arXiv:2304.05491v1 [math.ST] 11 Apr 2023

# 基于 R´enyi 伪距离的独立非同分布观测值模型选择

Angel Felipe[1], MariaJaenada[1], Pedro Miranda[1] 和 Leandro Pardo[11]统计与运算学系，马德里康普顿斯大学，西班牙

摘要

模型选择准则是用于在一组候选模型中选择最佳统计模型的规则，在拟合优度与模型复杂度之间寻求平衡。大多数常用的模型选择准则通过模型对数似然函数来衡量拟合优度，从而产生非稳健准则。本文提出了一类新的稳健模型选择准则，用于独立但不同分布的观测（i.n.i.d.o.），其基于R´enyi的伪距离（RP）。基于RP 的模型选择准则由调节参数 $\alpha$控制，该参数调节效率与稳健性之间的权衡。本文推导了关于 RP 准则的一些理论结果，并将该理论应用于多元线性回归模型，得到模型选择准则的显式表达式。此外，本文还考虑了受限模型，并相应推导出具有嵌套模型的多元线性回归模型下的显式表达式。最后，通过一项模拟研究，从经验上说明了该方法在稳健性方面的优势。

*Keywords:* R´ényi伪距离、稳健性、受限模型、多元线性回归模型。

## 1 引言

考虑一组来自未知分布、需要进行统计建模的实际观测。可以假设不同的候选模型能够拟合这些数据，因此自然会产生一个问题：如何选择最适合这些数据的模型。如果假设的模型过于简单、参数数量过少，就可能无法捕捉数据中的某些重要模式和关系。相反，如果假设的模型过于复杂、参数数量过多，估计得到的模型参数就可能过度拟合观测数据（包括可能存在的样本噪声），从而导致模型应用于新数据时表现不佳。模型选择准则是一种基于观测数据、从一组候选模型中选择统计模型的规则。它定义了一个客观的准则函数，用以量化这种折中。

在拟合优度与模型复杂度之间进行权衡，通常通过期望差异度或散度来衡量。随后，需要将差异度量最小化，以选择权衡效果最佳的模型。换言之，模型选择准则依赖于候选模型与真实模型（即生成数据的概率分布）之间公平性的度量。

赤池信息准则（AIC）是统计实践中最广为人知和使用的模型选择准则之一。它由赤池 [1, 2] 提出，是统计学文献中的第一个模型选择准则。AIC估计数据背后真实模型与拟合的候选模型之间的预期Kullback-Leibler散度 [20] ，并选择AIC最小的模型。当然，数据背后的真实模型通常是未知的，因此使用从观测数据中获得的经验估计。

沿袭与赤池信息准则相似的思路，文献中还提出了其他几种模型选择准则。例如，Schwarz 在 [24] 中提出了“贝叶斯信息准则”（BIC），其对模型复杂度的惩罚比赤池信息准则更强。Hurvich 和 Tsai [13, 14, 15] 也从赤池信息准则出发，研究了其偏差问题，并通过提出一种称为“修正赤池信息准则”（$\mathrm{AIC}_C$）的新准则对其进行校正。该准则试图解决这样一个问题：赤池信息准则仅渐近无偏，因此当样本量不够大、参数数量较多时，偏差可能会很显著。事实上，在小样本下，赤池信息准则往往会对观测数据过拟合。Konishi 和 Kitagawa [19] 将赤池信息准则所建立的框架扩展至一般框架，其中包括最大似然以外的其他估计方法，用于拟合假定的候选模型。由此得到的模型选择准则称为“广义信息准则”（GIC）。当拟合方法为最大似然时，GIC 的惩罚项可化为 Takeuchi 在 [25] 提出的“竹内信息准则”（TIC）的惩罚项。最后，Bozdogon [5] 提出了赤池信息准则的另一种变体，称为 CAIC，用以弥补其不具一致性的缺陷。关于模型选择准则的有趣综述可见 [23, 9]。

之前的大多数程序都根据 Kullback-Leibler 散度来衡量公平性。然而，人们也探索了其他一些散度度量，将这些方法扩展为具有更好稳健性的形式。例如，[22] 考虑了密度幂散度（DPD） [3] ，以定义稳健的模型选择准则。类似地，Toma 等人 [26] 引入了另一种基于 R´enyi 伪距离（RP） [18] 的稳健模型选择准则。

此前的所有准则都假设观测值相互独立且同分布。如果观测值相互独立但不具有相同分布（独立非同分布观测）(i.n.i.d.o.)，则会出现一个新的问题。在此背景下，Kurata和 Hamada [21] 考虑了一个基于 DPD 的准则，扩展了 [22] 的理论。本文的主要目的是在 i.n.i.d.o. 背景下，基于 RP 引入一种新的稳健模型选择准则，从而扩展 [26] 的方法。

本文其余部分安排如下。在第2节中，我们介绍适用于独立非同分布观测的RP，并给出后续各节所需的一些理论结果。

第3节讨论基于RP的准则，并给出了其在多元线性回归模型（多元线性回归模型）中的应用。第4节研究了受限情形，其中对参数空间施加了一些附加条件。推导出了多元线性回归模型中将具有许多参数的模型与另一个参数数量较少的模型进行比较时的相应显式表达式。第5节中的模拟研究展示了所提出准则的稳健性，并将其与其他模型选择准则进行了比较。第6节介绍了一个真实数据示例。最后的结论在第7节中给出。

## 2 R´enyi 的独立但非同分布观测值伪距离

设 $Y_1$，...，$Y_n$ 为独立非同分布观测，其中每个 $Y_i$ 分别具有真实概率分布函数 $G_i$、$i = 1$、...、$n$，以及概率密度函数 $g_i$、$i = 1$、...、$n$。出于推断目的，假定真实密度函数 $g_i$ 可能属于参数化密度族 $f_i(y, \boldsymbol{\theta})$ $i = 1$、...、$n$，其中 $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$ 是所有密度函数共用的模型参数。下文中，我们将用 $F_i(y, \boldsymbol{\theta})$ 表示与密度函数 $f_i(y, \boldsymbol{\theta})$ $i = 1$、...、$n$. 相对应的分布函数。

能够最佳拟合原始分布 $g_1$、...、$g_n$ 的 $\boldsymbol{\theta}$，自然会使真实密度和假设密度（$g_1(y)$ ...， $g_n(y)$）与（$f_1(y, \boldsymbol{\theta})$ ...， $f_n(y, \boldsymbol{\theta})$）之间的某种距离最小化。这里，我们将使用 [18] 中定义的 RP散度度量，作为衡量两组密度接近程度的度量。

定义 1 设 $f(\cdot, \boldsymbol{\theta})$ $g(\cdot)$ 为两个概率密度函数。调节参数为 $\alpha > 0$ 时， $f$ 和 $g$ 之间的 Rényi 伪距离（RP）定义为

$$R_\alpha\left(f(\cdot, \boldsymbol{\theta}), g(\cdot)\right) = \frac{1}{\alpha+1} \log\left(\int f(y, \boldsymbol{\theta})^{\alpha+1} dy\right) - \frac{1}{\alpha} \log\left(\int f(y, \boldsymbol{\theta})^{\alpha} g(y) dy\right) + \frac{1}{\alpha\left(\alpha+1\right)} \log\left(\int g(y)^{\alpha+1} dy\right). \tag{1}$$

调节参数 $\alpha$ 控制效率与稳健性之间的权衡。因此，当 $\alpha$ 取较小值时（在极限情况下 $\alpha = 0$），相应结果的效率较高，但稳健性较低。另一方面，当 $\alpha$ 取较大值时，结果将更稳健，但效率会有所损失。

式(1)中定义的RP散度始终为正，且仅当两个密度一致时才达到零。因此，近似底层分布的最佳模型参数值自然会使式(1)最小化在$\boldsymbol{\theta} \in \boldsymbol{\Theta}$. 事实上，如果真实分布 $g$ 属于所假设的参数化

参数为 $\boldsymbol{\theta}_0$ 的模型，其 RP 的全局极小值点必然是 $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

在 $\alpha = 0$ 中， $f$ 与 $g$之间对应的 R´enyi 伪距离可通过取连续极限定义如下

$$\begin{aligned} R_0\left(f(\cdot,\boldsymbol{\theta}), g(\cdot)\right) &= \lim_{\alpha\downarrow 0} R_\alpha\left(f(y,\boldsymbol{\theta}), g(y)\right) = \int g(y)\log\frac{g(y)}{f(y,\boldsymbol{\theta})}dy \\ &= \int g(y)\log g(y)dy - \int g(y) log f(y,\boldsymbol{\theta})dy. \end{aligned} \tag{2}$$

因此， $R_0\left(f(\cdot,\boldsymbol{\theta}),\ g(\cdot)\right)$ 与 $g$ 和 $f$之间的 Kullback-Leibler散度度量一致。RP散度已应用于许多不同的统计模型，并在仅损失少量效率的情况下展现出非常理想的稳健性。例如， [12] 以$\gamma$-交叉熵之名研究了 RP 散度。此外，Toma 和 Leoni-Auban [27] 基于 RP 定义了新的稳健且高效的度量。 [7], 在多元线性回归模型（MLRM）的背景下开发了基于 RP 的 Wald型检验，随后又在 [17] 将其扩展到广义多元回归模型。此外， [17] 采用基于 RP 的 Wald型检验，提出了一种比较两个相关正态总体的稳健方法。 [16] 研究了受限MRPE并考察了其渐近性质；此外，还提出了基于受限Rényi伪距离的 Rao型检验的应用。

注意，式 (1) 中的最后一项不依赖于 $\boldsymbol{\theta}$. 。因此，对于 $\alpha > 0$，可以通过最小化替代函数来获得 RP 测度的极小值点。

$$\frac{1}{\alpha+1}\log\left(\int f_i(y,\boldsymbol{\theta})^{\alpha+1}dy\right) - \frac{1}{\alpha}\log\left(\int f(y,\boldsymbol{\theta})^{\alpha} g(y)dy\right). \tag{3}$$

上述表达式可利用对数的性质改写为

$$-\frac{1}{\alpha}\log\frac{\int f(y,\boldsymbol{\theta})^{\alpha} g(y)dy}{\left(\int f(y,\boldsymbol{\theta})^{\alpha+1}dy\right)^{\frac{\alpha}{\alpha+1}}},$$

因此，在 $\boldsymbol{\theta}$ 中对 $\alpha > 0$最小化 $R_\alpha\left(f(\cdot,\boldsymbol{\theta})\ g(\cdot)\right)$ 等价于最小化

$$V_\alpha^*\left(\boldsymbol{\theta}\right) = -\frac{\int f(y,\boldsymbol{\theta})^{\alpha} g(y)dy}{\left(\int f(y,\boldsymbol{\theta})^{\alpha+1}dy\right)^{\frac{\alpha}{\alpha+1}}}. \tag{4}$$

同样，对于 $\alpha = 0$，我们有，式（2）中的第一项不依赖于 $\boldsymbol{\theta}$，因此，最小化 $R_0\left(f(\cdot,\boldsymbol{\theta})\ g(\cdot)\right)$ 等价于最小化

$$V_0^*\left(\boldsymbol{\theta}\right) = -\int g(y) log f(y,\boldsymbol{\theta})dy. \tag{5}$$

然而，此时当 $\alpha \to 0.$ 时，表达式 (4) 不会趋于表达式 (5)。为了恢复这种收敛性，并进一步推广基于

Kullback-Leibler散度的经典结果，我们对表达式 (4) 稍作修改如下：

$$V_\alpha(\boldsymbol{\theta}) = -\frac{\int f(y,\boldsymbol{\theta})^\alpha g(y)dy}{\alpha\left(\int f(y,\boldsymbol{\theta})^{\alpha+1}dy\right)^{\frac{\alpha}{\alpha+1}}} + \frac{1}{\alpha}, \tag{6}$$

其中，使（4）式最小化的 $\boldsymbol{\theta}$ 的取值与使（6）式最小化的取值相同。下一个引理证明目标函数所需的收敛性。

引理 2对于任意两个密度函数 $f(\cdot,\boldsymbol{\theta})$ 和 $g(\cdot)$，以下收敛关系成立

$$\lim_{\alpha\to 0} V_{i,\alpha}(\boldsymbol{\theta}) = V_{i,0}(\boldsymbol{\theta}).$$

证明。首先，注意到

$$\lim_{\alpha\to 0}\left(-\frac{\int f(y,\boldsymbol{\theta})^\alpha g(y)dy}{\alpha\left(\int f(y,\boldsymbol{\theta})^{\alpha+1}dy\right)^{\frac{\alpha}{\alpha+1}}} + \frac{1}{\alpha}\right) \tag{7}$$

导致不确定值（$0/0$）。我们记为

$$z(\alpha) = \left(\int f(y,\boldsymbol{\theta})^{\alpha+1}dy\right)^{\frac{\alpha}{\alpha+1}}.$$

对其对数求导数

$$\log z(\alpha) = \frac{\alpha}{\alpha+1}\log\left(\int f(y,\boldsymbol{\theta})^{\alpha+1}dy\right),$$

经过一些代数运算，我们得到 $\frac{\partial}{\partial\alpha}\log z(\alpha) = \frac{z'(\alpha)}{z(\alpha)} = \frac{1}{\alpha(\alpha+1)}\cdot$另一方面，函数 $\log z(\alpha)$ 的导数为

$$\frac{\partial \log z(\alpha)}{\partial\alpha} = \frac{1}{(\alpha+1)^2}\log\left(\int f(y,\boldsymbol{\theta})^{\alpha+1}dy\right) + \frac{\alpha}{\alpha+1}\frac{\left(\int f(y,\boldsymbol{\theta})^{\alpha+1}\log f(y,\boldsymbol{\theta})dy\right)}{\left(\int f(y,\boldsymbol{\theta})^{\alpha+1}dy\right)},$$

求解上述方程可得

$$\begin{aligned}\frac{\partial z(\alpha)}{\partial\alpha} = &\left[\frac{1}{(\alpha+1)^2}\log\left(\int f(y,\boldsymbol{\theta})^{\alpha+1}dy\right) + \frac{\alpha}{\alpha+1}\frac{\left(\int f(y,\boldsymbol{\theta})^{\alpha+1}\log f(y,\boldsymbol{\theta})dy\right)}{\left(\int f(y,\boldsymbol{\theta})^{\alpha+1}dy\right)}\right]\\ &\times\left(\int f(y,\boldsymbol{\theta})^{\alpha+1}dy\right)^{\frac{\alpha}{\alpha+1}}.\end{aligned}$$

因此，对式（7）应用洛必达法则，可得

$$\lim_{\alpha\to 0} -\frac{\int f(y,\boldsymbol{\theta})^\alpha g(y)dy}{\alpha\left(\int f(y,\boldsymbol{\theta})^{\alpha+1}dy\right)^{\frac{\alpha}{\alpha+1}}} + \frac{1}{\alpha} = \lim_{\alpha\to 0}\frac{-\int f(y,\boldsymbol{\theta})^\alpha g(y)\log f(y,\boldsymbol{\theta})dy + \frac{\partial z(\alpha)}{\partial\alpha}}{z - \alpha\frac{\partial z(\alpha)}{\partial\alpha}}.$$

最后，

- $\lim_{\alpha\to 0}\int f(y,\boldsymbol{\theta})^{\alpha}g(y)\log f(y,\boldsymbol{\theta})dy=\int g(y)\log f(y,\boldsymbol{\theta})dy.$
- $\lim_{\alpha\to 0}\frac{\partial z(\alpha)}{\partial\alpha}=\frac{1}{1}\log 1+\frac{0}{1}\frac{\int f(y,\boldsymbol{\theta})\log f(y,\boldsymbol{\theta})dy}{1}=0.$
- $\lim_{\alpha\to 0} z=1^{0}=1.$

因此，该结果成立。 ■

现在，令 $V_{i,\alpha}(\boldsymbol{\theta})$ 表示每一对分布 $(f_i(y,\boldsymbol{\theta})\ g_i(y))\ i=1,...,n$ 对应的目标函数，如式 (6) 所示。由于所有密度$f_i(y,\boldsymbol{\theta})$ 共享一个参数，最能近似不同真实密度的模型参数应使加权目标函数最小化，并对所有函数 $V_{i,\alpha}(\boldsymbol{\theta})$ 赋予相同的权重。因此，我们考虑

$$H_{\alpha}(\boldsymbol{\theta})=\frac{1}{n}\sum_{i=1}^{n}V_{i,\alpha}(\boldsymbol{\theta})=\frac{1}{n}\sum_{i=1}^{n}\left[-\frac{\int f_i(y,\boldsymbol{\theta})^{\alpha}g_i(y)dy}{\alpha\left(\int f_i(y,\boldsymbol{\theta})^{\alpha+1}dy\right)^{\frac{\alpha}{\alpha+1}}}+\frac{1}{\alpha}\right]. \tag{8}$$

定义3 考虑 $(g_1(y),...,g_n(y))$ 和 $(f_1(y,\boldsymbol{\theta}),...,f_n(y,\boldsymbol{\theta}))$ $n$ 对真实密度和假设密度，这些密度对应于独立非同分布观测随机变量 $Y_i, i=1,...,n.$ 对于任意 $\alpha\geq 0$，值 $\boldsymbol{\theta}_{\boldsymbol{g},\alpha}$ 满足

$$\boldsymbol{\theta}_{\boldsymbol{g},\alpha}=\arg\min_{\boldsymbol{\theta}}\frac{1}{n}\sum_{i=1}^{n}\left[-\frac{\int f_i(y,\boldsymbol{\theta})^{\alpha}g_i(y)dy}{\alpha\left(\int f_i(y,\boldsymbol{\theta})^{\alpha+1}dy\right)^{\frac{\alpha}{\alpha+1}}}+\frac{1}{\alpha}\right]=\arg\min_{\boldsymbol{\theta}}\frac{1}{n}\sum_{i=1}^{n}V_{i,\alpha}(\boldsymbol{\theta}).$$

称为根据 RP 得到的最佳拟合参数。

在下文中，我们假设存在一个开子集 $\boldsymbol{\Theta_0}\subset\boldsymbol{\Theta}$，其中包含最佳拟合参数 $\boldsymbol{\theta}_{\boldsymbol{g},\ \alpha}$.

