arXiv:2601.09874v1 [stat.ME] 14 Jan 2026

# 期望分位数线性回归中的交叉验证模型选择

Bilel BOUSSELMI[1] 和 Gabriela CIUPERCA[2]

**摘要**

对于误差可能呈非对称分布的线性模型，我们研究通过交叉验证进行变量选择的方法。数据被划分为训练集和验证集，其中验证集的观测值数量远大于训练集。针对模型系数，在训练集上计算期望分位数估计量或自适应 LASSO 期望分位数估计量。随后利用这些估计量计算验证集上的交叉验证均值评分（CVS）。我们证明，在两种情况下，使 CVS 最小的模型都是一致的：解释变量的数量固定时，以及该数量随观测值数量变化时。蒙特卡洛模拟验证了理论结果，并表明与文献中的另外两种方法相比，我们的估计方法更优。通过将交叉验证期望分位数模型选择技术应用于真实数据集，展示了该技术的实用性。

**关键词**：交叉验证；期望分位数估计；模型选择。**2020年数学主题分类**：62F12, 62J07, 62F07, 62J20。

## 1 引言

交叉验证（CV）是最广泛使用的模型选择方法之一。它因简单且灵活，适用于回归和分类框架而备受重视。Stone（1974年）提出了留一交叉验证技术，推动了交叉验证方法随着时间的推移不断发展并应用于模型

[1]ESME Lyon，法国里昂电子邮箱：bilel.bousselmi@esme.fr[2]里昂大学，里昂第一大学，法国国家科学研究中心，混合研究单位5208，卡米耶·若尔当研究所，布拉科尼耶大楼，1918年11月11日大道43号，F - 69622 维勒班 Cedex，法国电子邮箱：Gabriela.Ciuperca@univ-lyon1.fr

选择。该主题一直是多项研究广泛关注的焦点。例如，读者可参阅 Arlot（2010年）和 Jung（2018年）的论文，了解各种交叉验证方法及相关参考文献。交叉验证方法之一是训练-测试划分法。该方法也称留出法，是最早用于评估预测模型的策略之一。将可用数据分成两个独立子集——一个用于训练，另一个用于验证——这一思路产生于20世纪70年代和20世纪80年代的统计学习与模式识别领域（Devijver 和 Kittler（1982年）；Ripley（1996年））。这种方法简单而有效，能够估计模型对未见数据的泛化能力。后来，Kohavi（1995年）对留出法与交叉验证、自助法等更复杂的重采样技术进行了比较，既凸显了留出法在概念上的简洁性，也指出了它对具体数据划分方式的敏感性。此后，该方法被广泛采用为基准评估策略，至今仍是现代机器学习中的重要参照（更多详情见 Hastie 等人（2009年））。在条件均值回归中，Shao（1993年）首次从理论上证明了交叉验证的合理性，表明在适当条件下，基于平方误差损失的交叉验证能够实现变量选择一致性。Zhang（1993年）将简单的留一交叉验证方法扩展为多折交叉验证。Yates 等人（2018年）综述了交叉验证及其在生态学中的应用。关于在总样本量固定的情况下，如何选择训练集/测试集的最优样本量，Afendras 和 Markatou（2019年）研究了针对某些决策规则的重复训练-测试交叉验证和 k 折交叉验证。Bates 等人（2024年）的精彩论文讨论了交叉验证并提供了相关参考文献。这里不求面面俱到，我们再提及几项较新的交叉验证研究。Silva 和 Zanella（2024年）提出了一种稳健的新型混合估计量，用于计算贝叶斯留一交叉验证准则。Fang 等人（2022年）提出了一种交叉验证程序，用于选择最小二乘模型平均中的惩罚因子，并由此得到渐近最优估计量；Rabinowicz 和 Rosset（2022年）则研究了相关数据的交叉验证。
Wieczorek 和 Lei（2022年）提出了基于交叉验证序贯变体、采用数据驱动停止规则的前向选择实现模型选择一致性的充分条件。van de Wiel 等人（2021年）的论文提出了一个灵活框架，可在多重惩罚高维岭回归中处理多种响应类型。最近，Kim 等人（2025年）证明，当验证集远大于训练集时，采用检验分位数损失的交叉验证能够为惩罚估计量和非惩罚估计量一致地识别真实模型。我们将主要把本文结果与后者的理论和数值结果进行比较。关于交叉验证其他方面的更多信息，读者还可参阅 Wang 和 Cai（2021年）、Xiong 和 Shang 的论文

（2021年）、Wei 和 Wang（2021年）、Algamal（2020）、She 和 Tran（2019年）。Newey and Powell（1987年）提出的期望分位数回归，通过最小化加权二次损失，将最小二乘估计扩展到非对称情形。当误差分布呈偏态但具有有限矩时，期望分位数回归尤其具有吸引力。期望分位数回归与分位数回归有许多相似之处，例如能够适应分布的不对称性，并能探索均值以外的条件分布。然而，与分位数回归不同，期望分位数回归采用可微损失函数，这通常能带来更高效的估计，并有助于理论分析。尽管具有这些吸引人的特性，交叉验证在期望分位数回归中的理论作用（尤其是在变量选择方面）仍基本未得到探索。

本文旨在研究线性期望分位数回归模型中的交叉验证。我们考察固定维度和高维情形，其中协变量的数量可能随样本量增长。一个关键要求是，验证集的规模必须渐近地大于训练集；这一条件与通常优先保证训练样本量较大的做法形成对比。随后，我们基于训练集上得到的期望分位数 LASSO 估计量或自适应 LASSO 期望分位数估计量，证明在温和的正则性假设下，交叉验证能够一致地选择真实模型。模拟结果验证了理论结论，并展示了所提方法相较于文献中其他方法的优势。通过将交叉验证期望分位数方法应用于两个数据集，进一步展示了该方法的实际价值。

本文其余部分安排如下。第 2 节介绍模型，以及基于期望分位数和自适应期望分位数 LASSO 估计方法的交叉验证选择准则。针对模型参数个数的三种情形，我们在第 3 节证明模型的一致性。第 4 节展示模拟结果，第 5 节则介绍真实数据集上的应用。第 6 节总结交叉验证期望分位数方法在理论和数值方面的优势，并给出数值分析的结论。理论结果的证明见第 7 节。

# 2 模型与方法

首先介绍一些将在本文其余部分使用的符号。所有向量和矩阵均用粗体符号表示，且所有向量均为列向量。对于向量或矩阵，右上角的符号 $T$ 表示其转置。对于向量$\mathbf{v}$，我们用 $\|\mathbf{v}\|_1$ 表示其 $L_1$-范数，用 $\|\mathbf{v}\|_2$ 表示其欧几里得范数，并用$\|\mathbf{v}\|_\infty$ 表示其 $L_\infty$-范数。此外，对于正定矩阵，我们用 $\mu_{min}(.)$和 $\mu_{max}(.)$分别表示其最大和最小特征值。若 $E$ 是一个事件，则 $\mathbb{1}_E$ 表示该事件 $E$ 发生的示性函数。给定集合 $\mathcal{S}$，我们用 $T$ 表示其基数——

数。若 $(a_n)_{n\in\mathbb{N}}$和 $(b_n)_{n\in\mathbb{N}}$ 是两个正确定性序列，当 $\lim_{n\to\infty} a_n/b_n = \infty$时，我们使用记号 $a_n \gg b_n$。当 $n$ 趋于无穷时，依分布收敛、依概率收敛和几乎必然收敛分别记为 $\xrightarrow{\mathcal{L}}$、$\xrightarrow{\mathbb{P}}$、$\xrightarrow{a.s.}$。在本文中，始终用 $C$ 表示一个通用常数，该常数不依赖于规模 $n$，且其取值并不重要。

考虑以下线性模型：

$$Y_i = \mathbf{X}_i^t \boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, \cdots, n, \tag{1}$$

参数向量为 $\boldsymbol{\beta} \in \mathbb{R}^p$，其真值（未知）为 $\boldsymbol{\beta}^* = (\beta_1^*, \cdots, \beta_p^*)$。$\boldsymbol{\beta}^*$ 的维度可能取决于 $n$，但对于所有 $j = 1, \cdots, p$，$\beta_j^*$ 的取值不取决于$n$。向量$\mathbf{X}_i$ 包含 $p$ 个解释性确定变量的取值，以及观测 $i$的响应变量 $Y_i$ 的取值。对于 $i = 1, \cdots, n$，这些取值（$Y_i$，$\mathbf{X}_i$）已知，而模型误差（$\varepsilon_i$）$_{1\leqslant i\leqslant n}$ 是随机的。

令非零参数的指标集为，

$$\mathcal{M}^* \equiv \left\{ j \in \{1, \cdots, p\};\, \beta_j^* \neq 0 \right\}.$$

由于 $\boldsymbol{\beta}^*$ 未知，因此集合 $\mathcal{M}^*$ 也未知。

对于一个固定的实数 $\tau \in$（0，1），期望分位数函数 $\rho_\tau$（.）定义如下：

$$\rho_\tau(x) = |\tau - \mathbb{1}_{x<0}| x^2, \qquad \text{with} \quad x \in \mathbb{R}.$$

考虑以下函数：$g_\tau(x) \equiv \rho'_\tau(x-t)|_{t=0} = 2\tau x \mathbb{1}_{x\geq 0} + 2(1-\tau) x \mathbb{1}_{x<0}$，$h_\tau(x) \equiv \rho''_\tau(x-t)_{t=0} = 2\tau \mathbb{1}_{x\geq 0} + 2(1-\tau)\mathbb{1}_{x<0}$。
请注意，当 $\varepsilon$ 的矩存在但其分布不对称时，使用 Newey and Powell（1987年）提出的期望分位数估计方法。如果 $\tau = 0.5$，则可恢复经典最小二乘法。

模型 (1) 的模型误差 $(\varepsilon_i)_{1\leqslant i\leqslant n}$ 满足以下条件：

**(A1)** $(\varepsilon_i)_{1\leqslant i\leqslant n}$是独立同分布的，满足 $\mathbb{E}[\varepsilon_i^2] < \infty$，且 $\varepsilon$的 $\tau$ 阶期望分位数为零：$\mathbb{E}[g_\tau(\varepsilon)] = 0$。

确定性设计（$\mathbf{X}_i$）$_{1\leqslant i\leqslant n}$ 满足以下假设：

**(A2)** 存在两个常数 $0 < m_0 \leq M_0$，使得 $m_0 \leq \mu_{min}\left(n^{-1}\sum_{i=1}^n {}_i\,{}_i^t\right) \leq \mu_{max}\left(n^{-1}\sum_{i=1}^n {}_i\,{}_i^t\right) \leq M_0 < \infty$。

对于一组索引 $\mathcal{M} \subseteq \{1, \ldots, p\}$，我们关联一个如下形式的模型：

$$Y_i = \mathbf{X}_{i,\mathcal{M}}^\top \boldsymbol{\beta}_{\mathcal{M}} + \varepsilon_i, \qquad i = 1, \cdots, n,$$

其中，$\boldsymbol{\beta}_{\mathcal{M}}$ 是 $\boldsymbol{\beta} \in \mathbb{R}^p$ 中与 $\mathcal{M}$ 中的索引相对应的子向量，而$\mathbf{X}_{i,\ \mathcal{M}}$ 是$\mathbf{X}_i$中相应的子向量。为简便起见，我们将 $\mathcal{M}$ 称为模型。
如果 $\mathcal{M}^* \subsetneq \mathcal{M}_o$，则模型 $\mathcal{M}_o \subseteq \{1, \cdots, p\}$ 属于过度指定；如果 $\mathcal{M}^* \not\subset \mathcal{M}_u$，则模型 $\mathcal{M}_u$ 属于指定不足。
我们还需要对使用期望分位数方法得到的估计模型作出可识别性假设。对于满足 $\mathcal{M}^* \subsetneq \mathcal{M}_u$的任意子模型 $\mathcal{M}_u$ ，我们作出如下假设：

**(A3)** $\liminf_{n\to\infty} n^{-1} \sum_{i=1}^n \left(\rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{M}_u}^\top \boldsymbol{\beta}_{\mathcal{M}_u}^*) - \rho_\tau(\varepsilon_i)\right) = \liminf_{n\to\infty} k_n > 0.$

期望分位数模型通常需要假设（A1）（例如参见 Newey and Powell（1987年）、Ciuperca（2021）、Gu 和 Zou（2016年）、Liao等（2019年）、Zhao等（2018年））。假设（A2）是线性模型中的标准假设，也是参数可识别性的必要条件（参见 Ciuperca（2021）、Liao等（2019年））。

为了更好地理解假设（A3），令 $\tau = 1/2$，如果 $\mathcal{M}^* \subsetneq \mathcal{M}_u$。那么，假设（A3）右侧变为

$$n^{-1} \sum_{i=1}^n \left(\mathbf{X}_{i,\mathcal{M}^*\setminus\mathcal{M}_u}^\top \boldsymbol{\beta}_{\mathcal{M}^*\setminus\mathcal{M}_u}^*\right)^2 - 2n^{-1} \sum_{i=1}^n \varepsilon_i \mathbf{X}_{i,\mathcal{M}^*\setminus\mathcal{M}_u}^\top \boldsymbol{\beta}_{\mathcal{M}^*\setminus\mathcal{M}_u}^*.$$

另一方面，根据大数定律并利用假设（A1），我们有：

$$n^{-1} \sum_{i=1}^n \varepsilon_i \mathbf{X}_{i,\mathcal{M}^*\setminus\mathcal{M}_u}^\top \boldsymbol{\beta}_{\mathcal{M}^*\setminus\mathcal{M}_u}^* \xrightarrow[n\to\infty]{a.s.} 0.$$

接着，应用假设（A2），假设（A3）变为：

$$\begin{aligned} n^{-1} \sum_{i=1}^n \left(\mathbf{X}_{i,\mathcal{M}^*\setminus\mathcal{M}_u}^\top \boldsymbol{\beta}_{\mathcal{M}^*\setminus\mathcal{M}_u}^*\right)^2 &= \boldsymbol{\beta}^{*\top}_{\mathcal{M}^*\setminus\mathcal{M}_u} n^{-1} \sum_{i=1}^n \mathbf{X}_{i,\mathcal{M}^*\setminus\mathcal{M}_u} \mathbf{X}_{i,\mathcal{M}^*\setminus\mathcal{M}_u}^\top \boldsymbol{\beta}_{\mathcal{M}^*\setminus\mathcal{M}_u}^* \\ &\xrightarrow[n\to\infty]{} \boldsymbol{\beta}^{*\top}_{\mathcal{M}^*\setminus\mathcal{M}_u} \lim_{n\to\infty} \left(n^{-1} \sum_{i=1}^n \mathbf{X}_{i,\mathcal{M}^*\setminus\mathcal{M}_u} \mathbf{X}_{i,\mathcal{M}^*\setminus\mathcal{M}_u}^\top\right) \boldsymbol{\beta}_{\mathcal{M}^*\setminus\mathcal{M}_u}^*, \end{aligned}$$

这与最小二乘模型中 $\mathcal{M}^*$ 的定义一致。因此，假设 (A3) 是期望分位数方法的可识别性条件，而期望分位数方法是最小二乘法的一种推广。

接下来的两个小节将根据维度 $p$ 关于 $n$ 的函数关系，提出设计矩阵的其他假设。

根据训练-测试集划分交叉验证法的定义，我们考虑一个随机选取的训练集 $\mathcal{A} \subsetneq \{1, \cdots n\}$，其中包含 $n_{\mathcal{A}} = |\mathcal{A}|$ 个观测值，$n_{\mathcal{A}}$取决于 $n$。验证（测试）集为 $\mathcal{V} \equiv \{1, \cdots, n\} \setminus \mathcal{A}$，其中有 $n_{\mathcal{V}} \equiv |\mathcal{V}| = n - n_{\mathcal{A}}$ 个观测值。

本文提出的模型选择准则将在验证观测值数量远大于训练观测值数量的情形下进行考察。这需要满足以下假设：

**(A4)** $n_{\mathcal{A}} \underset{n\to\infty}{\longrightarrow} \infty$ and $n_{\mathcal{V}}/n_{\mathcal{A}} \underset{n\to\infty}{\longrightarrow} \infty$.

