0.014	0.012
		0.2	0.062	0.068	0.060	0.110	0.134	0.110	0.140	0.168	0.152	0.110	0.094	0.112
		0.4	0.572	0.652	0.560	0.606	0.678	0.598	0.616	0.654	0.616	0.522	0.626	0.540
	400	0	0.022	0.020	0.022	0.042	0.042	0.038	0.024	0.020	0.022	0.028	0.036	0.030
		0.2	0.346	0.380	0.272	0.392	0.420	0.352	0.278	0.282	0.260	0.268	0.282	0.208
		0.4	0.948	0.962	0.940	0.950	0.986	0.944	0.958	0.976	0.954	0.884	0.944	0.896
3	100	0	0.020	0.012	0.018	0.006	0.000	0.000	0.010	0.008	0.016	0.014	0.010	0.004
		0.2	0.072	0.082	0.080	0.170	0.142	0.156	0.120	0.080	0.120	0.100	0.096	0.086
		0.4	0.422	0.552	0.470	0.516	0.562	0.534	0.500	0.584	0.484	0.456	0.554	0.444
	200	0	0.038	0.024	0.040	0.014	0.020	0.018	0.020	0.020	0.018	0.022	0.022	0.024
		0.2	0.148	0.204	0.146	0.224	0.252	0.238	0.228	0.262	0.224	0.294	0.346	0.298
		0.4	0.838	0.906	0.860	0.896	0.944	0.918	0.890	0.972	0.910	0.870	0.948	0.880
	400	0	0.030	0.018	0.028	0.018	0.036	0.022	0.012	0.014	0.010	0.032	0.028	0.032
		0.2	0.546	0.662	0.564	0.580	0.642	0.584	0.538	0.664	0.566	0.592	0.644	0.560
		0.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			$\eta = 0.4$											
1	100	0	0.010	0.012	0.012	0.006	0.006	0.004	0.008	0.004	0.008	0.014	0.014	0.014
		0.2	0.044	0.030	0.036	0.024	0.016	0.024	0.014	0.022	0.014	0.022	0.030	0.032
		0.4	0.050	0.060	0.062	0.064	0.082	0.084	0.070	0.056	0.076	0.186	0.184	0.188
	200	0	0.010	0.018	0.010	0.022	0.012	0.016	0.026	0.016	0.016	0.026	0.014	0.016
		0.2	0.040	0.088	0.046	0.076	0.056	0.072	0.062	0.078	0.062	0.074	0.074	0.072
		0.4	0.350	0.416	0.342	0.252	0.274	0.220	0.260	0.406	0.306	0.328	0.340	0.334
	400	0	0.026	0.018	0.022	0.026	0.014	0.028	0.018	0.010	0.014	0.006	0.012	0.006
		0.2	0.242	0.242	0.240	0.146	0.200	0.164	0.218	0.282	0.214	0.270	0.212	0.252
		0.4	0.716	0.720	0.712	0.760	0.818	0.792	0.722	0.804	0.722	0.696	0.796	0.732
2	100	0	0.012	0.008	0.014	0.012	0.010	0.006	0.006	0.014	0.012	0.012	0.008	0.010
		0.2	0.044	0.022	0.034	0.042	0.042	0.038	0.026	0.040	0.028	0.026	0.020	0.022
		0.4	0.174	0.236	0.224	0.152	0.210	0.182	0.188	0.186	0.172	0.205	0.313	0.215
	200	0	0.030	0.022	0.032	0.010	0.010	0.010	0.014	0.008	0.018	0.028	0.010	0.028
		0.2	0.190	0.216	0.190	0.092	0.124	0.094	0.166	0.164	0.174	0.128	0.150	0.128
		0.4	0.690	0.808	0.720	0.674	0.752	0.648	0.802	0.840	0.778	0.696	0.764	0.724
	400	0	0.016	0.006	0.012	0.028	0.016	0.026	0.002	0.008	0.006	0.018	0.010	0.016
		0.2	0.464	0.534	0.488	0.422	0.434	0.432	0.430	0.480	0.430	0.394	0.420	0.400
		0.4	0.968	0.982	0.968	0.988	0.998	0.988	0.978	0.990	0.972	0.950	0.992	0.954
3	100	0	0.018	0.014	0.014	0.006	0.004	0.006	0.012	0.010	0.014	0.016	0.010	0.016
		0.2	0.094	0.122	0.080	0.116	0.154	0.118	0.098	0.110	0.106	0.084	0.070	0.092
		0.4	0.614	0.652	0.620	0.490	0.616	0.514	0.618	0.678	0.598	0.466	0.536	0.406
	200	0	0.012	0.028	0.012	0.032	0.020	0.022	0.038	0.030	0.036	0.012	0.010	0.012
		0.2	0.334	0.454	0.350	0.324	0.406	0.336	0.318	0.374	0.354	0.306	0.396	0.302
		0.4	0.952	0.984	0.960	0.960	0.994	0.974	0.958	0.990	0.974	0.938	0.970	0.938
	400	0	0.018	0.016	0.018	0.032	0.036	0.032	0.030	0.010	0.030	0.018	0.014	0.014
		0.2	0.682	0.782	0.692	0.700	0.792	0.688	0.688	0.798	0.682	0.690	0.814	0.714
		0.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

表 6：对受 20% 删失影响的两个配对样本的中位数（$\tau = 0.5$）进行比较的模拟结果。样本量为 $n = 100$，200，400，显著性水平为 $\alpha = 0.05$，Diff 为 $= \beta_{21} - \beta_{11}$，因此对于 Diff $= 0$，原假设成立。误差之间的协方差由 η 表示，见（6.1）。缩写 Bonf、T_{L_2} 和 T_{L_∞} 分别表示式（3.3）、（3.1）和（3.2）中定义的检验统计量。