对于任意固定的 $i=1$、...、 $n$，随机变量 $Y_i$ 的真实分布 $g_i$ 在实践中通常是未知的，因此 $\boldsymbol{\theta}_{\boldsymbol{g},\alpha}$ 必须通过经验估计。由于我们只有每个变量 $Y_i$ 的一个观测值，基于观测值 $y_i$ 进行 $g_i$估计的最佳方式是假设该分布在$y_i.$ 上是退化的。我们将这一退化分布记为 $\widehat{g}_{i.}$ 。因此，带有 $\alpha>0$ 的RP散度在公式（1）中给出，为

$$R_{\alpha}\left(f_i(Y_i,\boldsymbol{\theta}),\widehat{g}_i\right)=\frac{1}{\alpha+1}\log\left(\int f_i(y,\boldsymbol{\theta})^{\alpha+1}dy\right)-\frac{1}{\alpha}\log f_i(Y_i,\boldsymbol{\theta})^{\alpha}+k, \tag{9}$$

类似地，式 (2) 中给出的 $\alpha=0$ 的 RP 经验估计为

$$R_0\left(f_i(Y_i,\boldsymbol{\theta}),\widehat{g}_i\right)=-\log f_i(Y_i,\boldsymbol{\theta})+k, \tag{10}$$

其中，(9) 和 (10) 中的 $k$ 表示一个不依赖于 $\boldsymbol{\theta}.$ 的常数。如前所述，基于 RP 散度的模型参数 $\boldsymbol{\theta}$的最佳估计量，应使其经验估计最小化。不过，再次说明，对于 $\alpha>0$，最小化估计的 RP， $R_{\alpha}(f_i(Y_i,\boldsymbol{\theta})\ \widehat{g}_i)$，等价于最小化

$$\widehat{V}_{i,\alpha}\left(Y_i,\boldsymbol{\theta}\right)=-\frac{f_i(Y_i,\boldsymbol{\theta})^{\alpha}}{\alpha\left(\int f_i(y,\boldsymbol{\theta})^{\alpha+1}dy\right)^{\frac{\alpha}{\alpha+1}}}+\frac{1}{\alpha}. \tag{11}$$

对于 $\alpha = 0$，我们可以用同样的方法得出结论：在 $\boldsymbol{\theta}$ 中最小化$R_0\left(f_i(Y_i,\boldsymbol{\theta})\ \widehat{g}_i\right)$，等价于最小化

$$\widehat{V}_{0,\alpha}\left(Y_i,\boldsymbol{\theta}\right) = -\log f(Y_i,\boldsymbol{\theta}). \tag{12}$$

现在，关于参数真实值的所有可用信息都来自观测数据集，因此，为了得到能够联合拟合所有观测值的最佳估计，我们应考虑针对 $\alpha > 0$ 给出的加权目标函数:

$$\begin{aligned} H_{n,\alpha}(\boldsymbol{\theta}) =& \frac{1}{n}\sum_{i=1}^{n}\left[-\frac{f_i(Y_i,\boldsymbol{\theta})^{\alpha}}{\alpha L_{\alpha}^{i}\left(\boldsymbol{\theta}\right)}+\frac{1}{\alpha}\right] \\ =& \frac{1}{n}\sum_{i=1}^{n}\widehat{V}_{i,\alpha}(Y_i,\boldsymbol{\theta}). \end{aligned} \tag{13}$$

with

$$L_{\alpha}^{i}\left(\boldsymbol{\theta}\right) = \left(\int f_i(y,\boldsymbol{\theta})^{\alpha+1}dy\right)^{\frac{\alpha}{\alpha+1}},$$

相应地,

$$H_{n,0}(\boldsymbol{\theta}) = \lim_{\alpha\to 0} H_{n,\alpha}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\widehat{V}_{i,0}(Y_i,\boldsymbol{\theta}). \tag{14}$$

请注意，此时估计量的期望值确实就是理论目标函数。

$$V_{i,\alpha}(\boldsymbol{\theta}) = E_{Y_i}\left[\widehat{V}_{i,\alpha}(Y_i,\boldsymbol{\theta})\right], \quad H_{\alpha}(\boldsymbol{\theta}) = E_{Y_1,\ldots,Y_n}\left[H_{n,\alpha}(\boldsymbol{\theta})\right].$$

定义 4 设 $Y_1,\ldots,Y_n$ 为独立非同分布观测，且 $\alpha > 0$，则最小 RP $\widehat{}$估计量（MRPE） $\boldsymbol{\theta}_{\alpha}$定义为

$$\widehat{\boldsymbol{\theta}}_{\alpha} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} H_{n,\alpha}(\boldsymbol{\theta}), \tag{15}$$

其中，$H_{n,\ \alpha}(\boldsymbol{\theta})$在 $\alpha > 0$ 时由式（13）定义，在 $\alpha = 0$.时由式（14）定义。

请注意，当 $\alpha = 0$时，我们得到模型的最大似然估计量（MLE），因此 MRPE 族包含作为特例的经典估计量。

作为 MRPE，$\widehat{\boldsymbol{\theta}}_{\alpha}$是可微函数的极小值，则它必须使该函数的一阶导数$H_{n,\alpha}(\boldsymbol{\theta})$为零

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\widehat{V}_{i,\alpha}(Y_i;\boldsymbol{\theta})}{\partial\theta_j} = 0, \quad j = 1,\ldots,p.$$

也就是说，MRPE 的估计方程为

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\alpha L_{\alpha}^{i}\left(\boldsymbol{\theta}\right)^{2}}\left(\alpha f_i(Y_i,\boldsymbol{\theta})^{\alpha}u_j(Y_i,\boldsymbol{\theta})L_{\alpha}^{i}\left(\boldsymbol{\theta}\right) - \frac{\partial L_{\alpha}^{i}\left(\boldsymbol{\theta}\right)}{\partial\theta_j}f_i(Y_i,\boldsymbol{\theta})^{\alpha}\right) = 0, \quad j = 1,\ldots,p,$$

with

$$u_j(y,\boldsymbol{\theta}) = \frac{\partial \log(f_i(y,\boldsymbol{\theta}))}{\partial \theta_j},$$

and

$$\begin{aligned}\frac{\partial L_\alpha^i(\boldsymbol{\theta})}{\partial \theta_j} &= \frac{\alpha}{\alpha+1}\left(\int f_i(y,\boldsymbol{\theta})^{\alpha+1}dy\right)^{\frac{\alpha}{\alpha+1}-1}(\alpha+1)\int f_i(y,\boldsymbol{\theta})^{\alpha+1}u_j(y,\boldsymbol{\theta})dy \\ &= \alpha\left(\int f_i(y,\boldsymbol{\theta})^{\alpha+1}dy\right)^{\frac{\alpha}{\alpha+1}-1}\int f_i(y,\boldsymbol{\theta})^{\alpha+1}u_j(y,\boldsymbol{\theta})dy, i=1,...,n.\end{aligned}$$

有趣的是，如果 $Y_1$、...、$Y_n$ 是独立同分布（i.i.d.）的随机变量，那么 MRPE $\boldsymbol{\theta}_\alpha$ 与 [6] 中提出的估计量$\widehat{\boldsymbol{\theta}}_\alpha^*$ 相同。

接下来我们研究 MRPE 的渐近分布，$\widehat{\boldsymbol{\theta}}_\alpha$. 为简化记号，定义矩阵 $\boldsymbol{\Psi}_{n,\alpha}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})$ 和 $\Omega_{n,\alpha}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})$ 如下:

$$\boldsymbol{\Psi}_{n,\alpha}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{J}_\alpha^{(i)}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}), \tag{16}$$

with

$$\boldsymbol{J}_\alpha^{(i)}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) = \left(E_{Y_i}\left[\frac{\partial^2 \widehat{V}_{i,\alpha}(Y_i;\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}\right)_{j,k=1,...,p}, i=1,...,n,$$

and

$$\boldsymbol{\Omega}_{n,\alpha}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) = \frac{1}{n}\sum_{i=1}^{n}Var_{Y_i}\left[\left(\frac{\partial \widehat{V}_{i,\alpha}(Y_i;\boldsymbol{\theta})}{\partial\theta_j}\right)_{j=1,..,p}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}, i=1,...,n. \tag{17}$$

此外，设 $\lambda_1, ..., \lambda_n$ 为 $\Omega_{n,\alpha}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})$ 的特征值。从现在起，我们假设 $\inf_n \lambda_n > 0$，从而 $\Omega_{n,\alpha}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})$ 可逆。

我们考虑以下正则条件:

C1. 对所有 $i$ 而言，密度函数 $f_i(y,\boldsymbol{\theta})$ 的支撑集 $\mathcal{X}$ 都相同，且不依赖于 $\boldsymbol{\theta}$. 。此外，真实的概率密度函数$g_1, ..., g_n$ 具有相同的支撑集 $\mathcal{X}$。

C2. 对几乎所有 $y \in \mathcal{X}$，密度 $f_i(y,\boldsymbol{\theta})$ 关于 $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ 和 $i = 1, ..., n$. 存在所有三阶导数。

C3. 对于 $i = 1$、2,..., $n$，积分

$$\int f_i(y,\boldsymbol{\theta})^{1+\alpha}dy$$

关于 $\boldsymbol{\theta}$ 可进行三次微分，并且可以交换积分与微分。由此条件可得

$$\frac{\partial V_{i,\alpha}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = E_{Y_i}\left[\frac{\partial \widehat{V}_{i,\alpha}(Y_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right], \quad \frac{\partial^2 V_{i,\alpha}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T} = E_{Y_i}\left[\frac{\partial^2 \widehat{V}_{i,\alpha}(Y_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T}\right] = \boldsymbol{J}_{\alpha}^{(i)}(\boldsymbol{\theta}).$$

C4. 对于 $i = 1$、2、...、$n$，矩阵 $\boldsymbol{J}_{\alpha}^{(i)}(\boldsymbol{\theta}_{\boldsymbol{g},\ \alpha})$ 是正定的。

C5. 存在函数 $M_{jkl}^{(i)}$ 和常数 $m_{jkl}$，使得

$$\left|\frac{\partial^3 \widehat{V}_{i,\alpha}(y;\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k\partial\theta_l}\right| \leq M_{jkl}^{(i)}(y), \qquad \forall \boldsymbol{\theta} \in \boldsymbol{\Theta},\ \forall j,k,l$$

and

$$E_Y\left[M_{jkl}^{(i)}(Y)\right] = m_{jkl} < \infty, \qquad \forall \boldsymbol{\theta} \in \boldsymbol{\Theta},\ \forall j,k,l.$$

C6. 对所有 $j$、$k$、$l$和 $\boldsymbol{\theta} \in \boldsymbol{\Theta}$，序列 $\left\{\frac{\partial \widehat{V}_{i,\alpha}(Y_i,\boldsymbol{\theta})}{\partial\theta_j}\right\}_{j=1}$（，）、...、$p$、$\left\{\frac{\partial^2 \widehat{V}_{i,\alpha}(Y_i,\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right\}_j$（，）、$k$=1、...、$p$和 $\left\{\frac{\partial^3 \widehat{V}_{i,\alpha}(Y_i,\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k\partial l}\right\}_j$（，）、$k$、$l$=1、...、$p$在 Ces`aro 意义下一致可积，即

$$\lim_{n\to\infty}\left(\sup_{n>1}\frac{1}{n}\sum_{i=1}^{n} E_{Y_i}\left[\left|\frac{\partial \widehat{V}_{i,\alpha}(Y_i,\boldsymbol{\theta})}{\partial\theta_j}\right| I_{\left\{\frac{\partial V_{i,\alpha}(Y_i,\boldsymbol{\theta})}{\partial\theta_j} > n\right\}}(Y_i)\right]\right) = 0,$$

$$\lim_{n\to\infty}\left(\sup_{n>1}\frac{1}{n}\sum_{i=1}^{n} E_{Y_i}\left[\left|\frac{\partial^2 \widehat{V}_{i,\alpha}(Y_i,\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right| I_{\left\{\frac{\partial^2 V_{i,\alpha}(Y_i,\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k} > n\right\}}(Y_i)\right]\right) = 0,$$

$$\lim_{n\to\infty}\left(\sup_{n>1}\frac{1}{n}\sum_{i=1}^{n} E_{Y_i}\left[\left|\frac{\partial^3 \widehat{V}_{i,\alpha}(Y_i,\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k\partial\theta_l}\right| I_{\left\{\frac{\partial^3 V_{i,\alpha}(Y_i,\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k\partial\theta_l} > n\right\}}(Y_i)\right]\right) = 0.$$

C7. 对所有 $\varepsilon > 0$

$$\lim_{n\to\infty}\left\{\frac{1}{n}\sum_{i=1}^{n} E_{Y_i}\left[\left\|\boldsymbol{\Omega}_n^{-\frac{1}{2}}(\boldsymbol{\theta})\frac{\partial \widehat{V}_{i,\alpha}(Y_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|_2^2 I_{\left\{\left\|\boldsymbol{\Omega}_n^{-\frac{1}{2}}(\boldsymbol{\theta})\frac{\partial \widehat{V}_{i,\alpha}(Y_i,\theta)}{\partial\boldsymbol{\theta}}\right\|_2^2\right\}}(Y_i)\right] > \varepsilon\sqrt{n}\right\} = 0.$$

现在，下面的结果成立，其证明见 [7]，。

定理 5 假设前述正则条件 C1-C7 成立。则，

$$\sqrt{n}\boldsymbol{\Omega}_{n,\alpha}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^{-\frac{1}{2}}\boldsymbol{\Psi}_{n,\alpha}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\left(\widehat{\boldsymbol{\theta}}_{\alpha} - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right) \underset{n\to\infty}{\overset{L}{\to}} N(\mathbf{0}_p, \boldsymbol{I}_p), \qquad (18)$$

为 $\boldsymbol{I}_p$，即 p 维单位矩阵。

## 2.1 示例：MLRM 下的最小概率相对误差

设 $(Y_1, ..., Y_n)$ 为一组随机变量，它们通过多元线性回归模型与解释变量 $(\boldsymbol{X}_1, ..., \boldsymbol{X}_n)$ 相关，

$$Y_i = \boldsymbol{X}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{19}$$