值得注意的是，Kim 等人（2025）的论文也考虑了相同的假设，以通过交叉验证选择分位数模型中相关的解释变量。

因此，下面介绍一种模型选择准则，该准则基于在训练集上计算的期望分位数估计量，以及在验证集上计算的交叉验证均值。对于模型 $\mathcal{M}$，我们首先在大小为 $n_{\mathcal{A}}$ 的训练集 $\mathcal{A}$ 上计算 $\boldsymbol{\beta}_{\mathcal{M}}$的期望分位数估计量 $\boldsymbol{\beta}_{n_{\mathcal{A}},\ \mathcal{M}}$：

$$\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}} \equiv \underset{\boldsymbol{\beta}\in\mathbb{R}^{|\mathcal{M}|}}{\arg\min} \sum_{i\in\mathcal{A}} \rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{M}}^\top \boldsymbol{\beta}_{\mathcal{M}}).$$

期望分位数估计量的分量 $\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}}$ 记为 $\big(\widehat{\beta}_{(n_{\mathcal{A}},\mathcal{M});1}, \widehat{\beta}_{(n_{\mathcal{A}},\mathcal{M});2}, \cdots, \widehat{\beta}_{(n_{\mathcal{A}},\mathcal{M});p}\big)$。随后，基于该估计量，我们可以计算 $Y_i$ 在验证观测值上的预测，具体而言：$\widehat{Y}_i = \mathbf{X}_{i\ ,\mathcal{M}}^\top \widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}}$，其中 $i \in \mathcal{V}$。这些预测使我们能够计算测试集上的残差，从而定义期望分位数框架对应的交叉验证均值评分：

$$CVS(\mathcal{M}) \equiv \frac{1}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} \rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{M}}^\top \widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}}). \tag{2}$$

显然，关于 $\mathcal{M}$ 最小化关系 (2) 的 $CVS(\mathcal{M})$，相当于最小化以下差值：

$$CVS(\mathcal{M}) - CVS(\mathcal{M}^*) = \frac{1}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} \rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{M}}^\top \widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}}) - \frac{1}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} \rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{M}^*}^\top \widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}^*}) \tag{3}$$

本文提出的模型 $\mathcal{M}^*$ 估计准则如下:

$$\widehat{\mathcal{M}}_n \equiv \underset{\mathcal{M}\subseteq\{1,\cdots,p\}}{\arg\min}\ CVS(\mathcal{M}). \tag{4}$$

我们将此估计量称为交叉验证期望分位数估计量。
在接下来的两个小节中考察这一准则之前，无论 $p$ 是固定的还是依赖于 $n$，我们先考虑 $r_{n_{\mathcal{A}}}$ 是一个趋于无穷大的确定性序列，其中 $\infty$ ，当 $n \to \infty$时成立。该序列取决于观测值数量 $n_{\mathcal{A}}$ 和参数数量 $p$ 。事实上，对于给定模型 $\mathcal{M}$， $r_{n_{\mathcal{A}}}^{-1}$ 是 $\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\ \mathcal{M}}$ 趋近于$\boldsymbol{\beta}^*_{\mathcal{M}}$ 的收敛速率。
最后，对于在训练观测值 $\mathcal{A}$上估计的模型 $\mathcal{M}$ ，我们考虑$|\mathcal{M}|$维随机向量 $\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\ \mathcal{M}} \equiv r_{n_{\mathcal{A}}}\big(\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\ \mathcal{M}} - \boldsymbol{\beta}^*_{\mathcal{M}}\big)$。

我们还可以提出一种基于训练集上惩罚估计方法的模型选择准则。为此，引入以下带有自适应 LASSO 惩罚的随机过程:

$$R_{n_{\mathcal{A}},\mathcal{M}}(\boldsymbol{\beta}) \equiv \sum_{i=1}^{n_{\mathcal{A}}} \rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{M}}^t \boldsymbol{\beta}_{\mathcal{M}}) + n_{\mathcal{A}}\lambda_{n_{\mathcal{A}}} \sum_{j=1}^{p} \widehat{\omega}_{(n_{\mathcal{A}},\mathcal{M}),j}|\beta_j|,$$

其中

$$\widehat{\omega}_{(n_{\mathcal{A}},\mathcal{M}),j} = |\widehat{\beta}_{(n_{\mathcal{A}},\mathcal{M}),j}|^{-\gamma}, \quad j = 1,\cdots,p$$

且 $\gamma \in (0,1]$ 为已知常数。
利用这一最后的随机过程提出了自适应 LASSO 期望分位数估计量，该估计量能够自动识别 $\boldsymbol{\beta}$ 中为零和非零的参数（见 Liao 等人（ 2019 ）、 Ciuperca （ 2021 ）），其定义如下:

$$\widehat{\boldsymbol{\beta}}^{(aLASSO)}_{n_{\mathcal{A}},\mathcal{M}} \equiv \underset{\boldsymbol{\beta}\in\mathbb{R}}{\arg\min}\, R_{n_{\mathcal{A}},\mathcal{M}}(\boldsymbol{\beta}). \tag{5}$$

Liao 等人（2019）研究了估计量 $\widehat{\boldsymbol{\beta}}^{(aLASSO)}_{n_{\mathcal{A}}\ \mathcal{M}}$ ，在 $p$ 固定时的性质。其向 $\boldsymbol{\beta}^*$ 的收敛速率为 $n_{\mathcal{A}}^{-1/2}$ ，且 $\widehat{\boldsymbol{\beta}}^{(aLASSO)}_{n_{\mathcal{A}}\ \mathcal{M}}$ ，满足预言机性质，即以趋近于 1 的概率自动选择显著变量，并且非零估计量具有渐近正态性。如果 $p$ 取决于 $n$，Ciuperca（2021）则研究了估计量 $\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\ \mathcal{M}}$ 和 $\widehat{\boldsymbol{\beta}}^{(aLASSO)}_{n_{\mathcal{A}}\ \mathcal{M}}$ 。。

那么，针对自适应 LASSO 期望分位数估计量计算出的交叉验证均值分数如下:

$$aCVS(\mathcal{M}) \equiv \frac{1}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} \rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{M}}^\top \widehat{\boldsymbol{\beta}}^{(aLASSO)}_{n_{\mathcal{A}},\mathcal{M}}) \tag{6}$$

以及相应的模型估计量：

$$\widehat{\mathcal{M}}_n^{(a)} \equiv \underset{\mathcal{M}\subseteq\{1,\cdots,p\}}{\arg\min}\ aCVS(\mathcal{M}). \tag{7}$$

我们将 $\widehat{\mathcal{M}}_n^{(a)}$称为交叉验证自适应 LASSO 期望分位模型估计量。

模型估计量的一致性 $\widehat{\mathcal{M}}_n$ 和 $\widehat{\mathcal{M}}_n^{(a)}$将在下一节中进行研究，而主要结果的证明则推迟至第 7 节。由此产生的问题是：如果自适应 LASSO 期望分位数方法已经能选择模型，那么使用交叉验证均值评分准则来选择模型有何意义？另一个问题是，为什么使用式（7）中的自适应交叉验证得分准则，而不使用交叉验证得分？这些问题的答案见第 4 节，届时将介绍模拟结果。

模型估计量 $\widehat{\mathcal{M}}_n$ 和 $\widehat{\mathcal{M}}_n^{(a)}$将在第 4 节中进行数值研究，模拟结果将验证它们的一致性。

# 3 交叉验证模型估计量的渐近行为

本节中，我们先研究 $p$ 固定时交叉验证期望分位数估计量的一致性，再研究其依赖于 $n$时的情形。

## 3.1 $p$固定的情形

在本小节中，我们研究由 (4) 定义的模型估计量 $\widehat{\mathcal{M}}_n$ 和由 (7) 定义的模型估计量 $\widehat{\mathcal{M}}_n^{(a)}$的渐近行为，其中维度 $p$ 不依赖于 $n$。在这种情况下，我们考虑 $r_{n_{\mathcal{A}}} = n_{\mathcal{A}}^{1/2}$ 。

为了研究 $\widehat{\mathcal{M}}_n$，我们需要了解给定模型 $\mathcal{M}$时， $\boldsymbol{\beta}$ 的期望分位数估计量 $\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}}$ 的渐近行为。Ciuperca（2021）的定理 2.1(i)表明，对于给定模型 $\mathcal{M}$, 在假设 (A1) 和 (A2) 下，期望分位数估计量是一致的，即 $\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}} \xrightarrow[n\to\infty]{\mathbb{P}} \boldsymbol{\beta}^*_{\mathcal{M}}$，且收敛速率为 $r_{n_{\mathcal{A}}}^{-1} = n_{\mathcal{A}}^{-1/2}$ 阶。于是我们推得，对所有 $\mathcal{M} \subseteq \{1,\cdots,p\}$， $\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}} = O_{\mathbb{P}}(1)$。利用这一收敛速率，并采用与 Liao 等人（2019）定理 1 证明类似的方法，可以证明在相同的假设 (A1)、(A2) 下，随机向量 $\boldsymbol{\delta}_{n_{\mathcal{A}},\mathcal{M}}$ 渐近正态，即收敛 $\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_{|\mathcal{M}|},\ldots) \equiv \boldsymbol{\delta}_{\mathcal{M}}$ 成立。

在这些介绍性说明之后，我们可以通过以下定理证明由 (4) 定义的模型 $\widehat{\mathcal{M}}_n$ 的一致性。

⌒**定理 3.1** 在假设（*A1*）、（*A2*）、（*A3*）、（*A4*）下，我们有：
$\mathbb{P}[\mathcal{M}_n = \mathcal{M}^*] \underset{n\to\infty}{\longrightarrow} 1$。

在本节开头，我们看到可以基于期望分位数估计量构造自适应 LASSO 期望分位数估计量。该估计量利用训练集上的关系式（5）计算，并具有自动选择真实模型的性质。交叉验证均值评分可以根据利用训练集所作的模型预测，通过关系式（6）计算$\widehat{\boldsymbol{\beta}}^{(aLASSO)}_{n_{\mathcal{A}}\,\mathcal{M}}$ ，。

采用 Kim 等人（2025）中的相同论证，并考虑到 Liao 等人（2019）的定理 3 所确立的自适应 LASSO 期望分位数估计量的稀疏性，我们可以证明，定理 3.1 对由（7）定义的模型估计量$\widehat{\mathcal{M}}_n^{(a)}$同样成立。

## 3.2 情形 $p = p_n$, $p_n = O(n_{\mathcal{A}}^c)$

现在考虑解释变量的数量 $p$ 取决于 $n$，更确切地说，取决于训练集的大小 $n_{\mathcal{A}}$。我们假设训练集 $\mathcal{A}$和 $p$ 满足 $p = p_n = O(n_{\mathcal{A}}^c)$，其中常数 $c \in$ （$0, 1$] 且 $p < n_{\mathcal{A}}$）。这类依赖关系的一个例子是 $p = p_n = n^u$，其中 $u$ 为正数。考虑到假设（A4），常数 $u$严格大于 $c$。

注意，基数 $|\mathcal{M}^*|$ 本身可能依赖于 $n$。关于设计矩阵，还需考虑一个对（A2）的附加假设：

**(A5)** $\left(p_n/n_{\mathcal{A}}\right)^{1/2} \max_{1\leqslant i\leqslant n} \|\mathbf{X}_i\|_2 \underset{n\to\infty}{\longrightarrow} 0$, with the remark that $p_n/n_{\mathcal{A}} \underset{n\to\infty}{\longrightarrow} 0$.

Ciuperca（2021）中考虑了假设（A5），并对此作了评论。

另一方面，对于 $i \in \mathcal{A}$，我们有，$\text{tr}(\mathbf{X}_i^\top \mathbf{X}_i) = \text{tr}(\mathbf{X}_i \mathbf{X}_i^\top) = \|\mathbf{X}_i\|_2^2 \leq \max_{1\leqslant j\leqslant n_{\mathcal{A}}} \|\mathbf{X}_j\|_2^2$，结合假设（A2）可得 $\max_{1\leqslant j\leqslant n_{\mathcal{A}}} \|\mathbf{X}_j\|_2^2 \geq Cp$。最后这个关系与假设 (A5) 一起推出 $p_n n_{\mathcal{A}}^{-1/2} \to 0$，因为 $n \to \infty$，由此我们推得，若条件 $p_n = O(n_{\mathcal{A}}^c)$ 和假设 (A5) 成立，则 $c < 1/2$。因此，我们将分别讨论这两种情况：$0 \leq c < 1/2$ 和 $1/2 \leq c < 1$。

在接下来的两个小节中，根据 $c$ 小于还是大于 1/2，我们将给出具体的假设，并证明⌒模型估计量 $\mathcal{M}_n$ 的一致性。

### 3.2.1 情形 $0 < c < 1/2$

在此情况下，根据 Ciuperca（2021）中的定理 2.1，我们将以下内容视为序列 $(r_{n_{\mathcal{A}}})_{n_{\mathcal{A}} \in \mathbb{N}}$：$r_{n_{\mathcal{A}}} = n_{\mathcal{A}}^{1/2} p_n^{-1/2}$。
$p$ 依赖于 $n$ 这一事实，除了 (A2) 和 (A5) 外，还需要对 $(\mathbf{X}_i)_{i \in \mathcal{V}}$ 引入其他假设：

**(A6)** $n_{\mathcal{V}}^{-1/2} \max_{j \in \mathcal{V}} \|\mathbf{X}_j\|_2^2 \underset{n \to \infty}{\longrightarrow} 0.$

**(A7)** $r_{n_{\mathcal{A}}}^{-1} |\mathcal{M}^*| \max_{i \in \mathcal{V}} \left\|\mathbf{X}_{i,\mathcal{M}^*}\right\|_2^2 \underset{n \to \infty}{\longrightarrow} 0.$

当 (A5) 成立时，假设 (A6) 成立的一个例子是：当 $n$足够大时，$n_{\mathcal{V}}^{-1/2} n_{\mathcal{A}} / p_n$有界。事实上，可以写成：

$$n_{\mathcal{V}}^{-1/2} \max_{j \in \mathcal{V}} \|\mathbf{X}_j\|_2^2 \leq n_{\mathcal{V}}^{-1/2} \max_{1 \leqslant j \leqslant n} \|\mathbf{X}_j\|_2^2 = \left( \frac{p_n}{n_{\mathcal{A}}} \max_{1 \leqslant j \leqslant n} \|\mathbf{X}_j\|_2^2 \right) \frac{n_{\mathcal{A}}}{p_n} \frac{1}{n_{\mathcal{V}}^{1/2}}$$

考虑到假设 (A5)，如果 $n_{\mathcal{V}}^{-1/2} n_{\mathcal{A}} / p_n$ 有界，则假设 (A6) 成立。还要注意，当 $|\mathcal{M}^*|$有界时，假设 (A7) 可由 (A5) 推出。

在这些假设下，我们可以证明交叉验证期望分位数模型估计量的一致性。

**定理 3.2** 在假设 *(A1)-(A7)* 下，有：$\lim_{n \to \infty} \mathbb{P}[\mathcal{M}_n = \mathcal{M}^*] = 1$。

### 3.2.2 情形 $1/2 \leq c \leq 1$

假设 (A2) 和 (A5) 意味着 $c < 1/2$。如果 $p_n < n_{\mathcal{A}}$ 和 $c \in [1/2, 1]$，则需要用另一个假设替代假设 (A5)：

**(A8)** $\max_{1 \leqslant i \leqslant n} \|\mathbf{X}_i\|_\infty < \infty,$

例如 Ciuperca（2021）中考虑的假设。

在这种情况下，序列 $(r_{n_{\mathcal{A}}})_{n_{\mathcal{A}} \in \mathbb{N}}$ 是$r_{n_{\mathcal{A}}} = a_{n_{\mathcal{A}}}^{-1}$，其中 $(a_n)_{n \in \mathbb{N}}$ 是一个确定性序列，当 $n \to \infty$和 $n^{1/2} a_n \longrightarrow \infty$时收敛到零。在理论结果的$n \to \infty$证明中，使用了 H¨older 不等式 $|\mathbf{X}_{i,\mathcal{M}} \boldsymbol{\beta}_{\mathcal{M}}| \leq \|\mathbf{X}_{i,\mathcal{M}}\|_\infty \|\boldsymbol{\beta}_{\mathcal{M}}\|_1$。由于 $\|.\|_\infty \leq \|.\|_2$，假设（A6）、（A7）意味着：

**(A6bis)** $n_{\mathcal{V}}^{-1/2} \max_{j \in \mathcal{V}} \|\mathbf{X}_j\|_\infty^2 \underset{n \to \infty}{\longrightarrow} 0.$

**(A7bis)** $r_{n_{\mathcal{A}}}^{-1}|\mathcal{M}^*|$ 最大值$_{i\in\mathcal{V}}\left\|\mathbf{X}_{i,\ \mathcal{M}^*}\right\|_\infty^2 \underset{n\to\infty}{\longrightarrow} 0.$

假设（A8）将取代（A5）。不过，当 $c\in[1/2,1]$时，将考虑较弱的假设（A6bis）和（A7bis），而不是假设（A6）和（A7）。

需要强调的是，在 $c\in[1/2,1]$ 的情况下，对于模型 $\mathcal{M}$，在训练集 $\mathcal{A}$ 上计算得到的自适应 LASSO 惩罚权重具有如下形式：

$$\widehat{\omega}_{(n_{\mathcal{A}},\mathcal{M}),j}=\min\left(|\widetilde{\beta}_{(n_{\mathcal{A}},\mathcal{M}),j}|^{-\gamma},n_{\mathcal{A}}^{1/2}\right),\qquad \text{for } j\in\{1,\cdots,p\},$$

$\widetilde{\ }$ $\widehat{\ }$其中，$\boldsymbol{\beta}_{n_{\mathcal{A}},\ \mathcal{M}}$ 可以是期望分位数估计量$\boldsymbol{\beta}_{n_{\mathcal{A}},\ \mathcal{M}}$ 也可以是 LASSO 期望分位数估计量，后者是以下目标的最小化者：

$$n_{\mathcal{A}}^{-1}\sum_{i=1}^{n_{\mathcal{A}}}\rho_\tau(Y_i-\mathbf{X}_{i,\mathcal{M}}\boldsymbol{\beta}_{\mathcal{M}})+\nu_{n_{\mathcal{A}}}\|\boldsymbol{\beta}_{\mathcal{M}}\|_1,$$

其中，$\nu_{n_{\mathcal{A}}}$ 是一个严格为正的确定性序列，满足 $\nu_{n_{\mathcal{A}}}\underset{n\to\infty}{\longrightarrow}0$。Ciuperca（2021）证明，关于自适应 LASSO 期望分位数估计量的一致性，如果 lim $_{n_{\mathcal{A}}\to\infty}\lambda_{n_{\mathcal{A}}}|\mathcal{M}^*|^{1/2}b_{n_{\mathcal{A}}}^{-1}=0$ 且满足假设（A1）、（A2）、（A8），则 $\|\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}}}^{(aLASSO)}$ $_{\mathcal{M}}-\boldsymbol{\beta}_{\mathcal{M}}^*\|_1=O_{\mathbb{P}}(b_n)$，其中 $b_n$ 是一个确定性序列，当 $n\to\infty$ 时收敛于 0。