7 数据分析

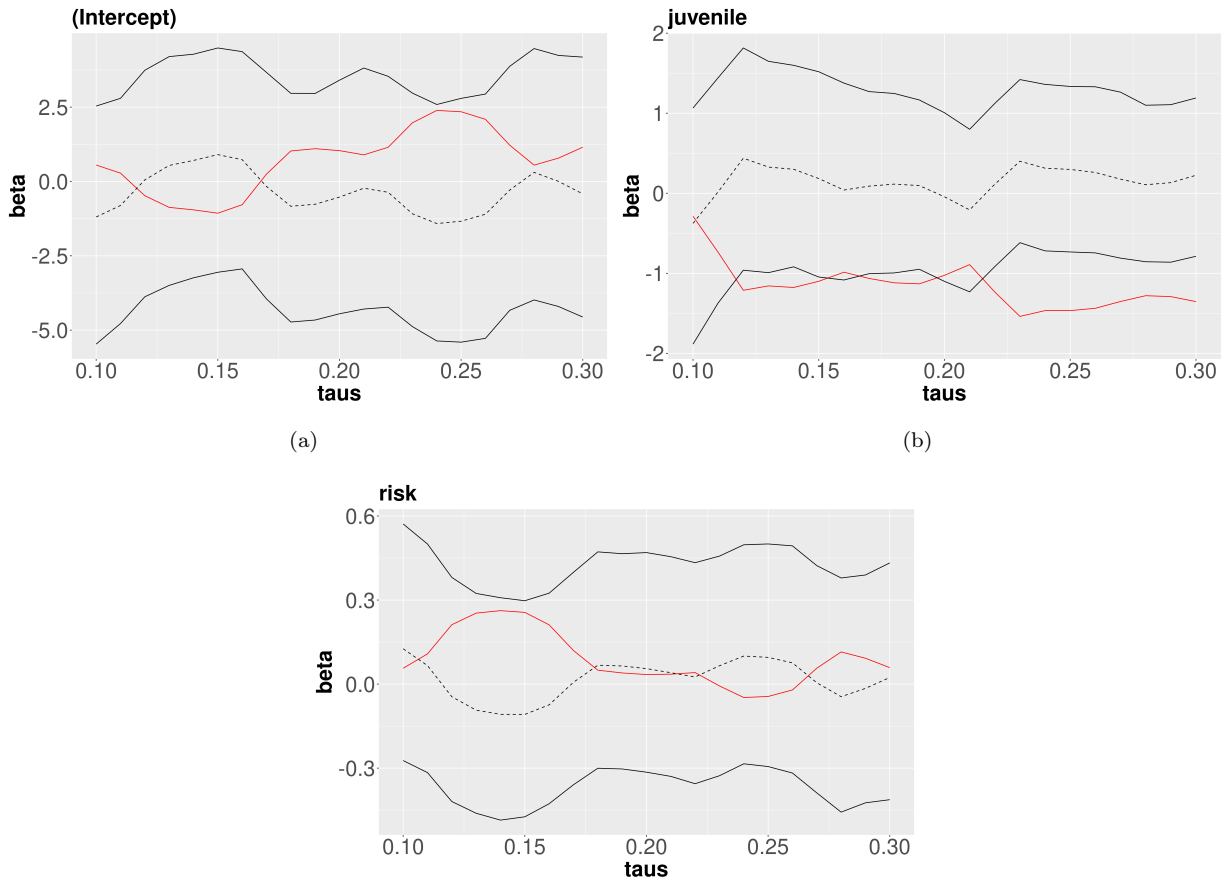
本节将使用一个关于糖尿病视网膜病变的数据集，展示所提出方法的一项应用，参见 Therneau（2022）。研究这个数据集尤其有意思，因为我们需要分别在独立样本和配对样本的情况下应用检验程序。此外，这也是文献中广为人知的一个案例，因此可以将我们的结果与以往结果进行比较。具体而言，我们将参考 Huster 等（1989）。

糖尿病视网膜病变是与糖尿病相关的一种并发症，其特征是眼睛视网膜内的微血管系统出现异常。糖尿病视网膜病变研究（DRS）于 1971 年启动，旨在研究激光光凝在延缓糖尿病视网膜病变患者失明方面的效果。双眼均患有糖尿病视网膜病变且视力为 20/100 的患者符合研究条件。随机选取每位患者的一只眼睛接受治疗，另一只眼睛则不予治疗并进行观察。对于每只眼睛，关注事件是从开始治疗到视力连续两次就诊均降至 5/200 以下所经历的时间。因此，存在约 6 个月的固有滞后时间（每 3 个月就诊一次）。因此，该数据集中的生存时间是实际失明时间（月），减去事件发生的最短可能时间（6.5 个月）。删失由死亡、退出研究或研究结束造成。该数据集包含 $n = 197$ 名高风险患者，依据 DRS 标准定义。

需要注意的是，糖尿病可按发病年龄分为两大类：青少年糖尿病和成人型糖尿病。这两种疾病的病程通常截然不同。DRS 研究的首要问题是评估激光光凝治疗的有效性。此外，我们还想了解治疗效果是否取决于糖尿病类型。最后，由于治疗组眼睛随机使用了氩 和氪 激光，我们也想了解哪种激光疗法最有效。

我们现在运用自己的方法来寻找这些问题的答案。为了评估激光治疗的有效性，我们将该方法应用于配对样本的情形。我们比较两个样本的分布，这两个样本分别包含 197 名患者治疗组眼睛和对照眼（未治疗眼）的相关数据。根据实验设计，两只眼睛的删失时间相同。样本中有 60% 的数据被删失，这限制了我们在分析中可使用的分位数水平范围。因此，我们比较 $\tau \in [0.1, 0.3]$ 的分位数，步长为 0.01。为了进行比较，我们选择了 Peng 和 Huang（2008）的估计量，并采用与模拟研究中相同的检验统计量标准化方法。

为了进行推断，我们采用与模拟研究中相同的百分位数自助法程序，其中自助重抽样次数为 $N = 10000$。我们采用第5节讨论的重新加权自助法。我们纳入青少年 协变量；青少年糖尿病患者的该变量取值为 1，否则取值为 0；还纳入 风险 协变量，该变量根据 RDS 的标准表示患者的风险等级，取值为 6 至 12 的离散值。我们分别通过图1和表7以图形和数值形式展示结果。



(a) (b) (c)

图1：每个系数（红色）的过程 $\{\hat{\beta}_{\text{treated}} - \hat{\beta}_{\text{control}}\}(\tau)$，以及自助法过程 $\{(\beta^*_{\text{treated}} - \beta^*_{\text{control}}) - (\beta_{\text{treated}} - \beta_{\text{control}})\}(\tau)$ 分布的 2.5% 和 97.5% 分位数（黑色实线）和自助法过程的均值（黑色虚线）。子图 (a)、(b) 和 (c) 分别显示估计截距以及 青少年和 风险 的估计系数，取值范围为 τ。

	$\tau = 0.10$	$\tau = 0.15$	$\tau = 0.20$	$\tau = 0.25$	$\tau = 0.30$
	$\hat{\beta}_{\text{treated}}$				
截距	4.64	3.78	5.71	7.36	6.61
青少年	-0.37	-0.67	-0.81	-1.27	-0.85
risk	-0.21	-0.05	-0.21	-0.29	-0.22
	$\hat{\beta}_{\text{control}}$				
截距	4.09	4.85	4.68	5.01	5.45
青少年	-0.08	0.43	0.22	0.20	0.50
risk	-0.27	-0.30	-0.24	-0.25	-0.28
	$\hat{\beta}_{\text{treated}} - \hat{\beta}_{\text{control}}$				
截距	0.55	-1.07	1.04	2.34	1.15
青少年	-0.29	-1.10	-1.02	-1.47	-1.35
risk	0.06	0.26	0.03	-0.04	0.06

表7：对于 $\tau \in \{0.10, 0.15, ..., 0.30\}$，过程 $\hat{\beta}_{\text{treated}}(\tau)$、$\hat{\beta}_{\text{control}}(\tau)$ 的分量截距、青少年 和风险 的取值及其差值。

我们可以使用第3节所述的检验统计量，对假设 H_0 : $\beta_{\text{treated}}(\tau) = \beta_{\text{control}}(\tau)$ for $\tau \in A = [0.1, 0.3]$ 进行检验，从而评估激光治疗的有效性。对于任何一种检验