其中，误差项 $\varepsilon_i' s$ 是均值为零、方差为 $\sigma^2$的独立同分布的正态随机变量，$\boldsymbol{X}_i^T = （X_{i1}, ..., X_{ip}）$是与第 $i$个条件相对应的自变量向量，而 $\boldsymbol{\beta} = \left(\beta_1, ..., \beta_p\right)^T$是待估计的回归系数向量。我们将认为，对于每个 $i$，$\boldsymbol{X}_i$ 是固定的，从而得到独立非同分布观测 $Y_i' s$，其中 $Y_i \sim \mathcal{N}（\boldsymbol{X}_i^T \boldsymbol{\beta}, \sigma^2）$。

接下来，我们推导参数$\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)$的MRPE 显式表达式。采用前面的记号，假设的密度函数为 $f_i(y, \boldsymbol{\beta}, \sigma) \equiv \mathcal{N}(\boldsymbol{X}_i^T \boldsymbol{\beta}, \sigma^2)$，然后利用公式（6），我们有，对于 $\alpha > 0$,

$$\begin{aligned}
\widehat{V}_{i,\alpha}(Y_i; \boldsymbol{\beta}, \sigma) &= -\frac{\frac{1}{(2\pi)^{\alpha/2}\sigma^{\alpha}} \exp\left(\frac{-\alpha(Y_i - \boldsymbol{X}_i^T \boldsymbol{\beta})^2}{2\sigma^2}\right)}{\alpha\left((2\pi)^{\alpha/2}\sigma^{\alpha}\sqrt{1+\alpha}\right)^{-\frac{\alpha}{\alpha+1}}} + \frac{1}{\alpha} \\
&= -\frac{1}{\alpha}\left(\frac{1+\alpha}{2\pi}\right)^{\frac{\alpha}{2(\alpha+1)}} \sigma^{-\frac{\alpha}{\alpha+1}} \exp\left(-\frac{\alpha}{2}\left(\frac{Y_i - \boldsymbol{X}_i^T \boldsymbol{\beta}}{\sigma}\right)^2\right) + \frac{1}{\alpha}.
\end{aligned} \tag{20}$$

因此，通过最小化平均目标函数，得到 $\alpha > 0$ 的 MRPE

$$\begin{aligned}
H_{n,\alpha}(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{i=1}^{n} \widehat{V}_{i,\alpha}(Y_i; \boldsymbol{\beta}, \sigma) \\
&= -\frac{1}{\alpha}\left(\frac{1+\alpha}{2\pi}\right)^{\frac{\alpha}{2(\alpha+1)}} \frac{1}{n}\sum_{i=1}^{n} \sigma^{-\frac{\alpha}{\alpha+1}} \exp\left(-\frac{\alpha}{2}\left(\frac{Y_i - \boldsymbol{X}_i^T \boldsymbol{\beta}}{\sigma}\right)^2\right) + \frac{1}{\alpha}.
\end{aligned}$$

忽略所有常数项，对于 $\alpha > 0$，多元线性回归模型的 MRPE 为

$$\left(\widehat{\boldsymbol{\beta}}_{\alpha}, \widehat{\sigma}_{\alpha}\right) = \arg\min_{\boldsymbol{\beta}, \sigma} \sum_{i=1}^{n} -\sigma^{-\frac{\alpha}{\alpha+1}} \exp\left(-\frac{\alpha}{2}\left(\frac{Y_i - \boldsymbol{X}_i^T \boldsymbol{\beta}}{\sigma}\right)^2\right).$$

此外，分别对 $\boldsymbol{\beta}$ 和 $\sigma$求导，得到 $\boldsymbol{\beta}_{\alpha}$ 和 $\widehat{\sigma}_{\alpha}$ 的估计方程：

$$\begin{array}{l}
\sum_{i=1}^{n} \exp\left(-\frac{\alpha}{2}\left(\frac{Y_i - \boldsymbol{X}_i^T \boldsymbol{\beta}}{\sigma}\right)^2\right)\left(\frac{Y_i - \boldsymbol{X}_i^T \boldsymbol{\beta}}{\sigma}\right)\boldsymbol{X}_i = \mathbf{0}_p \\
\sum_{i=1}^{n} \exp\left(-\frac{\alpha}{2}\left(\frac{Y_i - \boldsymbol{X}_i^T \boldsymbol{\beta}}{\sigma}\right)^2\right)\left\{\left(\frac{Y_i - \boldsymbol{X}_i^T \boldsymbol{\beta}}{\sigma}\right)^2 - \frac{1}{1+\alpha}\right\} = 0
\end{array}, \tag{21}$$

这与在 [7]中得到的方程组完全相同。对于 $\alpha = 0$，若记 $\mathbb{X} = (\boldsymbol{X}_1, ..., \boldsymbol{X}_n)^T_{n\times p}$ 和 $\boldsymbol{Y} = (Y_1, ..., Y_n)$，则可得到 $\widehat{\boldsymbol{\beta}}_0$ 和$\widehat{\sigma}_0$的最大似然估计，即:

$$\widehat{\boldsymbol{\beta}}_0 = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\mathbf{Y} \text{ and } \widehat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}_0\right)^2.$$

最后，根据 [7], 中的结果可知，矩阵 $\boldsymbol{\Psi}_{n,\ \alpha}(\boldsymbol{\beta}, \sigma)$ 和 $\Omega_{n,\ \alpha}(\boldsymbol{\beta}, \sigma)$ 可表示为

$$\begin{aligned}
\boldsymbol{\Psi}_{n,\alpha}(\boldsymbol{\beta}, \sigma) &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{J}^{(i)}\left(\boldsymbol{\beta}, \sigma^2\right) \\
&= k\sigma^{-\frac{3\alpha+2}{\alpha+1}}(\alpha+1)^{-\frac{3}{2}}\begin{bmatrix} \frac{1}{n}\mathbb{X}^T\mathbb{X} & 0 \\ 0 & \frac{2}{\alpha+1} \end{bmatrix} \\
&= K_1(\alpha+1)^{-\frac{3}{2}}\begin{bmatrix} \frac{1}{n}\mathbb{X}^T\mathbb{X} & 0 \\ 0 & \frac{2}{\alpha+1} \end{bmatrix},
\end{aligned}$$

and

$$\begin{aligned}
\boldsymbol{\Omega}_{n,\alpha}(\boldsymbol{\beta}, \sigma) &= \frac{1}{n}\sum_{i=1}^{n}Var_{Y_i}\left[\left(\frac{\partial V_{i,\alpha}(Y_i; \boldsymbol{\beta}, \sigma^2)}{\partial\theta_j}\right)_{j=1,..,k}\right] \\
&= K_1^2\sigma^2\frac{1}{(2\alpha+1)^{3/2}}\begin{bmatrix} \frac{1}{n}\mathbb{X}^T\mathbb{X} & \mathbf{0} \\ \mathbf{0} & \frac{(3\alpha^2+4\alpha+2)}{(\alpha+1)^2(2\alpha+1)} \end{bmatrix}.
\end{aligned}$$

with

$$k = \frac{1}{\alpha}\left(\frac{1+\alpha}{2\pi}\right)^{\frac{\alpha}{2(\alpha+1)}}, \quad K_1 = k\sigma^{-\frac{3\alpha+2}{\alpha+1}}. \tag{22}$$

因此，对于 $\alpha = 0$，我们可以得到两个矩阵中 $(\boldsymbol{\beta}, \sigma)$ 的 Fisher 信息矩阵，即

$$\boldsymbol{\Psi}_{n,0}(\boldsymbol{\beta}, \sigma) = \begin{bmatrix} \frac{1}{\sigma^2}\frac{1}{n}\mathbb{X}^T\mathbb{X} & 0 \\ 0 & \frac{2}{\sigma^2} \end{bmatrix},$$

and

$$\boldsymbol{\Omega}_{n,0}(\boldsymbol{\beta}, \sigma) = \begin{bmatrix} \frac{1}{\sigma^2}\frac{1}{n}\mathbb{X}^T\mathbb{X} & \mathbf{0} \\ \mathbf{0} & \frac{2}{\sigma^2} \end{bmatrix}.$$

# 3 基于 RP 的模型选择准则

本节介绍基于 RP 的模型选择准则。考虑一组 $l$ 候选模型

$$\left\{\boldsymbol{M}^{(s)} = \left(M_1^{(s)}, ..., M_n^{(s)}\right)\right\}_{s\in\{1,...,l\}} \tag{23}$$

使得每个 $\boldsymbol{M}^{(s)}$ 都由参数密度函数刻画。

$$\boldsymbol{f}(\cdot, \boldsymbol{\theta}_s) = (f_1(\cdot, \boldsymbol{\theta}_s), ..., f_n(\cdot, \boldsymbol{\theta}_s)),\ \boldsymbol{\theta}_s \in \boldsymbol{\Theta}_s \subset \mathbb{R}^{p_s},$$

其对应的分布函数为 $\boldsymbol{F}(.,\boldsymbol{\theta}_s) = (F_1(\boldsymbol{\theta}_s) ..., F_n(.,\boldsymbol{\theta}_s))$，其中$\boldsymbol{\theta}_s$ 是模型 $s.$ 中所有密度函数的公共参数。也就是说，每个候选模型都表示一个由公共参数定义的参数族，其中参数数量可能不同。根据随机样本 $Y_1, ..., Y_n$，我们需要依据某个合适的选择准则，从集合 $\{\boldsymbol{M}^{(s)}\}_{s\in\{1,...,l\}}$ 中选出最佳模型。为此，对于每个假设模型$\boldsymbol{M}^{(s)}$，我们首先应确定最能拟合该样本的最佳参数 $\boldsymbol{\theta}_s$，随后从该集合中选出拟合效果最佳的模型。然后，给定一组观测值，模型选择分两步进行：首先将所有候选模型拟合到数据，然后根据 RP 选择拟合优度与复杂度之间权衡最佳的模型。

接下来描述模型选择算法的第一步。考虑一个固定的参数模型 $\boldsymbol{M}^{(s)}$，用于刻画其背后的真实分布。如果已知真实分布，那么可以通过最大化式（8）中定义的、在 $s$-模型下的理论平均目标函数 $H_\alpha(\boldsymbol{\theta})$，得到最符合该模型的参数 $\boldsymbol{M}_{(s)}$，记为 $\boldsymbol{\theta}^s_{\boldsymbol{g},\alpha}$。

根据第2节中的讨论，如果其背后的真实分布未知，但我们有一个随机样本 $Y_1, ..., Y_n$，那么基于 RP方法、利用该样本对真实参数作出的最佳估计就是式（15）中定义的 MRPE。

将所有候选模型拟合到观测数据（或者在已知真实分布时拟合到真实分布）后，我们应选择拟合度与复杂度之间权衡最佳的模型。因此，我们需要一种衡量每个模型的最佳候选模型与真实分布之间差异的方法。具有相关密度 $\boldsymbol{f}(\cdot,\boldsymbol{\theta}^s_g)$ 的模型 $\boldsymbol{M}_{(s)}$，以及基于 RP 方法得到的最佳拟合参数 $\boldsymbol{\theta}^s_g$，其拟合优度可通过式（8）中给出的平均目标函数 $H_\alpha(\boldsymbol{\theta}^s_g)$ 来量化。

由于真实分布通常未知，$\boldsymbol{\theta}^s_g$ 由 $\widehat{\boldsymbol{\theta}}^s_\alpha$ 估计。因此，我们可以用 $H_\alpha(\widehat{\boldsymbol{\theta}}^s_\alpha)$估计 $H_\alpha(\boldsymbol{\theta}^s_g)$。但 $H_\alpha$ 仍需要估计，而自然的估计量是$H_{n,\ \alpha}(\widehat{\boldsymbol{\theta}}^s_\alpha)$。然而，由于同一个样本既用于估计参数，也用于估计 $H_\alpha$，因此以下关系并不成立彩票娱乐

$$E_{Y_1,...,Y_n}\left[H_{n,\alpha}\left(\widehat{\boldsymbol{\theta}}^s_\alpha\right)\right] \neq E_{Y_1,...,Y_n}\left[H_\alpha\left(\widehat{\boldsymbol{\theta}}^s_\alpha\right)\right].$$

此外，估计偏差会取决于模型，因此我们需要添加一项，以校正模型假设造成的偏差。

AIC准则选择使其最小的模型

$$-2\sum_{i=1}^{n}\log f_i(y_i,\boldsymbol{\theta}) + 2p = 2H_{n,0}\left(\boldsymbol{\theta}\right) + 2p,$$

其中 $2p$ 是用于修正偏差的项。遵循同样的思路，我们将$RP_{NH}$–准则定义如下：

定义 6 设 $\left\{\left(M_1^{(s)},...,M_{n}^{(s}\right)\right\}_{s\in\{1,...,l\}}$)是针对独立非同分布观测 $Y_1,...,Y_n$的 $l$个候选模型。根据 $RP_{NH}$–准则选出的模型 $(M_1^*,...,M_n^*)$ 满足

$$(M_1^*,...,M_n^*)=\min_{s\in\{1,...,l\}}RP_{NH}\left(M_1^{(s)},...,M_n^{(s)},\widehat{\boldsymbol{\theta}}_\alpha^s\right),$$

其中

$$RP_{NH}\left(M_1^{(s)},...,M_n^{(s)},\widehat{\boldsymbol{\theta}}_\alpha^s\right)=H_{n,\alpha}\left(\widehat{\boldsymbol{\theta}}_\alpha^s\right)+\frac{1}{n}trace\left(\boldsymbol{\Omega_n}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\alpha}}^{\boldsymbol{s}}\right)\boldsymbol{\Psi}_n^{-1}\left(\widehat{\boldsymbol{\theta}}_\alpha^s\right)\right). \tag{24}$$

我们可以观察到

$$\lim_{\alpha\to0}RP_{NH}\left(M_1^{(s)},...,M_n^{(s)},\widehat{\boldsymbol{\theta}}_\alpha^s\right)=-\frac{1}{n}\sum_{i=1}^{n}\log f_i(Y_i,\boldsymbol{\theta})+\frac{p}{n},$$

因此，除一个乘法常数 $2n.$外，我们重新得到AIC准则

为了证明 $RP_{NH}$–准则，我们将证明估计函数 $RP_{NH}\left(M_1^{(s)},...,M_{n}^{(s}\right)$）（用于量化选择模型所造成的损失）是其理论版本$E_{Y_1,...,Y_n}\left[H_\alpha\left(\widehat{\boldsymbol{\theta}}_\alpha^s\right)\right].$的无偏估计量。为此，我们将假设以下附加的正则性条件：

C8. 矩阵 $\boldsymbol{\Psi}_n^{-1}(\boldsymbol{\theta})$ 和 $\Omega_n(\boldsymbol{\theta})$ 对任意 $\boldsymbol{\theta}\in\boldsymbol{\Theta}.$都是连续的

定理 7 假设条件C1-C8成立。那么，

$$E_{Y_1,...,Y_n}\left[RP_{NH}\left(M_1^{(s)},...,M_n^{(s)},\widehat{\boldsymbol{\theta}}_\alpha^s\right)\right]=E_{Y_1,...,Y_n}\left[H_\alpha\left(\widehat{\boldsymbol{\theta}}_\alpha^s\right)\right],\forall s=1,...,l.$$

证明。考虑一个固定的 $s=1,...,l.$对在$\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s$ 附近定义于公式（6）中的$V_{i,\alpha}(\boldsymbol{\theta})$进行泰勒展开，并在 $\widehat{\boldsymbol{\theta}}_\alpha^s$ 处取值，得到

$$\begin{aligned}V_{i,\alpha}\left(\widehat{\boldsymbol{\theta}}_\alpha^s\right) &= V_{i,\alpha}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)+\left(\frac{\partial V_{i,\alpha}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s}\left(\widehat{\boldsymbol{\theta}}_\alpha^s-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\\ &\quad+\frac{1}{2}\left(\widehat{\boldsymbol{\theta}}_\alpha^s-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)^T\left(\frac{\partial^2V_{i,\alpha}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^T}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s}\left(\widehat{\boldsymbol{\theta}}_\alpha^s-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)+o\left(\left\|\widehat{\boldsymbol{\theta}}_\alpha^s-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right\|^2\right)\\ &= V_{i,\alpha}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)+\left(\frac{\partial V_{i,\alpha}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s}\left(\widehat{\boldsymbol{\theta}}_\alpha^s-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\\ &\quad+\frac{1}{2}\left(\widehat{\boldsymbol{\theta}}_\alpha^s-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)^T\boldsymbol{J}_\tau^{(i)}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\left(\widehat{\boldsymbol{\theta}}_\alpha^s-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)+o\left(\left\|\widehat{\boldsymbol{\theta}}_\alpha^s-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right\|^2\right).\end{aligned}$$