$\widehat{\ }$基于这些要素，可以通过以下定理证明模型估计量 $\mathcal{M}_n$ 的一致性。

**定理 3.3** 在假设（*A1*）—（*A4*）、（*A6bis*）、（*A7bis*）、（*A8*）下，我们有：

$$\lim_{n\to\infty}\mathbb{P}[\widehat{\mathcal{M}}_n=\mathcal{M}^*]=1.$$

**注 3.1** 对于 $c$ 的两种情况，当 $p=p_n=n^c$ 时，可以仿照固定 $p$ 的情形证明，通过对自适应 *LASSO* 期望分位数估计量采用交叉验证准则（*7*）即可实现模型选择。

**注 3.2** 在 *Kim* 等人（*2025*）的论文中，没有考虑 $p_n\to\infty$ 时的无惩罚情形。此外，在惩罚情形下，当 $p_n\to\infty$ 时，仅考虑了 *SCAD* 惩罚，并且要求满足非常强的条件，尤其是条件（*D3*）：$p_{n_{\mathcal{A}}}^3/n_{\mathcal{A}}\to\infty$。当 $|\mathcal{M}^*|$ 发散时，*Kim* 等人（*2025*）还施加了更强的条件：$|\mathcal{M}^*|^3p_n^3/n\to\infty$。

# 4 模拟研究

本节通过数值研究来说明理论结果，并将模型估计量与基于最小二乘法和分位数框架得到的估计量进行比较。模拟研究使用以下 R语言软件包：用于期望分位数和最小二乘法框架的包 $SALES$ ，以及用于分位数框架的包 $quantreg$。考虑的模型如下：

$$Y_i = \sum_{j=1}^{p} \beta_j^* X_{ji} + \varepsilon_i, \qquad i = 1, \cdots, n. \tag{8}$$

对于 $\varepsilon_i$ 的给定分布的实现值，我们可以在考虑假设（A1）的情况下，计算期望分位数指数 $\tau$ 的以下经验估计量：

$$\widehat{\tau}_{exp} = \frac{n^{-1}\sum_{i=1}^{n} \varepsilon_i \mathbb{1}_{\varepsilon_i<0}}{n^{-1}\big(\sum_{i=1}^{n} \varepsilon_i \mathbb{1}_{\varepsilon_i<0} - \sum_{i=1}^{n} \varepsilon_i \mathbb{1}_{\varepsilon_i>0}\big)}.$$

考虑四种模型误差： $\mathcal{N}$(0，1)、 $\mathcal{E}xp$(1)$-1.3$、($\mathcal{N}$(0，1))$^4-6$ median(($\mathcal{N}$(0，1))$^4$)、$\mathcal{E}xp$(1)–($\mathcal{N}$(0，1))$^4$,，其估计的期望分位数指数 $\hat{}\tau_{exp}$ 分别为 0。50，0。678，0。231，0。806（其直方图和密度估计见图1）。分位数指数固定为 0。5，因此这是中位数回归。($Y_i$，$\mathbf{X}_i$)$_{1\leqslant i\leqslant n}$ 的 $n$ 观测值根据固定百分比随机划分为训练集和测试集。

为了计算用于估计的模型参数 $\widehat{\mathcal{M}}_n$ 对于使交叉验证得分最小化的估计量，采用以下方法：期望分位数、分位数或最小二乘法。对于使交叉验证得分最小化的估计量 $\widehat{\mathcal{M}}_n^{(a)}$，模型参数通过自适应 LASSO 期望分位数、分位数或最小二乘估计量计算。对于每次蒙特卡洛重复，我们计算以下三个指标：

- 验证集上的均方误差（MSE）：

$$\mathrm{MSE}_{\mathcal{V}} = \frac{1}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} (Y_i - \hat{Y}_i)^2,$$

为了识别相关和不相关的解释变量，我们针对 $\widehat{\mathcal{M}}_n$ 和 $\widehat{\mathcal{M}}_n^{(a)}$计算了两个指标：

- 真正率（TPR$_{\mathcal{V}}$）：正确识别出的相关变量所占的比例，
- 真阴性率（真负率$_{\mathcal{V}}$）：被正确识别的无关变量所占的比率。

需要强调的是，完美的模型具有真正率$_{\mathcal{V}}$=真负率$_{\mathcal{V}}$=1。
该模型重复进行 $N$ 蒙特卡洛重复，其中随机变量 $(X_{ji})_{j\in\{1,\cdots,p\}}$ 和 $\varepsilon_i$ 在 $i=1,\cdots,n$ 中的实现值均重复生成。表中给出的 MSE、TPR 和 TNR 指标，是每次蒙特卡洛重复所得$MSE_{\mathcal{V}}$、TPR$_{\mathcal{V}}$和 TNR$_{\mathcal{V}}$的算术平均值。

表1考虑了 $p$ 的三个取值：2、4 和 7，其对应的真值 $\beta^*$分别为：(2,0)、(2,0,-1.5,-2) 和 (2,0,-1.5,-2,0,0,0)。为考察假设（A4）的影响，考虑比率 $s=n_{\mathcal{V}}/n$的三个取值：0.2、0.8 和 0.9。值得深入考察比率为 0.8，尤其是 0.9 时的模拟结果，因为当 $n_{\mathcal{V}}/n_{\mathcal{A}}$ 趋于无穷大时，理论基础才得以成立。因此，比率 $s$ 趋近于 1。从表1所列结果可以推断，当 $\varepsilon_i\sim\mathcal{N}$（0，1）时，期望分位数法和最小二乘法的结果相同，且真正率和真负率随比率 $s$的增加而提高。当 $\varepsilon_i\sim\mathcal{E}xp$（1）$-1.3$时，期望分位数法的表现不如最小二乘法（比较真正率和真负率时）。当分布 $(\mathcal{N}$（0，1))$^4-6$ 中位数((𝒩（0，1))$^4$)、$\mathcal{E}xp$（1）$-$(𝒩（0，1))$^4$,时，期望分位数法在均方误差、真正率和真负率方面的表现均优于最小二乘法。这可以理解为，模型误差分布必须具有很强的不对称性，期望分位数框架下的模型估计才会优于最小二乘法。另一方面，分位数法在 $s=0.2$ 和 $s=0.8$下所得的结果与 Kim 等人（2025）的结果一致。当 $s=0.9$时，分位数法的真负率趋近于 1 的速度较慢。因此，对于相同的模型误差分布，期望分位数法在给定的$n$下更为准确。

在表2和表3中，$\beta^*$ 的真实系数向量为 (2，0，$-1.5$，$-2$)，对于 $p=4$，为匹配维度，会附加 $p>4$ 个零。表2和表3所示结果，是针对 $\varepsilon$ 的相同四种分布、分位数指标等于 0.5 的情形，经过 200 次蒙特卡洛重复得到的。从这两张表可以推断，如果 $p$ 取决于 $n$，那么参数向量中的零越多，真负率就越高。对于 $n=100$，在均方误差和真正率方面，分位数（更确切地说，中位数）方法优于期望分位数方法，但真负率较低。对于 $n\in\{500,1000\}$，期望分位数方法和分位数方法在真正率方面的结果相近。然而，在真负率方面，期望分位数方法更善于识别无关变量。与固定 $p$的情形一样，我们发现，随着比率 $s$ 趋近于 1，期望分位数方法得到的真负率会更快地趋近于 1。随着样本量增加，零系数的识别效果会提高。对于非对称模型误差，最小二乘法总是识别出较少的无关变量，且无法识别所有相关变量，尤其是在 $n$较小时。

| $p$ | 误差 | 方法 | MSE | | | TPR | | | TNR | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 20% | 80% | 90% | 20% | 80% | 90% | 20% | 80% | 90% |
| 2 | $\mathcal{N}(0,\ 1)$ | 期望分位数 | 0.0020 | 0.0060 | 0.0106 | 1.000 | 1.000 | 1.000 | 0.555 | 0.750 | 0.790 |
| | | LS | 0.0020 | 0.0060 | 0.0106 | 1.000 | 1.000 | 1.000 | 0.565 | 0.745 | 0.775 |
| | | 分位数 | 0.0027 | 0.0091 | 0.0159 | 1.000 | 1.000 | 1.000 | 0.575 | 0.700 | 0.745 |
| | $\mathcal{E}xp(1)\ -1.3$ | 期望分位数 | 0.0023 | 0.0096 | 0.0162 | 1.000 | 1.000 | 1.000 | 0.590 | 0.685 | 0.775 |
| | | LS | 0.0014 | 0.0055 | 0.0100 | 1.000 | 1.000 | 1.000 | 0.565 | 0.730 | 0.800 |
| | | 分位数 | 0.0016 | 0.0056 | 0.0096 | 1.000 | 1.000 | 1.000 | 0.595 | 0.655 | 0.720 |
| | $(\mathcal{N}(0,\ 1))^4-6$中位数$(\mathcal{N}(0,\ 1)^4)$ | 期望分位数 | 0.0852 | 0.0765 | 0.2124 | 0.970 | 1.000 | 0.980 | 0.650 | 0.705 | 0.755 |
| | | LS | 0.4408 | 0.4128 | 1.0205 | 0.860 | 0.940 | 0.875 | 0.580 | 0.645 | 0.695 |
| | | 分位数 | 0.0016 | 0.0073 | 0.0157 | 1.000 | 1.000 | 1.000 | 0.635 | 0.675 | 0.680 |
| | $\mathcal{E}xp(1)$–$(\mathcal{N}(0,1))^4$ | 期望分位数 | 0.0852 | 0.0765 | 0.2124 | 0.970 | 1.000 | 0.980 | 0.650 | 0.705 | 0.755 |
| | | LS | 0.4408 | 0.4128 | 1.0205 | 0.860 | 0.940 | 0.875 | 0.580 | 0.645 | 0.695 |
| | | 分位数 | 0.0016 | 0.0073 | 0.0157 | 1.000 | 1.000 | 1.000 | 0.635 | 0.675 | 0.680 |
| 4 | $\mathcal{N}(0,\ 1)$ | 期望分位数 | 0.0021 | 0.0079 | 0.0155 | 1.000 | 1.000 | 1.000 | 0.610 | 0.665 | 0.790 |
| | | LS | 0.0021 | 0.0078 | 0.0154 | 1.000 | 1.000 | 1.000 | 0.615 | 0.645 | 0.790 |
| | | 分位数 | 0.0031 | 0.0124 | 0.0228 | 1.000 | 1.000 | 1.000 | 0.590 | 0.650 | 0.655 |
| | $\mathcal{E}xp(1)\ -1.3$ | 期望分位数 | 0.0037 | 0.0135 | 0.0272 | 1.000 | 1.000 | 1.000 | 0.610 | 0.715 | 0.735 |
| | | LS | 0.0021 | 0.0079 | 0.0169 | 1.000 | 1.000 | 1.000 | 0.615 | 0.740 | 0.740 |
| | | 分位数 | 0.0020 | 0.0070 | 0.0144 | 1.000 | 1.000 | 1.000 | 0.520 | 0.695 | 0.640 |
| | $(\mathcal{N}(0,\ 1))^4-6$中位数$(\mathcal{N}(0,\ 1)^4)$ | 期望分位数 | 0.0902 | 0.1436 | 0.2626 | 0.973 | 0.997 | 0.960 | 0.560 | 0.780 | 0.685 |
| | | LS | 0.4992 | 0.6757 | 0.8374 | 0.865 | 0.867 | 0.825 | 0.570 | 0.750 | 0.720 |
| | | 分位数 | 0.0015 | 0.0061 | 0.0161 | 1.000 | 1.000 | 1.000 | 0.515 | 0.645 | 0.700 |
| | $\mathcal{E}xp(1)$–$(\mathcal{N}(0,1))^4$ | 期望分位数 | 0.0851 | 0.1348 | 0.2619 | 0.977 | 0.998 | 0.967 | 0.525 | 0.745 | 0.750 |
| | | LS | 0.5098 | 0.6670 | 0.8919 | 0.857 | 0.880 | 0.803 | 0.560 | 0.720 | 0.765 |
| | | 分位数 | 0.0023 | 0.0136 | 0.0348 | 1.000 | 1.000 | 1.000 | 0.595 | 0.695 | 0.740 |
| 7 | $\mathcal{N}(0,\ 1)$ | 期望分位数 | 0.0016 | 0.0051 | 0.0107 | 1.000 | 1.000 | 1.000 | 0.599 | 0.721 | 0.782 |
| | | LS | 0.0016 | 0.0051 | 0.0107 | 1.000 | 1.000 | 1.000 | 0.600 | 0.730 | 0.781 |
| | | 分位数 | 0.0024 | 0.0073 | 0.0140 | 1.000 | 1.000 | 1.000 | 0.573 | 0.699 | 0.690 |
| | $\mathcal{E}xp(1)\ -1.3$ | 期望分位数 | 0.0029 | 0.0089 | 0.0150 | 1.000 | 1.000 | 1.000 | 0.573 | 0.725 | 0.784 |
| | | LS | 0.0017 | 0.0052 | 0.0095 | 1.000 | 1.000 | 1.000 | 0.584 | 0.739 | 0.787 |
| | | 分位数 | 0.0016 | 0.0041 | 0.0084 | 1.000 | 1.000 | 1.000 | 0.547 | 0.664 | 0.693 |
| | $(\mathcal{N}(0,\ 1))^4-6$中位数$(\mathcal{N}(0,\ 1)^4)$ | 期望分位数 | 0.0481 | 0.0950 | 0.1982 | 0.980 | 0.988 | 0.957 | 0.630 | 0.741 | 0.795 |
| | | LS | 0.3030 | 0.4434 | 0.6122 | 0.867 | 0.863 | 0.797 | 0.613 | 0.726 | 0.780 |
| | | 分位数 | 0.0011 | 0.0038 | 0.0078 | 1.000 | 1.000 | 1.000 | 0.591 | 0.650 | 0.689 |
| | $\mathcal{E}xp(1)$–$(\mathcal{N}(0,1))^4$ | 期望分位数 | 0.0667 | 0.0896 | 0.1716 | 0.967 | 0.992 | 0.962 | 0.608 | 0.735 | 0.789 |
| | | LS | 0.3531 | 0.4350 | 0.5453 | 0.840 | 0.878 | 0.813 | 0.601 | 0.709 | 0.772 |
| | | 分位数 | 0.0014 | 0.0064 | 0.0175 | 1.000 | 1.000 | 1.000 | 0.581 | 0.694 | 0.716 |

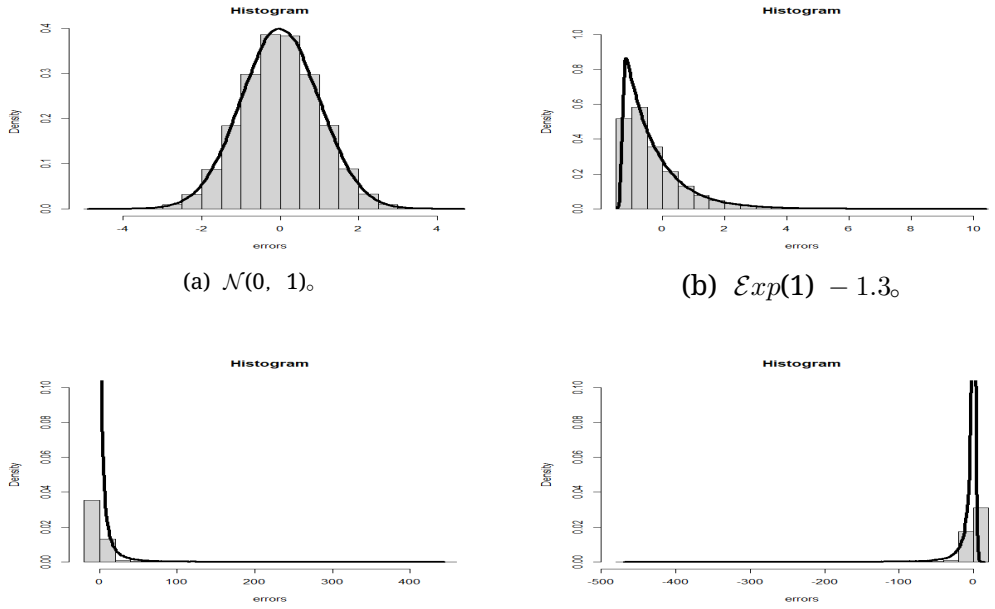
表1：对于 $\boldsymbol{\beta}^*$的不同维度：$p=2$（$\boldsymbol{\beta}^*=(2,\ 0)$），$p=4$（$\boldsymbol{\beta}^*=(2,\ 0,\ -1.5,\ -2)$），以及$p=7$（$\boldsymbol{\beta}^*=(2,\ 0,\ -1.5,\ -2,\ 0,\ 0,\ 0)$）下，通过 $\mathcal{M}_n$ 式(4)得到的均方误差、真正率和真负率结果。所有结果均基于 $N=200$ 次蒙特卡罗重复和 $n=500$。

