Statistics and Application 统计学与应用, 2024, 13(6), 2193-2203
Published Online December 2024 in Hans. https://www.hanspub.org/journal/sa
https://doi.org/10.12677/sa.2024.136213

Hans汉斯

基于γ散度的稳健小域估计方法

郭瑞新

青海师范大学数学与统计学院，青海 西宁

收稿日期：2024年11月4日；录用日期：2024年11月30日；发布日期：2024年12月5日

摘　要

被称为Fay-Herriot模型的两阶段正态分层模型被广泛应用于获得小范围内基于模型的均值估计。但是，当数据中出现异常值时，经验贝叶斯估计方法的性能可能很差，估计量会受到异常值的过度影响。本文提出了一种利用密度幂散度的修正方法，并提出了一种新的稳健经验贝叶斯小域估计方法，该方法对存在异常值的数据进行估计时有不错的效果。根据模型参数鲁棒估计量的渐近性质，推导了该估计量的均方误差并估计均方误差。通过数值模拟与实证分析，研究了该方法的数值性能，相比于其他估计量在处理异常值方面表现更为稳健。

关键词

Fay-Herriot模型，贝叶斯估计，小域估计，密度幂散度

Robust Small Area Estimation Method Based on γ-Divergence

Ruixin Guo

School of Mathematics and Statistics, Qinghai Normal University, Xining Qinghai

Received: Nov. 4th, 2024; accepted: Nov. 30th, 2024; published: Dec. 5th, 2024

Abstract

The two-stage normal hierarchical model known as the Fay-Herriot model is widely used to obtain model-based mean estimates for small areas. However, when outliers are present in the data, the performance of empirical Bayes estimation methods may be poor, and the estimators can be excessively influenced by outliers. This paper proposes a modification method utilizing density power divergence and introduces a new robust empirical Bayes small area estimation method, which performs well when estimating data with outliers. Based on the asymptotic properties of the robust estimator of model parameters, the mean squared error (MSE) of this estimator is derived and

文章引用: 郭瑞新. 基于 γ 散度的稳健小域估计方法[J]. 统计学与应用, 2024, 13(6): 2193-2203.
DOI: 10.12677/sa.2024.136213

estimated. Through numerical simulations and empirical analyses, the numerical performance of this method is studied, demonstrating its robustness in handling outliers compared to other estimators.

Keywords

Fay-Herriot Model, Bayes Estimation, Small Area Estimation, Density Power Divergence

1. 引言

在本文研究中，通常使用“小域”来表示无法产生足够精度的直接估计的任何领域。通常，领域样本大小倾向于随着领域的总体大小而增加，但情况并非总是如此。实践中，小域估计方法在政府统计、人口统计、医学统计、农业统计、贫困率估计等领域有广泛的应用[1]-[6]。在理论层面，小域估计也受到高度重视，从 Fay 提出的区域水平的 Fay-Herriot 模型[1]、Ghosh 等提出在 FH 模型下的稳健贝叶斯预测方法[2]、Sinha 和 Rao 等提出的稳健经验贝叶斯方法[3]、Jiang 和 Rao 等关于稳健小域估计的概述[4]、Morales 等提出的几类混合模型小区域估计理论[5]，小域估计已经形成了较为完整的理论体系。

为解决小域估计中 FH 模型中估计量受异常值影响的问题，Ghosh 等提出了针对区域水平模型的稳健贝叶斯预测方法，并用于克服由异常值造成的过度收缩问题[2]，但并没有考虑异常值对模型系数的影响。Sinha 和 Rao 利用 Huber-Ψ 函数，通过对由异常值引起的项减小权重的方式，建立了单元水平模型的小域稳健估计量[3]。目前，该方法普遍应用于小域稳健估计。在稳健估计的其他研究方面，Smith 等用稳健投影和分位数的方法研究了商业调查中的小域稳健估计[7]，Bertarelli 等通过偏差修正的方式研究了小域稳健估计问题[8]，Jiang 和 Rao 对稳健小域估计的研究历史进行系统总结后指出其他现代统计方法也可以用于小域估计[4]。

密度幂散度的稳健估计是一种在统计学中用于处理异常值或重尾现象，从而得到更加稳定和可靠的参数估计的方法。对于将密度幂散度与小域估计相结合的相关理论在 Sugasawa 和 Kubokawa 的使用混合模型进行小域估计的综述文章中进行了系统的整理[9]，实现方法与环境构建也得到整理完善[10]。在相关理论研究中，Kim 和 Lee 开发了一种稳健的估计方法，可以最小化散度族的指数多项式以达到简化计算的效果[11]，Basak 和 Basu 通过最小化合适的统计距离对 Bregaman 散度进行拓展优化[12]，Sugasawa 尝试了将散度应用到小域领域来改善区域均值的估计问题[13]，Nakagwa 和 Hashimoto 提出了基于 γ 散度的稳健贝叶斯推断，在处理严重污染的数据时有不错的估计效果[14]。调优参数的选取也至关重要，调优参数的选择会影响目标函数对异常值的敏感度。较大的参数会使得目标函数对异常值更加敏感，而较小的参数则会降低这种敏感度。而对于调优参数的选择 Sugasawa 和 Yonekura 提出一种选择调优参数的准则对于后验分布的选择[15]，Basak 等人以最小距离的选择方法处理调优参数的选择问题，但该方法对于污染严重的数据处理上没有很好的效果[16]，Yonekura 和 Sugasawa 对于所提出的选择准则在针对一般贝叶斯的背景下进行了改善[17]。本文将基于 γ 散度与小域估计中的 FH 模型相结合来改善数据因为样本量不足且异常值存在导致出现对区域均值估计效果不佳的情况，在下文中的数值模拟与实证分析可以证实本文的方法对于改善该问题有不错的效果。

2. 模型及结合散度的贝叶斯预测

2.1. Fay-Herriot 模型

Fay-Herriot 模型是一种区域层次模型，已知，如果特定区域的样本量较小，仅基于特定区域的样本数据的直接调查估计量会产生过大的标准误差。经验贝叶斯方法被广泛用于改进直接估计，通过缩小一些综合估计和借用强度。关于小域估计与 Fay-Herriot 模型的全面概述，请参考 Fay [1]和 Ghosh [2]的文章。

一个基本的区域级模型的两阶段的层次模型，称为 Fay-Herriot 模型，描述为

$$y_i\left|\theta_i \sim N\left(\theta_i, D_i\right),\ \theta_i \sim N\left(x_i^{\mathrm{T}}\beta, A\right)\ (i=1,\cdots,m)\right. \tag{1}$$