对 $i$ 求和并除以 $n$，考虑到$\boldsymbol{\theta}_{\boldsymbol{g},\,\alpha}^s$ 使$H_\alpha(\boldsymbol{\theta})$最大化，我们得到

$$H_\alpha(\widehat{\boldsymbol{\theta}}_\alpha^s) = H_\alpha\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right) - \frac{1}{2}\left(\widehat{\boldsymbol{\theta}}_\alpha^s - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)^T \boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\left(\widehat{\boldsymbol{\theta}}_\alpha^s - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right) + o\left(\left\|\widehat{\boldsymbol{\theta}}_\alpha^s - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right\|^2\right)$$

因此,

$$E_{Y_1,\ldots,Y_n}\left[nH_\alpha\left(\widehat{\boldsymbol{\theta}}_\alpha^s\right)\right] = nH_\alpha\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right) - \frac{1}{2}E_{Y_1,\ldots,Y_n}\left[\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_\alpha^s - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)^T \boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_\alpha^s - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\right] + o_p(1). \tag{25}$$

但根据式（18），并在 [10], 中应用推论 2.1，我们有

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_\alpha^s - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)^T \boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_\alpha^s - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right) \xrightarrow[n\longrightarrow\infty]{\mathcal{L}} \sum_{i=1}^k \lambda_i(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s) Z_i^2,$$

其中， $\lambda_1(\boldsymbol{\theta}_{\boldsymbol{g},\ \alpha}^s)$ ..., $\lambda_n(\boldsymbol{\theta}_{\boldsymbol{g},\ \alpha}^s)$ 是矩阵的特征值

$$\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)^{-1}\boldsymbol{\Omega}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)^{-1} = \boldsymbol{\Omega}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)^{-1}$$

且 $Z_1$,..., $Z_k$ 是相互独立的正态随机变量，其均值为零，方差为 1。因此,

$$\begin{aligned} E_{Y_1,\ldots,Y_n} & \left[\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_\alpha^s - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)^T \boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_\alpha^s - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\right] \\ = & \quad \sum\nolimits_{i=1}^k \lambda_i(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s) + o_P(1) \\ = & \quad trace\left(\boldsymbol{\Omega}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)^{-1}\right) + o_P(1). \end{aligned}$$

另一方面，考虑到 $\widehat{\boldsymbol{\theta}}_\alpha^s$ 使 $H_{n,\ \alpha}(\boldsymbol{\theta})$ 最大，对 $H_{n,\ \alpha}(\boldsymbol{\theta})$ 在 $\widehat{\boldsymbol{\theta}}_\alpha^s$ 处作泰勒展开，并在 $\boldsymbol{\theta}_{\boldsymbol{g},\ \alpha}^s$ 处取值，可得

$$H_{n,\alpha}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right) = H_{n,\alpha}\left(\widehat{\boldsymbol{\theta}}_\alpha^s\right) + \frac{1}{2}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s - \widehat{\boldsymbol{\theta}}_\alpha^s\right)^T \left(\frac{\partial^2 H_{n,\alpha}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\ \partial\boldsymbol{\theta}^T}\right)_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_\alpha^s}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s - \widehat{\boldsymbol{\theta}}_\alpha^s\right) + o\left(\left\|\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s - \widehat{\boldsymbol{\theta}}_\alpha^s\right\|^2\right).$$

但随后，乘以 $n$ 并考虑期望值,

$$\begin{aligned} nH_\alpha\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right) \quad = \quad & E_{Y_1,\ldots,Y_n}\left[nH_{n,\alpha}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\right] = E_{Y_1,\ldots,Y_n}\left[nH_{n,\alpha}\left(\widehat{\boldsymbol{\theta}}_\alpha^s\right)\right] \\ & + \frac{1}{2}E_{Y_1,\ldots,Y_n}\left[\sqrt{n}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s - \widehat{\boldsymbol{\theta}}_\alpha^s\right)^T \left(\frac{\partial^2 H_{n,\alpha}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\ \partial\boldsymbol{\theta}^T}\right)_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_\alpha^s}\sqrt{n}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s - \widehat{\boldsymbol{\theta}}_\alpha^s\right)\right] + o_p(1). \end{aligned}$$

此外,

$$\left(\frac{\partial^2 H_{n,\alpha}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}^T}\right)_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_\alpha^s} \xrightarrow[n\longrightarrow\infty]{\mathcal{P}} -\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right). \tag{26}$$

由 $\boldsymbol{\Psi}_n.$ 的连续性可知。因此，代入（25）式中

$$\begin{aligned}
&E_{Y_1,\ldots,Y_n}\left[nH_\alpha\left(\widehat{\boldsymbol{\theta}}_\alpha^s\right)\right]\\
&\quad=nH_\alpha\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)-\frac{1}{2}E_{Y_1,\ldots,Y_n}\left[\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_\alpha^s-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)^T\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_\alpha^s-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\right]+o_p(1)\\
&\quad=E_{Y_1,\ldots,Y_n}\left[nH_{n,\alpha}\left(\widehat{\boldsymbol{\theta}}_\alpha^s\right)\right]-\frac{1}{2}E_{Y_1,\ldots,Y_n}\left[\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_\alpha^s-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)^T\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_\alpha^s-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\right]+o_p(1)\\
&\qquad-\frac{1}{2}E_{Y_1,\ldots,Y_n}\left[\sqrt{n}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s-\widehat{\boldsymbol{\theta}}_\alpha^s\right)^T\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\sqrt{n}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s-\widehat{\boldsymbol{\theta}}_\alpha^s\right)\right]+o_p(1)\\
&\quad=E_{Y_1,\ldots,Y_n}\left[nH_{n,\alpha}\left(\widehat{\boldsymbol{\theta}}_\alpha^s\right)\right]-E_{Y_1,\ldots,Y_n}\left[\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_\alpha^s-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)^T\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_\alpha^s-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\right]+o_p(1),
\end{aligned}$$

因此,

$$\begin{aligned}
E_{Y_1,\ldots,Y_n}\left[H_\alpha\left(\widehat{\boldsymbol{\theta}}_\alpha^s\right)\right] &= E_{Y_1,\ldots,Y_n}\left[H_{n,\alpha}\left(\widehat{\boldsymbol{\theta}}_\alpha^s\right)\right]\\
&\quad-\frac{1}{n}E_{Y_1,\ldots,Y_n}\left[\sqrt{n}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s-\widehat{\boldsymbol{\theta}}_\alpha^s\right)^T\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\sqrt{n}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s-\widehat{\boldsymbol{\theta}}_\alpha^s\right)\right]+o_p(1)\\
&= E_{Y_1,\ldots,Y_n}\left[H_{n,\alpha}\left(\widehat{\boldsymbol{\theta}}_\alpha^s\right)\right]-\frac{1}{n}trace\left(\boldsymbol{\Omega}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\boldsymbol{\Psi}_n^{-1}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^s\right)\right).
\end{aligned}$$

因此，结论成立。 ■

接下来，我们推导在多元线性回归模型下 $RP_{NH}$ 准则的显式表达式。

## 3.1 示例：多元线性回归模型下基于 RP 的模型选择

我们考虑第2.1节中定义的多元线性回归模型。

$$Y_i=\boldsymbol{X}_i^T\boldsymbol{\beta}+\varepsilon_i,\quad i=1,\ldots,n. \tag{27}$$

我们考虑几个模型 $\{(M_1^{(s)},...,M_{(s)n})\}_{s=1,\ldots,l}$，其中每个模型在所考虑的参数 $\boldsymbol{\beta}$ 上有所不同。例如，考虑四个解释变量（$X_1,X_2,X_3,X_4$）以及由下式给出的四个不同模型

$$\begin{aligned}
(M_1^{(1)},...,M_n^{(1)})&\equiv Y_i=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\epsilon_i,\\
(M_1^{(2)},...,M_n^{(2)})&\equiv Y_i=\beta_0+\beta_1X_1+\beta_2X_2+\beta_4X_4+\epsilon_i\\
(M_1^{(3)},...,M_n^{(3)})&\equiv Y_i=\beta_0+\beta_1X_1+\beta_3X_3+\beta_4X_4+\epsilon_i,\\
(M_1^{(4)},...,M_n^{(4)})&\equiv Y_i=\beta_0+\beta_2X_2+\beta_3X_3+\beta_4X_4+\epsilon_i.
\end{aligned}$$

每个模型都有五个需要估计的参数。接下来确定 $RP_{NH}\left(M_1^{(s)},...,M_{(s)n},\widehat{\boldsymbol{\theta}}_\alpha^s\right)$对于$s=$1,2,3,4所对应的值。

如第2节2.1所述，对于每个 $s=$1,2,3,4， $\widehat{\boldsymbol{\beta}}_\alpha^s$ 和 $\widehat{\sigma}_\alpha^s$的估计量是以下方程组的解

$$\begin{array}{l}\sum_{i=1}^{n}\exp\left(-\frac{\alpha}{2}\left(\frac{Y_i-\boldsymbol{X}_{s,i}^T\boldsymbol{\beta}}{\sigma}\right)^2\right)\left(\frac{Y_i-\boldsymbol{X}_{s,i}^T\boldsymbol{\beta}}{\sigma}\right)\boldsymbol{X}_{s,i}=\mathbf{0}_4\\ \sum_{i=1}^{n}\exp\left(-\frac{\alpha}{2}\left(\frac{Y_i-\boldsymbol{X}_{s,i}^T\boldsymbol{\beta}}{\sigma}\right)^2\right)\left\{\left(\frac{Y_i-\boldsymbol{X}_{s,i}^T\boldsymbol{\beta}}{\sigma}\right)^2-\frac{1}{1+\alpha}\right\}=0\end{array}, \tag{28}$$

其中 $X_{s,i}$ 对应于观测值 $i$ 中限定在模型 $s$. 所包含变量上的取值。注意，尽管 $\boldsymbol{\beta}$ 在不同模型中含义不同，但 $\sigma$. 并非如此。不过， $\sigma$ 的估计因模型而异，因此对于第 $s$ 个模型，该估计记为 $\widehat{\sigma}_\alpha^s$ 。

在 $\alpha=0$处，模型参数可以显式地表示为

$$\widehat{\boldsymbol{\beta}}_0^s=(\mathbb{X}_s^T\mathbb{X}_s)^{-1}\mathbb{X}_s^T\mathbf{Y}\ \text{ and }\ (\widehat{\sigma}_0^s)^2=\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\boldsymbol{X}_{s,i}^T\widehat{\boldsymbol{\beta}}_0\right)^2.$$

因此，根据公式（13），

$$H_{n,\alpha}(\widehat{\boldsymbol{\beta}},\widehat{\sigma})=\frac{1}{\alpha}\frac{1}{n}\sum_{i=1}^{n}-k\widehat{\sigma}^{-\frac{\alpha}{\alpha+1}}\exp\left(-\frac{\alpha}{2}\left(\frac{Y_i-\boldsymbol{X}_i^T\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}}\right)^2\right)+\frac{1}{\alpha},$$

其中 $k$ 如（22）中所定义。

接下来，我们来求出 $\boldsymbol{\Psi}_{s,n}(\boldsymbol{\beta}^s,\sigma)$ 和 $\boldsymbol{\Omega}_{s,n}(\boldsymbol{\beta}^s,\sigma)$ 的表达式。注意，这些矩阵也取决于模型 $s$。再次应用上一节的结果，可得

$$\begin{array}{l}\boldsymbol{\Psi}_{s,n}(\boldsymbol{\beta}^s,\sigma)=K_1(\alpha+1)^{-\frac{3}{2}}\begin{bmatrix}\frac{1}{n}\mathbb{X}_s^T\mathbb{X}_s & 0\\ 0 & \frac{2}{\alpha+1}\end{bmatrix},\\ \boldsymbol{\Omega}_{s,n}(\boldsymbol{\beta}^s,\sigma)=K_1^2\sigma^2\frac{1}{(2\alpha+1)^{3/2}}\begin{bmatrix}\frac{1}{n}\mathbb{X}_s^T\mathbb{X}_s & 0\\ 0 & \frac{(3\alpha^2+4\alpha+2)}{2(\alpha+1)(2\alpha+1)}\end{bmatrix},\end{array} \tag{29}$$

其中， $K_1$ 在式（22）中定义。请注意，这些矩阵的维度为（$p+1$）$\times$（$p+1$），其中 $p$是每个模型中向量 $\boldsymbol{\beta}$的维数。在我们的例子中，$p=4$，因此，

$$\boldsymbol{\Omega}_n\left(\widehat{\boldsymbol{\theta}}_\alpha^s\right)\boldsymbol{\Psi}_{s,n}^{-1}\left(\widehat{\boldsymbol{\beta}}_\alpha^s,\widehat{\sigma}_\alpha^s\right)=(\widehat{\sigma}_\alpha^s)^2K_1\frac{(\alpha+1)^{\frac{3}{2}}}{(2\alpha+1)^{\frac{3}{2}}}\begin{bmatrix}\boldsymbol{I}_{p\times p} & \mathbf{0}\\ \mathbf{0}^T & \frac{3\alpha^2+4\alpha+2}{(\alpha+1)^2(2\alpha+1)}\end{bmatrix},$$

因此，

$$trace\left(\boldsymbol{\Omega}_n\left(\widehat{\boldsymbol{\theta}}_\alpha^s\right)\boldsymbol{\Psi}_{s,n}^{-1}\left(\widehat{\boldsymbol{\beta}}_\alpha^s,\widehat{\sigma}_\alpha^s\right)\right)=(\widehat{\sigma}_\alpha^s)^2K_1\left(p\frac{(\alpha+1)^{\frac{3}{2}}}{(2\alpha+1)^{\frac{3}{2}}}+\frac{(\alpha+1)^{\frac{1}{2}}\left(3\alpha^2+4\alpha+2\right)}{2(2\alpha+1)^{5/2}}\right).$$

因此，应用式（6）中定义的 $RP_{NH}$–准则

$$
\begin{aligned}
&RP_{NH}(M_1^{(s)},...,M_n^{(s)},\widehat{\boldsymbol{\beta}}_\alpha^s,\widehat{\sigma}_\alpha^2) \\
&\qquad = -\frac{1}{\alpha}\left(\frac{1+\alpha}{2\pi}\right)^{\frac{\alpha}{2(\alpha+1)}}\frac{1}{n}\sum_{i=1}^{n}(\widehat{\sigma}_\alpha^s)^{-\frac{\alpha}{\alpha+1}}\exp\left(-\frac{\alpha}{2}\left(\frac{Y_i-\boldsymbol{X}_{s,i}^T\widehat{\boldsymbol{\beta}}_\alpha^s}{\widehat{\sigma}_\alpha^s}\right)^2\right) \\
&\qquad\quad +\frac{1}{\alpha}+\frac{1}{n}(\widehat{\sigma}_\alpha^s)^2K_1\left(p\frac{(\alpha+1)^{\frac{3}{2}}}{(2\alpha+1)^{\frac{3}{2}}}+\frac{(\alpha+1)^{\frac{1}{2}}\left(3\alpha^2+4\alpha+2\right)}{2\left(2\alpha+1\right)^{5/2}}\right).
\end{aligned}
\tag{30}
$$

最后，我们选择在 $s$, $RP_{NH}(M_{(s)\,1},...,M_{(s)n},\widehat{\boldsymbol{\beta}}_\alpha^s,\widehat{\sigma}_\alpha^s)$ 中取最小值的模型，作为四个候选模型中最合适的模型。