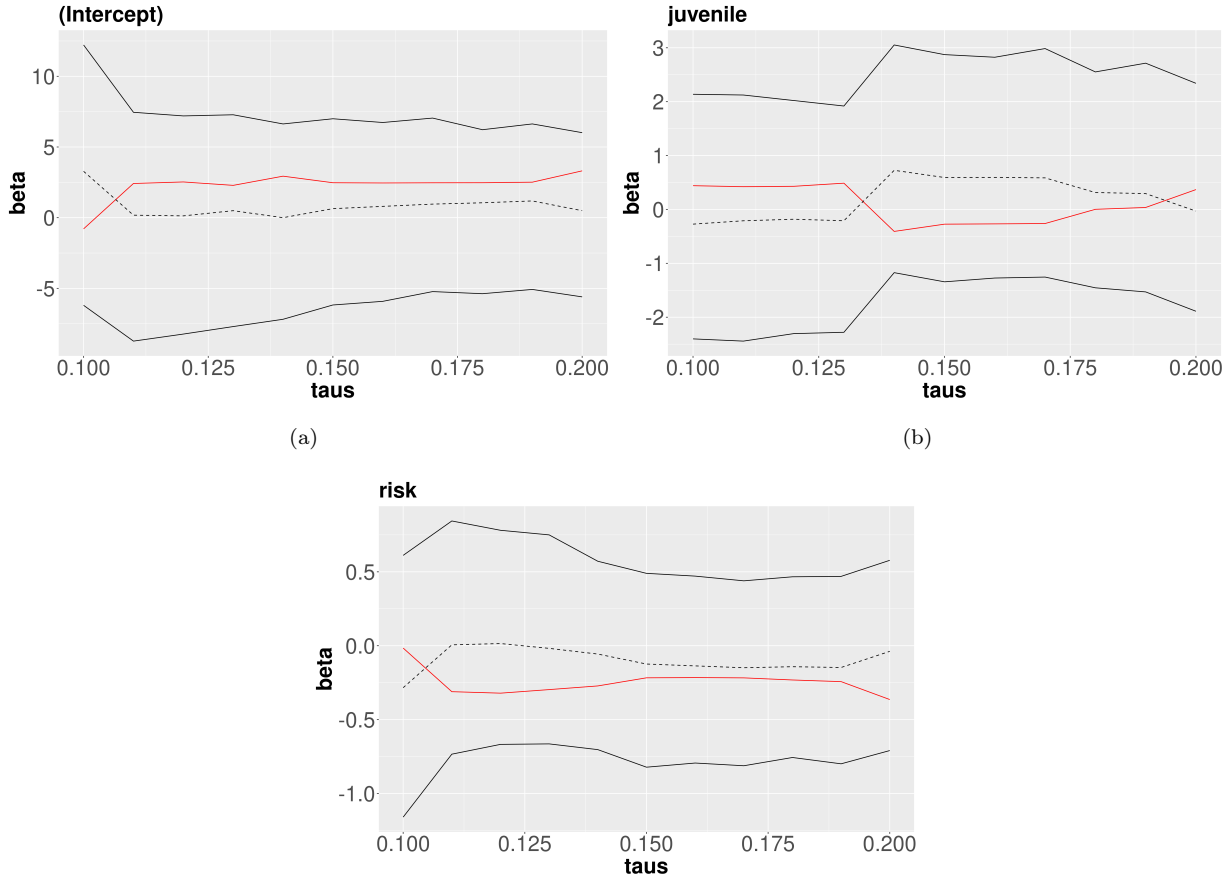
统计学 T_{L_2}，T_{L_∞} 和 Bonferroni校正后，相应的 p值小于 0.05。因此，我们可以得出结论（置信水平为 0.95）：有统计学证据表明治疗组与对照组之间存在差异。我们还可以得出结论，治疗组眼睛的分位数大于对照眼的分位数。事实上，可以观察到，对于每个固定的协变量向量 $Z = (Z_{\text{Intercept}}, Z_{\text{juvenile}}, Z_{\text{risk}}) \in \{1\} \times \{0,1\} \times \{6,...,12\}$，都有 $\beta_{\text{treated}}(\tau)Z \geq \beta_{\text{control}}(\tau)Z$。

使用德尔塔方法，我们还可以检验是否存在一个分量 $k = 1$，2，3，使得原假设 $H_0: \beta_{\text{treated, k}}(\tau) = \beta_{\text{control, k}}(\tau)$ 对于 $\tau \in A$ 被拒绝（显著性水平为 0.05）。只有青少年分量的此原假设被拒绝（p值等于 0.02），因此我们可以得出结论，治疗对年轻型和成人型糖尿病的疗效存在显著差异。由于 $\hat{\beta}_{\text{treated, juvenile}}(\tau) < 0$，我们还可以指出，治疗对成人型糖尿病更有效。我们的所有结论都与 Huster 等（1989）的研究一致。

最后，我们可以检验接受氩 激光治疗与接受氙 激光治疗的患者，其分位数是否存在差异。为此，我们仅考虑治疗组眼睛的数据。共有 $n_{\text{argon}} = 100$ 名患者接受了氩 激光治疗，另有 $n_{\text{xenon}} = 97$ 名患者接受了氙 激光治疗。对于此类比较，我们在独立样本（样本量不同）的框架下使用我们的方法，并针对 $\tau \in A$检验 H_0： $\beta_{\text{argon}}(\tau) = \beta_{\text{xenon}}(\tau)$，其中包括与之前相同的协变量。为避免出现未定义值，我们设定 $A = [0.1, 0.2]$，步长设为0.01。自助重抽样次数及估计量的选择与之前相同。图2和表8分别展示了估计量的图形结果和数值结果。无论选择哪种检验统计量，p值都大于0.05。因此，我们得出结论：在0.95的置信水平下，无法发现两种激光治疗的分位数之间存在显著差异。这些结果与有关激光治疗的具体研究一致，例如Plumb等（1982）报告的研究。

	$\tau = 0.10$	$\tau = 0.15$	$\tau = 0.20$
	$\hat{\beta}_{\text{argon}}$		
截距	2.57	6.11	5.92
青少年	-0.48	-0.94	-0.82
risk	-0.03	-0.25	-0.23
	$\hat{\beta}_{\text{xenon}}$		
截距	3.36	3.64	2.60
青少年	-0.92	-0.66	-1.18
risk	-0.01	-0.04	0.14
	$\hat{\beta}_{\text{argon}} - \hat{\beta}_{\text{xenon}}$		
截距	-0.79	2.47	3.31
青少年	0.44	-0.27	0.37
risk	-0.02	-0.22	-0.36

表8：过程 $\hat{\beta}_{\text{argon}}(\tau)$、$\hat{\beta}_{\text{xenon}}(\tau)$及其差值中分量截距、青少年 和风险 的值，适用于$\tau \in \{0.10, 0.15, 0.20\}$。



(a) (b)

(c)

图2：每个系数的过程 $\{\hat{\beta}_{\text{argon}} - \hat{\beta}_{\text{xenon}}\}(\tau)$（红色），以及自助法过程 $\{(\beta^*_{\text{argon}} - \beta^*_{\text{xenon}}) - (\beta_{\text{argon}} - \beta_{\text{xenon}})\}(\tau)$ 的$\hat{}\,\hat{}\,\hat{}$分布的 2.5% 和 97.5% 分位数（黑色实线）及均值自助法过程（黑色虚线）。子图 (a)、(b) 和 (c) 分别展示了在一系列 τ值下估计的截距以及 青少年和 风险的估计系数。