(1)式中，y_i 为小域均值 θ_i 的直接估计量，D_i 为抽样方差，假设抽样方差为已知，x_i 和 β 分别为协变量和回归系数的向量，A 为未知方差。设 $f=\left(b^{\mathrm{T}}, A\right)^{\mathrm{T}}$ 为(1)式中的未知参数的向量。由于 $y_i \sim N\left(x_i^{\mathrm{T}}\beta, A+D_i\right)$ 在(1)的模型下，ϕ 可以通过最大化对数边际似然来估计。

$$\log f\left(y;\phi\right)=-\frac{m}{2}\log\left(2\pi\right)-\frac{1}{2}\sum_{i=1}^{m}\log\left(A+D_i\right)-\frac{1}{2}\sum_{i=1}^{m}\frac{\left(y_i-x_i^{\mathrm{T}}\beta\right)^2}{A+D_i} \tag{2}$$

其中 $y=\left(y_i,\cdots,y_m\right)^{\mathrm{T}}$.在平方误差损失条件下，$\theta_i$ 的贝叶斯预测量为：

$$\tilde{\theta}_i\left(y_i;\phi\right)=y_i-\frac{D_i}{A+D_i}\left(y_i-x_i^{\mathrm{T}}\beta\right) \tag{3}$$

θ_i 的经验贝叶斯估计量 $\hat{\theta}_i\left(y_i;\phi\right)=\tilde{\theta}_i\left(y_i;\hat{\phi}_\alpha\right)$。

当 y_i 可以用辅助信息 x_i 很好地解释表达时，经验贝叶斯估计量可以估计出很好的效果。但是，x_i 并不一定在所有区域都可以很好的解释表达直接估计量 y_i；也就是说，在某些地区 y_i 可能离 $x_i^{\mathrm{T}}\beta$ 很远，本文称之为异常观测。对于这类的观测，相应的 θ_i 的可以从与(1)式中与假设分布不同的分布中生成；也就是说，假设的 θ_i 的分布可能是错误的。在本文中考虑了一种异常观测存在的情况，并关注在这种情况下会出现以下两个问题。

问题 1：贝叶斯预测量(3)可能将 y_i 向 $x_i^{\mathrm{T}}\beta$ 过度收缩。

问题 2：基于(2)的估计量 $\hat{\phi}$ 会受到异常观测值的过度影响。

这些问题已经在 Fay [1]和 Ghosh [2]等人的作品中得到了解决，但本文将通过使用密度幂散度扩展了这一主题的知识体系。

本文的中心思想是基于使用边际似然函数(2)的贝叶斯预测量(3)的替代表达式。贝叶斯预测量(3)可以写成：

$$\tilde{\theta}_i\left(y_i;\phi\right)=y_i-D_i\frac{\partial}{\partial y_i}\log f\left(y_i;\phi\right) \tag{4}$$

如果 $y_i\left|\theta_i \sim N\left(\theta_i, D_i\right)\right.$ 上述表达式成立；也就是说，只要边际似然函数 $\log f\left(y_i;\phi\right)$ 在 θ_i 分布不服从如上式(1)中的正态分布时，我们应该改变 ϕ。从(4)式中可以看出，经典经验贝叶斯估计量 θ_i 可以通过最大化边际似然函数的方法然后求导数来确定。因此，本文将用密度幂散度代替对数边际似然的方法，其中包括 Kullback-Leibler 散度作为一种特殊情况。(1)式下的密度幂散度具有封闭形式，而本文的稳健贝叶斯预测量具有更简单的形式。本文还考虑了模型参数的稳健估计，并给出了它们的渐近性质。此外，本文构造了基于参数自举法的稳健经验贝叶斯估计法的均方误差估计量，并证明了其渐近有效性。

已有很多学者将广义似然用于贝叶斯推断，解决了观测值假设分布出现错误的情况，但本文处理的是(1)式中未直接观测到的区域均值 θ_i 的假设分布出现错误的情况。Ghosh 等人[7]在 Fay-Herriot 模型(1)中

提出了一种稳健的贝叶斯预测量，使用 β 的影响函数来处理属性 1，但没有处理属性 2。Sinha 和 Rao [3] 提出利用 Huber 的函数推导出一般线性混合模型中的贝叶斯预测方法和参数估计的方法，从而解决了性质 1 和 2；然而，所得的贝叶斯预测方法在处理属性 1 时具有局限性。

2.2. 与 γ 散度相结合的稳健贝叶斯估计量

对于 γ 散度，在统计学和数据分析中，γ 散度通常用于构建稳健估计方法，以处理非正态性或存在异常值的数据。通过引入 γ 散度和 γ 似然函数，可以构建基于单元水平模型的小域稳健估计方法，从而得到更加稳健的估计结果。

设 $X_1,\cdots,X_n$ 是独立同分布的(*i.i.d.*)该随机变量是从概率密度函数 $g(x)$ 中生成。设 $f(x)$ 为基本的概率密度函数，$\delta(x)$ 为污染概率密度函数。假设 $g(x)$ 是由 $g(x)=(1-\varepsilon)f(x)+\varepsilon\delta(x)$ 给出的污染概率密度函数，其中 ε 是污染的比值。在本文中，γ 散度通常考虑污染密度函数 $\delta(x)$ 主要位于密度函数 $f(x)$ 尾部的情况。换句话说，对于一个异常值 x_0，γ 散度将认为 $f(x_0)\approx 0$。本文考虑参数模型 $f_\theta(x)=f(x;\theta)(\theta\in\Theta)$ 作为一个候选模型，其中 $\Theta\subset\mathbb{R}^P$ 是 $\theta=(\theta_1,\cdots,\theta_n)^{\mathrm{T}}$ 的一个参数空间。

本文假设目标密度包含在候选模型 $\{f_\theta|\theta\in\Theta\}$ 中，也就是说，$f(x)$ 可以用参数 $f(x)=f_{\theta_0}(x)$ 来表示。在下文中，为了简单起见，本文将省略函数的参数。Nakagawa 和 Hashimoto 所研究的关于概率密度 f_θ 和 g 之间的散度如下：

$$\begin{aligned}D_\gamma(g,f_\theta)&=\frac{1}{\gamma(\gamma+1)}\log\int g^{1+\gamma}\mathrm{d}x-\frac{1}{\gamma}\log\int gf_\theta^{\gamma}\mathrm{d}x+\frac{1}{\gamma+1}\log\int f_\theta^{1+\lambda}\mathrm{d}x\\&=-d_\gamma(g,g)+d_\gamma(g,f_\theta)\end{aligned}$$