# 4 受限模型

我们来考虑模型选择问题的一个特例。在某些情况下，有必要将基于 $\boldsymbol{\theta}\in\boldsymbol{\Theta}\subset\mathbb{R}^p$、具有$p$ 个参数的完整模型，与参数还必须满足以下形式线性约束的其他受限模型进行比较。

$$
\left\{\boldsymbol{\theta}\in\boldsymbol{\Theta}/\ \boldsymbol{m}(\boldsymbol{\theta})=\mathbf{0}_r\right\},
\tag{31}
$$

其中， $\mathbf{0}_r$ 表示维数为 $r$ 且具有 $r<p$ 和 $\boldsymbol{m}$ 的零向量： $\mathbb{R}^p\to\mathbb{R}^r$是一个向量值函数，使得该 $p\times r$ 矩阵

$$
\mathbf{M}\left(\boldsymbol{\theta}\right)=\frac{\partial\boldsymbol{m}^T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}
\tag{32}
$$

存在且在 $\boldsymbol{\theta}$ 中连续，并且 $\mathrm{rank}(\mathbf{M}\left(\boldsymbol{\theta}\right))=r,\forall\boldsymbol{\theta}\in\boldsymbol{\Theta}$. 与基于散度的受限估计相关， [4] 定义了受限最小密度幂散度估计量。随后， [8] 给出了适用于一般总体的受限MRPE。

给定一个候选模型，我们已经确定，基于 RP 的该模型最佳拟合参数定义为

$$
\boldsymbol{\theta}_{\boldsymbol{g},\alpha}=\arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}\subset\mathbb{R}^p}H_\alpha(\boldsymbol{\theta}),
$$

其中， $H_\alpha(\boldsymbol{\theta})$ 在式 (8) 中定义。另一方面，将相同的准则应用于受限模型，可得受限模型的最佳拟合参数为

$$
\boldsymbol{\theta}_{\boldsymbol{g},\alpha}^R=\arg\min_{\boldsymbol{\theta}\in\Theta/\ \boldsymbol{m}(\boldsymbol{\theta})=\mathbf{0}_r}H_\alpha(\boldsymbol{\theta}).
$$

按照与第2节类似的论证，我们定义受限MRPE如下。

定义8 给定 $Y_1, ..., Y_n$ 为独立非同分布观测，受限MRPE（RMRPE）$\sim\boldsymbol{\theta}_\alpha$ 定义为

$$\widetilde{\boldsymbol{\theta}}_\alpha = \arg \min_{\boldsymbol{\theta}\in\Theta/\boldsymbol{m}(\boldsymbol{\theta})=\mathbf{0}_r} H_{n,\alpha}(\boldsymbol{\theta}), \tag{33}$$

其中 $H_{n,\ \alpha}(\boldsymbol{\theta})$ 对于 $\alpha > 0$ 的定义见式 (13)，对于 $\alpha = 0.$ 的定义见式 (14)

注意

$$H_{n,\alpha}(\widehat{\boldsymbol{\theta}}_\alpha) \leq H_{n,\alpha}(\widetilde{\boldsymbol{\theta}}_\alpha).$$

下面的定理给出了 RMPRE 的一种表示。

定理9 假设满足条件C1-C8，且 $\boldsymbol{\theta}_{\boldsymbol{g},\ \alpha}$ 满足受限模型的条件。则，

$$n^{1/2}(\widetilde{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}) = \boldsymbol{P}^*(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) n^{1/2} \left(\frac{\partial H_{n,\alpha}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}} + o_p(1),$$

为

$$\boldsymbol{P}^*(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) = \boldsymbol{Q}_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) \boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^T \boldsymbol{\Psi}_n \left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)^{-1} - \boldsymbol{\Psi}_n \left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)^{-1}, \tag{34}$$

*with*

$$\boldsymbol{Q}_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) = \boldsymbol{\Psi}_n \left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)^{-1} \boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) \left[\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^T \boldsymbol{\Psi}_n \left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)^{-1} \boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\right]^{-1}. \tag{35}$$

~证明。 $\boldsymbol{\theta}$, $\boldsymbol{\theta}_\alpha$ 的 RMRPE估计量必须满足

$$\left\{ \begin{array}{r} \left(\frac{\partial H_{n,\alpha}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}_\alpha} + \boldsymbol{M}(\widetilde{\boldsymbol{\theta}}_\alpha)\boldsymbol{\lambda}_n = \mathbf{0}_p, \\ \boldsymbol{m}(\widetilde{\boldsymbol{\theta}}_\alpha) = \mathbf{0}_r, \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{r} \left(\frac{\partial H_{n,\alpha}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}_\alpha} = -\boldsymbol{M}(\widetilde{\boldsymbol{\theta}}_\alpha)\boldsymbol{\lambda}_n \\ \boldsymbol{m}(\widetilde{\boldsymbol{\theta}}_\alpha) = \mathbf{0}_r \end{array} \right., \tag{36}$$

其中， $\boldsymbol{\lambda}_n$ 是拉格朗日乘子向量。现在，应用式（18），可得 $\widetilde{\boldsymbol{\theta}}_\alpha = \boldsymbol{\theta}_{\boldsymbol{g},\ \alpha} + \boldsymbol{t}n^{-1/2}$，其中 $||\boldsymbol{t}|| < c$，对于某个 $0 < c < \infty$。应用泰勒展开可得，

$$\begin{aligned} \left(\frac{\partial H_{n,\alpha}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}_\alpha} = &\left(\frac{\partial H_{n,\alpha}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}} + \left(\frac{\partial^2 H_{n,\alpha}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}} (\widetilde{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}) \\ &+ o(||\widetilde{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}||^2), \end{aligned}$$

因此

$$\begin{aligned} n^{1/2} \left(\frac{\partial H_{n,\alpha}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}_\alpha} = &n^{1/2} \left(\frac{\partial H_{n,\alpha}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}} \\ &+ \left(\frac{\partial^2 H_{n,\alpha}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}} n^{1/2}(\widetilde{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}) + o(n^{1/2}||\widetilde{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}||^2). \end{aligned}$$

然而，

$$o(n^{1/2}||\widetilde{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}||^2) = o(n^{1/2}||\boldsymbol{t}||^2/n) = o(n^{-1/2}||\boldsymbol{t}||^2) = o(O_p(1)) = o_p(1).$$

Now,

$$\left(\frac{\partial^2 H_{n,\alpha}(\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial^2\hat{V}_i(Y_i;\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}$$
$$\xrightarrow{P} \frac{1}{n}\sum_{i=1}^{n} E_{Y_i}\left[\left(\frac{\partial^2\hat{V}_i(Y;\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}\right] = \left(\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)\right)_{jk}.$$

因此，

$$n^{1/2}\left(\frac{\partial H_{n,\alpha}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}_\alpha} = n^{1/2}\left(\frac{\partial H_{n,\alpha}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}} + \boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right) n^{1/2}(\widetilde{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}) + o_p(1). \tag{37}$$

~由于 RMRPE $\boldsymbol{\theta}_\alpha$ 必须满足（36）中的条件，并结合（37），我们有

$$n^{1/2}\left(\frac{\partial H_{n,\alpha}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}} = -\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right) n^{1/2}(\widetilde{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}) - \boldsymbol{M}(\widetilde{\boldsymbol{\theta}}_\alpha) n^{1/2}\boldsymbol{\lambda}_n + o_p(1).$$

利用 $\boldsymbol{M}$ 的连续性，可将其写为

$$-\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right) n^{1/2}(\widetilde{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}) - \boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) n^{1/2}\boldsymbol{\lambda}_n = n^{1/2}\left(\frac{\partial H_{n,\alpha}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}} + o_p(1). \tag{38}$$

另一方面，对 $\boldsymbol{m}$ 作泰勒展开，可得

$$n^{1/2}\boldsymbol{m}(\widetilde{\boldsymbol{\theta}}_\alpha) = n^{1/2}\boldsymbol{m}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) + \boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^T n^{1/2}(\widetilde{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}) + o_p(1). \tag{39}$$

~由（39）并利用 $\boldsymbol{m}\left(\boldsymbol{\theta}_\alpha\right) = \mathbf{0}_r$、$\boldsymbol{m}\left(\boldsymbol{\theta}_{\boldsymbol{g},\ \alpha}\right) = \mathbf{0}_r$，可得

$$\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^T n^{1/2}(\widetilde{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}) + o_p(1) = \mathbf{0}_r. \tag{40}$$

现在，我们可以将方程（38）和（40）写成矩阵形式:

$$\begin{pmatrix} -\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right) & -\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) \\ \boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^T & \mathbf{0}_{r\times r} \end{pmatrix} \begin{pmatrix} n^{1/2}(\widetilde{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}) \\ n^{1/2}\boldsymbol{\lambda}_n \end{pmatrix} = \begin{pmatrix} n^{1/2}\left(\frac{\partial H_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}} \\ \mathbf{0}_r \end{pmatrix} + o_p(1).$$

因此，

$$\begin{pmatrix} n^{1/2}(\widetilde{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}) \\ n^{1/2}\boldsymbol{\lambda}_n \end{pmatrix} = \begin{pmatrix} -\boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) & -\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) \\ \boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^T & \mathbf{0}_{r\times r} \end{pmatrix}^{-1} \begin{pmatrix} n^{1/2}\left(\frac{\partial H_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}} \\ \mathbf{0}_r \end{pmatrix} + o_p(1).$$

But

$$\begin{pmatrix} -\boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) & -\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) \\ \boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^T & \mathbf{0} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{P}^*_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\boldsymbol{\alpha}}) & -\boldsymbol{Q}_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) \\ -\boldsymbol{Q}_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^T & \boldsymbol{R}_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) \end{pmatrix},$$

其中，$\boldsymbol{P}^*_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})$ 和 $\boldsymbol{Q}_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})$ 分别由式 (34) 和 (35) 给出。矩阵 $\boldsymbol{R}_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})$ 是使上述方程右端等于所示逆矩阵所需的矩阵。然后，

$$n^{1/2}(\widetilde{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_{\boldsymbol{g},\alpha}) = \boldsymbol{P}^*(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) n^{1/2}\left(\frac{\partial H_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}} + o_p(1) \qquad (41)$$

且结论成立。 ■

在以下引理中，我们将建立矩阵 $\boldsymbol{P}^*_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\ \alpha})$ 的一个性质，下一定理将用到该性质。

引理 10 给定 $\boldsymbol{P}^*_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})$ 和 $\boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})$，可得

$$\boldsymbol{P}^*_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{P}^*_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) = -\boldsymbol{P}^*_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}).$$

证明。应用定义并记作

$$\boldsymbol{A}^{-1}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) = \left[\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^T \boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^{-1} \boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\right]^{-1},$$

我们得到

$$\begin{aligned}
&\boldsymbol{P}^*_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{P}^*_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) \\
&= \left[\boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^{-1}\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{A}^{-1}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^T\boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^{-1} - \boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^{-1}\right] \\
&\quad \boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{P}^*_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) \\
&= \left[\boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^{-1}\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{A}^{-1}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^T - Id\right]\boldsymbol{P}^*_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) \\
&= \boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^{-1}\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{A}^{-1}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^T\boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^{-1}\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{A}^{-1}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^{-1} \\
&\quad - \boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^{-1}\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{A}^{-1}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^T\boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^{-1} - \boldsymbol{P}^*_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}) \\
&= -\boldsymbol{P}^*_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}).
\end{aligned}$$

因此，结论成立。 ■

假设现在我们已经选定了一个拟合效果最佳的模型，并想知道该模型是否对数据过拟合，以及受限模型是否更准确。那么，我们可以将这个问题表述为一个包含两个模型的模型选择问题：

一个较大的模型和一个受限模型，并应用上一节的结果。因此，只需针对两个模型分别计算 $RP_{NH}((M_{(s)1},...,M_{(s)n},\boldsymbol{\theta})$，并选择取值最小的模型。假设受限模型是正确的，在下述定理中，我们将建立其渐近分布

$$2n\left[RP_{NH}\left(M_1^{(s)},...,M_n^{(s)},\widehat{\boldsymbol{\theta}}_\alpha\right)-RP_{NH}\left(M_1^{(s)},...,M_n^{(s)},\widetilde{\boldsymbol{\theta}}_\alpha\right)\right],$$

其中 $RP_{NH}\left((M_1^{(s)},...,M_{(s)n},\widehat{\boldsymbol{\theta}}_\alpha\right)$ 已在（24）中给出，并

$$RP_{NH}\left((M_1^{(s)},...,M_n^{(s)},\widetilde{\boldsymbol{\theta}}_\alpha\right)=H_{n,\alpha}\left(\widetilde{\boldsymbol{\theta}}_\alpha\right)+\frac{1}{n}trace\left(\boldsymbol{\Omega}_n^R\left(\widetilde{\boldsymbol{\theta}}_\alpha\right)\boldsymbol{\Psi}_n^R\left(\widetilde{\boldsymbol{\theta}}_\alpha\right)^{-1}\right),$$

其中，$\boldsymbol{\Psi}_n^R\left(\widetilde{\boldsymbol{\theta}}_\alpha\right)$ 和$\Omega_n^R\left(\widetilde{\boldsymbol{\theta}}_\alpha\right)$ 是在（16）和（17）中定义的、但针对受限模型的矩阵。

请注意，选择受限模型的概率是

$$\Pr\left(RP_{NH}\left(M_1^{(k)},...,M_n^{(k)},\widehat{\boldsymbol{\theta}}_\alpha\right)-RP_{NH}\left(M_1^{(k)},...,M_n^{(k)},\widetilde{\boldsymbol{\theta}}_\alpha\right)>0\right).$$

定理11 假设条件C1-C8成立，并且假设拟合参数 $\boldsymbol{\theta}_{\boldsymbol{g},\alpha}$ 属于受限模型。那么，...的渐近分布

$$2n\left(RP_{NH}\left((M_1^{(s)},...,M_n^{(s)},\widehat{\boldsymbol{\theta}}_\alpha\right)-RP_{NH}\left((M_1^{(s)},...,M_n^{(s)},\widetilde{\boldsymbol{\theta}}_\alpha\right)\right)$$

与随机变量的分布一致

$$\sum_{j=1}^{r}\lambda_j(\boldsymbol{\theta}_{g,\alpha})(\boldsymbol{\theta})Z_j^2+2trace(\boldsymbol{\Omega}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{\Psi}_n^{-1}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}))-2trace(\boldsymbol{\Omega}_n^R(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})(\boldsymbol{\Psi}_n^R)^{-1}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})),$$

其中 $Z_1,...,Z_k$ 是相互独立的标准正态变量，$\lambda_1(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})...,\lambda_r(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})$ 是 $-\boldsymbol{Q}_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^T\boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^{-1}\Omega_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})$ 的非零特征值，并且

$$r=rank\left(\boldsymbol{\Omega}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{Q}_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^T\boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^{-1}\boldsymbol{\Omega}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\right).$$

证明。令

$$L=2n\left[RP_{NH}\left(M_1^{(s)},...,M_n^{(s)},\widehat{\boldsymbol{\theta}}_\alpha\right)-RP_{NH}\left(M_1^{(s)},...,M_n^{(s)},\widetilde{\boldsymbol{\theta}}_\alpha\right)\right].$$

那么，

$$L=2n\left[H_{n,\alpha}\left(\widehat{\boldsymbol{\theta}}_\alpha\right)-H_{n,\alpha}\left(\widetilde{\boldsymbol{\theta}}_\alpha\right)\right]+2trace\left[\boldsymbol{\Omega}_n\left(\widehat{\boldsymbol{\theta}}_\alpha\right)\boldsymbol{\Psi}_n\left(\widehat{\boldsymbol{\theta}}_\alpha\right)^{-1}\right]-2trace\left[\boldsymbol{\Omega}_n^R\left(\widetilde{\boldsymbol{\theta}}_\alpha\right)\boldsymbol{\Psi}_n^R\left(\widetilde{\boldsymbol{\theta}}_\alpha\right)^{-1}\right].$$