8 结论

本文通过允许响应变量为右删失，并借助假定为协变量线性的分位数曲线比较两个总体，从而为存在协变量时两个总体的比较研究作出了贡献。本文同时考虑独立样本和配对样本两种情形。尽管这一设定在实践中有许多应用，但此前的文献尚未对此进行研究。

本文可以从许多方向进行扩展。首先，将该方法扩展到包含两个以上总体的情形将很有意义。这种扩展应该是可行的，尽管绝非易事。例如，参见 Pardo-Fernández 和 Van Keilegom（2006），他们在不同背景下考虑了两个以上总体的情形。其次，我们在模拟研究中没有使用现有的所有分位数回归估计量。如果开展一项全面的模拟研究，对所有现有估计量进行比较，将有助于了解基于这些估计量的检验的相对表现。最后，将方法扩展到其他类型的不完整数据，也将是一个有益的发展方向。

参考文献

De Backer, M.、El Ghouch, A. 和 Van Keilegom, I.（2019）。删失分位数回归的改进损失函数，《美国统计学会会刊》**114**（527）：1126–1137。De Backer, M.、El Ghouch, A. 和 Van Keilegom, I.（2020）。线性删失分位数回归：一种新颖的最小距离方法，《斯堪的纳维亚统计学杂志》**47**（4）：1275–1306。Delgado, M. A. 和 Escanciano, J. C.（2013）。条件随机占优检验，《商业与经济统计学杂志》**31**（1）：16–28。Dette, H.、Wagener, J. 和 Volgushev, S.（2011）。条件分位数曲线的比较，《斯堪的纳维亚统计学杂志》**38**（1）：63–88。Dette, H.、Wagener, J. 和 Volgushev, S.（2013）。分位数曲线的非参数比较：一种随机过程方法，《非参数统计学杂志》**25**（1）：243–260。Giacomini, R.、Politis, D. N. 和 White, H.（2013）。一种用于开展涉及自助法估计量的蒙特卡罗实验的加速方法，《计量经济学理论》**29**（3）：567–589。Giné, E.（1996）。经验过程及其应用：综述，《伯努利》**2**（1）：1–28。Huster, W. J.、Brookmeyer, R. 和 Self, S. G.（1989）。带协变量的配对生存数据建模，《生物统计学》**45**：145–156。Koul, H. L. 和 Schick, A.（1997）。非参数回归曲线相等性的检验，《统计规划与推断杂志》**65**（2）：293–314。Lee, M.-J.（2009）。随机删失响应的分布处理效应的非参数检验，《皇家统计学会会刊*B*辑》**71**（1）：243–264。Li, G.、Tiwari, R. C. 和 Wells, M. T.（1996）。两样本问题中的分位数比较函数及其在诊断标志物比较中的应用，《美国统计学会会刊》**91**（434）：689–698。Neocleous, T.、Vanden Branden, K. 和 Portnoy, S.（2006）。对 Portnoy 的《删失回归分位数》（98（2003），1001–1012）的勘误，《美国统计学会会刊》**101**（474）：860–861。Pardo-Fernández, J. C. 和 Van Keilegom, I.（2006）。删失响应回归曲线的比较，《斯堪的纳维亚统计学杂志》**33**（3）：409–434。Peng, L.（2012）。关于删失回归分位数自洽估计的注记，《多元分析杂志》**105**：368–379。Peng, L.（2021）。生存数据的分位数回归，《统计学及其应用年度评论》**8**：413–437。Peng, L. 和 Huang, Y.（2008）。基于分位数回归模型的生存分析，《美国统计学会会刊》**103**（482）：637–649。

Plumb, A., Swan, A., Chignell, A. 和 Shilling, J. (1982). 氩弧与氩激光光凝术治疗增殖性糖尿病视网膜病变的比较试验，英国眼科学杂志**66**(4): 213–218. Portnoy, S. (2003). 删失回归分位数，《美国统计学会会刊》**98**(464): 1001–1012. Portnoy, S. and Lin, G. (2010). 删失回归分位数的渐近理论，《非参数统计学杂志》**22**(1): 115–130. Sant'Anna, P. H. (2014). 持续时间结果的条件处理效应非参数检验，技术报告，Citeseer. Sun, Y. (2006). 条件分位数函数的一致非参数相等性检验，《计量经济学理论》**22**(4): 614–632. Therneau, T. M. (2022). *R*中的生存分析。R 软件包版本 3.3-1。**网址：** *https://CRAN.R-project.org/package=survival*Van der Vaart, A. W. 和 Wellner, J. A. (1996). 弱收敛与经验过程，施普林格统计学系列，施普林格出版社，纽约。Vilar-Fernández, J. 和 González-Manteiga, W. (2004). 相关误差下回归曲线的非参数比较，统计学**38**(2): 81–99. Vilar-Fernández, J. M., Vilar-Fernández, J. A. 和 González-Manteiga, W. (2007). 相关误差下回归曲线非参数比较的自助法检验，*Test* **16**(1): 123–144. Wang, H. J. 和 Wang, L. (2009). 局部加权删失分位数回归，《美国统计学会会刊》**104**(487): 1117–1128. Yang, X., Narisetty, N. N. 和 He, X. (2018). 删失分位数回归估计的新方法，计算与图形统计学杂志**27**(2): 417–425.

A 证明

定理 3.1 的证明。由假设 (A) 和 Donsker 定理可知，$n_j^{1/2}$ ($\hat{}\beta_j - \beta_j$)(τ) 弱收敛到一个均值为 0、协方差函数为 $\Sigma_j(\tau_1,\tau_2) = E[g_j(\tau_1, X_j, \Delta_j, Z_j)g_j(\tau_2, X_j, \Delta_j, Z_j)^T]$ 的紧高斯过程 $G_j(\tau)$。然后，由 Slutsky 定理和假设 (EG)，可得 $n^{1/2}$($\hat{}\beta_j - \beta_j$)(τ) 弱收敛到 $p_j^{-1/2}G_j(\tau)$，因此其极限协方差函数为 $p_j^{-1}\Sigma_j(\tau_1,\tau_2)$。最后，由两个样本的独立性可知，在 H_0 下，$n^{1/2}$($\hat{}\beta_1 - \hat{}\beta_2$)($\tau$) $= n^{1/2}$($\hat{}\beta_1 - \beta_1$)(τ) $- n^{1/2}$($\hat{}\beta_2 - \beta_2$)(τ) 弱收敛到 $p_1^{-1/2}G_1(\tau) - p_2^{-1/2}G_2(\tau)$，这是一个均值为 0、协方差函数为 $p_1^{-1}\Sigma_1(\tau_1,\tau_2) + p_2^{-1}\Sigma_2(\tau_1,\tau_2)$ 的紧高斯过程。□