其中 $\gamma>0$ 是一个关于稳健性的调优参数，而 $d_\gamma(g,f_\theta)$ 被称为 γ-交叉熵。这个散度也被称为“γ-发散”。为了推导出 θ 的最小 γ 散度估计量，本文可以考虑最小化问题：

$$\min_{\theta\in\Theta}d_\gamma(g,f_\theta)=\min_{\theta\in\Theta}\left\{-\frac{1}{\gamma}\log\int gf_\theta^{\gamma}\mathrm{d}x+\frac{1}{\gamma+1}\log\int f_\theta^{1+\gamma}\mathrm{d}x\right\}$$

对于 θ。虽然真实的密度函数 $g(x)$ 是未知的，但本文注意到 γ-交叉熵可以通过对 $d_\gamma(\bar{g},f_\theta)$ 进行经验估计，其中 $\bar{g}$ 是 $x_1,\cdots,x_n$ 的经验概率密度。然后，通过 $\arg\min_{\theta\in\Theta}d_\gamma(\bar{g},f_\theta)$ 得到 θ 的稳健估计量。

现在，本文考虑如下的 γ-交叉熵的单调变换：

$$\tilde{d}_\gamma(g,f_\theta)=-\frac{1}{\gamma}\left\{\exp\left(-\gamma d_\gamma(g,f_\theta)\right)-1\right\}=-\frac{\frac{1}{\gamma}\int gf_\theta^{\gamma}\mathrm{d}x}{\left\{\int f_\theta^{1+\gamma}\mathrm{d}x\right\}^{\gamma/(1+\gamma)}}+\frac{1}{\gamma}\tag{5}$$

注意到，(5)等式右侧的第二项，即“$+1/\gamma$”是下文证明命题 1 中的 $\tilde{d}_\gamma(g,f_\theta)$ 的必要条件。但在本文计算后验分布时，这一项在分母和分子中将被抵消不会对计算造成影响。这种转换在本文中是必要的。然后本文将给出了变换后的 γ-交叉熵的一些性质。

命题 1. 设 g 和 f 是概率密度函数且设 κ_1,κ_2 和 κ 是常数。那么 $\tilde{d}_\gamma(g,f)$ 具有以下性质：

(i) $\tilde{d}_\gamma(\kappa_1 g,\kappa_2 f)=\kappa_1\tilde{d}_\gamma(g,f)+(1-\kappa_1)/\gamma$。

(ii) $\tilde{d}_\gamma(g,f)=\tilde{d}_\gamma(g,g)$，当且仅当 $f=\kappa g$ 时成立。

(iii) $\lim_{\gamma\to 0}\tilde{d}_\gamma(g,f)=-\int g\log f\mathrm{d}x=d_{KL}(g,f)$，其中 $d_{KL}(g,f)$ 是 g 和 f 之间的 Kullback-Leibler 交叉熵。

证明(i)和(ii)的证明在这里省略，因为这两条性质是由 γ-交叉熵的定义直接证明。本文主要对性质(iii)

进行证明，(iii)的证明可以通过泰勒展开式 $f^{\gamma}=1+\gamma\log f+O\left(\gamma^{2}\right)\left(\gamma\to 0\right)$ 来证明。

然后

$$\frac{1}{\gamma}\int gf^{\gamma}\mathrm{d}x-1=\frac{1}{\gamma}\int g\left(f^{\gamma}-1\right)\mathrm{d}x=\int g\log f\mathrm{d}x+O\left(\gamma\right)\left(\gamma\to 0\right),$$

$$\left(\int f^{1+\gamma}\mathrm{d}x\right)^{\gamma/(1+\gamma)}=1+O\left(\gamma^{2}\right)\left(\gamma\to 0\right).$$

因此，有：

$$\tilde{d}_{\gamma}\left(g,f\right)=-\frac{1}{\gamma}\left[\frac{\int gf^{\gamma}\mathrm{d}x}{\left\{\int f^{1+\gamma}\mathrm{d}x\right\}^{\gamma/1+\gamma}}-1\right]\to-\int g\log f\mathrm{d}x\left(\gamma\to 0\right)$$

这样就完成了证明。

由经验密度函数 $\bar{g}$ 代替真实密度 g，于是可以得到：

$$\begin{aligned}-n\tilde{d}_{\gamma}\left(\bar{g},f_{\theta}\right)&=\sum_{i-1}^{n}\frac{f_{\theta}\left(X_{i}\right)^{\gamma}}{\gamma\left\{\int f_{\theta}\left(X_{i}\right)^{1+\gamma}\mathrm{d}x\right\}^{\gamma/(1+\gamma)}}-\frac{n}{\gamma}\\&=\sum_{i=1}^{n}q_{\theta}^{(\gamma)}\left(X_{i}\right)-\frac{n}{\gamma}=Q_{n}^{(\gamma)}\left(\theta\right)\end{aligned}\tag{6}$$

其中

$$q_{\theta}^{(\gamma)}\left(x\right)=q^{(\gamma)}\left(\theta;x\right)=\frac{1}{\gamma}f_{\theta}\left(x\right)^{\gamma}\left\{\int f_{\theta}\left(x\right)^{1+\gamma}\mathrm{d}x\right\}^{-\gamma/(1+\gamma)}$$

本文称成 $Q_{n}^{(\gamma)}\left(\theta\right)$ 为 γ-似然，这是一种拟似然(或称加权似然)。根据命题 1 的性质(iii)可以得到：

$$\lim_{\gamma\to 0}Q_{n}^{(\gamma)}\left(\theta\right)=\sum_{i=1}^{n}\log f_{\theta}\left(x_{i}\right)$$

所以，γ-似然是对数似然的一种泛化。本文将用 γ-似然代替 FH 模型区域均值估计的对数边际似然。因为在确定模型的前提下使用 γ-似然所以可以得到：

$$\lim_{\gamma\to 0}Q^{\gamma}\left(y;\phi\right)=\log f\left(y;\phi\right)$$

在模型(1)下，y_{i} 是独立的且服从 $y_{i}\sim N\left(x_{i}^{\mathrm{T}}\beta,A+D_{i}\right)$ 的正态分布，因此(6)式可以用该形式表达

$$Q^{\gamma}\left(y;\phi\right)=\sum_{i=1}^{m}\left\{\frac{s_{i}\left(y_{i};\phi\right)}{\gamma}-\frac{i}{\gamma}\right\}\tag{7}$$