首先，请注意，

$$\begin{aligned} H_n\left(\widetilde{\boldsymbol{\theta}}_\alpha\right) &= H_{n,\alpha}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)+\left(\frac{\partial H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}\left(\widetilde{\boldsymbol{\theta}}_\alpha-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right) \\ &\quad+\frac{1}{2}\left(\widetilde{\boldsymbol{\theta}}_\alpha-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)^T\left(\frac{\partial^2 H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}\left(\widetilde{\boldsymbol{\theta}}_\alpha-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)+o(||\widetilde{\boldsymbol{\theta}}_\alpha-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}||^2). \end{aligned}$$

因此，

$$\begin{aligned} 2n\left[H_n\left(\widetilde{\boldsymbol{\theta}}_\alpha\right)-H_{n,\alpha}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)\right] &= 2\sqrt{n}\left(\frac{\partial H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}\sqrt{n}\left(\widetilde{\boldsymbol{\theta}}_\alpha-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right) \\ &\quad+\sqrt{n}\left(\widetilde{\boldsymbol{\theta}}_\alpha-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)^T\left(\frac{\partial^2 H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}\sqrt{n}\left(\widetilde{\boldsymbol{\theta}}_\alpha-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)+o_p(1). \end{aligned}$$

现在，考虑到

$$\sqrt{n}\left(\widetilde{\boldsymbol{\theta}}_\alpha-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)=\boldsymbol{P}_\alpha^*(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\sqrt{n}\left(\frac{\partial H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}+o_p(1),$$

and

$$\left(\frac{\partial^2 H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}\to\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right),$$

由式（26）可知，得出结论:

$$\begin{aligned} &2n\left[H_n\left(\widetilde{\boldsymbol{\theta}}_\alpha\right)-H_{n,\alpha}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)\right] \\ &=2\sqrt{n}\left(\frac{\partial H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right)^T_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}\boldsymbol{P}_\alpha^*(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\sqrt{n}\left(\frac{\partial H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}} \\ &\quad+\sqrt{n}\left(\frac{\partial H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right)^T_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}\boldsymbol{P}_\alpha^*(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)\boldsymbol{P}_\alpha^*(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\sqrt{n}\left(\frac{\partial H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}+o_p(1). \end{aligned}$$

现在，应用前述引理，我们知道

$$\boldsymbol{P}_\alpha^*(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)\boldsymbol{P}_\alpha^*(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})=-\boldsymbol{P}_\alpha^*(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}),$$

因此，

$$2n\left[H_n\left(\widetilde{\boldsymbol{\theta}}_\alpha\right)-H_{n,\alpha}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)\right]=\sqrt{n}\left(\frac{\partial H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right)^T_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}\boldsymbol{P}_\alpha^*(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\sqrt{n}\left(\frac{\partial H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}+o_p(1).$$

另一方面,

$$\begin{aligned} H_{n,\alpha}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right) &= H_n\left(\widehat{\boldsymbol{\theta}}_\alpha\right)+\left(\frac{\partial H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_\alpha}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}-\widehat{\boldsymbol{\theta}}_\alpha\right) \\ &\quad+\frac{1}{2}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}-\widehat{\boldsymbol{\theta}}_\alpha\right)^T\left(\frac{\partial^2 H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right)_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_\alpha}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}-\widehat{\boldsymbol{\theta}}_\alpha\right)+o(||\boldsymbol{\theta}_{\boldsymbol{g},\alpha}-\widehat{\boldsymbol{\theta}}_\alpha||^2). \end{aligned}$$

现在，考虑到

$$\left(\frac{\partial^2 H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right)_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_\alpha}\longrightarrow\left(\frac{\partial^2 H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}\longrightarrow\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right),$$

and

$$\left(\frac{\partial H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_\alpha}=0,$$

我们得出结论:

$$2n\left[H_n\left(\widehat{\boldsymbol{\theta}}_\alpha\right)-H_{n,\alpha}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)\right]=-\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_\alpha-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)^T\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_\alpha-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)+o_p(1).$$

Applying $\left(\frac{\partial H_{n,\alpha}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_\alpha}=\mathbf{0}$, we have by Taylor

$$\mathbf{0}=n^{1/2}\left(\frac{\partial H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}^T}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}+\boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})n^{1/2}\left(\widehat{\boldsymbol{\theta}}_\alpha-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)+o_p(1),$$

以便

$$n^{1/2}\left(\widehat{\boldsymbol{\theta}}_\alpha-\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)=-n^{1/2}\boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^{-1}\left(\frac{\partial H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}^T}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}+o_p(1).$$

因此,

$$2n\left[H_n\left(\widehat{\boldsymbol{\theta}}_\alpha\right)-H_{n,\alpha}\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)\right]=-\sqrt{n}\left(\frac{\partial H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right)^T_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)^{-1}\sqrt{n}\left(\frac{\partial H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}+o_p(1).$$

但是由于 $\boldsymbol{P}^*(\boldsymbol{\theta}_{\boldsymbol{g},\ \alpha})=\boldsymbol{Q}_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\ \alpha})\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\ \alpha})^T\boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\ \alpha})^{-1}-\boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\ \alpha})^{-1}$，我们得到

$$\begin{aligned} 2n\left[H_{n,\alpha}\left(\widehat{\boldsymbol{\theta}}_\alpha\right)-H_{n,\alpha}\left(\widetilde{\boldsymbol{\theta}}_\alpha\right)\right]=&-\sqrt{n}\left(\frac{\partial H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}^T}\right)^T_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}\boldsymbol{Q}_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^T\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)^{-1} \\ &\times\sqrt{n}\left(\frac{\partial H_{n,\alpha}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}+o_p(1). \end{aligned}$$

最后，我们得到,

$$\sqrt{n}\left(\frac{\partial H_{n,\alpha}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}^T}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\boldsymbol{g},\alpha}}\xrightarrow[n\to\infty]{L}N(\mathbf{0}_k,\boldsymbol{\Omega}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)),$$

因此， $2n\left[H_{n,\alpha}\left(\widehat{\boldsymbol{\theta}}_\alpha\right)-H_{n,\alpha}\left(\widetilde{\boldsymbol{\theta}}_\alpha\right)\right]$ 的渐近分布与随机变量的分布一致

$$\sum_{i=1}^{r}\lambda_i(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})Z_i^2,$$

其中 $Z_1$、…、 $Z_r$ 是相互独立的标准正态变量， $\lambda_1(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\ldots$、 $\lambda_r(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})$ 是 $-\boldsymbol{Q}_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^T\boldsymbol{\Psi}_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^{-1}\Omega_n(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})$ 的非零特征值，并且

$$r=\operatorname{rank}\left(\boldsymbol{Q}_\alpha(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})\boldsymbol{M}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})^T\boldsymbol{\Psi}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)^{-1}\boldsymbol{\Omega}_n\left(\boldsymbol{\theta}_{\boldsymbol{g},\alpha}\right)\right).$$

更多详细信息请参见 [10] 中的推论 2.1。证明至此完成。 ■

上述结果提供了一种渐近计算过拟合概率的方法，这在模型选择理论中具有重要意义。

## 4.1 示例：受限参数空间下多元线性回归模型中的基于 RP 的模型选择。

我们将考虑第3.1节中定义的多元线性回归模型，并且我们感兴趣的是在这些约束下比较完整模型与受限模型。

$$\beta_{p-r+1}=\ldots=\beta_p=0.$$

在此情况下，模型参数为 $\boldsymbol{\theta}=\left(\beta_0,\ldots,\beta_p,\sigma\right)$，定义约束的函数 $\boldsymbol{m}(\boldsymbol{\theta})$ 是

$$\boldsymbol{m}(\boldsymbol{\theta})=\boldsymbol{m}\left(\beta_0,\ldots,\beta_p,\sigma\right)=(\beta_{p-r+1},\ldots,\beta_p).$$

因此，其导数为

$$\boldsymbol{M}(\boldsymbol{\theta})=\frac{\partial\boldsymbol{m}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=\begin{pmatrix}\mathbf{0}_{(p-r+1)\times r}\\ \boldsymbol{I}_{r\times r}\\ \mathbf{0}_{1\times r}\end{pmatrix}.$$

将设计矩阵 $\mathbb{X}$ 表示为

$$\mathbb{X}=\left(\mathbb{X}_1,\mathbb{X}_2\right),$$

其中， $\mathbb{X}_1$ 为 $n\times(p-r+1)$矩阵， $\mathbb{X}_2$ 为 $n\times r$ 矩阵。显然， $\mathbb{X}_1$ 是受限模型的设计矩阵，而 $\mathbb{X}_2$ 对应于完整模型中不属于小模型的参数所对应的设计矩阵。式（29）中给出的矩阵$\boldsymbol{\Psi}_n(\boldsymbol{\beta},\sigma)$和$\Omega_n(\boldsymbol{\beta},\sigma)$，使用记号$\mathbb{X}_1$ 和 $\mathbb{X}_2$，可改写为

$$\boldsymbol{\Psi}_n\left(\boldsymbol{\beta},\sigma\right)=K_1\left(\alpha+1\right)^{-\frac{3}{2}}\begin{bmatrix}\frac{1}{n}\mathbb{X}_1^T\mathbb{X}_1 & \frac{1}{n}\mathbb{X}_1^T\mathbb{X}_2 & 0\\ \frac{1}{n}\mathbb{X}_2^T\mathbb{X}_1 & \frac{1}{n}\mathbb{X}_2^T\mathbb{X}_2 & 0\\ 0 & 0 & \frac{2}{\alpha+1}\end{bmatrix},$$

其中 $K_1$ 如（22）所定义，以及

$$\boldsymbol{\Omega}_n(\boldsymbol{\beta},\sigma) = K_1^2\sigma^2\frac{1}{(2\alpha+1)^{3/2}}\begin{bmatrix} \frac{1}{n}\mathbb{X}_1^T\mathbb{X}_1 & \frac{1}{n}\mathbb{X}_1^T\mathbb{X}_2 & 0 \\ \frac{1}{n}\mathbb{X}_2^T\mathbb{X}_1 & \frac{1}{n}\mathbb{X}_2^T\mathbb{X}_2 & 0 \\ 0 & 0 & \frac{(3\alpha^2+4\alpha+2)}{(\alpha+1)^2(2\alpha+1)} \end{bmatrix}.$$

现在，矩阵 $\boldsymbol{\Psi}_n(\boldsymbol{\beta},\sigma)$ 的逆为

$$\boldsymbol{\Psi}_n^{-1}(\boldsymbol{\beta},\sigma) = K_1(\alpha+1)^{3/2}\begin{bmatrix} n\boldsymbol{A}_{11} & n\boldsymbol{A}_{12} & 0 \\ n\boldsymbol{A}_{21} & n\boldsymbol{A}_{22} & 0 \\ 0 & 0 & \frac{\alpha+1}{2} \end{bmatrix},$$

with

$$\begin{aligned} \boldsymbol{A}_{11} &= \left(\mathbb{X}_1^T\mathbb{X}_1\right)^{-1} + \left(\mathbb{X}_1^T\mathbb{X}_1\right)^{-1}\mathbb{X}_1^T\mathbb{X}_2\boldsymbol{D}^{-1}\mathbb{X}_2^T\mathbb{X}_1\left(\mathbb{X}_1^T\mathbb{X}_1\right)^{-1}, \\ \boldsymbol{A}_{12} &= -\left(\mathbb{X}_1^T\mathbb{X}_1\right)^{-1}\mathbb{X}_1^T\mathbb{X}_2\boldsymbol{D}^{-1}, \\ \boldsymbol{A}_{21} &= -\boldsymbol{D}^{-1}\mathbb{X}_2^T\mathbb{X}_1\left(\mathbb{X}_1^T\mathbb{X}_1\right)^{-1}, \\ \boldsymbol{A}_{22} &= \boldsymbol{D}^{-1}, \end{aligned}$$

正在
быть

$$\boldsymbol{D} = \mathbb{X}_2^T\mathbb{X}_2 - \mathbb{X}_2^T\mathbb{X}_1\left(\mathbb{X}_1^T\mathbb{X}_1\right)^{-1}\mathbb{X}_1^T\mathbb{X}_2.$$

因此，我们有矩阵 $\boldsymbol{\Psi}_n^{-1}(\boldsymbol{\beta},\sigma)$ 可以计算为

$$\begin{aligned} \boldsymbol{\Psi}_n^{-1}(\boldsymbol{\beta},\sigma)\boldsymbol{M}(\boldsymbol{\beta},\sigma) &= K_1^{-1}(\alpha+1)^{3/2}\begin{bmatrix} n\boldsymbol{A}_{11} & n\boldsymbol{A}_{12} & 0 \\ n\boldsymbol{A}_{21} & n\boldsymbol{A}_{22} & 0 \\ 0 & 0 & \frac{\alpha+1}{2} \end{bmatrix}\begin{pmatrix} \mathbf{0}_{(p-r)\times r} \\ \boldsymbol{I}_{r\times r} \\ \mathbf{0}_r \end{pmatrix} \\ &= K_1^{-1}(\alpha+1)^{3/2}n\begin{bmatrix} -\left(\mathbb{X}_1^T\mathbb{X}_1\right)^{-1}\mathbb{X}_1^T\mathbb{X}_2\boldsymbol{D}^{-1} \\ \boldsymbol{D}^{-1} \\ 0 \end{bmatrix}. \end{aligned}$$

另一方面,

$$\left(\boldsymbol{M}(\boldsymbol{\beta},\sigma)^T\boldsymbol{\Psi}_n^{-1}(\boldsymbol{\beta},\sigma)\boldsymbol{M}(\boldsymbol{\beta},\sigma)\right)^{-1} = \frac{K_1(\alpha+1)^{-3/2}}{n}\boldsymbol{D},$$

and

$$\boldsymbol{M}(\boldsymbol{\beta},\sigma)^T\boldsymbol{\Psi}_n^{-1}(\boldsymbol{\beta},\sigma)\boldsymbol{\Omega}_n(\boldsymbol{\beta},\sigma) = (\alpha+1)^{3/2}\frac{K_1\sigma^2}{(2\alpha+1)^{3/2}}(\mathbf{0},\boldsymbol{I}_{\boldsymbol{r}\times\boldsymbol{r}},\mathbf{0}),$$

因此，将上述表达式相乘可得

$$\boldsymbol{Q}_\alpha(\boldsymbol{\beta},\sigma) = \boldsymbol{\Psi}_n^{-1}(\boldsymbol{\beta},\sigma)\boldsymbol{M}(\boldsymbol{\beta},\sigma)\left[\boldsymbol{M}(\boldsymbol{\beta},\sigma)^T\boldsymbol{\Psi}_n^{-1}(\boldsymbol{\beta},\sigma)\boldsymbol{M}(\boldsymbol{\beta},\sigma)\right]^{-1} = \begin{bmatrix} -\left(\mathbb{X}_1^T\mathbb{X}_1\right)^{-1}\mathbb{X}_1^T\mathbb{X}_2 \\ \boldsymbol{I}_{r\times r} \\ 0 \end{bmatrix},$$

and

$$\boldsymbol{Q}_{\alpha}(\boldsymbol{\beta},\sigma)\boldsymbol{M}(\boldsymbol{\beta},\sigma)^{T}\boldsymbol{\Psi}_{n}\left(\boldsymbol{\beta},\sigma\right)^{-1}\boldsymbol{\Omega}_{n}\left(\boldsymbol{\beta},\sigma\right)=\left(\alpha+1\right)^{3/2}\frac{K_{1}\sigma_{\boldsymbol{g},\alpha}^{2}}{\left(2\alpha+1\right)^{3/2}}\left(\begin{array}{ccc}\mathbf{0} & \left(\mathbb{X}_{1}^{T}\mathbb{X}_{1}\right)^{-1}\mathbb{X}_{1}^{T}\mathbb{X}_{2} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{r\times r} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0}\end{array}\right).$$

因此，在这种情况下，我们可以计算出前 $r$–个特征值:

$$\lambda_{1}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})=\ldots=\lambda_{r}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})=\left(\alpha+1\right)^{3/2}\frac{K_{1}\sigma_{\boldsymbol{g},\alpha}^{2}}{\left(2\alpha+1\right)^{3/2}},$$

因此,

$$\sum_{i=1}^{r}\lambda_{i}(\boldsymbol{\theta}_{\boldsymbol{g},\alpha})Z_{i}^{2}=-\left(\alpha+1\right)^{3/2}\frac{K_{1}\sigma_{\boldsymbol{g},\alpha}^{2}}{\left(2\alpha+1\right)^{3/2}}\chi_{r}^{2}.$$