推论 3.2 的证明。在局部备择假设 H_1 下，该假设也可写为 $\beta_2(\tau) = \beta_1(\tau) - n^{-1/2}b(\tau)$，过程 $n^{1/2}$($\hat{}\beta_1 - \hat{}\beta_2$)($\tau$) 可写为 $n^{1/2}$($\hat{}\beta_1 - \beta_1$)(τ) $- n^{1/2}$($\hat{}\beta_2 - \beta_2$)(τ) $+ b(\tau)$。由于 $b(\tau)$ 是确定性的，极限过程是一个非中心高斯过程，其协方差函数与定理 3.1 中给出的一致。这意味着

$$n^{1/2}(\hat{\beta}_1 - \hat{\beta}_2)(\tau) \to_d W_I(\tau) + b(\tau),$$

其中，W_I 是与定理 3.1 中相同的高斯过程。由于 T_{L_2}、T_{L_∞} 和 $T_{B,\ k}$ 是连续映射，根据连续映射定理可得结果。□

定理 3.2 的证明。由假设 (A) 可知，在 H_0 下，

$$\begin{aligned}
&n^{1/2}(\hat{\beta}_1(\tau) - \hat{\beta}_2(\tau)) \\
&= n^{1/2}(\hat{\beta}_1(\tau) - \beta_1(\tau)) - n^{1/2}(\hat{\beta}_2(\tau) - \beta_2(\tau)) \\
&= n^{-1/2}\sum_{i=1}^{n}[g_1(\tau, X_{1i}, \Delta_{1i}, Z_{1i}) - g_2(\tau, X_{2i}, \Delta_{2i}, Z_{2i})] + o_P(1),
\end{aligned}$$

对 $\tau \in A$ 一致。由于两个 Donsker 类之差仍是 Donsker 类（参见 Van der Vaart 和 Wellner（1996）中的例 2.10.7），我们可以得出结论：$n^{1/2}$($\hat{}\beta_1(\tau) - \hat{}\beta_2(\tau)$) 弱收敛于一个均值为零的高斯过程，其协方差函数如定理陈述所示。□

定理 5.1 的证明。首先，我们分别考虑两个样本。根据经验过程理论的经典结果（例如，参见 Giné（1996）中的定理 3.2），条件 (A*) 保证 $n^{1/2}$($\hat{}\beta_j^* - \hat{}\beta_j$) 收敛到一个紧高斯过程；此外，由于 η_{ji} 的均值为零且方差等于 1，给定数据后，过程 $n^{1/2}$($\hat{}\beta_j^* - \hat{}\beta_j$) 的渐近协方差算子为

$$\begin{aligned}
&E(\eta_j^2)\frac{n}{n_j^2}\sum_{i=1}^{n_j} g_j(\tau_1, X_{ji}, \Delta_{ji}, Z_{ji})g_j(\tau_2, X_{ji}, \Delta_{ji}, Z_{ji})^T \\
&\overset{a.s.}{\to} p_j^{-1}E[g_j(\tau_1, X_j, \Delta_j, Z_j)g_j(\tau_2, X_j, \Delta_j, Z_j)^T],
\end{aligned}$$

这就是过程 $n^{1/2}$($\hat{}\beta_j - \beta_j$) 的渐近协方差。

向两个样本的推广与独立样本的定理 3.1 证明以及配对样本的定理 3.2 证明中的做法类似。现在需要考虑的两个过程是 $n^{1/2}$($\hat{}\beta_j^* - \hat{}\beta_j$)($\tau$)，其中$j = 1$，2。请注意，在独立样本的情况下，两个自助过程相互独立，因为

这些 η_{ji}都是独立同分布的。因此，容易证明 $n^{1/2}(\hat{}\beta_1^*(\tau) -\hat{}\hat{}\hat{}\beta_2^*(\tau) - \beta_1(\tau) + \beta_2(\tau))$ 的渐近协方差收敛于原过程的渐近协方差。而在配对样本的情况下，给定数据的自助法渐近协方差等于（其中 $V_{ji} = (X_{ji}, \Delta_{ji}, Z_{ji}), \quad j = 1, 2$）

$$
\begin{aligned}
& E(\eta_1^2)\frac{1}{n}\sum_{i=1}^{n}[g_1(\tau_1, V_{1i}) - g_2(\tau_1, V_{2i})][g_1(\tau_2, V_{1i}) - g_2(\tau_2, V_{2i})]^T \\
& \xrightarrow{a.s.} E\{[g_1(\tau_1, V_1) - g_2(\tau_1, V_2)][g_1(\tau_2, V_1) - g_2(\tau_2, V_2)]^T\},
\end{aligned}
$$

这就是 $n^{1/2}(\hat{}\beta_1 - \hat{}\beta_2 - \beta_1 + \beta_2)$ 的渐近协方差，由此可得该断言。 □

B 模拟结果

我们给出了删失比例为 40% 时的其他模拟结果。

			邦弗朗尼。	T_{L_2}	T_{L_∞}	邦弗朗尼。	T_{L_2}	T_{L_∞}	邦弗朗尼。	T_{L_2}	T_{L_∞}	邦弗朗尼	T_{L_2}	T_{L_∞}	
Mod.	n_2	Diff.	De Backer 等（2019）			Peng 和 Huang（2008）			Portnoy（2003）			Wang 和 Wang（2009）			SA
1	100	0	0.038	0.014	0.020	0.034	0.022	0.024	0.010	0.006	0.004	0.030	0.018	0.022	0.066
		0.2	0.118	0.110	0.122	0.135	0.131	0.110	0.039	0.031	0.033	0.110	0.064	0.092	0.148
		0.4	0.506	0.502	0.456	0.586	0.609	0.582	0.433	0.435	0.352	0.570	0.568	0.524	0.516
	200	0	0.046	0.038	0.036	0.050	0.042	0.034	0.028	0.024	0.024	0.048	0.026	0.028	0.032
		0.2	0.248	0.204	0.238	0.236	0.212	0.200	0.180	0.182	0.160	0.220	0.224	0.218	0.254
		0.4	0.772	0.824	0.740	0.836	0.857	0.859	0.764	0.810	0.762	0.828	0.876	0.786	0.702
	400	0	0.046	0.044	0.038	0.030	0.022	0.020	0.018	0.002	0.006	0.032	0.018	0.026	0.022
		0.2	0.376	0.366	0.348	0.298	0.292	0.286	0.124	0.086	0.080	0.292	0.270	0.252	0.358
		0.4	0.932	0.952	0.904	0.878	0.910	0.880	0.762	0.830	0.782	0.882	0.906	0.884	0.792
2	100	0	0.032	0.026	0.024	0.030	0.020	0.024	0.024	0.008	0.012	0.036	0.004	0.018	0.032
		0.2	0.216	0.230	0.184	0.362	0.284	0.270	0.112	0.137	0.102	0.204	0.172	0.160	0.106
		0.4	0.888	0.938	0.896	0.888	0.914	0.876	0.830	0.865	0.804	0.884	0.892	0.856	0.556
	200	0	0.060	0.034	0.038	0.038	0.028	0.022	0.038	0.036	0.026	0.034	0.028	0.032	0.050
		0.2	0.556	0.446	0.454	0.470	0.450	0.396	0.442	0.476	0.422	0.458	0.426	0.402	0.222
		0.4	0.998	0.996	0.994	0.984	0.998	0.990	0.986	0.990	0.986	0.978	0.990	0.982	0.752
	400	0	0.036	0.022	0.026	0.056	0.046	0.048	0.018	0.008	0.010	0.060	0.036	0.038	0.050
		0.2	0.564	0.684	0.638	0.640	0.642	0.618	0.474	0.464	0.436	0.596	0.662	0.578	0.350
		0.4	0.998	1.000	1.000	1.000	1.000	1.000	0.996	0.996	0.994	1.000	1.000	1.000	0.844
3	100	0	0.016	0.012	0.016	0.036	0.016	0.036	0.016	0.004	0.012	0.030	0.018	0.026	0.048
		0.2	0.476	0.470	0.480	0.512	0.510	0.510	0.261	0.261	0.180	0.520	0.538	0.530	0.162
		0.4	0.996	0.998	0.996	0.996	1.000	1.000	0.982	0.990	0.978	1.000	1.000	1.000	0.752
	200	0	0.032	0.028	0.026	0.046	0.026	0.024	0.028	0.024	0.026	0.034	0.030	0.026	0.046
		0.2	0.790	0.796	0.762	0.774	0.780	0.736	0.664	0.754	0.674	0.748	0.772	0.740	0.306
		0.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000	0.934
	400	0	0.052	0.034	0.038	0.048	0.030	0.030	0.016	0.008	0.006	0.036	0.018	0.016	0.040
		0.2	0.910	0.916	0.892	0.918	0.926	0.916	0.716	0.794	0.752	0.878	0.874	0.860	0.490
		0.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.970