其中 $V_{i}=\left\{2\pi\left(A+D_{i}\right)\right\}^{-1/2}$ 且

$$s_{i}\left(y_{i};\phi\right)=\left\{V_{i}^{-\gamma}\left(1+\gamma\right)^{1/2}\right\}^{\gamma/(1+\gamma)}V_{i}^{\gamma}\exp\left(-\frac{\gamma\left(y_{i}-x_{i}^{\mathrm{T}}\beta\right)^{2}}{2\left(A+D_{i}\right)}\right)。$$

于是本文将 γ-似然函数(7)式中的 $Q^{\gamma}\left(y;\phi\right)$ 来代替(4)中对数边际似然 $\log f\left(y;\phi\right)$。来定义 θ_{i} 的鲁棒贝叶斯预测因子 $\tilde{\theta}_{i}^{R}$。可以得到：

$$\frac{\partial}{\partial y_{i}}Q^{\gamma}\left(y;\phi\right)=\frac{\partial}{\partial y_{i}}\frac{s_{i}\left(y_{i};\phi\right)}{\gamma}=\frac{1}{A+D_{i}}\left(y_{i}-x_{i}^{\mathrm{T}}\beta\right)s_{i}\left(y_{i};\phi\right)$$

则稳健贝叶斯预测为：

$$\tilde{\theta}_{i}^{R}=y_{i}-\frac{D_{i}}{A+D_{i}}\left(y_{i}-x_{i}^{\mathrm{T}}\beta\right)s_{i}\left(y_{i};\phi\right)\tag{8}$$

在(7)式中的收缩因子为 $s_i(y_i;\phi)/(A+D_i)$，它依赖于 y_i，而经典的贝叶斯预测(3)中的收缩因子是 $D_i/(A+D_i)$，它不依赖于 y_i。而且，当 $\gamma=0$ 时，$s_i(y_i;\phi)=1$ 则 $\tilde{\theta}_i^R$ 将变为 θ_i。此外，当 $D_i\to 0$，稳健贝叶斯预测器 $\tilde{\theta}_i^R$ 将简化为直接估计量 y_i，与经典的贝叶斯与测量 θ_i 一样。

3. 稳健经验贝叶斯估计量与均方误差

3.1. 稳健的参数估计

本文定义 ϕ 的稳健估计量 $\hat{\phi}_\gamma$，$\hat{\phi}_\gamma=\arg\max Q^\gamma(y;\phi)$，其中 $Q^\gamma(y;\phi)$ 在(7)中给出。则稳健估计量 $\hat{\phi}_\gamma$ 满足

$$\begin{aligned}\frac{\partial Q^\gamma}{\partial\beta}&=\sum_{i=1}^m\frac{x_i s_i(y_i;\phi)(y_i-x_i^{\mathrm{T}}\beta)}{A+D_I}=0,\\ \frac{\partial Q^\gamma}{\partial A}&=\sum_{i=1}^m\left\{\frac{\gamma(y_i-x_i^{\mathrm{T}}\beta)^2 s_i(y_i;\phi)-\gamma s_i(y_i;\phi)/\sqrt{2\pi}}{2(A+D_i)^2}-\frac{\gamma^2 s_i(y_i;\phi)/\sqrt{2\pi}}{2(1+\gamma)(A+D_i)^{3/2}}\right\}.\end{aligned}\tag{9}$$

本文使用 Newton-Raphson 来求解这些方程，求解过程在此省略详细求导过程请参考小域估计的文章[1]。本文将稳健估计量 $\hat{\phi}_\gamma$ 替换(7)中的贝叶斯估计量 ϕ，最终得到稳健经验贝叶斯估计量 $\hat{\theta}_i^R=\tilde{\theta}_i^R(y_i;\hat{\phi}_\gamma)$。

3.2. 稳健的参数估计

调优参数 γ 与稳健性有关，但很难给出选择过程的解释。根据 Ghosh [2]的文章，本文考虑基于稳健贝叶斯预测量的均方误差来选择(7)中的 γ，这使下文能够以可解释的方式指定调优参数 γ。均方误差的公式由以下定理给出。

定理 1. 在模型(1)中，$E\{(\tilde{\theta}_i^R-\theta_i)^2\}=g_{1i}(A)+g_{2i}(A)$，其中

$$g_{1i}(A)=\frac{AD_i}{A+D_i},\ g_{2i}(A)=\frac{D_i^2}{A+D_i}\left\{\frac{V_i^{2\gamma}}{(2\gamma+1)^{3/2}}-\frac{2V_i^{\gamma}}{(\gamma+1)^{3/2}}+1\right\}$$

其中 $g_{2i}(A)$ 随 γ 的增大而增大。

经典贝叶斯预测量(3)的均方误差对应于 $g_{1i}(A)$，因此 $\tilde{\theta}_i^R$ 相对于 θ_i 的超额均方误差为 $g_{2i}(A)$，当 $\gamma=0$ 时 $g_{2i}(A)$ 接近于零。因此，$\tilde{\theta}_i^R$ 的稳健性与均方误差之间模型(1)下存在着一个权衡。为了得到更好的调优参数 γ，在此本文定义 $Exc(\gamma)=100\%\times\sum_{i=1}^m g_{2i}(\hat{A}_\gamma)/\sum_{i=1}^m g_{1i}(\hat{A}_\gamma)$ 作为相对超额均方误差的一个百分比，其中 $\hat{A}_\gamma$ 是 A 在式(11)中的稳健估计所得。本文对调优参数 γ 的选择遵循使 $Exc(\gamma)$ 不超过用户指定的 $c\%$；换句话说我们寻找 γ 满足 $Exc(\gamma^*)=c\%$。该等式存在唯一解，因为根据定理 1 可知 $Exc(\gamma)$ 随 γ 的增大而增大。在实证分析中，我们首先计算指定 $c\%$ 的调优参数 γ，并将所有估计过程中的 γ 替换为 γ 进行的。而在下文中为了理论证明的简单性，我们假设 γ 是已知的，而在数值模拟与实证分析中我们将使用该调优参数的选择方法来研究它可能造成的影响。