| $n$ | 误差 | 方法 | MSE | | TPR | | TNR | |
|---|---|---|---|---|---|---|---|---|
| | | | 80% | 90% | 80% | 90% | 80% | 90% |
| 100 | $\mathcal{N}(0,\ 1)$ | 期望分位数 | 0.0508 | 0.144 | 1.000 | 0.995 | 0.735 | 0.715 |
| | | LS | 0.0500 | 0.1421 | 1.000 | 0.995 | 0.725 | 0.735 |
| | | 分位数 | 0.0670 | 0.1912 | 1.000 | 0.992 | 0.685 | 0.700 |
| | $\mathcal{E}xp(1)\ -1.3$ | 期望分位数 | 0.0750 | 0.1802 | 0.998 | 0.987 | 0.725 | 0.745 |
| | | LS | 0.0538 | 0.1473 | 0.998 | 0.992 | 0.725 | 0.750 |
| | | 分位数 | 0.0496 | 0.1353 | 1.000 | 0.995 | 0.695 | 0.675 |
| | $(\mathcal{N}(0,\ 1))^4 - 6$中位数$(\mathcal{N}(0,\ 1)^4)$ | 期望分位数 | 0.7952 | 0.9239 | 0.852 | 0.805 | 0.760 | 0.735 |
| | | LS | 1.6470 | 1.7049 | 0.685 | 0.642 | 0.710 | 0.665 |
| | | 分位数 | 0.1204 | 0.4463 | 0.982 | 0.918 | 0.740 | 0.730 |
| | $\mathcal{E}xp(1)\ -(\mathcal{N}(0,1))^4$ | 期望分位数 | 0.7454 | 1.0503 | 0.837 | 0.753 | 0.730 | 0.760 |
| | | LS | 1.6139 | 1.9022 | 0.623 | 0.645 | 0.660 | 0.710 |
| | | 分位数 | 0.2377 | 0.6993 | 0.972 | 0.860 | 0.695 | 0.780 |
| 500 | $\mathcal{N}(0,1)$ | 期望分位数 | 0.0078 | 0.0155 | 1.000 | 1.000 | 0.693 | 0.790 |
| | | LS | 0.0078 | 0.0154 | 1.000 | 1.000 | 0.698 | 0.790 |
| | | 分位数 | 0.0101 | 0.0228 | 1.000 | 1.000 | 0.660 | 0.655 |
| | $\mathcal{E}xp(1)\ -1.3$ | 期望分位数 | 0.0108 | 0.0272 | 1.000 | 1.000 | 0.762 | 0.735 |
| | | LS | 0.0065 | 0.0169 | 1.000 | 1.000 | 0.785 | 0.740 |
| | | 分位数 | 0.0060 | 0.0144 | 1.000 | 1.000 | 0.652 | 0.640 |
| | $(\mathcal{N}(0,\ 1))^4 - 6$中位数$(\mathcal{N}(0,\ 1)^4)$ | 期望分位数 | 0.1094 | 0.2626 | 0.993 | 0.960 | 0.748 | 0.685 |
| | | LS | 0.4931 | 0.8374 | 0.883 | 0.825 | 0.725 | 0.720 |
| | | 分位数 | 0.0047 | 0.0161 | 1.000 | 1.000 | 0.650 | 0.700 |
| | $\mathcal{E}xp(1)\ -(\mathcal{N}(0,1))^4$ | 期望分位数 | 0.1095 | 0.2619 | 0.990 | 0.967 | 0.750 | 0.750 |
| | | LS | 0.4794 | 0.8919 | 0.897 | 0.803 | 0.735 | 0.765 |
| | | 分位数 | 0.0085 | 0.0348 | 1.000 | 1.000 | 0.693 | 0.740 |
| 1000 | $\mathcal{N}(0,\ 1)$ | 期望分位数 | 0.0024 | 0.0053 | 1.000 | 1.000 | 0.749 | 0.792 |
| | | LS | 0.0024 | 0.0053 | 1.000 | 1.000 | 0.748 | 0.792 |
| | | 分位数 | 0.0031 | 0.0073 | 1.000 | 1.000 | 0.706 | 0.653 |
| | $\mathcal{E}xp(1)\ -1.3$ | 期望分位数 | 0.0042 | 0.0090 | 1.000 | 1.000 | 0.757 | 0.767 |
| | | LS | 0.0024 | 0.0054 | 1.000 | 1.000 | 0.744 | 0.788 |
| | | 分位数 | 0.0020 | 0.0047 | 1.000 | 1.000 | 0.655 | 0.660 |
| | $(\mathcal{N}(0,\ 1))^4 - 6$中位数$(\mathcal{N}(0,\ 1)^4)$ | 期望分位数 | 0.0351 | 0.0961 | 1.000 | 0.990 | 0.767 | 0.782 |
| | | LS | 0.2181 | 0.3841 | 0.957 | 0.910 | 0.756 | 0.760 |
| | | 分位数 | 0.0016 | 0.0039 | 1.000 | 1.000 | 0.669 | 0.690 |
| | $\mathcal{E}xp(1)\ -(\mathcal{N}(0,1))^4$ | 期望分位数 | 0.0348 | 0.0975 | 1.000 | 0.995 | 0.764 | 0.787 |
| | | LS | 0.2199 | 0.4228 | 0.962 | 0.892 | 0.729 | 0.778 |
| | | 分位数 | 0.0026 | 0.0085 | 1.000 | 1.000 | 0.690 | 0.718 |

表 2：当维度$p$ 随 $n$, $p = O(n_{\mathcal{A}}^c)$变化时， $c = 1/3$和 $s \in \{0.8, 0.9\}$的 (4) 的 $\mathcal{M}_n$ 所得均方误差、真正率和真负率结果。 $n$ 和 $p$的取值：当 $n = 100$ 时，我们考虑 $p = 4$ （对于 s=0.8）和 $p = 4$ （对于 s=0.9）；当 $n = 500$时，我们考虑 $p = 5$ （对于 s=0.8）和 $p = 4$ （对于 s=0.9）；当 $n = 1000$ 时，我们考虑$p = 7$ （对于 s=0.8）和 $p = 6$ （对于 s=0.9）。

| $n$ | 误差 | 方法 | MSE | | TPR | | TNR | |
|---|---|---|---|---|---|---|---|---|
| | | | 80% | 90% | 80% | 90% | 80% | 90% |
| 100 | $\mathcal{N}(0,\ 1)$ | 期望分位数 | 0.0199 | 0.0600 | 1.000 | 0.997 | 0.727 | 0.770 |
| | | LS | 0.0196 | 0.0586 | 1.000 | 1.000 | 0.730 | 0.768 |
| | | 分位数 | 0.0260 | 0.0652 | 1.000 | 1.000 | 0.716 | 0.708 |
| | $\mathcal{E}xp(1)\ -1.3$ | 期望分位数 | 0.0311 | 0.0861 | 1.000 | 0.987 | 0.710 | 0.718 |
| | | LS | 0.0229 | 0.0714 | 1.000 | 0.990 | 0.734 | 0.752 |
| | | 分位数 | 0.0152 | 0.0468 | 1.000 | 0.997 | 0.673 | 0.730 |
| | $(\mathcal{N}(0,\ 1))^4 - 6$ 中位数$(\mathcal{N}(0,\ 1)^4)$ | 期望分位数 | 0.4046 | 0.4873 | 0.840 | 0.770 | 0.729 | 0.774 |
| | | LS | 1.0411 | 0.8863 | 0.617 | 0.613 | 0.649 | 0.724 |
| | | 分位数 | 0.0424 | 0.2036 | 0.987 | 0.913 | 0.703 | 0.782 |
| | $\mathcal{E}xp(1)\ -(\mathcal{N}(0,1))^4$ | 期望分位数 | 0.3672 | 0.5840 | 0.840 | 0.733 | 0.741 | 0.784 |
| | | LS | 0.9102 | 1.1841 | 0.660 | 0.563 | 0.693 | 0.722 |
| | | 分位数 | 0.0984 | 0.2816 | 0.963 | 0.877 | 0.723 | 0.774 |
| 500 | $\mathcal{N}(0,\ 1)$ | 期望分位数 | 0.0067 | 0.0088 | 1.000 | 1.000 | 0.667 | 0.740 |
| | | LS | 0.0067 | 0.0087 | 1.000 | 1.000 | 0.683 | 0.740 |
| | | 分位数 | 0.0079 | 0.0072 | 1.000 | 1.000 | 0.550 | 0.680 |
| | $\mathcal{E}xp(1)\ -1.3$ | 期望分位数 | 0.0042 | 0.012 | 1.000 | 1.000 | 0.800 | 0.720 |
| | | LS | 0.0029 | 0.0073 | 1.000 | 1.000 | 0.767 | 0.720 |
| | | 分位数 | 0.0032 | 0.0041 | 1.000 | 1.000 | 0.617 | 0.800 |
| | $(\mathcal{N}(0,\ 1))^4 - 6$ 中位数$(\mathcal{N}(0,\ 1)^4)$ | 期望分位数 | 0.0837 | 0.0510 | 1.000 | 1.000 | 0.800 | 0.760 |
| | | LS | 0.3440 | 0.2155 | 0.800 | 0.867 | 0.717 | 0.760 |
| | | 分位数 | 0.0017 | 0.0025 | 1.000 | 1.000 | 0.650 | 0.720 |
| | $\mathcal{E}xp(1)\ -(\mathcal{N}(0,1))^4$ | 期望分位数 | 0.0393 | 0.0835 | 1.000 | 0.933 | 0.700 | 0.820 |
| | | LS | 0.1822 | 0.3265 | 0.933 | 0.733 | 0.717 | 0.700 |
| | | 分位数 | 0.0030 | 0.0061 | 1.000 | 1.000 | 0.767 | 0.640 |
| 1000 | $\mathcal{N}(0,\ 1)$ | 期望分位数 | 0.0012 | 0.0048 | 1.000 | 1.000 | 0.80 | 0.820 |
| | | LS | 0.0012 | 0.0048 | 1.000 | 1.000 | 0.80 | 0.820 |
| | | 分位数 | 0.0019 | 0.0061 | 1.000 | 1.000 | 0.710 | 0.794 |
| | $\mathcal{E}xp(1)\ -1.3$ | 期望分位数 | 0.0031 | 0.0048 | 1.000 | 1.000 | 0.799 | 0.828 |
| | | LS | 0.0023 | 0.0035 | 1.000 | 1.000 | 0.796 | 0.848 |
| | | 分位数 | 0.0011 | 0.0019 | 1.000 | 1.000 | 0.755 | 0.801 |
| | $(\mathcal{N}(0,\ 1))^4 - 6$ 中位数$(\mathcal{N}(0,\ 1)^4)$ | 期望分位数 | 0.0234 | 0.0305 | 1.000 | 0.990 | 0.867 | 0.959 |
| | | LS | 0.0162 | 0.1787 | 1.000 | 1.000 | 0.802 | 0.839 |
| | | 分位数 | 0.0012 | 0.0021 | 1.000 | 1.000 | 0.799 | 0.803 |
| | $\mathcal{E}xp(1)\ -(\mathcal{N}(0,1))^4$ | 期望分位数 | 0.0302 | 0.0436 | 1.000 | 1.000 | 0.988 | 1.000 |
| | | LS | 0.1267 | 0.1396 | 1.000 | 1.000 | 0.935 | 0.954 |
| | | 分位数 | 0.0002 | 0.0009 | 1.000 | 1.000 | 0.870 | 0.889 |

表3：当维度$p$ 随 $n$, $p = O(n_{\mathcal{A}}^c)$变化时，(4)的 $\mathcal{M}_n$ 所得到的均方误差、真正率和真负率结果，适用于 $c = 4/5$和 $s \in \{0.8, 0.9\}$。 $n$ 和 $p$的取值：当 $n = 100$ 时，我们考虑 $p = 10$ (对于 s=0.8) 和 $p = 8$ (对于 s=0.9)；当 $n = 500$时，我们考虑 $p = 20$ (对于 s=0.8) 和 $p = 15$ (对于 s=0.9)；当 $n = 1000$ 时，我们考虑$p = 50$ (对于 s=0.8) 和 $p = 30$ (对于 s=0.9)。




(a) $\mathcal{N}(0,\ 1)$。

(b) $\mathcal{E}xp(1)\ -1.3$。

(c) $(\mathcal{N}(0,\ 1))^4 - 6$ 中位数$((\mathcal{N}(0,\ 1))^4)$。

(d) $\mathcal{E}xp(1)\ -(\mathcal{N}(0,1))^4$.

图1： $\varepsilon$的四种分布的直方图和密度估计。

现在，让我们考察为自适应 LASSO 期望分位数估计量计算的交叉验证均值评分（6）的模型估计量。经过 200 次蒙特卡洛重复后，将估计量（7）与同样四种模型误差下相应的自适应 LASSO 中位数估计量和最小二乘估计量进行比较。表4中所考虑的模型包含七个解释变量，其中三个是相关变量。
比较 $p = 7$ 对应的表4和表1中的结果可以看出，对于给定的误差分布和技术，期望分位数、分位数和最小二乘法这三种方法的交叉验证模型估计量具有几乎相同的真正率。不过，由于自适应 LASSO 满足稀疏性，模型 $\widehat{\mathcal{M}}_n^{(a)}$识别无关变量的能力远优于未惩罚估计量的模型 $\mathcal{M}_n$ 。 就准确性而言，两种模型估计量 $\mathcal{M}_n$ 和 $\widehat{\mathcal{M}}_n^{(a)}$的均方误差值非常接近。

表 5 展示了当 $p = 10$ 时， $\boldsymbol{\beta}^*$ 取两个值的结果： $\boldsymbol{\beta}^0$ =(2，0，$-1.5$，0，0，0，0，0，0，0) 和 $\boldsymbol{\beta}^1$ =(0。2，0，$-0.5$，0，0，0，0，0，0，0)，后者的非零值更接近零。在此表中，我们使用真正率和真负率这两项标准，对由 $\widehat{\mathcal{M}}_n^{(a)}$ 得到的结果进行比较，其结果记为交叉验证指数、交叉验证最小二乘、交叉验证二次型，其中 $n = 100$和 $n_{\mathcal{A}} = 15$。学习观测值的数量与参数数量非常接近。此外，对于三种估计方法，比较还涵盖了全部 $n$ 观测值上的自适应 LASSO 估计量（记为前向选择指数分布、前向选择最小二乘、前向选择分位数）。交叉验证估计量能够识别零系数

| | | | | MSE | | | TPR | | | TNR | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | $n$ | 误差 | 方法 | 20% | 80% | 90% | 20% | 80% | 90% | 20% | 80% | 90% |
| | 500 | 正态分布 (0,1) | 期望分位数 | 0.0012 | 0.0051 | 0.0127 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | | LS | 0.0012 | 0.0051 | 0.0126 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | | 中位数 | 0.0018 | 0.0081 | 0.0188 | 1.000 | 1.000 | 1.000 | 1.000 | 0.993 | 0.985 |
| | 500 | 指数分布 Exp(1)-1.3 | 期望分位数 | 0.0022 | 0.0096 | 0.0180 | 1.000 | 1.000 | 1.000 | 0.994 | 0.999 | 1.000 |
| | | | LS | 0.0013 | 0.0058 | 0.0116 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 |
| | | | 中位数 | 0.0011 | 0.0038 | 0.008 | 1.000 | 1.000 | 1.000 | 1.000 | 0.994 | 0.989 |
| 7 | 500 | $(N(0,1))^4 - 6$ 中位数$(N(0,1)^4)$ | 期望分位数 | 0.0450 | 0.0968 | 0.2085 | 0.980 | 0.988 | 0.945 | 0.887 | 0.946 | 0.964 |
| | | | LS | 0.2991 | 0.4497 | 0.6209 | 0.863 | 0.843 | 0.768 | 0.772 | 0.858 | 0.886 |
| | | | 中位数 | 0.0011 | 0.0043 | 0.009 | 1.000 | 1.000 | 1.000 | 1.000 | 0.989 | 0.991 |
| | 500 | 指数分布 Exp(1) – ($N$(0, 1))$^4$ | 期望分位数 | 0.0631 | 0.0906 | 0.1852 | 0.967 | 0.992 | 0.945 | 0.876 | 0.929 | 0.963 |
| | | | LS | 0.3489 | 0.4452 | 0.5618 | 0.838 | 0.862 | 0.787 | 0.761 | 0.835 | 0.886 |
| | | | 中位数 | 0.0012 | 0.0097 | 0.0372 | 1.000 | 1.000 | 1.000 | 1.000 | 0.994 | 0.971 |

表 4：当 $p = 7$ 时，均方误差、真正率和真负率的结果，其中 $\beta^* =$(2，0，$-1.5$，$-2$，0，0，0)。结果基于针对关系 (7) 的模型估计量 $\widehat{\mathcal{M}}_n^{(a)}$ 进行的 $N = 200$ 次蒙特卡罗重复。

在全样本规模下，表现更好，非零系数的表现则略逊于自适应 LASSO 估计量。值得注意的是，当 $\beta_j^*$ 的非零值接近零时，交叉验证期望分位数估计量能够很好地识别无关变量，而交叉验证最小二乘法、交叉验证分位数法和自适应 LASSO 方法则不能做到这一点。

现在我们来考察各方法执行一次蒙特卡罗循环所需的计算时间，包括通过交叉验证或在整个数据库上进行计算，以及采用惩罚估计量或未惩罚估计量的情况。所考虑的模型包含 $p \in \{10, 18, 20, 30\}$ 个解释变量，其中只有三个是相关变量。从表 6 可以看出，无论采用惩罚估计量还是非惩罚估计量，使用交叉验证方法时，每个期望分位数所需的计算时间都短于每个分位数所需的时间。随着 $p$的增大，这种计算时间差异也会扩大。我们还可以看出，对于期望分位数，在所有观测值上进行计算所需的时间少于交叉验证，但随着 $p$增大，这种差异会缩小。表中的结果进一步凸显了交叉验证期望分位数方法相较于交叉验证分位数方法的优势，尤其是在解释变量数量较多的模型中。