表 9：删失比例等于 40% 时，两个独立样本的分位数曲线比较模拟结果（τ 值的范围）。第一个样本的大小为 $n_1 = 200$，显著性水平为 $\alpha = 0.05$，且 Diff $= \beta_{21} - \beta_{11}$，因此对于 Diff $= 0$，原假设成立。缩写 SA 表示 Sant'Anna (2014) 的检验统计量，Bonf、T_{L_2} 和 T_{L_∞} 分别表示式 (3.3)、(3.1) 和 (3.2) 中定义的检验统计量。

			Bonf	T_{L_2}	T_{L_∞}	Bonf	T_{L_2}	T_{L_∞}	Bonf	T_{L_2}	T_{L_∞}	Bonf	T_{L_2}	T_{L_∞}
Mod.	n_2	Diff.	De Backer 等（2019）			Peng 和 Huang（2008）			Portnoy（2003）			Wang 和 Wang（2009）		
1	100	0	0.016	0.008	0.014	0.028	0.022	0.030	0.004	0.004	0.004	0.020	0.018	0.022
		0.2	0.054	0.070	0.056	0.044	0.044	0.034	0.014	0.016	0.016	0.040	0.030	0.040
		0.4	0.344	0.358	0.316	0.200	0.234	0.206	0.202	0.166	0.200	0.234	0.288	0.262
	200	0	0.022	0.024	0.020	0.018	0.012	0.018	0.020	0.012	0.020	0.024	0.016	0.024
		0.2	0.118	0.146	0.106	0.108	0.122	0.106	0.122	0.104	0.118	0.080	0.082	0.076
		0.4	0.640	0.694	0.628	0.470	0.516	0.458	0.586	0.616	0.600	0.530	0.586	0.442
	400	0	0.010	0.014	0.012	0.012	0.020	0.014	0.010	0.014	0.010	0.004	0.004	0.006
		0.2	0.252	0.288	0.264	0.122	0.122	0.110	0.068	0.058	0.068	0.168	0.182	0.150
		0.4	0.660	0.828	0.718	0.674	0.772	0.676	0.498	0.666	0.488	0.756	0.826	0.756
2	100	0	0.024	0.018	0.024	0.010	0.010	0.010	0.004	0.002	0.006	0.028	0.022	0.028
		0.2	0.086	0.102	0.088	0.084	0.114	0.086	0.068	0.078	0.074	0.114	0.144	0.116
		0.4	0.754	0.842	0.800	0.566	0.658	0.606	0.569	0.645	0.577	0.736	0.804	0.744
	200	0	0.028	0.014	0.024	0.030	0.028	0.026	0.026	0.022	0.024	0.022	0.014	0.016
		0.2	0.326	0.380	0.286	0.232	0.266	0.222	0.326	0.360	0.308	0.306	0.328	0.284
		0.4	0.952	0.980	0.954	0.892	0.944	0.880	0.930	0.962	0.926	0.812	0.914	0.828
	400	0	0.022	0.024	0.022	0.022	0.010	0.022	0.002	0.002	0.000	0.026	0.036	0.026
		0.2	0.418	0.468	0.424	0.310	0.450	0.368	0.224	0.296	0.232	0.294	0.450	0.366
		0.4	0.992	0.998	0.988	0.984	0.996	0.984	0.948	0.976	0.940	0.974	0.996	0.982
3	100	0	0.010	0.004	0.010	0.024	0.016	0.020	0.000	0.000	0.000	0.012	0.014	0.012
		0.2	0.342	0.394	0.322	0.273	0.333	0.248	0.126	0.106	0.118	0.214	0.276	0.210
		0.4	0.960	0.970	0.962	0.958	0.980	0.954	0.870	0.940	0.868	0.910	0.966	0.930
	200	0	0.030	0.028	0.024	0.026	0.020	0.022	0.022	0.012	0.020	0.042	0.028	0.050
		0.2	0.514	0.628	0.506	0.504	0.572	0.512	0.440	0.612	0.482	0.472	0.622	0.472
		0.4	0.996	0.998	0.994	0.998	1.000	0.998	0.998	1.000	0.998	0.998	1.000	0.996
	400	0	0.014	0.012	0.016	0.006	0.004	0.006	0.004	0.008	0.004	0.008	0.010	0.008
		0.2	0.660	0.724	0.648	0.716	0.750	0.704	0.512	0.526	0.474	0.672	0.754	0.632
		0.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

表 10：在两个受 40% 删失影响的独立样本中，对中位数（$\tau = 0.5$）进行比较的模拟结果。第一个样本的大小为 $n_1 = 200$，显著性水平为 $\alpha = 0.05$，Diff 为 $= \beta_{21} - \beta_{11}$，因此对于 Diff $= 0$，原假设成立。缩写 Bonf、T_{L_2} 和 T_{L_∞} 分别表示式（3.3）、（3.1）和（3.2）中定义的检验统计量。