3.3. 稳健估计量的渐进性质

我们考虑在模型(1)下稳健估计量的渐近性质。为此，下文需要假设以下正则性条件。

条件 1：$0<D_*\le\min_{1\le i\le m}D_i\le\max_{1\le i\le m}D_i\le D^*<\infty$，其中 D_* 与 D_* 不依赖于 m。

条件 2：$\max_{1\le i\le m}x_i^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}x_i=O(m^{-1})$，其中 $X=(x_1,\cdots,x_m)^{\mathrm{T}}$。

条件 3：当 $m\to\infty$ 时，$X^{\mathrm{T}}X/m$ 将收敛于一个正定矩阵。

Rao 的文章中也假设了类似的条件。因为(9)式的导数在模型(1)下期望为 0，我们将得到以下的结果。

定理2. 在条件1~3情况下，$\hat{\beta}_\gamma$ 与 $\hat{A}_\gamma$ 是渐进独立的且分布分别为 $N\left(\beta, m^{-1}J_\beta^{-1}K_\beta J_\beta^{-1}\right)$ 与 $N\left(A, K_A/mJ_A^2\right)$，其中

$$J_\beta = \frac{1}{m(\gamma+1)^{3/2}}\sum_{i=1}^{m}\frac{V_i^\beta x_i x_i^{\mathrm{T}}}{A+D_i},\quad J_A = \frac{1}{2m}\sum_{i=1}^{m}\frac{V_i^\gamma\left(\gamma^2+2\right)}{\left(A+D_i\right)^2(\gamma+1)^{5/2}},$$

$$K_\beta = \frac{1}{m(2\gamma+1)^{3/2}}\sum_{i=1}^{m}\frac{V_i^{2\beta}x_i x_i^{\mathrm{T}}}{A+D_i},\quad K_A = \frac{1}{m}\sum_{i=1}^{m}\frac{V_i^{2\gamma}}{\left(A+D_i\right)^2}\left\{\frac{2\left(2\gamma^2+1\right)}{(2\gamma+1)^{5/2}}-\frac{\gamma^2}{(\gamma+1)^3}\right\}。$$

当 $\gamma=0$，$\hat{\beta}_\gamma$ 的渐进协方差与 $\hat{A}_\gamma$ 的

渐进方差将简化为 Datta 与 Lahiri 文章中所给出的 β 与 A 的极大似然估计量的渐进协方差和方差，因为(7)将简化为对数似然(2)。

3.4. 稳健经验贝叶斯估计量的均方误差

为了评估估计量 $\hat{\theta}_i^R$ 的风险，下文考虑均方误差 $M_i = E\left\{\left(\hat{\theta}_i^R-\theta_i\right)^2\right\}$，其中的期望是关于 θ_i 和 $y_i\left(i=1,\cdots,m\right)$ 的联合分布的并遵循假设模型(1)。均方误差可视为综合贝叶斯风险，是小域估计中风险的标准度量。

由于 $\hat{\theta}_i^R$ 依赖于估计量 $\hat{\phi}_\gamma$，均方误差 M_i 考虑了由于 $\hat{\phi}_\gamma$ 引起的额外波动。因此很难对 M_i 进行直接计算，所以使用了 M_i 的二阶近似。按照这个思路，下文用以下的定理给出了 M_i 的近似。

定理 3. 在条件 1~3 下，

$$M_i = g_{1i}(A)+g_{2i}(A)+\frac{g_{3i}(A)}{m}+\frac{g_{4i}(A)}{m}+\frac{2g_{5i}(A)}{m}+o\left(m^{-1}\right) \tag{10}$$

其中 $g_{1i}(A)$ 和 $g_{2i}(A)$ 已经在定理 1 中给出且

$$g_{3i}(A) = \frac{D_i^2V_i^{2\gamma}}{B_i^2(2\gamma+1)^{3/2}}x_i^{\mathrm{T}}J_\beta^{-1}K_\beta J_\beta^{-1}x_i,\quad g_{4i}(A) = \frac{D_i^2V_i^{2\gamma}K_A}{B_i^3(2\gamma+1)^{7/2}J_A^2}\left(\gamma^4-\frac{1}{2}\gamma^2+1\right),$$

$$\begin{aligned} g_{5i}(A) = {} & \frac{\gamma D_i^2x_i^{\mathrm{T}}J_\beta^{-1}K_\beta J_\beta^{-1}x_i}{2B_i^4}\left(3B_iC_{11}-\gamma C_{21}\right) \\ & +\frac{D_i^2K_A}{24B_i^6J_A^2}\left\{3\gamma B_i^2C_{21}+(\gamma-2)(3\gamma+8)C_{11}\right\} \\ & +\frac{D_i^2x_i^TJ_\beta^{-1}x_i}{B_i^4}\left(B_iC_{12}-\gamma C_{22}\right)+\frac{D_i^2}{2B_i^6J_A}\left\{\gamma C_{32}-2B_iC_{22}+(2-\gamma)B_i^2C_{12}\right\} \\ & +\frac{D_i^2}{2B_i^4}\left\{b_A-\frac{\gamma V_i^\gamma}{(\gamma+1)^{3/2}B_iJ_A}\right\}\left\{(2-\gamma)B_iC_{11}-\gamma C_{21}\right\} \end{aligned}。$$

其中 $B_i = A+D_i, b_A = \lim_{m\to\infty} mE\left(\hat{A}_\gamma-A\right)$ 是 $\hat{A}_\gamma$ 的一阶偏差，且

$$C_{jk} = (2j-1)!!B_i^j\left\{V_i^{k\gamma}(k\gamma+1)^{-j-1/2}-V_i^{k\gamma+\gamma}(k\gamma+\gamma+1)^{-j-1/2}\right\}。$$

其中 $(2j-1)!!=(2j-1)(2j-3)\cdots(1)$。

这个定理的证明在 Datta 和 Lahiri 的补充材料中有类似的推导。近似公式(10)是基于已知的调优参数 γ，因此在估计得到 γ 的情况下可能会有所不同。(10)中的表达式简化为 Datta 和 Lahiri 文章中给出的经典经验贝叶斯估计量的均方误差，当 $\gamma=0$ 时 $C_{jk}=0$，$g_{5i}(A)=0$。