# 5 真实数据上的应用

为了体现我们的模型选择方法在实际中的价值，本节介绍两个基于真实数据库的应用。我们将结果与中位数和最小二乘交叉验证所得结果进行比较。

Be在介绍这些应用之前，请注意期望分位数指数的取值，$\tau$ ，

| 方法 | 当 $\beta^* = \beta^0$ 时的结果 | | 当 $\beta^* = \beta^1$ 时的结果 | |
|---|---|---|---|---|
| | TPR | TNR | TPR | TNR |
| 交叉验证指数_ | 0.9 | 1.000 | 0.8 | 0.992 |
| 前向选择指数分布_ | 1.0 | 0.901 | 0.88 | 0.900 |
| 交叉验证最小二乘_ | 0.8 | 0.900 | 0.76 | 0.875 |
| 前向选择最小二乘_ | 0.9 | 0.765 | 0.9 | 0.702 |
| $CV_Q$ | 1.0 | 0.950 | 0.87 | 0.910 |
| $FS_Q$ | 1.0 | 1.000 | 0.87 | 0.991 |

表 5：交叉验证估计量（记为 "交叉验证 ..."）和自适应 LASSO 估计量在所有观测值上的真正率和真负率（记为 "前向选择最小二乘 ..."）。基于 10 次蒙特卡罗重复得到的结果，其中 $n = 100$, $n_{\mathcal{A}} = 15$, $n_{\mathcal{V}} = 85$, $p = 10$,$\beta^0 =$(2，0，$-1.5$，0，0，0，0，0，0，0) 和 $\beta^1 =$ (0。2，0，$-0.5$，0，0，0，0，0，0，0)。

| 维度 $p$ | 方法 | 计算时间（秒 / 小时） | |
|---|---|---|---|
| | | 未惩罚 | 惩罚 |
| 10 | $CVQuantile$ | 7.68 (≈ 0.002 小时) | 8.05 (≈ 0.002 小时) |
| | $CVExpectile$ | 6.38 (≈ 0.002 小时) | 6.38 (≈ 0.002 小时) |
| | $fullExpectile$ | 3.54 (≈ 0.0008 小时) | |
| 18 | $CVQuantile$ | 1819（≈ 0.505 小时） | 1858（≈ 0.516 小时） |
| | $CVExpectile$ | 1454（≈ 0.403 小时） | 1977（≈ 0.549 小时） |
| 20 | $CVQuantile$ | 10740（≈ 2.983 小时） | 12776（≈ 3.548 小时） |
| | $CVExpectile$ | 7288（≈ 2.024 小时） | 11613（≈ 3.225 小时） |
| | $fullExpectile$ | 6488（≈ 1.802 小时） | |
| 30 | $CVQuantile$ | 57000（≈ 15.833 小时） | 63201（≈ 17.555 小时） |
| | $CVExpectile$ | 29650（≈ 8.236 小时） | 35120（≈ 9.722 小时） |

表 6：未惩罚和惩罚估计量的交叉验证分位数方法与期望分位数方法的计算时间。使用未惩罚期望分位数方法对全部 $n$ 个观测值（完整数据）进行模型估计的计算时间。我们考虑 $n = 100$、一次蒙特卡洛重复、 $n_{\mathcal{A}} = 20$以及 $p$ 的不同取值。

在这些情况下， $\varepsilon_i$ 是未知的，因为它们无法观测。由于可观测到（$Y_i$，$\mathbf{X}_i$）$_{1\leqslant i\leqslant n}$，观测值记为（$y_i$，$\mathbf{x}_i$）$_{1\leqslant i\leqslant n}$，因此可以采用一种方法估计期望分位数指数的值。

基于假设（A7）以及 $\mathbf{X}$ 对 $Y$ 的依赖关系，我们提出以下方案，根据假设（A1）计算 $\tau$ 的经验估计。
步骤 *0*。 首先，我们考虑 $\eta^{(0)}(y_i) \equiv y_i - median(y_1, \cdots, y_n)$，用于计算：

$$\widehat{\tau}^{(0)} = \frac{n^{-1}\sum_{i=1}^{n}\eta^{(0)}(y_i)\mathbb{1}_{\eta^{(0)}(y_i)<0}}{n^{-1}\big(\sum_{i=1}^{n}\eta^{(0)}(y_i)\mathbb{1}_{\eta^{(0)}(y_i)<0} - \sum_{i=1}^{n}\eta^{(0)}(y_i)\mathbb{1}_{\eta^{(0)}(y_i)>0}\big)}. \tag{9}$$

步骤 *1*。 利用指数 $\widehat{\tau}^{(0)}$，计算模型参数的期望估计，据此计算 $y_i$ 的预测值 $\widehat{y}_i^{(1)}$ 及相应的残差 $e_i^{(1)} \equiv y_i - \widehat{y}_i^{(1)}$。我们计算：

$$\widehat{\tau}^{(1)} = \frac{n^{-1}\sum_{i=1}^{n}e_i^{(1)}\mathbb{1}_{e_i^{(1)}<0}}{n^{-1}\big(\sum_{i=1}^{n}e_i^{(1)}\mathbb{1}_{e_i^{(1)}<0} - \sum_{i=1}^{n}e_i^{(1)}\mathbb{1}_{e_i^{(1)}>0}\big)}. \tag{10}$$

步骤 *2.* 我们将把 $\widehat{\tau} = \widehat{\tau}^{(1)}$ 视为 $\tau$ 的估计值。

## 5.1 急性水生毒性

我们考虑以下数据： *http://archive.ics.uci.edu/dataset/505/qsar+aquatic+toxic* 性。该数据集包含 546 个观测值，用于预测大型蚤的急性水生毒性，以 LC50 表示。分子描述符的八个连续解释变量为：总极性表面积 TPSA(Tot)、氢键受体原子数 SAacc、以原子为中心的片段 H-050、Moriguchi log P 分配系数 MLOGP、连接性指数 连接性指数 RDCHI、二维自相关 GATS1p、构造指数 nN 和原子中心片段C-040。数据的完整说明见 Cassotti 等人（2014年）的论文。数据被划分为训练集（20%）和验证集（80%）。使用训练集，通过三种参数估计方法（期望分位数、最小二乘法和中位数）进行交叉验证（4），以执行变量选择。表 7 汇总的结果显示，期望分位数模型和最小二乘模型保留了变量 TPSA(Tot)、SAacc、MLOGP、RDCHI、GATS1p、nN，而分位数模型选择了 TPSA(Tot)、SAacc、RDCHI、GATS1p、nN。通过公式（10）计算的期望分位数指数 $\tau$ 估计值为 0.41，表明残差不对称。请注意，三种估计方法所选模型的残差均不服从正态分布。每种方法的 Shapiro-Wilk 检验 p 值为

| 方法 | TPSA(Tot) | SAacc | H-050 | MLOGP | RDCHI | GATS1p | nN | C-040 |
|---|---|---|---|---|---|---|---|---|
| 期望分位数 | 0.0277 | -0.0141 | 0 | 0.5301 | 0.3848 | -0.5353 | -0.2107 | 0 |
| 最小二乘法 | 0.0282 | -0.0144 | 0 | 0.5268 | 0.3873 | -0.5235 | -0.1917 | 0 |
| 分位数 | 0.0283 | -0.0108 | 0 | 0.7077 | 0 | -0.5631 | -0.1718 | 0 |

表7：急性水生毒性数据、通过交叉验证期望分位数、交叉验证最小二乘法和交叉验证中位数方法选定的变量及估计系数。

在训练集上小于 $10^{-4}$，在验证集上小于 $10^{-7}$（交叉验证期望分位数方法见图 2）。

现在，让我们比较通过交叉验证选定的中位数模型和最小二乘模型与基于整个 $n$ 观测值数据库构建的两个相应模型。这些后者模型得到的结果如表8所示。需要注意的是，R SALES 软件包不提供关于系数期望估计值的任何假设检验。因此，本文所提出方法的另一个重要方面是，它能够直接获得非零系数和零系数的估计值。比较表7和表8可以发现，尽管该变量在整个数据集中具有显著性，但中位数交叉验证方法并未保留变量RDCHI 。交叉验证期望分位数方法正确识别出了显著变量，而交叉验证最小二乘法无法使用，因为残差不服从高斯分布。

| **解释变量** | **LS** | | | | **中位数** | | | |
|---|---|---|---|---|---|---|---|---|
| **变量** | 估计值 | 标准误 | t值 | Pr(> $\|t\|$) | 估计值 | 标准误 | Z值 | p 值 |
| 截距 | 2.70 | 0.24 | 11.01 | $< 2 \times 10^{-16}$ | 2.84 | 0.18 | 15.54 | 0.000 |
| 总拓扑极性表面积 | 0.02 | 0.0027 | 10.21 | $< 2 \times 10^{-16}$ | 0.02 | 0.002 | 14.4 | 0.000 |
| SAacc | -0.015 | 0.002 | -7.19 | $2 \times 10^{-12}$ | -0.016 | 0.001 | -10.72 | 0.000 |
| H_050 | 0.04 | 0.06 | 0.67 | 0.50 | 0.087 | 0.04 | 1.95 | 0.05 |
| MLOGP | 0.44 | 0.06 | 7.04 | $5 \times 10^{-12}$ | 0.45 | 0.047 | 9.71 | 0.000 |
| RDCHI | 0.51 | 0.13 | 3.770 | 0.0001 | 0.47 | 0.10 | 4.71 | 0.000 |
| GATS1p | -0.57 | 0.15 | -3.7 | 0.0002 | -0.71 | 0.11 | -6.23 | 0.000 |
| nN | -0.22 | 0.048 | -4.65 | $4 \times 10^{-6}$ | -0.27 | 0.036 | -7.56 | 0.000 |
| C_040 | 0.003 | 0.078 | 0.043 | 0.96 | -0.027 | 0.058 | -0.468 | 0.64 |

表8：完整数据上采用最小二乘法和中位数回归得到的急性水生毒性数据估计值、标准误、检验统计量值和p值。最小二乘法残差和中位数残差不服从正态分布。（夏皮罗-威尔克p值分别为 $1.06 \times 10^{-8}$ 和 $3.39 \times 10^{-10}$）。

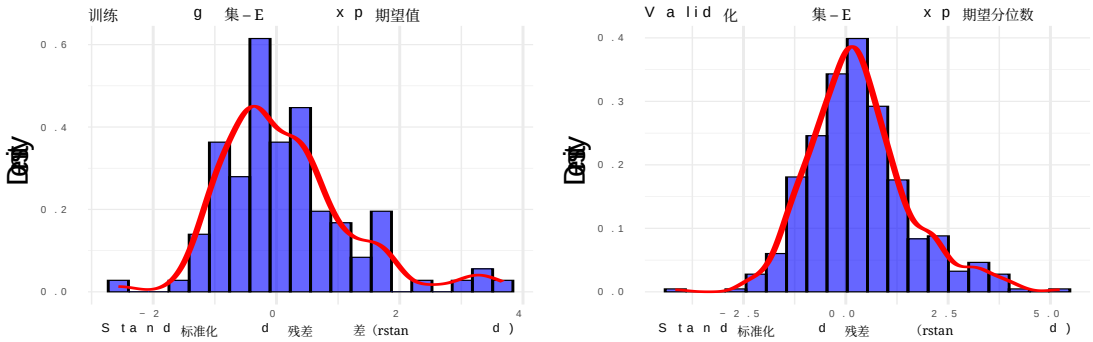



图 2：急性水生毒性数据：训练集和验证集上与交叉验证期望分位数方法对应的标准化残差。

## 5.2 能源消耗

我们考虑以下数据：*https://archive.ics.uci.edu/dataset/851/steel+*工业*+*能源*+*消耗。该数据集包含35040个观测值，用于预测钢铁工业能源消耗，以 `Usage kWh` 表示。描述生产条件、时间背景和工业环境特征的六个连续解释变量分别为：滞后无功电能 `Lagging kVarh`、预期无功电能 `Leading kVarh`、与生产相关的二氧化碳排放 `CO2`、延迟模式下的功率因数 `Lagging Factor`、提前模式下的功率因数 `Leading Factor` 以及自午夜起的秒数 `NSM`。期望分位数指数的估计值 $\hat{\tau}$ 为 0.11，这表明模型误差具有高度不对称性。

| 方法 | 滞后千乏时 | 超前千乏时 _ | CO2 | 滞后功率因数 | 超前功率因数 | NSM |
|---|---|---|---|---|---|---|
| 交叉验证期望分位数 | 0.1613 | -0.206 | 1749.074 | 0 | -0.0678 | 0 |
| 交叉验证最小二乘 | 0.4001 | 0.1675 | 1560.351 | 0.1656 | 0.0992 | 0 |
| 交叉验证中位数 | 0.279 | 0.1769 | 1748.261 | 0.0655 | 0.0535 | 0 |

表10：能源消耗数据：交叉验证期望分位数、交叉验证最小二乘法和交叉验证中位数方法选定的变量及估计系数。

| 解释变量 | 期望分位数 | LS | | | | 中位数 | | | |
|---|---|---|---|---|---|---|---|---|---|
| 变量 | 估计值 | 估计值 | 标准误 | t值 | Pr(>\|t\|) | 估计值 | 标准误 | Z值 | p 值 |
| 截距 | 12.57 | -13.46 | 0.36 | -37.40 | $< 2 \times 10^{-16}$ | -7.16 | 0.052 | -136.04 | 0.000 |
| 滞后千乏时 | 0.141 | 0.30 | 0.0038 | 78.050 | $< 2 \times 10^{-16}$ | 0.277 | 0.00057 | 486.609 | 0.000 |
| 超前千乏时 | -0.263 | 0.13 | 0.01 | 13 | $< 2 \times 10^{-16}$ | 0.17 | 0.001 | 113.91 | 0.000 |
| CO2 | 1768.23 | 1685.0 | 4.86 | 346.53 | $< 2 \times 10^{-16}$ | 1757.28 | 0.71 | 2469.54 | 0.000 |
| 滞后功率因数 | -0.004 | 0.12 | 0.0025 | 49.39 | $< 2 \times 10^{-16}$ | 0.065 | 0.00038 | 175.357 | 0.000 |
| 超前功率因数 | -0.075 | 0.075 | 0.0027 | 28.03 | $< 2 \times 10^{-16}$ | 0.053 | 0.0004 | 136.153 | 0.000 |
| NSM | $7 \times 10^{-10}$ | 0.000009 | 0.000001 | 7.42 | $1.17 \times 10^{-13}$ | 0.0000018 | 0.00000018 | 9.82 | $< 2 \times 10^{-16}$ |

表10：能源消耗数据：使用完整数据集时，期望分位数、最小二乘法和中位数方法的估计值、标准误、t值和p值。

使用交叉验证期望分位数方法时，训练集和验证集上的标准化残差不服从正态分布（见图3的直方图）。交叉验证期望分位数、交叉验证最小二乘和交叉验证中位数方法均未选中变量 NSM，但在整个数据集上，最小二乘法和中位数法均认为该变量具有显著性（见表9和表10）。这可以解释为：最小二乘法得到的残差不服从正态分布，而且使用中位数法得到的相应系数估计值非常接近 0。

对于使用交叉验证期望分位数方法得到的系数估计值与使用交叉验证中位数和交叉验证最小二乘得到的估计值符号相反这一现象，我们给出以下解释。交叉验证最小二乘的估计值并不可靠，因为相应的残差不服从正态分布。使用交叉验证分位数方法得到的估计值对应于中位数回归；在误差高度偏态的情况下，这些估计值可能与使用交叉验证期望分位数得到的估计值不同。

# 6 结论与交叉验证期望分位数方法的优势 s

本文提出并从理论和数值角度研究了采用训练-测试划分法的期望分位数方法。据作者所知，文献中此前尚未研究过这一主题。此外，与通常情况不同，训练集和验证集的特点是验证集远大于训练集。模型参数的期望分位数或自适应期望分位数 LASSO 估计量，是利用从整个数据库中随机选取的训练集计算得到的。随后，通过最小化验证集上的交叉验证均值评分来选择最优模型。我们在三种情形下证明了模型的一致性，这些情形基于解释变量的数量，而该数量可能与观测值数量相关，也可能无关。

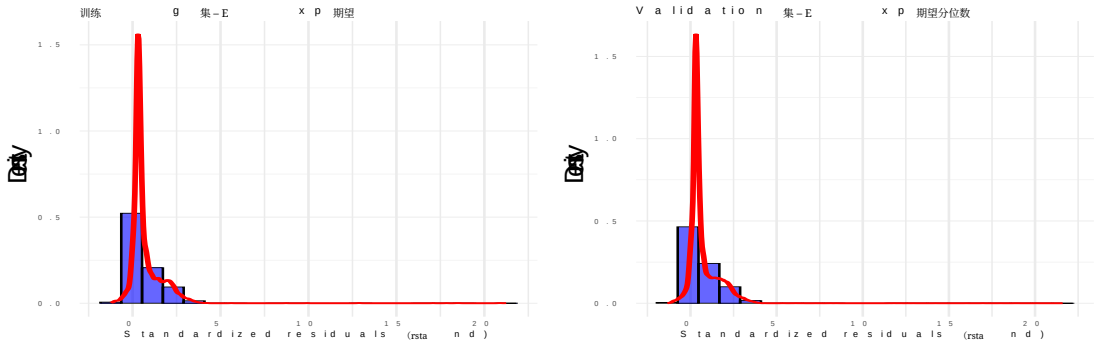



图3：能源消耗数据：训练集和验证集上对应于交叉验证期望分位数方法的标准化残差。

交叉验证期望分位数模型估计量的一致性意味着，我们可以直接识别真实模型中系数为零的变量，并估计系数非零的变量；随着观测值数量趋于无穷，这一结果成立的概率趋于 1。因此，这种模型选择方法适用于大型模型，因为它无需进行假设检验，就能直接识别不显著变量。基于所有观测值的自适应 LASSO 期望分位数方法也会产生直接收缩至零的估计量。不过，在大型验证集上进行交叉验证，能够进一步提高模型估计量的稳健性。在数值分析的语境下，这一结果尤为值得关注，因为 R$ernet$ 软件包中的$SALES$ 函数不提供线性模型参数的假设检验。此外，当模型误差呈非对称分布时，适合采用交叉验证期望分位数方法；此时正态性这一经典假设不再适用，因此经典最小二乘法也不再合适。上一节中的模拟结果表明，与同样适用于非对称误差情形的交叉验证分位数方法相比，交叉验证期望分位数方法更准确、运行时间更短，尤其是在解释变量数量较多的模型中。模拟还揭示了另一个非常重要的方面：当 $\beta_j^*$ 的非零值接近零时，与交叉验证最小二乘、交叉验证分位数和自适应 LASSO 方法不同，交叉验证期望分位数估计量能够很好地识别无关变量。