			Bonf	T_{L_2}	T_{L_∞}	Bonf	T_{L_2}	T_{L_∞}	Bonf	T_{L_2}	T_{L_∞}	Bonf	T_{L_2}	T_{L_∞}
			De Backer 等（2019）			Peng 和 Huang（2008）			Portnoy（2003）			Wang 和 Wang（2009）		
Mod.	n	Diff.	$\eta = 0.2$											
1	100	0	0.026	0.018	0.016	0.013	0.007	0.009	0.026	0.011	0.009	0.044	0.014	0.018
		0.2	0.098	0.058	0.080	0.108	0.068	0.066	0.069	0.041	0.037	0.090	0.046	0.058
		0.4	0.196	0.204	0.184	0.156	0.098	0.111	0.245	0.182	0.200	0.238	0.170	0.172
	200	0	0.032	0.018	0.018	0.051	0.018	0.028	0.047	0.029	0.029	0.034	0.014	0.016
		0.2	0.118	0.128	0.108	0.119	0.092	0.098	0.095	0.082	0.086	0.088	0.072	0.078
		0.4	0.428	0.468	0.386	0.483	0.496	0.448	0.383	0.358	0.286	0.370	0.382	0.358
	400	0	0.038	0.032	0.032	0.042	0.018	0.014	0.032	0.014	0.008	0.036	0.026	0.022
		0.2	0.202	0.212	0.198	0.291	0.222	0.194	0.232	0.222	0.222	0.276	0.224	0.226
		0.4	0.820	0.842	0.816	0.802	0.838	0.796	0.831	0.880	0.835	0.750	0.756	0.732
2	100	0	0.024	0.010	0.016	0.030	0.015	0.017	0.025	0.021	0.025	0.018	0.008	0.006
		0.2	0.080	0.036	0.044	0.077	0.060	0.045	0.048	0.035	0.050	0.084	0.040	0.048
		0.4	0.274	0.292	0.236	0.375	0.330	0.292	0.336	0.351	0.347	0.261	0.246	0.204
	200	0	0.036	0.030	0.026	0.050	0.018	0.018	0.026	0.024	0.030	0.030	0.026	0.024
		0.2	0.230	0.258	0.202	0.240	0.236	0.192	0.160	0.148	0.130	0.200	0.168	0.130
		0.4	0.808	0.778	0.770	0.670	0.796	0.686	0.782	0.739	0.731	0.748	0.732	0.702
	400	0	0.042	0.024	0.020	0.052	0.048	0.042	0.036	0.024	0.024	0.034	0.020	0.024
		0.2	0.356	0.362	0.354	0.533	0.561	0.509	0.511	0.511	0.481	0.554	0.486	0.462
		0.4	0.992	0.992	0.992	0.982	0.990	0.984	0.982	0.986	0.984	0.996	0.996	0.990
3	100	0	0.030	0.022	0.018	0.042	0.034	0.034	0.016	0.016	0.014	0.024	0.010	0.008
		0.2	0.170	0.136	0.136	0.182	0.117	0.115	0.139	0.139	0.129	0.226	0.150	0.172
		0.4	0.540	0.636	0.492	0.705	0.600	0.575	0.648	0.650	0.616	0.568	0.580	0.514
	200	0	0.042	0.024	0.026	0.056	0.038	0.040	0.042	0.034	0.034	0.030	0.018	0.014
		0.2	0.406	0.408	0.380	0.368	0.364	0.308	0.321	0.379	0.303	0.368	0.390	0.358
		0.4	0.956	0.982	0.960	0.976	0.984	0.972	0.952	0.974	0.948	0.964	0.986	0.978
	400	0	0.072	0.064	0.066	0.070	0.042	0.042	0.026	0.008	0.018	0.030	0.012	0.018
		0.2	0.786	0.806	0.758	0.810	0.824	0.818	0.758	0.846	0.738	0.746	0.780	0.762
		0.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
			$\eta = 0.4$											
1	100	0	0.046	0.022	0.016	0.023	0.011	0.006	0.017	0.004	0.011	0.012	0.004	0.008
		0.2	0.090	0.058	0.060	0.079	0.051	0.056	0.035	0.012	0.021	0.086	0.058	0.064
		0.4	0.234	0.230	0.184	0.257	0.161	0.176	0.215	0.185	0.180	0.226	0.158	0.146
	200	0	0.036	0.032	0.032	0.044	0.036	0.038	0.052	0.047	0.043	0.020	0.010	0.016
		0.2	0.162	0.144	0.148	0.116	0.089	0.081	0.164	0.130	0.134	0.124	0.102	0.110
		0.4	0.656	0.604	0.582	0.518	0.471	0.438	0.693	0.663	0.633	0.502	0.494	0.490
	400	0	0.030	0.022	0.020	0.052	0.030	0.042	0.036	0.018	0.022	0.028	0.016	0.022
		0.2	0.392	0.306	0.320	0.359	0.399	0.347	0.395	0.393	0.391	0.330	0.274	0.256
		0.4	0.918	0.936	0.910	0.921	0.943	0.921	0.931	0.951	0.945	0.926	0.950	0.924
2	100	0	0.008	0.008	0.008	0.019	0.008	0.012	0.032	0.022	0.028	0.024	0.012	0.012
		0.2	0.090	0.058	0.044	0.117	0.077	0.092	0.092	0.058	0.066	0.092	0.056	0.042
		0.4	0.530	0.458	0.386	0.471	0.348	0.342	0.464	0.457	0.365	0.449	0.377	0.315
	200	0	0.048	0.020	0.022	0.032	0.036	0.016	0.024	0.020	0.020	0.066	0.022	0.030
		0.2	0.312	0.228	0.230	0.236	0.234	0.201	0.242	0.202	0.169	0.276	0.206	0.176
		0.4	0.890	0.950	0.896	0.918	0.924	0.892	0.912	0.929	0.904	0.940	0.912	0.880
	400	0	0.036	0.016	0.016	0.048	0.008	0.022	0.048	0.032	0.032	0.042	0.030	0.034
		0.2	0.582	0.562	0.524	0.588	0.606	0.560	0.673	0.707	0.669	0.584	0.598	0.548
		0.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998
3	100	0	0.020	0.014	0.012	0.028	0.006	0.006	0.022	0.006	0.010	0.022	0.012	0.012
		0.2	0.214	0.178	0.172	0.181	0.161	0.159	0.226	0.188	0.162	0.168	0.132	0.132
		0.4	0.820	0.814	0.746	0.819	0.832	0.799	0.754	0.813	0.770	0.710	0.742	0.726
	200	0	0.024	0.014	0.012	0.040	0.018	0.030	0.018	0.010	0.020	0.036	0.022	0.022
		0.2	0.644	0.646	0.594	0.560	0.542	0.500	0.646	0.588	0.566	0.626	0.614	0.576
		0.4	0.998	0.996	0.996	0.994	0.996	0.994	0.998	0.998	0.998	1.000	1.000	0.998
	400	0	0.024	0.024	0.020	0.050	0.034	0.034	0.056	0.034	0.036	0.036	0.014	0.012
		0.2	0.886	0.914	0.904	0.942	0.936	0.904	0.888	0.920	0.898	0.936	0.930	0.898
		0.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

表 11：删失比例为 40% 时，两个配对样本的分位数曲线比较模拟结果（ τ 值范围）。样本量为 $n = 100$，200，400，显著性水平为 $\alpha = 0.05$ 且 Diff $= \beta_{21} - \beta_{11}$，因此 Diff $= 0$ 满足原假设。误差之间的协方差由 η 表示，见（6.1）。缩写 Bonf、T_{L_2} 和 T_{L_∞} 分别表示式（3.3）、（3.1）和（3.2）中定义的检验统计量。