3.5. 均方误差的估计

由于定理 3 给出的均方误差的近似取决于未知参数 A 的选择，因此不能在实证分析中使用；所以使用均方误差的二阶无偏估计量。如果 $E\left(\hat{T}\right)=T+o\left(m^{-1}\right)$，则称估计量 $\hat{T}$ 是二阶无偏的。由定理 3 可知，$g_{3i}(A)$，$g_{4i}(A)$ 和 $g_{5i}(A)$ 是光滑的，因此 $g_{3i}(A)$，$g_{4i}(A)$ 和 $g_{5i}(A)$ 是二阶无偏的。相比之下，$g_{1i}\left(\hat{A}_\gamma\right)$ 和 $g_{2i}\left(\hat{A}_\gamma\right)$ 可能有相对大的偏差,因为 $g_{1i}\left(\hat{A}_\gamma\right)$ 和 $g_{2i}\left(\hat{A}_\gamma\right)$ 是 $O(1)$。由于这些项的偏差和偏差校正估计的推导需要比较繁琐的代数计算，因此下文使用参数自举的方法，方法与 Butar 和 Lahiri 的文章中的类似。本文定义了自举估计量

$$\hat{M}_i=2g_{12i}\left(\hat{A}_\gamma\right)-\frac{1}{B}\sum_{b=1}^{B}g_{12i}\left(\hat{A}_\gamma^{(b)}\right)+\frac{g_{3i}\left(\hat{A}_\gamma\right)}{m}+\frac{g_{4i}\left(\hat{A}_\gamma\right)}{m}+\frac{2g_{5i}\left(\hat{A}_\gamma\right)}{m} \tag{11}$$

其中 $g_{12i}\left(\hat{A}_\gamma\right)=g_{1i}\left(\hat{A}_\gamma\right)+g_{i2}\left(\hat{A}_\gamma\right)$ 且 $\hat{A}_\gamma^{(b)}$ 是基于自举样本 $y_i^{(b)},\cdots,y_m^{(b)}$ 的自举估计量，自举样本是由下式产生 $y_i^{(b)}=x_i^{\mathrm{T}}\hat{\beta}_\gamma+v_i^{(b)}+\varepsilon_i^{(b)}$，$v_i^{(b)}\sim N\left(0,\hat{A}_\gamma\right)$，$\varepsilon_i^{(b)}\sim N\left(0,D_i\right)$。根据 Chang 和 Hall 的文章，得到以下定理：

定理 4. 在条件 1~3 的情况下设 $\hat{M}_i^{\dagger}$ 为 $\hat{M}_i^{\dagger}$ 在 $B=0$ 时得到的理想版本。定义 $\hat{M}_i=\hat{M}_i^{\dagger}+U_i$，其中 U_i 表示仅执行有限数量的引导复制而产生的错误项。那么 $E\left(\hat{M}_i^{\dagger}\right)-M_i=o\left(m^{-1}\right)$ 且 $U_i=O_p\left\{(mB)^{-1/2}\right\}$。

定理 4 表示，估计量(11)的理想版本 $\hat{M}_i^{\dagger}$ 是二阶无偏的。此外，该定理说明如果 $B=O\left(m^{1+\delta}\right)$ 对于较小的 $\delta>0$，则由有限次数的自举迭代引起的误差为 $o_p\left(m^{-1}\right)$，因此估计量 $\hat{M}_i^{\dagger}$ 的表现应与理想估计量 $\hat{M}_i^{\dagger}$ 相似。然而，理论情况下是基于已知的 γ，因此自举估计量(11)不一定在估计所得的 γ 下被证明。

尽管在(10)中采用了一种附加形式的偏差校正，但在以 Butar 和 Lahiri 为基础也提出了其他形式的偏差校正。对于 $g_{5i}(A)$，本文必须计算 $\hat{A}_\gamma$ 的一阶偏差 b_A 的估计，它可以从参数自举样本中计算得到。换句话说，本文可以使用参数自举法计算 $g_{5i}(A)$。如定理 3 所示，$E\left\{\left(\hat{\theta}_i^R-\tilde{\theta}_i^R\right)\left(\tilde{\theta}_i^R-\tilde{\theta}_i\right)\right\}=m^{-1}g_{5i}(A)+o\left(m^{-1}\right)$，所以下文可以用

$$B^{-1}\sum_{b=1}^{B}\left\{\hat{\theta}_i^R\left(y_i^{(b)},\hat{\phi}_\gamma^{(b)}\right)-\tilde{\theta}_i^R\left(y_i^{(b)},\hat{\phi}_\gamma\right)\right\}\left\{\tilde{\theta}_i^R\left(y_i^{(b)},\hat{\phi}_\gamma\right)-\tilde{\theta}_i\left(y_i^{(b)},\hat{\phi}_\gamma\right)\right\},$$

替代了 $g_{5i}(A)$，其中 $\hat{\phi}_\gamma^{(b)}$ 是参数自举法估计量。

4. 数值模拟与实证分析

4.1. 数值模拟

首先本文评估了所提出的稳健估计方法的估计精度，并将其与现有估计方法进行了比较。本文考虑 Fay-Herriot 模型：

$$y_i=\theta_i+\varepsilon_i,\quad y_i=\theta_i+\varepsilon_i\theta_i=\beta_0+\beta_1x_i+A^{1/2}u_i\ (i=1,\cdots,m)。$$

当 $m=30$，$\beta_0=0$，$\beta_1=2$，$A=0.5$，且 $\varepsilon_i\sim N\left(0,D_i\right)$。辅助变量 x_i 由 $(0,1)$ 上的均匀分布生成。接下来，将 m 个区域分成 5 组，每组包含相同数量的区域，并在每组中设置相同的 D_i 值。各组的 D_i 值分别为 $(0.2,0.4,0.6,0.8,1.0)$。对于 u_i 的分布，下文假设结构为 $u_i\sim(1-\xi)N(0,1)+\xi N\left(0,10^2\right)$，其中 ξ 决定了假设分布受污染的程度。本文考虑以下三种情况：(I) $\xi=0$，(II) $\xi=0.15$，(III) $\xi=0.3$。在情景(II)和(III)中，一些观测值有非常大的残差，辅助信息 x_i 对于这些异常观测值是有用的。

本文使用提出的具有密度幂散度的稳健经验贝叶斯估计方法估计 θ_i。将使用两个不同的膨胀率分别为 $c=1$ 和 $c=5$，并根据上文中描述的选择准则来选择 γ。比较了几种可供选择的方法：经典经验贝叶斯估计量、稳健贝叶斯估计量与使用 Sinha 和 Rao 提出的稳健估计方程使用极大似然方法估计的模型参数，以及 Ghosh 提出的稳健经验贝叶斯估计量使用模型参数的极大似然估计量。