# 7 主要结果的证明

本节将给出第 2 节中所述三个定理的证明。每个定理的证明都将通过证明以下两个

关系来确立，适用于任意过度指定模型 $\mathcal{M}_o$ 和任意欠指定模型 $\mathcal{M}_u$:

$$\lim_{n\to\infty} \mathbb{P}[CVS(\mathcal{M}_o) > CVS(\mathcal{M}^*)] = 1 \tag{11}$$

and

$$\lim_{n\to\infty} \mathbb{P}[CVS(\mathcal{M}_u) > CVS(\mathcal{M}^*)] = 1. \tag{12}$$

为证明 (11)，我们将把 $CVS(\mathcal{M}_o) - CVS(\mathcal{M}^*)$ 分解为:

$$\begin{aligned} CVS(\mathcal{M}_o) - CVS(\mathcal{M}^*) &= \frac{1}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} \rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{M}_o}^\top \widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}_o}) - \frac{1}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} \rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{M}^*}^\top \widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}^*}) \\ &= \frac{1}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} \Big( \rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{M}_o}^\top \widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}_o}) - \rho_\tau(\varepsilon_i) \Big) \\ &\quad - \frac{1}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} \Big( \rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{M}^*}^\top \widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}^*}) - \rho_\tau(\varepsilon_i) \Big) \\ &\equiv A_n(\mathcal{M}_o) - B_n(\mathcal{M}^*). \end{aligned} \tag{13}$$

为证明关系 (12)，我们将研究以下分解:

$$CVS(\mathcal{M}_u) - CVS(\mathcal{M}^*) \equiv A_n(\mathcal{M}_u) - B_n(\mathcal{M}^*). \tag{14}$$

为了研究 $A_n(\mathcal{M}_o)$，我们将其写成如下形式:

$$\begin{aligned} A_n(\mathcal{M}_o) &= \frac{1}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} \big( \rho_\tau(\varepsilon_i - \mathbf{X}_{i,\mathcal{M}_o}^\top \frac{\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_o}}{r_{n_{\mathcal{A}}}}) - \rho_\tau(\varepsilon_i) \big) \\ &= \Bigg( \frac{1}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} \big( \rho_\tau(\varepsilon_i - \mathbf{X}_{i,\mathcal{M}_o}^\top \frac{\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_o}}{r_{n_{\mathcal{A}}}}) - \rho_\tau(\varepsilon_i) \big) \\ &\quad - \frac{1}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} \mathbb{E}\big[ \rho_\tau(\varepsilon_i - \mathbf{X}_{i,\mathcal{M}_o}^\top \frac{\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_o}}{r_{n_{\mathcal{A}}}}) - \rho_\tau(\varepsilon_i) \big] \Bigg) \\ &\quad + \frac{1}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} \mathbb{E}\big[ \rho_\tau(\varepsilon_i - \mathbf{X}_{i,\mathcal{M}_o}^\top \frac{\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_o}}{r_{n_{\mathcal{A}}}}) - \rho_\tau(\varepsilon_i) \big] \equiv \Delta_1 + \Delta_2. \end{aligned} \tag{15}$$

**定理 3.1 的证明**。
我们首先证明关系 (11)。
根据泰勒展开，对于 $t \to 0$: $\mathbb{E}[\rho_\tau(\varepsilon - t)] - \rho_\tau(\varepsilon)] = -\mathbb{E}[g_\tau(\varepsilon)]t + 2^{-1}\mathbb{E}[h_\tau(\varepsilon)]t^2 + o(t^2)$
$= 2^{-1}\mathbb{E}[h_\tau(\varepsilon)]t^2 + o(t^2)$。另一方面，根据假设（A1），由此我们知道 $(\varepsilon_i)_{i\in\mathcal{A}}$ 与 $(\varepsilon_i)_{i\in\mathcal{V}}$ 相互独立，因此可推得

$\widehat{}(\varepsilon_i)_{i\in\mathcal{V}}$ 与 $\boldsymbol{\delta}_{n_\mathcal{A},\,\mathcal{M}_o}$ 相互独立。然后，考虑到假设（A2）$\widehat{}$以及 $\boldsymbol{\beta}_{n_\mathcal{A},\,\mathcal{M}_o}$ 的收敛速率，我们得到：

$$\Delta_2 = \frac{1}{2n_\mathcal{V}}\sum_{i\in\mathcal{V}}\mathbb{E}[h_\tau(\varepsilon)]\frac{\mathbb{E}\big[\big(\mathbf{X}_{i,\mathcal{M}_o}^\top\widehat{\boldsymbol{\delta}}_{n_\mathcal{A},\mathcal{M}_o}\big)^2\big]}{r_{n_\mathcal{A}}^2}\big(1+o_\mathbb{P}(1)\big) = O\bigg(\frac{\big\|\widehat{\boldsymbol{\delta}}_{n_\mathcal{A},\mathcal{M}_o}\big\|_2^2}{r_{n_\mathcal{A}}^2}\bigg) = O(r_{n_\mathcal{A}}^{-2}). \tag{16}$$

为了研究 $\Delta_1$，我们将其写成如下形式：

$$\Delta_1 = \frac{1}{n_\mathcal{V}}\sum_{i\in\mathcal{V}}\bigg(-\frac{\widehat{\boldsymbol{\delta}}_{n_\mathcal{A},\mathcal{M}_o}^\top}{r_{n_\mathcal{A}}}g_\tau(\varepsilon_i)\mathbf{X}_{i,\mathcal{M}_o} + D_{1,i} - \mathbb{E}[D_{1,i}]\bigg), \tag{17}$$

with

$$D_{1,i} \equiv \rho_\tau(\varepsilon_i - \mathbf{X}_{i,\mathcal{M}_o}^\top\frac{\widehat{\boldsymbol{\delta}}_{n_\mathcal{A},\mathcal{M}_o}}{r_{n_\mathcal{A}}}) - \rho_\tau(\varepsilon_i) + \frac{\widehat{\boldsymbol{\delta}}_{n_\mathcal{A},\mathcal{M}_o}^\top}{r_{n_\mathcal{A}}}g_\tau(\varepsilon_i)\mathbf{X}_{i,\mathcal{M}_o}.$$

对于关系（17）右侧的第一项，根据假设 (A1) 和 (A2) 以及 $r_n = n_\mathcal{A}^{1/2}$，由中心极限定理（CLT）可得：

$$\frac{1}{n_\mathcal{V}}\sum_{i\in\mathcal{V}}\frac{\widehat{\boldsymbol{\delta}}_{n_\mathcal{A},\mathcal{M}_o}^\top}{r_{n_\mathcal{A}}}g_\tau(\varepsilon_i)\mathbf{X}_{i,\mathcal{M}_o} = O_\mathbb{P}\bigg(\frac{\|\widehat{\boldsymbol{\delta}}_{n_\mathcal{A},\mathcal{M}_o}\|_2}{n_\mathcal{V}^{1/2}n_\mathcal{A}^{1/2}}\bigg) = O_\mathbb{P}(n_\mathcal{A}^{-1/2}n_\mathcal{V}^{-1/2}). \tag{18}$$

为了分析（17）右侧的最后两项，我们应用 Gu and Zou（2016）的引理1和柯西–施瓦茨不等式。由此，以概率1可得：

$$|D_{1,i}| = \frac{1}{r_{n_\mathcal{A}}^2}\big|\mathbf{X}_{i,\mathcal{M}_o}^\top\widehat{\boldsymbol{\delta}}_{n_\mathcal{A},\mathcal{M}_o}\big|^2V_i \le \|\mathbf{X}_{i,\mathcal{M}_o}\|_2^2\|\widehat{\boldsymbol{\beta}}_{n_\mathcal{A},\mathcal{M}_o} - \boldsymbol{\beta}^*_{\mathcal{M}_o}\|_2^2V_i, \tag{19}$$

其中，$V_i$是一个取值于区间 $[\min(\tau,1-\tau),\ \max(\tau,1-\tau)]$内的随机变量。然后，利用假设 (A1) 和 (A2)，以及随机变量$V_i$ 取值属于区间 (0，1) 这一事实，我们得到：

$$\begin{aligned}
\frac{1}{n_\mathcal{V}}\sum_{i\in\mathcal{V}}\big(D_{1,i}-\mathbb{E}[D_{1,i}]\big) &= \frac{1}{n_\mathcal{V}}O_\mathbb{P}\bigg(\sqrt{\sum_{i\in\mathcal{V}}\mathbb{V}\mathrm{ar}\,[D_{1,i}]}\bigg) \le \frac{1}{n_\mathcal{V}}O_\mathbb{P}\bigg(\sqrt{\sum_{i\in\mathcal{V}}\mathbb{E}[D_{1,i}^2]}\bigg)\\
&\le \frac{1}{n_\mathcal{V}}O_\mathbb{P}\bigg(\sqrt{\sum_{i\in\mathcal{V}}\mathbb{E}[\|\mathbf{X}_{i,\mathcal{M}_o}\|_2^2\|\widehat{\boldsymbol{\beta}}_{n_\mathcal{A},\mathcal{M}_o} - \boldsymbol{\beta}^*_{\mathcal{M}_o}\|_2^2V_i]^2}\bigg)\\
&= \frac{1}{n_\mathcal{V}}O_\mathbb{P}\bigg(\sqrt{n_\mathcal{V}\mathbb{E}\big[\|\widehat{\boldsymbol{\beta}}_{n_\mathcal{A},\mathcal{M}_o} - \boldsymbol{\beta}^*_{\mathcal{M}_o}\|_2^4\big]}\bigg) = O_\mathbb{P}\big(n_\mathcal{V}^{-1/2}\mathbb{E}\big[\|\widehat{\boldsymbol{\beta}}_{n_\mathcal{A},\mathcal{M}_o} - \boldsymbol{\beta}^*_{\mathcal{M}_o}\|_2^2\big]\big)\\
&= O_\mathbb{P}(n_\mathcal{V}^{-1/2}n_\mathcal{A}^{-1}).
\end{aligned}$$

将这一关系与关系 (16) 和 (18) 结合起来，我们得到： $\Delta_1 = O_{\mathbb{P}}(n_{\mathcal{V}}^{-1/2} n_{\mathcal{A}}^{-1/2})$。 $\Delta_1$ 的阶结合关系 (16)，意味着： $\Delta_2/\Delta_1 = O_{\mathbb{P}}(n_{\mathcal{V}}^{1/2} n_{\mathcal{A}}^{-1/2})$。因此，由假设（A4）可知 $n_{\mathcal{V}} \gg n_{\mathcal{A}}$，我们推得 $\Delta_2 \gg |\Delta_1|$以趋于 1 的概率成立，进而

$$A_n(\mathcal{M}_o) = \Delta_2(1 + o_{\mathbb{P}}(1)) = O_{\mathbb{P}}(n_{\mathcal{A}}^{-1}). \tag{20}$$

现在我们研究 $B_n(\mathcal{M}^*)$。对于任意 $i \in \mathcal{V}$，有：

$$Y_i - \mathbf{X}_{i,\mathcal{M}^*}^\top \widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}^*} = \varepsilon_i - \mathbf{X}_{i,\mathcal{M}^*}^\top \big(\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}^*} - \boldsymbol{\beta}^*_{\mathcal{M}^*}\big) = \varepsilon_i - \mathbf{X}_{i,\mathcal{M}^*}^\top \frac{\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}^*}^\top}{r_{n_{\mathcal{A}}}}.$$

同样，利用 $\mathbb{E}[h_\tau(\varepsilon)] > 0$ 这一事实，我们可以证明：

$$B_n(\mathcal{M}^*) = \frac{1}{n_{\mathcal{V}}} \sum_{i \in \mathcal{V}} \mathbb{E}[h_\tau(\varepsilon_i)] \frac{\big(\mathbf{X}_{i,\mathcal{M}^*}^\top \widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}^*}\big)^2}{r_{n_{\mathcal{A}}}^2} \big(1 + o_{\mathbb{P}}(1)\big) = O_{\mathbb{P}}\left( \frac{\big\|\boldsymbol{\delta}_{n_{\mathcal{A}},\mathcal{M}^*}\big\|_2^2}{r_{n_{\mathcal{A}}}^2} \right) = O_{\mathbb{P}}(n_{\mathcal{A}}^{-1}) > 0. \tag{21}$$

于是，由关系式 (13)、(15)、(20) 和 (21)，可得：

$$\begin{aligned} A_n(\mathcal{M}_o) - B_n(\mathcal{M}^*) = & \Bigg( \frac{1}{n_{\mathcal{V}}} \sum_{i \in \mathcal{V}} \mathbb{E}[h_\tau(\varepsilon_i)] \widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_o} \frac{\mathbf{X}_{i,\mathcal{M}_o}^\top \mathbf{X}_{i,\mathcal{M}_o}}{r_{n_{\mathcal{A}}}^2} \widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_o}^\top \\ & - \frac{1}{n_{\mathcal{V}}} \sum_{i \in \mathcal{V}} \mathbb{E}[h_\tau(\varepsilon_i)] \widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}^*} \frac{\mathbf{X}_{i,\mathcal{M}^*}^\top \mathbf{X}_{i,\mathcal{M}^*}}{r_{n_{\mathcal{A}}}^2} \widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}^*}^\top \Bigg) (1 + o_{\mathbb{P}}(1)). \end{aligned}$$

对于模型 $\mathcal{M}$，令维度为 $|\mathcal{M}|$的方阵为：

$$\mathbb{S}_{\mathcal{M}} \equiv \lim_{n \to \infty} n_{\mathcal{V}}^{-1} \sum_{i \in \mathcal{V}} \mathbb{E}[h_\tau(\varepsilon_i)] \mathbf{X}_{i,\mathcal{M}}^\top \mathbf{X}_{i,\mathcal{M}}.$$

于是，由斯卢茨基引理以及模型误差$(\varepsilon_i)_{i \in \mathcal{V}}$与 $\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\,\mathcal{M}_o}^\top$ 和 $\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\,\mathcal{M}^*}^\top$ 相互独立这一事实，可得：

$$A_n(\mathcal{M}_o) - B_n(\mathcal{M}^*) \xrightarrow[n \to \infty]{\mathcal{L}} \boldsymbol{\delta}_{\mathcal{M}_o}^\top \mathbb{S}_{\mathcal{M}_o} \boldsymbol{\delta}_{\mathcal{M}_o} - \boldsymbol{\delta}_{\mathcal{M}^*}^\top \mathbb{S}_{\mathcal{M}^*} \boldsymbol{\delta}_{\mathcal{M}^*} > 0.$$

极限分布严格为正，因为 $\mathcal{M}^* \subsetneq \mathcal{M}_o$。关系式 (11) 的证明至此完成。

现在我们证明关系 (12)。
根据关系 (21)，我们有 $B_n(\mathcal{M}^*) = O_{\mathbb{P}}(n_{\mathcal{A}}^{-1}) > 0$。关于

$$A_n(\mathcal{M}_u) = \frac{1}{n_{\mathcal{V}}} \sum_{i \in \mathcal{V}} \Big( \rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{M}_u}^\top \widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}_u}) - \rho_\tau(\varepsilon_i) \Big)$$

我们将 $\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\,\mathcal{M}_u}$ 写成 $r_{n_{\mathcal{A}}}^{-1}\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\,\mathcal{M}_u}+\boldsymbol{\beta}^*_{\mathcal{M}_u}$ 的形式，这意味着：

$$A_n(\mathcal{M}_u)=\frac{1}{n_{\mathcal{V}}}\sum_{i\in\mathcal{V}}\Bigg(\rho_\tau(Y_i-r_{n_{\mathcal{A}}}^{-1}\mathbf{X}_{i,\mathcal{M}_u}^\top\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}-\mathbf{X}_{i,\mathcal{M}_u}^\top\boldsymbol{\beta}^*_{\mathcal{M}_u})-\rho_\tau(\varepsilon_i)\Bigg).$$

在这种情况下，对于模型 $\mathcal{M}_u$，我们得到相合估计量 $\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}_u}$，其收敛速率为 $r_{n_{\mathcal{A}}}^{-1}$ 阶。于是，$r_{n_{\mathcal{A}}}^{-1}\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}=O_{\mathbb{P}}(r_{n_{\mathcal{A}}}^{-1})$。
我们将 $\rho_\tau(Y_i-r_{n_{\mathcal{A}}}^{-1}\mathbf{X}_{i,\mathcal{M}_u}^\top\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\,\mathcal{M}_u}-\mathbf{X}_{i,\,\mathcal{M}_u}^\top\boldsymbol{\beta}^*_{\mathcal{M}_u})$，定义为：

$$\rho_\tau(Y_i-r_{n_{\mathcal{A}}}^{-1}\mathbf{X}_{i,\mathcal{M}_u}^\top\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}-\mathbf{X}_{i,\mathcal{M}_u}^\top\boldsymbol{\beta}^*_{\mathcal{M}_u})-\rho_\tau(Y_i-\mathbf{X}_{i,\mathcal{M}_u}^\top\boldsymbol{\beta}^*_{\mathcal{M}_u})+\rho_\tau(Y_i-\mathbf{X}_{i,\mathcal{M}_u}^\top\boldsymbol{\beta}^*_{\mathcal{M}_u}).\tag{22}$$