			Bonf	T_{L_2}	T_{L_∞}	Bonf	T_{L_2}	T_{L_∞}	Bonf	T_{L_2}	T_{L_∞}	Bonf	T_{L_2}	T_{L_∞}
			De Backer 等（2019）			Peng 和 Huang（2008）			Portnoy（2003）			Wang 和 Wang（2009）		
Mod.	n	Diff.	$\eta = 0.2$											
1	100	0	0.010	0.000	0.010	0.012	0.006	0.012	0.004	0.008	0.004	0.006	0.002	0.006
		0.2	0.038	0.040	0.034	0.012	0.008	0.010	0.016	0.012	0.014	0.028	0.030	0.028
		0.4	0.086	0.102	0.088	0.025	0.046	0.042	0.052	0.073	0.058	0.042	0.070	0.046
	200	0	0.020	0.018	0.026	0.016	0.022	0.022	0.026	0.020	0.018	0.016	0.018	0.016
		0.2	0.042	0.044	0.054	0.024	0.022	0.026	0.101	0.064	0.091	0.048	0.048	0.058
		0.4	0.216	0.238	0.198	0.157	0.131	0.141	0.243	0.225	0.211	0.170	0.214	0.162
	400	0	0.062	0.038	0.046	0.012	0.018	0.010	0.022	0.026	0.014	0.042	0.028	0.036
		0.2	0.092	0.090	0.088	0.110	0.132	0.102	0.078	0.086	0.072	0.126	0.154	0.146
		0.4	0.496	0.586	0.486	0.466	0.518	0.456	0.544	0.606	0.540	0.410	0.506	0.448
2	100	0	0.014	0.008	0.016	0.012	0.006	0.014	0.018	0.016	0.016	0.012	0.006	0.012
		0.2	0.050	0.048	0.052	0.032	0.032	0.045	0.028	0.024	0.030	0.040	0.042	0.040
		0.4	0.128	0.166	0.122	0.090	0.157	0.136	0.166	0.186	0.180	0.140	0.194	0.158
	200	0	0.026	0.014	0.022	0.028	0.012	0.030	0.020	0.028	0.014	0.024	0.032	0.024
		0.2	0.100	0.138	0.106	0.122	0.130	0.120	0.090	0.102	0.084	0.126	0.122	0.130
		0.4	0.384	0.488	0.418	0.444	0.478	0.448	0.354	0.444	0.352	0.504	0.550	0.474
	400	0	0.022	0.032	0.030	0.036	0.040	0.038	0.026	0.026	0.022	0.024	0.030	0.030
		0.2	0.234	0.260	0.220	0.276	0.290	0.276	0.268	0.322	0.260	0.260	0.360	0.254
		0.4	0.806	0.900	0.846	0.834	0.904	0.842	0.902	0.948	0.892	0.882	0.950	0.888
3	100	0	0.018	0.012	0.014	0.028	0.032	0.028	0.008	0.014	0.012	0.018	0.014	0.016
		0.2	0.022	0.028	0.038	0.076	0.082	0.090	0.026	0.034	0.036	0.042	0.036	0.046
		0.4	0.360	0.496	0.392	0.204	0.270	0.208	0.306	0.396	0.348	0.326	0.384	0.370
	200	0	0.022	0.026	0.020	0.010	0.004	0.010	0.016	0.016	0.022	0.020	0.010	0.020
		0.2	0.190	0.282	0.202	0.176	0.216	0.206	0.296	0.288	0.274	0.194	0.264	0.202
		0.4	0.778	0.828	0.818	0.792	0.868	0.798	0.780	0.894	0.768	0.838	0.866	0.834
	400	0	0.020	0.022	0.024	0.022	0.012	0.018	0.020	0.016	0.022	0.016	0.020	0.018
		0.2	0.352	0.520	0.394	0.374	0.476	0.402	0.508	0.580	0.504	0.546	0.566	0.548
		0.4	0.994	0.998	0.994	0.992	0.994	0.988	0.988	0.998	0.992	0.992	0.994	0.990
			$\eta = 0.4$											
1	100	0	0.032	0.022	0.032	0.006	0.016	0.010	0.016	0.008	0.016	0.016	0.006	0.012
		0.2	0.026	0.022	0.024	0.004	0.016	0.006	0.017	0.012	0.017	0.020	0.018	0.022
		0.4	0.034	0.060	0.068	0.050	0.033	0.042	0.039	0.039	0.028	0.058	0.046	0.056
	200	0	0.010	0.012	0.010	0.016	0.006	0.008	0.018	0.012	0.016	0.004	0.000	0.006
		0.2	0.078	0.076	0.076	0.060	0.058	0.072	0.020	0.034	0.026	0.032	0.046	0.036
		0.4	0.254	0.256	0.232	0.175	0.236	0.171	0.233	0.337	0.275	0.190	0.246	0.168
	400	0	0.022	0.022	0.022	0.032	0.030	0.030	0.024	0.016	0.022	0.024	0.032	0.020
		0.2	0.168	0.206	0.170	0.104	0.116	0.136	0.080	0.082	0.084	0.140	0.154	0.144
		0.4	0.528	0.676	0.582	0.532	0.612	0.518	0.462	0.632	0.494	0.506	0.576	0.504
2	100	0	0.012	0.010	0.016	0.008	0.004	0.006	0.006	0.010	0.006	0.008	0.004	0.008
		0.2	0.066	0.066	0.060	0.026	0.049	0.030	0.040	0.051	0.047	0.030	0.036	0.030
		0.4	0.202	0.232	0.188	0.147	0.183	0.139	0.162	0.193	0.162	0.124	0.160	0.126
	200	0	0.026	0.016	0.022	0.012	0.014	0.016	0.016	0.018	0.014	0.012	0.010	0.010
		0.2	0.136	0.142	0.120	0.108	0.128	0.106	0.080	0.100	0.088	0.092	0.094	0.108
		0.4	0.594	0.732	0.644	0.504	0.660	0.494	0.654	0.732	0.676	0.498	0.674	0.506
	400	0	0.028	0.028	0.026	0.048	0.030	0.052	0.026	0.016	0.018	0.020	0.016	0.020
		0.2	0.338	0.384	0.306	0.312	0.352	0.300	0.216	0.274	0.236	0.356	0.428	0.344
		0.4	0.934	0.954	0.946	0.944	0.984	0.964	0.972	0.988	0.966	0.974	0.988	0.972
3	100	0	0.014	0.000	0.014	0.012	0.006	0.014	0.004	0.000	0.006	0.008	0.002	0.010
		0.2	0.078	0.062	0.088	0.104	0.062	0.110	0.036	0.054	0.050	0.040	0.084	0.066
		0.4	0.454	0.548	0.472	0.325	0.431	0.341	0.324	0.439	0.360	0.320	0.444	0.364
	200	0	0.010	0.008	0.012	0.032	0.012	0.022	0.012	0.012	0.016	0.010	0.006	0.008
		0.2	0.326	0.352	0.320	0.284	0.320	0.284	0.240	0.332	0.216	0.206	0.214	0.210
		0.4	0.882	0.914	0.892	0.872	0.916	0.848	0.818	0.924	0.826	0.716	0.858	0.730
	400	0	0.008	0.008	0.008	0.016	0.016	0.014	0.020	0.016	0.020	0.014	0.008	0.016
		0.2	0.622	0.714	0.618	0.578	0.668	0.502	0.708	0.782	0.700	0.446	0.618	0.432
		0.4	0.996	1.000	0.996	0.996	0.998	0.996	1.000	1.000	1.000	1.000	1.000	1.000

表 12：在 40% 删失条件下，对两个配对样本的中位数（$\tau = 0.5$）进行比较的模拟结果。样本量为 $n = 100$，200，400，显著性水平为 $\alpha = 0.05$，Diff $= \beta_{21} - \beta_{11}$，因此原假设对于 Diff $= 0$ 成立。误差之间的协方差由 η 表示，见（6.1）。缩写 Bonf、T_{L_2} 和 T_{L_∞} 分别表示（3.3）、（3.1）和（3.2）中定义的检验统计量。