表 1 报告了同一组内平均的均方误差值。由于经典经验贝叶斯方法中的正态性假设在情景(I)中是正确的，因此经验贝叶斯方法产生的均方误差比其他方法小这是自然的；然而，包括本文提出的方法在内的一些稳健方法的估计效果与经典方法相当。另一方面，在情景(II)和(III)中，当存在异常观测时，本文所提出的方法产生的均方误差往往小于其他方法，特别是对于抽样方差较大的群体。在这些场景中，当 $c=5$ 时所提出的方法的性能要优于 $c=1$ 时的方法，因为 c 越大，方法的稳健性越强。然而，本文所提出的方法的估计效果在情景(I)中，$c=5$ 比 $c=1$ 更差，这可以被视为为情景(II)和(III)中估计效果有更强的稳健性所付出的合理代价。

4.2. 实证分析

我们考虑应用美国劳工统计局的鲜奶消费数据，该数据在 Arora 和 Lahiri 的文章与 You 和 Chapman 的文章中都进行了研究。在该数据集中，有 43 个地区图 1 显示了 y_i 的散点图以及 $\beta_1,\cdots,\beta_4$ 的最大似然估计量。表明在 R_1 和 R_2 区域存在一些异常值。为了证实这一点，计算本文了标准化残差

$$r_i=\left(\hat{A}+D_i\right)^{-1/2}\left\{y_i-\sum_{j=1}^{4}\hat{\beta}_j I\left(i\in M_j\right)\right\}\left(i=1,\cdots,m\right)。$$

当正确指定 Fay-Herriot 模型(1)时，r_i 的分布接近标准正态分布。然而，如表 2 所示，在某些地区，r_i 的绝对值很高。本文使用 Sinha 和 Rao 的稳健估计方法和提出的密度幂散度方法作为(9)在 1%和 5%膨胀率下的解来估计参数。

从表 2 可以看出，在 r_i 绝对值较大的区域，$\hat{\theta}_i^{EB}$ 与 $\hat{\theta}_i^{R}$ 的差异较大。$\hat{M}_i^{EB}$ 和 $\hat{M}_i$ 之间的关系也出现了类似的情况。相反，$\hat{\theta}_i^{SR}$ 和 M_i^{SR} 的值与其他值不同，这可能是由于表 2 中 A 的估计值较低。

Table 1. The average mean square error of different estimation methods within the same group; All values are multiplied by 1000

表 1. 不同估计方法的均方误差在同一组内的平均值；所有的值都乘以 1000

情景	组别	DEB1	DEB2	EB	REB1	REB2	GEB
(I)	1	141	142	156	174	158	158
	2	218	225	251	281	259	306
	3	269	282	316	338	325	368
	4	307	326	350	376	359	393
	5	337	364	387	404	388	417
(II)	1	190	186	193	329	189	189
	2	360	345	372	1021	362	364
	3	513	482	604	1836	528	533
	4	662	620	879	2710	668	684
	5	798	745	860	3598	813	828
(III)	1	197	195	210	227	195	195
	2	389	383	391	515	385	385
	3	575	561	579	953	565	569
	4	758	737	773	1522	770	767
	5	937	905	947	2249	911	913

DEB1 为膨胀率为 1%时密度幂散度，即 $c=1$；DEB2 为膨胀率为 5%时密度幂散度，即 $c=5$；EB 为经验贝叶斯方法；REB1 为 Sinha 和 Rao 的稳健经验贝叶斯方法；REB2，为 Sinha 和 Rao 的最大似然稳健贝叶斯预测方法；GEB，为 Ghosh 等人的稳健经验贝叶斯方法。

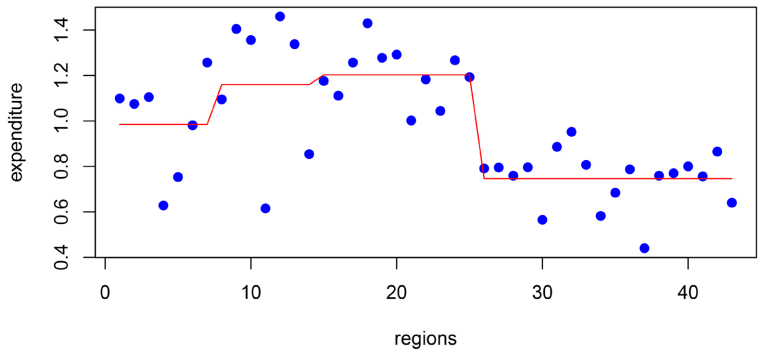

Figure 1. Scatter plot of average expenditure on fresh milk in 43 regions, along with the maximum likelihood estimation value
图 1. 43 个地区鲜奶平均支出的散点图，以及最大似然估计值

Table 2. Empirical Bayesian estimators, the robust empirical Bayesian estimators proposed in this paper, the robust empirical Bayesian estimators of Sinha and Rao, and the mean squared error estimates of these three methods; Multiply the values of $\hat{M}_i^{EB}$, $\hat{M}_i$ and $\hat{M}_i^{SR}$ by 100
表 2. 经验贝叶斯估计量，本文提出的稳健经验贝叶斯估计量与 Sinha 和 Rao 的稳健经验贝叶斯估计量，以及该三种方法对均方误差的估计；将 $\hat{M}_i^{EB}$，$\hat{M}_i$ 和 $\hat{M}_i^{SR}$ 的值乘以 100

地区	区域	y_i	r_i	$\hat{\theta}_i^{EB}$	$\hat{\theta}_i^{R}$	$\hat{\theta}_i^{SR}$	$\hat{M}_i^{EB}$	$\hat{M}_i$	$\hat{M}_i^{SR}$
3	1	1.10	0.63	1.02	1.02	1.02	1.35	1.35	0.81
4	1	0.63	−2.05	0.78	0.76	0.91	0.85	0.85	4.70
5	1	0.75	−1.25	0.86	0.87	0.92	0.96	0.96	4.44
10	2	1.36	1.38	1.19	1.24	1.23	1.39	1.37	3.79
11	2	0.62	−3.01	0.80	0.73	1.07	0.77	0.78	5.06
12	2	1.46	1.54	1.20	1.24	1.23	1.63	1.62	5.66
19	3	1.28	0.47	1.22	1.22	1.22	1.30	1.32	0.77
22	3	1.18	−0.01	1.19	1.19	1.19	0.80	0.84	0.56
31	4	0.89	0.63	0.76	0.76	0.75	1.54	1.63	0.78
37	4	0.44	−1.84	0.54	0.54	0.61	0.64	0.65	3.59