根据 Gu and Zou（2016）的引理1，对于任意 $i\in\mathcal{V}$，有：

$$\begin{aligned}&\rho_\tau(Y_i-r_{n_{\mathcal{A}}}^{-1}\mathbf{X}_{i,\mathcal{M}_u}^\top\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}-\mathbf{X}_{i,\mathcal{M}_u}^\top\boldsymbol{\beta}^*_{\mathcal{M}_u})-\rho_\tau(Y_i-\mathbf{X}_{i,\mathcal{M}_u}^\top\boldsymbol{\beta}^*_{\mathcal{M}_u})\\&=r_{n_{\mathcal{A}}}^{-1}g_\tau(Y_i-\mathbf{X}_{i,\mathcal{M}_u}^\top\boldsymbol{\beta}^*_{\mathcal{M}_u})\mathbf{X}_{i,\mathcal{M}_u}^\top\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}+O_{\mathbb{P}}\big(r_{n_{\mathcal{A}}}^{-1}\mathbf{X}_{i,\mathcal{M}_u}^\top\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}\big)^2,\end{aligned}\tag{23}$$

对于该情形，应用与（19）相同的推理可得$O_{\mathbb{P}}\big(r_{n_{\mathcal{A}}}^{-1}\mathbf{X}_{i,\mathcal{M}_u}^\top\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}\big)^2=O_{\mathbb{P}}(r_{n_{\mathcal{A}}}^{-2})=O_{\mathbb{P}}(n_{\mathcal{A}}^{-1})$。

我们研究 $n_{\mathcal{V}}^{-1}\sum_{i\in\mathcal{V}}r_{n_{\mathcal{A}}}^{-1}g_\tau(Y_i-\mathbf{X}_{i,\mathcal{M}_u}^\top\boldsymbol{\beta}^*_{\mathcal{M}_u})\mathbf{X}_{i,\mathcal{M}_u}^\top\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}$。由于 $\mathcal{M}^*\not\subseteq\mathcal{M}_u$，因此 $Y_i=\mathbf{X}_{i,\mathcal{M}_u}^\top\boldsymbol{\beta}^*_{\mathcal{M}_u}+\mathbf{X}_{i,\mathcal{M}^*\setminus\mathcal{M}_u}^\top\boldsymbol{\beta}^*_{\mathcal{M}^*\setminus\mathcal{M}_u}+\varepsilon_i$，这意味着：

$$\begin{aligned}&n_{\mathcal{V}}^{-1}\sum_{i\in\mathcal{V}}r_{n_{\mathcal{A}}}^{-1}g_\tau(Y_i-\mathbf{X}_{i,\mathcal{M}_u}^\top\boldsymbol{\beta}^*_{\mathcal{M}_u})\mathbf{X}_{i,\mathcal{M}_u}^\top\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}\\&=n_{\mathcal{V}}^{-1}\sum_{i\in\mathcal{V}}r_{n_{\mathcal{A}}}^{-1}\Bigg(g_\tau(\varepsilon_i+\mathbf{X}_{i,\mathcal{M}^*\setminus\mathcal{M}_u}^\top\boldsymbol{\beta}^*_{\mathcal{M}^*\setminus\mathcal{M}_u})-g_\tau(\varepsilon_i)\Bigg)\mathbf{X}_{i,\mathcal{M}_u}^\top\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}\\&\quad+n_{\mathcal{V}}^{-1}\sum_{i\in\mathcal{V}}r_{n_{\mathcal{A}}}^{-1}g_\tau(\varepsilon_i)\mathbf{X}_{i,\mathcal{M}_u}^\top\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}\end{aligned}\tag{24}$$

关系（24）的最后一项根据中心极限定理为 $O_{\mathbb{P}}$（$n_{\mathcal{A}}^{-1/2}n_{\mathcal{V}}^{-1/2}$）。
现在我们使用 Gu 和 Zou（2016）的引理 2，并可写为：

$$\begin{aligned}&n_{\mathcal{V}}^{-1}\sum_{i\in\mathcal{V}}r_{n_{\mathcal{A}}}^{-1}\Bigg(g_\tau(\varepsilon_i+\mathbf{X}_{i,\mathcal{M}^*\setminus\mathcal{M}_u}^\top\boldsymbol{\beta}^*_{\mathcal{M}^*\setminus\mathcal{M}_u})-g_\tau(\varepsilon_i)\Bigg)\mathbf{X}_{i,\mathcal{M}_u}^\top\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}\\&=n_{\mathcal{V}}^{-1}\sum_{i\in\mathcal{V}}r_{n_{\mathcal{A}}}^{-1}\mathbf{X}_{i,\mathcal{M}^*\setminus\mathcal{M}_u}^\top\boldsymbol{\beta}^*_{\mathcal{M}^*\setminus\mathcal{M}_u}\mathbf{X}_{i,\mathcal{M}_u}^\top\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}\widetilde{V}_i\\&=O_{\mathbb{P}}(\|r_{n_{\mathcal{A}}}^{-1}\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}\|_2)=O_{\mathbb{P}}(r_{n_{\mathcal{A}}}^{-1})=O_{\mathbb{P}}(n_{\mathcal{A}}^{-1/2}),\end{aligned}$$

~其中， $V_i$是一个取值于区间 $[-2\max(\tau,1-\tau), 2\max(\tau,1-\tau)]$的随机变量。

考虑关系 (22)、(23)、(24)，可得 $A_n(\mathcal{M}_u)$ 变为

$$A_n(\mathcal{M}_u) = \frac{1}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} \big(\rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{M}_u}^\top \boldsymbol{\beta}^*_{\mathcal{M}_u}) - \rho_\tau(\varepsilon_i)\big) + O_{\mathbb{P}}(n_{\mathcal{A}}^{-1/2}).$$

然而，对于任意 $i \in \mathcal{V}$，我们可以写成 $\varepsilon_i = Y_i - \mathbf{X}_i^\top \boldsymbol{\beta}^*_{\mathcal{M}^*} = Y_i - \mathbf{X}_{i,\mathcal{M}_u}^\top \boldsymbol{\beta}^*_{\mathcal{M}_u} - \mathbf{X}_{i,\mathcal{M}^*\setminus\mathcal{M}_u}^\top \boldsymbol{\beta}^*_{\mathcal{M}^*\setminus\mathcal{M}_u}$ 并且我们依赖假设（A3）。
于是，当 $n$ 足够大时，可得 $CVS(\mathcal{M}_u) - CVS(\mathcal{M}^*) > 0$，由此可推出关系 (12)。■

**定理 3.2 的证明。**
我们首先利用分解 $CVS(\mathcal{M}_o) - CVS(\mathcal{M}^*) \equiv A_n(\mathcal{M}_o) - B_n(\mathcal{M}^*)$，建立关系 (11)。

我们首先研究 $A_n(\mathcal{M}_o) \equiv \Delta_1 + \Delta_2$。对于 $\Delta_2$，按照与定理 3.1 证明中类似的推理，并利用假设（A2），我们得到：

$$\Delta_2 = \frac{1}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} \mathbb{E}\big[\rho_\tau(\varepsilon_i - \mathbf{X}_{i,\mathcal{M}_o}^\top \frac{\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_o}}{r_{n_{\mathcal{A}}}}) - \rho_\tau(\varepsilon_i)\big] = O_{\mathbb{P}}\left(\frac{\|\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_o}\|_2^2}{r_{n_{\mathcal{A}}}^2}\right) = O_{\mathbb{P}}(r_{n_{\mathcal{A}}}^{-2}) = O_{\mathbb{P}}\left(\frac{p_n}{n_{\mathcal{A}}}\right).$$

对于最后一个关系，我们结合 Ciuperca（2021）的定理 2.1 以及假设（A1）、（A2）、（A5）得到了该结论。
项 $\Delta_1 \equiv n_{\mathcal{V}}^{-1} \sum_{i\in\mathcal{V}} \big(\rho_\tau(\varepsilon_i - \mathbf{X}_{i,\mathcal{M}_o}^\top r_{n_{\mathcal{A}}}^{-1} \widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_o}) - \rho_\tau(\varepsilon_i)\big) - n_{\mathcal{V}}^{-1} \sum_{i\in\mathcal{V}} \mathbb{E}\big[\rho_\tau(\varepsilon_i - \mathbf{X}_{i,\mathcal{M}_o}^\top r_{n_{\mathcal{A}}}^{-1} \widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_o}) - \rho_\tau(\varepsilon_i)\big]$ 可以写成式（17）的形式。
利用假设（A1），我们得到：$\mathbb{E}[g_\tau(\varepsilon_i)\mathbf{X}_{i,\mathcal{M}_o}^\top \widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_o}] = \mathbf{0}_{p_n}$；利用假设（A2），我们有：$\mathbb{V}\mathrm{ar}\,[\sum_{i\in\mathcal{V}} g_\tau(\varepsilon_i)\mathbf{X}_{i,\mathcal{M}_o}^\top \widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_o}] = O_{\mathbb{P}}(n_{\mathcal{V}}\|\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_o}\|_2^2)$。然后，利用中心极限定理，我们得到：

$$\frac{1}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} \frac{\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_o}^\top}{r_{n_{\mathcal{A}}}} g_\tau(\varepsilon_i)\mathbf{X}_{i,\mathcal{M}_o} = O_{\mathbb{P}}(n_{\mathcal{V}}^{-1/2}\|\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}_o} - \boldsymbol{\beta}^*_{\mathcal{M}_o}\|_2) = O_{\mathbb{P}}(n_{\mathcal{V}}^{-1/2} r_{n_{\mathcal{A}}}^{-1}). \quad (25)$$

一方面，我们有：$n_{\mathcal{V}}^{-1} \sum_{i\in\mathcal{V}} \big(D_{1,i} - \mathbb{E}[D_{1,i}]\big) \le n_{\mathcal{V}}^{-1} O_{\mathbb{P}}\left(\sqrt{\sum_{i\in\mathcal{V}} \mathbb{E}[D_{1,i}^2]}\right)$。

此外，利用 Gu 和 Zou（2016）的引理1，我们得到：
$D_{1,i} = \frac{1}{r_{n_{\mathcal{A}}}^2} \big|\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_o}^\top \mathbf{X}_{i,\mathcal{M}_o}\big|^2 V_i$，其中随机变量 $V_i \in [\min(\tau, 1-\tau), \max(\tau, 1-\tau)]$
$\lesssim \|\mathbf{X}_{i,\mathcal{M}_o}\|_2^2 \|\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}_o} - \boldsymbol{\beta}^*_{\mathcal{M}_o}\|_2^2$，其概率为1，$\le r_{n_{\mathcal{A}}}^{-2} \max_{i\in\mathcal{V}} \|\mathbf{X}_{i,\mathcal{M}_o}\|_2^2$.

问题源于 $\big|\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\ \mathcal{M}_o}^{\top}\mathbf{X}_{i,\ \mathcal{M}_o}\big|^4$ 中的四次方，这使得在考虑总和 $\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\ \mathcal{M}_o}^{\top}\sum_{i\in\mathcal{V}}\mathbf{X}_{i,\ \mathcal{M}_o}\mathbf{X}_{i\ ,\ \mathcal{M}_o}^{\top}\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\ \mathcal{M}_o}$ 时无法应用假设（A2）。因此，

$$\frac{1}{n_{\mathcal{V}}}\sqrt{\sum_{i\in\mathcal{V}}\mathbb{E}[D_{1,i}^2]}\leq\frac{1}{n_{\mathcal{V}}}\sqrt{r_{n_{\mathcal{A}}}^{-4}\sum_{i\in\mathcal{V}}\max_{j\in\mathcal{V}}\|\mathbf{X}_j\|_2^4}=r_{n_{\mathcal{A}}}^{-2}\frac{\max_{j\in\mathcal{V}}\|\mathbf{X}_j\|_2^2}{n_{\mathcal{V}}^{1/2}},$$

由此

$$\frac{1}{n_{\mathcal{V}}}\sum_{i\in\mathcal{V}}\big(D_{1,i}-\mathbb{E}[D_{1,i}]\big)\leq O_{\mathbb{P}}\left(r_{n_{\mathcal{A}}}^{-2}\frac{\max_{j\in\mathcal{V}}\|\mathbf{X}_j\|_2^2}{n_{\mathcal{V}}^{1/2}}\right).\tag{26}$$

因此，考虑到关系（25）和（26），对于 $\Delta_1$，我们有：

$$\Delta_1=O_{\mathbb{P}}\left(\max\left(r_{n_{\mathcal{A}}}^{-2}\frac{\max_{j\in\mathcal{V}}\|\mathbf{X}_j\|_2^2}{n_{\mathcal{V}}^{1/2}},n_{\mathcal{V}}^{-1/2}r_{n_{\mathcal{A}}}^{-1}\right)\right).$$

我们比较 $\Delta_1$ 和 $\Delta_2$。利用假设（A6），可得：

$$r_{n_{\mathcal{A}}}^{-2}\frac{\max_{j\in\mathcal{V}}\|\mathbf{X}_j\|_2^2}{n_{\mathcal{V}}^{1/2}}r_{n_{\mathcal{A}}}^2=\frac{\max_{j\in\mathcal{V}}\|\mathbf{X}_j\|_2^2}{n_{\mathcal{V}}^{1/2}}\underset{n\to\infty}{\longrightarrow}0.$$

根据假设（A4）中的 $n_{\mathcal{A}}\ll n_{\mathcal{V}}$ 这一事实，可得

$$n_{\mathcal{V}}^{-1/2}r_{n_{\mathcal{A}}}^{-1}r_{n_{\mathcal{A}}}^2=n_{\mathcal{V}}^{-1/2}r_{n_{\mathcal{A}}}=\left(\frac{n_{\mathcal{A}}}{p_n}\right)^{1/2}\frac{1}{n_{\mathcal{V}}^{1/2}}\ll\frac{n_{\mathcal{V}}^{1/2}}{p_n^{1/2}}\frac{1}{n_{\mathcal{V}}^{1/2}}\underset{n\to\infty}{\longrightarrow}0.$$

于是我们得到 $\Delta_2\gg\Delta_1$，这意味着

$$A_n(\mathcal{M}_o)=\Delta_2\big(1+o_{\mathbb{P}}(1)\big)=O_{\mathbb{P}}\big(r_{n_{\mathcal{A}}}^{-2}\big).$$

现在我们研究 $B_n(\mathcal{M}^*)$。与定理 3.1 的证明中关系（21）的处理类似，利用假设（A2），我们可以写出：

$$B_n(\mathcal{M}^*)=O_{\mathbb{P}}\left(\frac{\|\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}^*}\|_2^2}{r_{n_{\mathcal{A}}}^2}\right)=O_{\mathbb{P}}\big(\|\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}^*}-\boldsymbol{\beta}^*_{\mathcal{M}^*}\|_2^2\big)=O_{\mathbb{P}}\big(r_{n_{\mathcal{A}}}^{-2}\big).\tag{27}$$

结论：

$$r_{n_{\mathcal{A}}}^2\big(A_n(\mathcal{M}_o)-B_n(\mathcal{M}^*)\big)-\frac{1}{n_{\mathcal{V}}}\sum_{i\in\mathcal{V}}\mathbb{E}[h_\tau(\varepsilon_i)]\Big(\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_o}\mathbf{X}_{i,\mathcal{M}_o}^{\top}\mathbf{X}_{i,\mathcal{M}_o}\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_o}^{\top}$$
$$-\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}^*}\mathbf{X}_{i,\mathcal{M}^*}^{\top}\mathbf{X}_{i,\mathcal{M}^*}\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}^*}^{\top}\Big)\underset{n\to\infty}{\overset{\mathbb{P}}{\longrightarrow}}>0,\qquad\text{since }\mathcal{M}^*\not\subseteq\mathcal{M}_o$$

and relation $^{(11)\ \mathrm{i}}$ 的证明d

。现在我们证明关系（12），为此，我们结合（15）研究分解（14）。由关系（27）可得 $B_n(\mathcal{M}^*) = O_{\mathbb{P}}(r_{n_{\mathcal{A}}}^{-2})$。

为研究 $A_n(\mathcal{M}_u)$，我们考虑分解 $\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\ \mathcal{M}_u} = \big(\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\ \mathcal{M}_u} - \boldsymbol{\beta}^*_{\mathcal{M}_u}\big) + \boldsymbol{\beta}^*_{\mathcal{M}_u}$。因此，我们得到：

$$CVS(\mathcal{M}_u) = \frac{1}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} \left( \rho_\tau(Y_i - \mathbf{X}^\top_{i,\mathcal{M}_u} \frac{\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}}{r_{n_{\mathcal{A}}}} - \mathbf{X}^\top_{i,\mathcal{M}_u} \boldsymbol{\beta}^*_{\mathcal{M}_u}) - \rho_\tau(\varepsilon_i) \right).$$

我们写作：

$$\rho_\tau(Y_i - \mathbf{X}^\top_{i,\mathcal{M}_u} \frac{\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}}{r_{n_{\mathcal{A}}}} - \mathbf{X}^\top_{i,\mathcal{M}_u} \boldsymbol{\beta}^*_{\mathcal{M}_u})$$

$$= \rho_\tau(Y_i - \mathbf{X}^\top_{i,\mathcal{M}_u} \frac{\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}}{r_{n_{\mathcal{A}}}} - \mathbf{X}^\top_{i,\mathcal{M}_u} \boldsymbol{\beta}^*_{\mathcal{M}_u}) - \rho_\tau(Y_i - \mathbf{X}^\top_{i,\mathcal{M}_u} \boldsymbol{\beta}^*_{\mathcal{M}_u}) + \rho_\tau(Y_i - \mathbf{X}^\top_{i,\mathcal{M}_u} \boldsymbol{\beta}^*_{\mathcal{M}_u})$$