另一方面，我们有

$$\boldsymbol{\Omega}_{n}\left(\widehat{\boldsymbol{\beta}}_{\alpha},\widehat{\sigma}_{\alpha}\right)\boldsymbol{\Psi}_{n}^{-1}\left(\widehat{\boldsymbol{\beta}}_{\alpha},\widehat{\sigma}_{\alpha}\right)=K_{1}\sigma_{\boldsymbol{g},\alpha}^{2}\frac{\left(\alpha+1\right)^{3/2}}{\left(2\alpha+1\right)^{3/2}}\left[\begin{array}{ccc}\boldsymbol{I} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \frac{\left(3\alpha^{2}+4\alpha+2\right)}{2\left(2\alpha+1\right)\left(\alpha+1\right)}\end{array}\right],$$

因此，上述矩阵的迹为

$$trace\left(\boldsymbol{\Omega}_{n}\left(\widehat{\boldsymbol{\beta}}_{\alpha},\widehat{\sigma}_{\alpha}\right)\boldsymbol{\Psi}_{n}^{-1}\left(\widehat{\boldsymbol{\beta}}_{\alpha},\widehat{\sigma}_{\alpha}\right)\right)\rightarrow\sigma_{\boldsymbol{g},\alpha}^{2}K_{1}\frac{\left(\alpha+1\right)^{3/2}}{\left(2\alpha+1\right)^{3/2}}\left(\left(p+1\right)+\frac{\left(3\alpha^{2}+4\alpha+2\right)}{2\left(2\alpha+1\right)\left(\alpha+1\right)}\right),$$

and

$$trace\left(\boldsymbol{\Omega}_{n}^{R}\left(\widetilde{\boldsymbol{\beta}}_{\alpha},\widetilde{\sigma}_{\alpha}\right)\right)\left(\boldsymbol{\Psi}_{n}^{R}\right)^{-1}\left(\widetilde{\boldsymbol{\beta}}_{\alpha},\widetilde{\sigma}_{\alpha}\right)\rightarrow\sigma_{\boldsymbol{g},\alpha}^{2}K_{1}\frac{\left(\alpha+1\right)^{3/2}}{\left(2\alpha+1\right)^{3/2}}\left(\left(p-r+1\right)+\frac{\left(3\alpha^{2}+4\alpha+2\right)}{2\left(2\alpha+1\right)\left(\alpha+1\right)}\right).$$

因此,

$$trace\left(\boldsymbol{\Omega}_{n}\left(\widehat{\boldsymbol{\beta}}_{\alpha},\widehat{\sigma}_{\alpha}\right)\boldsymbol{\Psi}_{n}^{-1}\left(\widehat{\boldsymbol{\beta}}_{\alpha},\widehat{\sigma}_{\alpha}\right)\right)-trace\left(\boldsymbol{\Omega}_{n}^{R}\left(\widetilde{\boldsymbol{\beta}}_{\alpha},\widetilde{\sigma}_{\alpha}\right)\left(\boldsymbol{\Psi}_{n}^{R}\right)^{-1}\left(\widetilde{\boldsymbol{\beta}}_{\alpha},\widetilde{\sigma}_{\alpha}\right)\right)\rightarrow\sigma_{\boldsymbol{g},\alpha}^{2}K_{1}\frac{\left(\alpha+1\right)^{3/2}}{\left(2\alpha+1\right)^{3/2}}r.$$

最后，当受限模型正确时，选择受限模型的渐近概率是

$$\begin{aligned}&\Pr\left(2n\left(RP_{NH}(M_{1}^{(s)},...,M_{n}^{(s)},\widehat{\boldsymbol{\theta}}_{\alpha})-RP_{NH}(M_{1}^{(s)},...,M_{n}^{(s)},\widetilde{\boldsymbol{\theta}}_{\alpha})\right)>0\right)\rightarrow\\ &\quad\Pr\left(\left(-\alpha+1\right)^{3/2}\frac{K_{1}\sigma_{\boldsymbol{g},\alpha}^{2}}{\left(2\alpha+1\right)^{3/2}}\chi_{r}^{2}+2\left(\alpha+1\right)^{3/2}\frac{K_{1}\sigma_{\boldsymbol{g},\alpha}^{2}}{\left(2\alpha+1\right)^{3/2}}r>0\right)=\\ &\quad=\Pr\left(\left(\alpha+1\right)^{3/2}\frac{K_{1}\sigma_{\boldsymbol{g},\alpha}^{2}}{\left(2\alpha+1\right)^{3/2}}\left(2r-\chi_{r}^{2}\right)>0\right)=\Pr\left(\chi_{r}^{2}<2r\right).\end{aligned}$$

# 5 模拟研究

为评估本文引入的 $RP_{NH}$准则的性能，我们考虑多项式回归模型的情形。我们取该模型

$$Y_i = X_i + 2X_i^2 - X_i^3 + X_i^4 + \epsilon_i, i = 1, ..., n,$$

其中 $\epsilon_i \sim \mathcal{N}(0, 1)$，变量 $X_i$ 固定，并在区间 [-2, 2] 上均匀选取。接下来，取 $n = 100$，使得

$$X_i = -2 + \frac{4}{102}(i + 1), i = 1, ..., 100.$$

我们考虑了若干旨在拟合这些数据的理论模型。这些模型由定义模型的多项式次数决定。注意，回归系数的表达式与多元线性回归模型（MLRM）中的相同，只需令 $X^i$ 为 $X_i$，因此我们可以使用前几节推导出的公式。在本例中，我们考虑了六种不同的模型，从常数（0 次）到 5 次多项式不等。如此定义的每个模型都由其次数表征，记为 $p$。

我们选取 1000 组不同的样本数据（$Y^s, X^s$） $s = 1$，...，1000，并针对每组样本，根据若干准则选出拟合效果最佳的模型。我们考虑了 $AIC$、$BIC$、$AIC_c$ 和 $RP_{NH}$ 准则，并采用不同的调节参数取值，即 $\alpha = 0.01$、0.02、0.04、0.07、0.1、0.2、0.4、0.5、0.7 和 1。

在表 1 中，我们列出了每种模型选择准则选中各模型的次数。从这些结果可以看出，$BIC$似乎是拟合效果最好的模型选择准则，其他模型选择准则的表现则相近。

| $p$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $AIC$ | 0 | 0 | 0 | 0 | 836 | 164 |
| $BIC$ | 0 | 0 | 0 | 0 | 967 | 33 |
| $AIC_c$ | 0 | 0 | 0 | 0 | 864 | 136 |
| $RPNH_{0.01}$ | 0 | 0 | 0 | 0 | 822 | 178 |
| $RPNH_{0.02}$ | 0 | 0 | 0 | 0 | 822 | 178 |
| $RPNH_{0.04}$ | 0 | 0 | 0 | 0 | 826 | 174 |
| $RPNH_{0.1}$ | 0 | 0 | 0 | 0 | 822 | 178 |
| $RPNH_{0.2}$ | 0 | 0 | 0 | 0 | 834 | 166 |
| $RPNH_{0.4}$ | 0 | 0 | 0 | 0 | 842 | 158 |
| $RPNH_{0.5}$ | 0 | 0 | 0 | 0 | 841 | 159 |
| $RPNH_{0.7}$ | 0 | 0 | 0 | 0 | 838 | 162 |
| $RPNH_{1.0}$ | 0 | 0 | 0 | 0 | 837 | 163 |

表 1：未受污染数据的结果。

正如全文所阐述的，我们预计 $RP_{NH}$ 将是一项稳健的选择准则。为检验这一假设，我们考察了这样一种情形：

污染。因此，我们考虑前述模型，但在部分数据中引入污染。更具体地说，我们定义

$$\epsilon_i \sim \mathcal{U}(\min_i(X_i + 2X_i^2 - X_i^3 + X_i^4) - r, \max_i(X_i + 2X_i^2 - X_i^3 + X_i^4) + r),$$

对于随机选取的部分数据而言。在这里， $r$ 是衡量污染强度的常数，即 $r$越大，污染越严重。我们考虑了三个数值 $r = 1$、5、10。此外，我们改变了受污染影响的数据比例。在本研究中，我们将污染比例设为0.05、0.10、0.20、0.30。

再次，我们根据不同的模型选择准则选出了拟合效果最佳的模型，并进行了1000次实验。对于每种污染情况和污染强度组合 $r$ ，各模型被选中的次数见表2、表3、表4和表5。每张表的左侧部分对应 $r = 1$，中间部分对应 $r = 5$ ，右侧部分对应$r = 10$.

| | $r=1$ | | | | | | $r=5$ | | | | | | $r=10$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | 0 | 1 | 2 | 3 | 4 | 5 | 0 | 1 | 2 | 3 | 4 | 5 | 0 | 1 | 2 | 3 | 4 | 5 |
| $AIC$ | 0 | 0 | 1 | 19 | 659 | 321 | 0 | 0 | 10 | 27 | 622 | 341 | 0 | 0 | 9 | 38 | 616 | 337 |
| $BIC$ | 0 | 0 | 16 | 64 | 802 | 118 | 0 | 0 | 39 | 84 | 751 | 126 | 0 | 0 | 57 | 96 | 715 | 132 |
| $AIC_c$ | 0 | 0 | 1 | 22 | 694 | 283 | 0 | 0 | 12 | 33 | 651 | 304 | 0 | 0 | 12 | 43 | 634 | 311 |
| $RPNH_{0.01}$ | 0 | 0 | 3 | 14 | 844 | 139 | 0 | 0 | 5 | 18 | 830 | 147 | 0 | 0 | 9 | 22 | 812 | 157 |
| $RPNH_{0.02}$ | 0 | 0 | 0 | 13 | 866 | 121 | 0 | 0 | 2 | 47 | 833 | 118 | 0 | 0 | 6 | 62 | 810 | 122 |
| $RPNH_{0.04}$ | 0 | 0 | 2 | 20 | 844 | 134 | 0 | 0 | 1 | 18 | 833 | 148 | 0 | 0 | 1 | 18 | 833 | 148 |
| $RPNH_{0.1}$ | 0 | 0 | 0 | 0 | 835 | 165 | 0 | 0 | 0 | 0 | 836 | 164 | 0 | 0 | 0 | 0 | 830 | 170 |
| $RPNH_{0.2}$ | 0 | 0 | 0 | 0 | 829 | 171 | 0 | 0 | 0 | 0 | 833 | 167 | 0 | 0 | 0 | 0 | 833 | 167 |
| $RPNH_{0.4}$ | 0 | 0 | 0 | 0 | 837 | 163 | 0 | 0 | 0 | 0 | 835 | 165 | 0 | 0 | 0 | 0 | 839 | 161 |
| $RPNH_{0.5}$ | 0 | 0 | 0 | 0 | 837 | 163 | 0 | 0 | 0 | 0 | 848 | 152 | 0 | 0 | 0 | 0 | 834 | 166 |
| $RPNH_{0.7}$ | 0 | 0 | 0 | 0 | 842 | 158 | 0 | 0 | 0 | 0 | 836 | 164 | 0 | 0 | 0 | 0 | 831 | 169 |
| $RPNH_{1.0}$ | 0 | 0 | 0 | 0 | 838 | 162 | 0 | 0 | 0 | 0 | 836 | 164 | 0 | 0 | 0 | 0 | 830 | 170 |

表2：污染程度为5%的结果

| | $r=1$ | | | | | | $r=5$ | | | | | | $r=10$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | 0 | 1 | 2 | 3 | 4 | 5 | 0 | 1 | 2 | 3 | 4 | 5 | 0 | 1 | 2 | 3 | 4 | 5 |
| $AIC$ | 0 | 0 | 17 | 64 | 591 | 328 | 0 | 0 | 18 | 82 | 575 | 325 | 0 | 0 | 47 | 106 | 538 | 309 |
| $BIC$ | 0 | 0 | 94 | 155 | 629 | 122 | 0 | 0 | 95 | 180 | 611 | 114 | 0 | 0 | 153 | 178 | 558 | 111 |
| $AIC_c$ | 0 | 0 | 21 | 75 | 621 | 283 | 0 | 0 | 24 | 93 | 601 | 282 | 0 | 0 | 55 | 123 | 556 | 266 |
| $RPNH_{0.01}$ | 0 | 0 | 24 | 37 | 770 | 169 | 0 | 0 | 26 | 48 | 750 | 176 | 0 | 0 | 37 | 61 | 705 | 197 |
| $RPNH_{0.02}$ | 0 | 0 | 19 | 30 | 800 | 151 | 0 | 0 | 24 | 40 | 780 | 156 | 0 | 0 | 30 | 60 | 747 | 163 |
| $RPNH_{0.04}$ | 0 | 0 | 16 | 60 | 809 | 115 | 0 | 0 | 19 | 100 | 764 | 117 | 0 | 0 | 23 | 94 | 770 | 113 |
| $RPNH_{0.1}$ | 0 | 0 | 0 | 5 | 845 | 150 | 0 | 0 | 0 | 1 | 839 | 160 | 0 | 0 | 0 | 1 | 851 | 148 |
| $RPNH_{0.2}$ | 0 | 0 | 0 | 0 | 829 | 171 | 0 | 0 | 0 | 0 | 835 | 165 | 0 | 0 | 0 | 0 | 844 | 156 |
| $RPNH_{0.4}$ | 0 | 0 | 0 | 0 | 829 | 171 | 0 | 0 | 0 | 0 | 840 | 160 | 0 | 0 | 0 | 0 | 850 | 150 |
| $RPNH_{0.5}$ | 0 | 0 | 0 | 0 | 824 | 176 | 0 | 0 | 0 | 0 | 845 | 155 | 0 | 0 | 0 | 0 | 841 | 159 |
| $RPNH_{0.7}$ | 0 | 0 | 0 | 0 | 831 | 169 | 0 | 0 | 0 | 0 | 833 | 167 | 0 | 0 | 0 | 0 | 834 | 166 |
| $RPNH_{1.0}$ | 0 | 0 | 0 | 0 | 841 | 159 | 0 | 0 | 0 | 0 | 835 | 165 | 0 | 0 | 0 | 0 | 833 | 167 |

表3：污染程度为10%时的结果。

| | $r=1$ | | | | | | $r=5$ | | | | | | $r=10$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | 0 | 1 | 2 | 3 | 4 | 5 | 0 | 1 | 2 | 3 | 4 | 5 | 0 | 1 | 2 | 3 | 4 | 5 |
| $AIC$ | 0 | 0 | 52 | 134 | 509 | 305 | 0 | 0 | 76 | 176 | 464 | 284 | 0 | 0 | 148 | 169 | 397 | 286 |
| $BIC$ | 0 | 0 | 238 | 210 | 440 | 112 | 0 | 0 | 278 | 234 | 398 | 90 | 0 | 0 | 367 | 223 | 325 | 85 |
| $AIC_c$ | 0 | 0 | 64 | 154 | 512 | 270 | 0 | 0 | 91 | 191 | 476 | 242 | 0 | 0 | 179 | 180 | 400 | 241 |
| $RPNH_{0.01}$ | 0 | 0 | 41 | 92 | 596 | 271 | 0 | 0 | 43 | 93 | 561 | 303 | 0 | 0 | 52 | 95 | 514 | 339 |
| $RPNH_{0.02}$ | 0 | 0 | 37 | 85 | 625 | 253 | 0 | 0 | 39 | 92 | 589 | 280 | 0 | 0 | 47 | 90 | 561 | 302 |
| $RPNH_{0.04}$ | 0 | 0 | 29 | 75 | 676 | 220 | 0 | 0 | 32 | 88 | 661 | 219 | 0 | 0 | 43 | 93 | 648 | 216 |
| $RPNH_{0.1}$ | 0 | 0 | 42 | 214 | 631 | 113 | 0 | 0 | 41 | 138 | 693 | 128 | 0 | 0 | 20 | 64 | 810 | 106 |
| $RPNH_{0.2}$ | 0 | 0 | 0 | 0 | 836 | 164 | 0 | 0 | 0 | 0 | 831 | 169 | 0 | 0 | 0 | 0 | 849 | 151 |
| $RPNH_{0.4}$ | 0 | 0 | 0 | 0 | 837 | 163 | 0 | 0 | 0 | 0 | 833 | 167 | 0 | 0 | 0 | 0 | 840 | 160 |
| $RPNH_{0.5}$ | 0 | 0 | 0 | 0 | 836 | 164 | 0 | 0 | 0 | 0 | 829 | 171 | 0 | 0 | 0 | 0 | 840 | 160 |
| $RPNH_{0.7}$ | 0 | 0 | 0 | 0 | 845 | 155 | 0 | 0 | 0 | 0 | 827 | 173 | 0 | 0 | 0 | 0 | 849 | 151 |
| $RPNH_{1.0}$ | 0 | 0 | 0 | 0 | 834 | 166 | 0 | 0 | 0 | 0 | 823 | 177 | 0 | 0 | 0 | 0 | 836 | 164 |