1989 年鲜奶平均支出的估计值 y_i，以及抽样方差 D_i。继 Arora 和 Lahiri 的研究之后，我们考虑如果第 i 个区域属于第 j 个区域的话，我们有 $x_i^{\mathrm{T}}\beta=\beta_j\ (j=1,\cdots,4)$ 的 Fay-Herriot 模型(1)。地区分成四个区域分别是 $R_1=\{1,\cdots,7\}$，$R_2=\{8,\cdots,14\}$，$R_3=\{15,\cdots,25\}$ 和 $R_4=\{26,\cdots,43\}$。

5. 结论与展望

通过数值模拟与实证分析表明，与现有方法相比，本文提出的方法在数据中存在异常观测的情况下特别适用。本研究聚焦于稳健小域估计问题，主要的包含稳健小域估计方法的比较、密度幂散度在小域估计模型中的应用的研究。提出了一种针对存在数据存在异常观测值的 Fay-Herriot 模型的稳健小域估计

方法。本文通过引入γ散度，给出了解决具有异常观测的稳健估计方法。对主要结论进行归纳，包括如下两个方面：

第一，对基本的小域估计方法、小域稳健估计方法进行了比较研究。通过对现有小域估计方法的分析发现，小域估计量对区域随机效应的方差较为敏感。当观测数据存在异常值或者随机效应是有偏分布时，现有估计量具有较大的偏差，且容易受异常值影响。而现有的稳健估计方法中，一部分方法是针对于特殊情形下的小域模型进行估计的方法，不具有普适性。另一部分稳健估计方法容易出现过度收缩的情形。因此本文中引入了密度幂散度族的估计方法，给出了γ散度的定义、相关的性质以及如何进行估计等。

第二，本文研究了区域水平模型的稳健小域估计。首先将 DPD 散度引入了区域水平模型，通过推广似然函数的表达，给出了稳健估计量的具体公式。结合估计量的表达式，给出了估计量的 MSE 及其估计。其次，将γ散度应用于区域水平模型，给出了估计量的表达式和 MSE。最后，给出了模型调整参数的选择算法和估计量 MSE 估计的程序。通过模拟数据和实际数据，验证了本研究中提出方法的优越性。结果一致表明在适当的调整参数中，本文提出的方法对于模型的系数的估计和小域目标的估计均具有较小的 MSE，表现出了稳健性。

参考文献

[1] Fay, R.E. (1987) Application of Multivariate Regression to Small Domain Estimation. In: Platek, R., Rao, J.N.K. and Särndal, C.E., *et al*., *Small Area Statistics*, Wiley, 91-102.

[2] Ghosh, M., Maiti, T. and Roy, A. (2008) Influence Functions and Robust Bayes and Empirical Bayes Small Area Estimation. *Biometrika*, **95**, 573-585. https://doi.org/10.1093/biomet/asn030

[3] Sinha, S.K. and Rao, J.N.K. (2009) Robust Small Area Estimation. *Canadian Journal of Statistics*, **37**, 381-399. https://doi.org/10.1002/cjs.10029

[4] Jiang, J. and Rao, J.S. (2020) Robust Small Area Estimation: An Overview. *Annual Review of Statistics and Its Application*, **7**, 337-360. https://doi.org/10.1146/annurev-statistics-031219-041212

[5] Morales, D., Esteban, M.D., Perez, A., *et al*. (2021) A Course on Small Area Estimation and Mixed Models: Methods, Theory and Applications in R. Springer, 239-266.

[6] 王小宁. 小域估计在问卷分割中的应用研究[J]. 统计与信息论坛, 2021, 36(9): 3-10.

[7] Smith, P.A., Bocci, C., Tzavidis, N., Krieg, S. and Smeets, M.J.E. (2021) Robust Estimation for Small Domains in Business Surveys. *Journal of the Royal Statistical Society Series C*: *Applied Statistics*, **70**, 312-334. https://doi.org/10.1111/rssc.12460

[8] Bertarelli, G., Chambers, R. and Salvati, N. (2020) Outlier Robust Small Domain Estimation via Bias Correction and Robust Bootstrapping. *Statistical Methods & Applications*, **30**, 331-357. https://doi.org/10.1007/s10260-020-00514-w

[9] Sugasawa, S. and Kubokawa, T. (2020) Small Area Estimation with Mixed Models: A Review. *Japanese Journal of Statistics and Data Science*, **3**, 693-720. https://doi.org/10.1007/s42081-020-00076-x

[10] Team, R.C. (2020) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

[11] Kim, B. and Lee, S. (2023) Robust Estimation for General Integer-Valued Autoregressive Models Based on the Exponential-Polynomial Divergence. *Journal of Statistical Computation and Simulation*, **94**, 1300-1316. https://doi.org/10.1080/00949655.2023.2283764

[12] Basak, S. and Basu, A. (2022) The Extended Bregman Divergence and Parametric Estimation. *Statistics*, **56**, 699-718. https://doi.org/10.1080/02331888.2022.2070622

[13] Sugasawa, S. (2020) Robust Empirical Bayes Small Area Estimation with Density Power Divergence. *Biometrika*, **107**, 467-480. https://doi.org/10.1093/biomet/asz075

[14] Nakagawa, T. and Hashimoto, S. (2019) Robust Bayesian Inference via γ-Divergence. *Communications in Statistics—Theory and Methods*, **49**, 343-360. https://doi.org/10.1080/03610926.2018.1543765

[15] Sugasawa, S. and Yonekura, S. (2021) On Selection Criteria for the Tuning Parameter in Robust Divergence. *Entropy*, **23**, Article ID: 1147. https://doi.org/10.3390/e23091147

[16] Basak, S., Basu, A. and Jones, M.C. (2020) On the "Optimal" Density Power Divergence Tuning Parameter. *Journal of Applied Statistics*, **48**, 536-556. https://doi.org/10.1080/02664763.2020.1736524

[17] Yonekura, S. and Sugasawa, S. (2023) Adaptation of the Tuning Parameter in General Bayesian Inference with Robust Divergence. *Statistics and Computing*, **33**, Article No. 39. https://doi.org/10.1007/s11222-023-10205-7