因此，按照定理 3.1 的证明，我们研究：

$$\frac{1}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} \left( \rho_\tau(Y_i - \mathbf{X}^\top_{i,\mathcal{M}_u} \frac{\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}}{r_{n_{\mathcal{A}}}} - \mathbf{X}^\top_{i,\mathcal{M}_u} \boldsymbol{\beta}^*_{\mathcal{M}_u}) - \rho_\tau(Y_i - \mathbf{X}^\top_{i,\mathcal{M}_u} \boldsymbol{\beta}^*_{\mathcal{M}_u}) \right).$$

在假设（A1）、（A2）、（A5）下，$\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\ \mathcal{M}_u}$ 向 $\boldsymbol{\beta}^*_{\mathcal{M}_u}$ 的收敛速率是 $r_{n_{\mathcal{A}}}^{-1}$，即：$\|\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}_u} - \boldsymbol{\beta}^*_{\mathcal{M}_u}\|_2 = O_{\mathbb{P}}(r_{n_{\mathcal{A}}}^{-1})$。那么，对于关系 (23)，我们有：

$$O_{\mathbb{P}}\big(r_{n_{\mathcal{A}}}^{-1} \big| \mathbf{X}^\top_{i,\mathcal{M}_u} \widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u} \big|\big) \le O_{\mathbb{P}}\big(r_{n_{\mathcal{A}}}^{-1} \|\mathbf{X}_{i,\mathcal{M}_u}\|_2 \|\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}\|_2\big) \le O_{\mathbb{P}}\big(r_{n_{\mathcal{A}}}^{-1} \max_{1\le i\le n} \|\mathbf{X}_{i,\mathcal{M}_u}\|_2\big) = o_{\mathbb{P}}(1).$$

对于最后一个关系，我们使用了假设 (A5)。

现在我们研究以下关系，根据柯西-施瓦茨不等式和 Gu 和 Zou（2016）的引理 2，可得：

$$\begin{aligned}
&\frac{1}{n_{\mathcal{V}}} \left| \sum_{i\in\mathcal{V}} r_{n_{\mathcal{A}}}^{-1} \left( g_\tau(\varepsilon_i + \mathbf{X}^\top_{i,\mathcal{M}^*\setminus\mathcal{M}_u} \boldsymbol{\beta}^*_{\mathcal{M}^*\setminus\mathcal{M}_u}) - g_\tau(\varepsilon_i) \right) \mathbf{X}^\top_{i,\mathcal{M}_u} \widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u} \right| \\
&\le \frac{1}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} r_{n_{\mathcal{A}}}^{-1} \left| g_\tau(\varepsilon_i + \mathbf{X}^\top_{i,\mathcal{M}^*\setminus\mathcal{M}_u} \boldsymbol{\beta}^*_{\mathcal{M}^*\setminus\mathcal{M}_u}) - g_\tau(\varepsilon_i) \right| \cdot \left| \mathbf{X}^\top_{i,\mathcal{M}_u} \widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u} \right| \\
&\le \frac{2}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} r_{n_{\mathcal{A}}}^{-1} \left| \mathbf{X}^\top_{i,\mathcal{M}^*\setminus\mathcal{M}_u} \boldsymbol{\beta}^*_{\mathcal{M}^*\setminus\mathcal{M}_u} \right| \cdot \left| \mathbf{X}^\top_{i,\mathcal{M}_u} \widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u} \right| \\
&\le \frac{2}{n_{\mathcal{V}}} \sum_{i\in\mathcal{V}} r_{n_{\mathcal{A}}}^{-1} \big\|\mathbf{X}_{i,\mathcal{M}^*\setminus\mathcal{M}_u}\big\|_2 \big\|\boldsymbol{\beta}^*_{\mathcal{M}^*\setminus\mathcal{M}_u}\big\|_2 \big\|\mathbf{X}_{i,\mathcal{M}_u}\big\|_2 \big\|_2 \widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}\big\|_2 \\
&\le 2 r_{n_{\mathcal{A}}}^{-1} \max_{i\in\mathcal{V}} \big\|\mathbf{X}_{i,\mathcal{M}^*}\big\|_2^2 \big\|\boldsymbol{\beta}^*_{\mathcal{M}^*\setminus\mathcal{M}_u}\big\|_2 \big\|_2 \widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}\big\|_2 \\
&= o_{\mathbb{P}}(1). \qquad (28)
\end{aligned}$$

对于最后一个关系，我们使用了假设 (A7)。于是， $A_n(\mathcal{M}_u)$ 变为：

$$A_n(\mathcal{M}_u) = \frac{1}{n_{\mathcal{V}}} \sum_{i \in \mathcal{V}} \left( \rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{M}_u}^\top \boldsymbol{\beta}^*_{\mathcal{M}_u}) - \rho_\tau(\varepsilon_i) \right) + o_{\mathbb{P}}(1).$$

另一方面，我们有恒等式 $\rho_\tau(\varepsilon_i) = \rho_\tau(Y_i - \mathbf{X}_{i,\,\mathcal{M}^*}^\top \boldsymbol{\beta}^*_{\mathcal{M}^*}) = \rho_\tau(Y_i - \mathbf{X}_{i,\,\mathcal{M}_u}^\top \boldsymbol{\beta}^*_{\mathcal{M}_u} - \mathbf{X}_{i,\,\mathcal{M}^* \setminus \mathcal{M}_u}^\top \boldsymbol{\beta}^*_{\mathcal{M}^* \setminus \mathcal{M}_u})$。因此，利用假设（A3），我们得到：

$$A_n(\mathcal{M}_u) = \liminf_{n \to \infty} k_n + o_{\mathbb{P}}(1) > 0.$$

关系（12）得证。 ■

**定理 3.3 的证明。**
我们仅给出与定理 3.2 的证明相关的重要内容。关系（11）的证明。根据 Ciuperca（2021）中引理 3.1 的证明，我们有：

$\mathbb{V}\mathrm{ar}\left[\sum_{i \in \mathcal{V}} g_\tau(\varepsilon_i) \mathbf{X}_{i,\,\mathcal{M}_o}^\top \widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\,\mathcal{M}_o}\right] = O_{\mathbb{P}}\big(n_{\mathcal{V}} \|\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\,\mathcal{M}_o}\|_1^2\big)$。利用假设（A1）、（A2）和（A8），关系（25）变为，

$$\frac{1}{n_{\mathcal{V}}} \sum_{i \in \mathcal{V}} \frac{\mathbf{X}_{i,\mathcal{M}_o}^\top \widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_o}}{r_{n_{\mathcal{A}}}} g_\tau(\varepsilon_i) = O_{\mathbb{P}}(n_{\mathcal{V}}^{-1/2} \|\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}_o} - \boldsymbol{\beta}^*_{\mathcal{M}_o}\|_1) = O_{\mathbb{P}}\big(n_{\mathcal{V}}^{-1/2} a_{n_{\mathcal{A}}}\big). \quad (29)$$

此外，根据假设（A1）、（A2）和（A8），与关系（19）类似，可得：

$$\begin{aligned} D_{1,i}^2 &\le 4 \|\mathbf{X}_{i,\mathcal{M}_o}^\top\|_\infty^4 \|\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}_o} - \boldsymbol{\beta}^*_{\mathcal{M}_o}\|_1^4 \le 4 \max_{i \in \mathcal{V}} \|\mathbf{X}_{i,\mathcal{M}_o}^\top\|_\infty^4 \|\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}_o} - \boldsymbol{\beta}^*_{\mathcal{M}_o}\|_1^4 \\ &= \max_{i \in \mathcal{V}} \|\mathbf{X}_{i,\mathcal{M}_o}^\top\|_\infty^4 O_{\mathbb{P}}(a_n^4). \end{aligned}$$

于是，

$$n_{\mathcal{V}}^{-1} \sqrt{\sum_{i \in \mathcal{V}} \mathbb{E}[D_{1,i}^2]} \le r_{n_{\mathcal{A}}}^{-2} \frac{\max_{j \in \mathcal{V}} \|\mathbf{X}_{j,\mathcal{M}_o}^\top\|_\infty^2}{n_{\mathcal{V}}^{1/2}}.$$

因此，关系（26）变为

$$\frac{1}{n_{\mathcal{V}}} \sum_{i \in \mathcal{V}} \big(D_{1,i} - \mathbb{E}[D_{1,i}]\big) \le \frac{1}{n_{\mathcal{V}}} O_{\mathbb{P}}\left( r_{n_{\mathcal{A}}}^{-2} \frac{\max_{j \in \mathcal{V}} \|\mathbf{X}_{j,\mathcal{M}_o}^\top\|_\infty^2}{n_{\mathcal{V}}^{1/2}} \right). \quad (30)$$

考虑到关系 (29) 和 (30)，我们有：

$$\Delta_1 = \left( -\frac{1}{n_{\mathcal{V}}} \sum_{i \in \mathcal{V}} \frac{\mathbf{X}_{i,\mathcal{M}_o}^\top \widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_o}}{r_{n_{\mathcal{A}}}} g_\tau(\varepsilon_i) \right) \big(1 + o_{\mathbb{P}}(1)\big) = O_{\mathbb{P}}\left( \max\left( r_{n_{\mathcal{A}}}^{-2} \frac{\max_{j \in \mathcal{V}} \|\mathbf{X}_j\|_\infty^2}{n_{\mathcal{V}}^{1/2}}, n_{\mathcal{V}}^{-1/2} r_{n_{\mathcal{A}}}^{-1} \right) \right)$$

和 $\Delta_2 = O_{\mathbb{P}}\big(r_{n_{\mathcal{A}}}^{-2}\|\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_o}\|_1^2\big) = O_{\mathbb{P}}(r_{n_{\mathcal{A}}}^{-2}) = O_{\mathbb{P}}\big(a_{n_{\mathcal{A}}}^2\big)$。

利用假设 (A6bis)，与情形 $0 < c < 1/2$类似，我们得到 $\Delta_2 \gg \Delta_1$，其成立的概率趋于 1。因此，我们推得 $A_n(\mathcal{M}_o) = \Delta_2\big(1 + o_{\mathbb{P}}(1)\big) = O_{\mathbb{P}}(a_{n_{\mathcal{A}}}^2)$。现在我们研究 $B_n(\mathcal{M}^*) = O_{\mathbb{P}}\big(r_{n_{\mathcal{A}}}^{-2}\|\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}^*}\|_1^2\big) = O_{\mathbb{P}}\big(\|\widehat{\boldsymbol{\beta}}_{n_{\mathcal{A}},\mathcal{M}^*} - \boldsymbol{\beta}^*_{\mathcal{M}^*}\|_1^2\big) = O_{\mathbb{P}}(a_{n_{\mathcal{A}}})$。最后，同样地，我们有

$$\mathbb{P}\big[A_n(\mathcal{M}_o) - B_n(\mathcal{M}^*) > 0\big] \underset{n\to\infty}{\longrightarrow} 1.$$

关系式 (12) 的证明。
我们只给出与定理 3.2 证明相关的关键关系式。因此，我们有：

$$\begin{aligned}
&\frac{1}{n_{\mathcal{V}}}\sum_{i\in\mathcal{V}}\left(\rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{M}_u}^\top \frac{\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}}{r_{n_{\mathcal{A}}}} - \mathbf{X}_{i,\mathcal{M}_u}^\top \boldsymbol{\beta}^*_{\mathcal{M}_u}) - \rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{M}_u}^\top \boldsymbol{\beta}^*_{\mathcal{M}_u})\right) \\
&= \frac{1}{n_{\mathcal{V}}}\sum_{i\in\mathcal{V}} \mathbf{X}_{i,\mathcal{M}^*\setminus\mathcal{M}_u}^\top \boldsymbol{\beta}^*_{\mathcal{M}^*\setminus\mathcal{M}_u} \mathbf{X}_{i,\mathcal{M}_u}^\top \frac{\widehat{\boldsymbol{\delta}}_{n_{\mathcal{A}},\mathcal{M}_u}}{r_{n_{\mathcal{A}}}} \widetilde{V}_i = o_{\mathbb{P}}(1), \qquad (31)
\end{aligned}$$

~此处应用了假设 (A7)，其中 $V_i$ 是一个取值于区间 $[-2\max(\tau, 1-\tau), 2\max(\tau, 1-\tau)]$的随机变量。因此，$A_n(\mathcal{M}_u)$变为：

$$A_n(\mathcal{M}_u) = \frac{1}{n_{\mathcal{V}}}\sum_{i\in\mathcal{V}}\big[\rho_\tau(Y_i - \mathbf{X}_{i,\mathcal{M}_u}^\top \boldsymbol{\beta}^*_{\mathcal{M}_u}) - \rho_\tau(\varepsilon_i)\big] + o_{\mathbb{P}}(1) = \liminf_{n\to\infty} k_n + a_n > 0.$$

■

# 参考文献


Afendras, G., Markatou, M., 2019年. 交叉验证中训练集/测试集规模的最优性与重抽样的有效性. J. Statist. Plann. Inference 199, 286–301.

Algamal, Z. Y., 2020. 岭回归模型的收缩参数选择：一种基于改进交叉验证的方法. Comm. Statist. Simulation Comput. 49(7), 1922–1930.

Arlot, S., 2010. 模型选择的交叉验证程序综述，《统计调查》4, 40–79.

Bates, S., Hastie, T., Tibshirani, R., 2024年。交叉验证：它估计什么，以及估计效果如何？《美国统计学会会刊》119(546)，1434–1445。

Cassotti, M., Ballabio, D., Consonni, V., Mauri, A., Tetko, I.V., Todeschini R., 2014年。使用 GA-kNN 方法预测大型蚤的急性水生毒性，《实验动物替代方法》（ATLA）42，31–41。Ciuperca G., 2021年。具有可能非对称误差的高维线性模型中的变量选择。《计算统计与数据分析》155，论文编号 107112，19 页。Devijver, P. A., Kittler, J., 1982年。《模式识别：统计方法》。Prentice-Hall，伦敦。Fang, F., Yang, Q., Tian, W., 2022年。用于选择最小二乘模型平均中惩罚因子的交叉验证。《经济学快报》217，论文编号 110683，5 页。Gu, Y., Zou, H., 2016年。非对称最小二乘回归的高维推广及其应用。《统计年鉴》44(6)，2661–2694。Hastie, T., Tibshirani, R., & Friedman, J., 2009年。《统计学习基础：数据挖掘、推断与预测》（第2版）。Springer，纽约。Jhung, Y., 2018年。用于模型选择的多重预测 $K$折交叉验证，《非参数统计杂志》30(1)，197–215。Kim, S., Lee, S., Shin, N., Jung, Y., 2025年。通过验证集方法实现线性分位数回归的变量选择一致性。《统计与概率快报》223，论文编号 110431，5 页。Kohavi, R., 1995年。关于用于准确率估计和模型选择的交叉验证与自助法的研究，《第14届国际人工智能联合会议（IJCAI）论文集》，1137–1143。Liao, L., Park, C., Choi, H., 2019年。惩罚期望回归：惩罚分位数回归的替代方法。《统计数学年刊》71(2)，409–438。Newey, W.K., Powell, J.L., 1987年。非对称最小二乘估计与检验。《计量经济学》55(4)，818–847。Rabinowicz, A., Rosset, S., 2022年。相关数据的交叉验证。《美国统计学会会刊》117(538)，718–731。Ripley, B. D., 1996年。《模式识别与神经网络》。剑桥大学出版社，剑桥。

Shao, J., 1993年. 基于交叉验证的线性模型选择, 《美国统计学会会刊》 88(422), 486–494. She, Y., Tran, H., 2019年. 关于稀疏降秩回归的交叉验证. 英国皇家统计学会会刊B辑：统计方法论, 81(1), 145–161. Silva, L.A., Zanella, G., 2024年. 高维贝叶斯模型的稳健留一交叉验证. 《美国统计学会会刊》 119(547), 2369–2381. Stone, M., 1974年. 统计预测的交叉验证选择与评估, 《英国皇家统计学会会刊B辑》36, 111–147. Wang, Q., Cai, X., 2021年. 分区抽样下交叉验证的高效方差估计量. 《统计学》55(3), 660–681. Wei, Y., Wang, Q., 2021年. 响应变量随机缺失时线性模型中基于交叉验证的模型平均. 《统计与概率快报》171, 文章编号108990, 6. Wieczorek, J., Lei, J., 2022年. 前向选择和序贯交叉验证在高维回归中的模型选择性质. 《加拿大统计学杂志》50(2), 454–470. van de Wiel, M.A., van Nee, M.M., Rauschenberger, A., 2021年. 多重惩罚高维岭回归的快速交叉验证. 《计算与图形统计学杂志》30(4), 835–847. Xiong, J., Shang, J., 2021年. 一种通过交叉验证进行混合模型选择的惩罚方法. 《统计学通讯：理论与方法》50(11), 2481–2507. Yates, L.A., Aandahl, Z., Richards, S.A., Brook, B.W., 2023年. 模型选择的交叉验证：生态学实例综述. 《生态学专论》93(1), e1557. Zhang, P., 1993年. 通过多重交叉验证进行模型选择. 《统计年鉴》21(1), 299–313. Zhao, J., Chen, Y., Zhang, Y., 2018年. 用于分析高维异方差性的期望分位数回归. 《统计与概率快报》137, 304–311.