表 4：污染程度为 20% 时的结果。

| | $r=1$ | | | | | | $r=5$ | | | | | | $r=10$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | 0 | 1 | 2 | 3 | 4 | 5 | 0 | 1 | 2 | 3 | 4 | 5 | 0 | 1 | 2 | 3 | 4 | 5 |
| $AIC$ | 0 | 0 | 112 | 178 | 433 | 277 | 0 | 0 | 114 | 209 | 373 | 304 | 0 | 0 | 192 | 183 | 374 | 251 |
| $BIC$ | 0 | 0 | 327 | 256 | 327 | 90 | 0 | 0 | 385 | 240 | 276 | 99 | 0 | 0 | 457 | 212 | 253 | 78 |
| $AIC_c$ | 0 | 0 | 136 | 191 | 436 | 237 | 0 | 0 | 137 | 233 | 368 | 262 | 0 | 0 | 219 | 189 | 371 | 221 |
| $RPNH_{0.01}$ | 0 | 0 | 51 | 79 | 519 | 351 | 0 | 0 | 55 | 90 | 488 | 367 | 0 | 0 | 58 | 90 | 428 | 424 |
| $RPNH_{0.02}$ | 0 | 0 | 46 | 77 | 540 | 337 | 0 | 0 | 48 | 84 | 520 | 348 | 0 | 0 | 52 | 87 | 472 | 389 |
| $RPNH_{0.04}$ | 0 | 0 | 44 | 81 | 573 | 302 | 0 | 0 | 46 | 80 | 555 | 319 | 0 | 0 | 53 | 78 | 533 | 336 |
| $RPNH_{0.1}$ | 0 | 0 | 70 | 187 | 550 | 193 | 0 | 0 | 63 | 221 | 537 | 179 | 0 | 0 | 55 | 139 | 628 | 178 |
| $RPNH_{0.2}$ | 0 | 0 | 17 | 68 | 774 | 141 | 0 | 0 | 11 | 13 | 817 | 159 | 0 | 0 | 2 | 8 | 854 | 136 |
| $RPNH_{0.4}$ | 0 | 0 | 0 | 0 | 856 | 144 | 0 | 0 | 0 | 0 | 833 | 167 | 0 | 0 | 0 | 0 | 841 | 159 |
| $RPNH_{0.5}$ | 0 | 0 | 0 | 0 | 845 | 155 | 0 | 0 | 0 | 0 | 830 | 170 | 0 | 0 | 0 | 0 | 832 | 168 |
| $RPNH_{0.7}$ | 0 | 0 | 0 | 0 | 834 | 166 | 0 | 0 | 0 | 0 | 815 | 185 | 0 | 0 | 0 | 0 | 826 | 174 |
| $RPNH_{1.0}$ | 0 | 0 | 0 | 0 | 828 | 172 | 0 | 0 | 1 | 1 | 813 | 185 | 0 | 0 | 0 | 0 | 841 | 159 |

表 5：污染程度为 30% 时的结果。

从这些表中的结果可以看出，$AIC$、$BIC$ 和 $AIC_c$ 的表现显著下降，具体而言，存在污染时，选出真实阶数 $p=4$ 的比例会降低。正如预期，受污染数据的比例越高，表现就越差。不过请注意，它们受 $r$不同取值的影响并不大。

另一方面，对于 $RP_{NH}$ 以及较大的调节参数值，结果与未受污染的情况相当相似。这是预期的结果，与其他考虑 RP 的情形表现出相同的行为。对于 $\alpha = 0.4$和 $\alpha = 0.5$，表现最佳，效率良好，稳健性也非常出色。

最后，为了检验这是否是这些方法的通常行为，我们针对多项式回归的不同取值重复了这项研究，其中每个系数在 $\{-2,-1,0,1,2\}$. 这为每个 $r=1$、5、10 的取值带来了3125 个不同的模型，因此我们共有9375 种不同的情形。对于所有这些情形，我们都可以得出相同的结论。

# 6 真实数据示例

在本节中，我们根据基于RP的新模型选择工具分析一组真实数据。我们考虑了 [11] 中提出、后来在 [26]中研究的问题。因变量 $Y$ 衡量的是由四种成分——铝酸三钙（$X_1$）、硅酸三钙（$X_2$）、铝铁酸四钙（$X_3$）和硅酸二钙（$X_4$）——决定的每克卡路里形式的热量。数据见表6。假设 $Y$ 可以作为$X_1$、 $X_2$、 $X_3$、 $X_4$ 的多元线性回归模型（MLRM）来表示。对于调节参数的不同取值，我们采用了 $RP_{NH}$ 程序来选择最佳模型。

| $X_1$ | $X_2$ | $X_3$ | $X_4$ | $Y$ |
|---|---|---|---|---|
| 7 | 26 | 6 | 60 | 78.5 |
| 1 | 29 | 15 | 52 | 74.3 |
| 11 | 56 | 8 | 20 | 104.3 |
| 11 | 31 | 8 | 47 | 87.6 |
| 7 | 52 | 6 | 33 | 95.9 |
| 11 | 55 | 9 | 22 | 109.2 |
| 3 | 71 | 17 | 6 | 102.7 |
| 1 | 31 | 22 | 44 | 72.5 |
| 2 | 54 | 18 | 22 | 93.1 |
| 21 | 47 | 4 | 26 | 115.9 |
| 1 | 40 | 23 | 34 | 83.8 |
| 11 | 66 | 9 | 12 | 113.3 |
| 10 | 68 | 8 | 12 | 109.4 |

表6：Hald水泥数据。

考虑自变量的不同子集，我们得到15个不同的多元线性模型，目标是选出最佳模型。不过，已知至少需要两个自变量，因为水泥需要至少两种反应物的组合。因此，我们可以排除四个简单线性回归模型，最终考虑11个可能的模型。

我们针对不同的调节参数取值，应用式（30）中定义的 $RP_{NH}$-准则来选择最合适的模型。结果见表7。如表中所示， $X_1$、 $X_2$、 $X_3$和 $X_1$、 $X_2$、 $X_4$ 这两种组合似乎是最佳候选，彼此之间差异很小。这些结果与 [26]中得到的结论相似。另请注意模型 $X_1$、 $X_2$.的良好表现

# 7 结论

本文提出了一种针对相互独立但分布不完全相同的观测值进行模型选择的新方法，旨在效率方面与其他基于最大似然的方法相竞争，同时

| | $RPNH_{0.01}$ | $RPNH_{0.02}$ | $RPNH_{0.04}$ | $RPNH_{0.05}$ | $RPNH_{0.07}$ |
|---|---|---|---|---|---|
| $X_1,\ X_2$ | 2.4179 | 2.3655 | 2.2642 | 2.2170 | 2.1280 |
| $X_1,\ X_3$ | 3.8824 | 4.3871 | 4.1321 | 4.0138 | 3.7933 |
| $X_1,\ X_4$ | 2.5408 | 2.4827 | 2.3738 | 2.3226 | 2.2261 |
| $X_2,\ X_3$ | 3.3638 | 3.2836 | 3.1140 | 3.0349 | 2.8868 |
| $X_2,\ X_4$ | 3.7164 | 3.8865 | 3.6579 | 3.5523 | 3.3565 |
| $X_3,\ X_4$ | 2.9519 | 2.8785 | 2.7413 | 2.6771 | 2.5564 |
| $X_1, X_2, X_3$ | 2.4024 | 2.3493 | 2.2495 | 2.2026 | 2.1141 |
| $X_1, X_2, X_4$ | 2.4013 | 2.3484 | 2.2490 | 2.2023 | 2.1142 |
| $X_1, X_3, X_4$ | 2.4295 | 2.3759 | 2.2752 | 2.2279 | 2.1386 |
| $X_2, X_3, X_4$ | 2.6091 | 2.5490 | 2.4363 | 2.3834 | 2.2837 |
| $X_1, X_2, X_3, X_4$ | 2.4747 | 2.4197 | 2.3164 | 2.2679 | 2.1765 |
| 最佳模型 | $(X_1, X_2, X_4)$ | $(X_1, X_2, X_4)$ | $(X_1, X_2, X_4)$ | $(X_1, X_2, X_4)$ | $(X_1, X_2, X_3)$ |
| | | | | | |
| | $RPNH_{0.1}$ | $RPNH_{0.2}$ | $RPNH_{0.4}$ | $RPNH_{0.5}$ | $RPNH_{0.7}$ |
| $X_1, X_2$ | 2.0064 | 1.6822 | 1.2656 | 1.1249 | 0.9197 |
| $X_1,\ X_3$ | 3.4984 | 2.7476 | 1.8439 | 1.5662 | 1.2019 |
| $X_1,\ X_4$ | 2.0946 | 1.7454 | 1.3004 | 1.1517 | 0.9411 |
| $X_2,\ X_3$ | 2.6873 | 2.1710 | 1.5474 | 1.3482 | 1.0702 |
| $X_2,\ X_4$ | 3.0967 | 2.4472 | 1.7055 | 1.4771 | 1.1616 |
| $X_3,\ X_4$ | 2.3930 | 1.9647 | 1.4322 | 1.2638 | 1.0347 |
| $X_1,\ X_2,\ X_3$ | 1.9933 | 1.6716 | 1.2598 | 1.1264 | 0.9173 |
| $X_1, X_2, X_4$ | 1.9939 | 1.6735 | 1.2623 | 1.1240 | 0.9228 |
| $X_1, X_3, X_4$ | 2.0167 | 1.6921 | 1.2756 | 1.1352 | 0.9985 |
| $X_2, X_3, X_4$ | 2.1482 | 1.7891 | 1.3340 | 1.1824 | 1.0089 |
| $X_1, X_2, X_3, X_4$ | 2.0521 | 1.7221 | 1.3014 | 1.1601 | 0.9548 |
| 最佳模型 | $(X_1, X_2, X_3)$ | $(X_1, X_2, X_3)$ | $(X_1, X_2, X_3)$ | $(X_1, X_2, X_4)$ | $(X_1, X_2, X_3)$ |

表7：Hald水泥数据的结果。

对离群数据更加稳健。为此，我们采用了 RP，这一工具已被证明能在许多统计问题中提供稳健估计。我们提出了一种模型选择准则，即 $RP_{NH}$ 准则，并对著名的赤池信息准则进行了扩展。此外，我们还证明样本估计量是无偏估计量。接下来，我们考虑受限模型的情形，并提出了一种方法，用以判断大型模型是否更适合对现有数据建模。作为应用示例，当我们旨在为一组数据找到最佳拟合模型时，我们构建了多元线性回归模型。我们进行了模拟研究，结果表明这种新方法在存在污染时表现良好，也就是说，模拟结果显示该方法具有稳健性。此外，其效率损失似乎较小。最后，我们将这种新方法应用于一个真实数据场景。

## 致谢

本研究由西班牙资助项目 PID2021-124933NB-I00 提供支撑集。

# 参考文献

[1] Akaike,H. (1973)。信息论与最大似然原理的扩展。载于 B. N. Petrov与F. Cs´aki（编），第二届国际信息论研讨会论文集（第 267–281页）。布达佩斯，匈牙利：Akad´emia Kiad´o。[2] Akaike,H. (1974)。重新审视统计模型识别。IEEE 自动控制学报，AC-19，716–723。[3] Basu, A.、Harris I. R.、Hjort,N.L.和 Jones,M. C. (1998)。通过最小化密度幂散度进行稳健且高效的估计。Biometrika，85（3），549–559。[4] Basu, A.、Mandal, A.、Mart´ın,N.和Pardo, L. (2018)。基于密度幂散度的复合假设检验。Sankhya, 80（13），222–262。[5] Bozdogan, H.(1987)。模型选择与赤池信息准则（赤池信息准则）：一般理论及其解析扩展。Psychometrika，52，345–370。[6] Broniatowski, M.、Toma,A.和 Vajda,I.(2012)。统计估计中的可分解伪散度及其应用。统计规划与推断杂志，142，2574–2585。[7] Castilla,E.、Jaenada, M. 和Pardo,L.(2022)。基于 R´enyi 伪散度的独立非同分布观测值的估计与检验。IEEE 信息论学报，68，7，4588–4609。[8] Castilla, E.、Jaenada, M.、Mart´ın, N. 和 Pardo,L. (2023)。通过基于 r´enyi伪散度估计量的 Wald 型检验比较两个相关正态总体的稳健方法。Statistics and Computing，DOI：10.1007/s11222-022-10162-7。[9] Cavanaugh, J.E.和 Neath,A. A. (2011)。赤池信息准则：背景、推导、性质与改进。统计科学国际百科全书，26–29。doi:10.1007/978-3-642-04898-2 111。[10] Dik,J. J. 和 Gunst, M. C. M. (1985)。正态变量中一般二次型的分布。Statistica Neerlandica，39，14–26。[11] Draper, N.R.和 Smith, H. (1981)。应用回归分析，第2 版。Wiley Blackwell。美国新泽西州霍博肯。[12] Fujisawa,H.和 Eguchi,S. (2008)。针对严重污染的小偏差稳健参数估计。多元分析杂志，99，2053–2081。[13] Hurvich，C.M.和 Tsai, C. L. (1989)。小样本中的回归与时间序列模型选择。Biometrika，76，297–307。

[14] Hurvich，C. M.和 Tsai，C. L.（1993）。用于向量自回归模型选择的修正 Akaike 信息准则。《时间序列分析杂志》，14，271–279。[15] Hurvich，C. M. 和 Tsai，C.L.（1995）。小样本中扩展拟似然模型的模型选择。《生物统计学》，51，1077–1084。[16] Jaenada，M.、Miranda，P. 和Pardo，L.（2022）。基于受限最小R´enyi 伪散度估计量的稳健检验统计量。《熵》，24，616。[17] Jaenada，M.和Pardo，L.（2022）。基于最小R´enyi伪散度估计量的广义线性模型稳健统计推断。《熵》，24，123。[18] Jones，M. C.、Hjort，N. L.、Harris，I.R.和 Basu，A.（2001）。相关的基于密度的最小散度估计量之比较。《生物计量学》，88，865–873。[19] Konishi，S.和 Kitagawa，G.（1996）。模型选择中的广义信息准则。《生物计量学》，83，875–890。[20] Kullback，S. 和 Leibler，R.A.（1951）。论信息与充分性。《数学统计年刊》，22（1），79–86。[21] Kurata，S.和 Hamada，E.（2018）。模型选择准则族的稳健推广与渐近性质。《统计学通信》（理论与方法），47，3，532-547。[22] Mattheou，K.、Lee，S.和 Karagrigoriou，A.（2009）。基于 BHHJ散度度量的模型选择准则。《统计规划与推断杂志》，139，228–235。[23] Rao，C.R. 和 Wu，Y.（2001）。论模型选择。《IMS讲义：专著系列》，312，1-57。[24] Schwarz，G. E.（1978）。估计模型的维数。《统计学年刊》，6（2），461–-464。[25] Takeuchi，K.（1976）。信息统计量的分布与模型适切性准则。《数学科学》，153，12-18（日文）。[26] Toma，A.、Karagrigoriou，A.和 Trentou，P.（2020）。基于伪距离的稳健模型选择准则。《熵》，22(3)，304。[27] Toma，A.和Leoni-Auban，S.（2010）。基于对偶散度估计量和鞍点近似的稳健检验。《多元分析杂志》，101，1143–1155。