arXiv:2302.03795v1 [stat.ME] 7 Feb 2023

基于GAL分布的贝叶斯半参数标量对函数分位数回归（含测量误差）

Roger S. Zoh、Annie Yu、Carmen Tekwe

2023年2月9日

摘要

分位数回归为研究协变量与响应变量分布中均值以外各方面之间的关联提供了一种一致的方法。当回归协变量的测量存在误差时，需要进行测量误差（ME）校正，以获得有效推断。这一点对于标量协变量和函数型协变量都成立。在此，我们提出扩展贝叶斯测量误差和贝叶斯分位数回归领域的研究，使其能够处理现有协变量中可能存在的复杂测量误差。我们的方法采用广义非对称拉普拉斯（GAL）分布作为工作似然。与相应的非对称拉普拉斯（AL）分布相比，GAL分布族近来已成为贝叶斯分位数回归建模中更灵活的分布族。随后，我们通过一系列模拟情景，对测量误差校正步骤中的两种方法进行了比较和对照。最后，我们将该方法应用于2013—2014年NHANES数据集，分析50岁及以上成年人群中，以分钟级设备测得的身体活动为函数的体质指数（BMI）分位数。

1 引言

分位数回归方法为探索响应变量 Y 的各种分布特征与一组协变量$\mathbf{X}$之间的关联提供了一个框架，超越了传统的均值回归。分位数回归方法的相关文献不断增多，其应用涵盖经济学Fitzenberger 等 (2001)；教育学 Martins 和 Pereira (2004)；医学 Wei 等 (2006)；Hong 等(2019)，仅举几例。（Koenker，2017；Koenker 等，2017）对分位数回归方法的进展作了很好的综述。

分位数回归参数的估计主要是从频率学派的角度进行探讨Koenker (2017)。这在一定程度上是由于分位数回归具有非参数特性，其中

并未明确假设任何参数分布。参数估计则基于最小化目标函数（检查／针球函数）。

本文主要关注贝叶斯方法。贝叶斯方法能够传播模型各个方面的不确定性，并在推断中考虑所有模型不确定性。贝叶斯分位数回归仍是一个十分活跃的研究领域。已有多种方法，包括基于狄利克雷过程混合的贝叶斯非参数方法 Gelfand 和 Kottas (2002)；半参数方法 Reich 和 Smith (2013)；Lancaster 和 Jae Jun (2010)；以及采用非对称拉普拉斯分布（ALD）族进行全参数建模，并将其作为贝叶斯分位数回归的一种候选似然 Yu 和 Moyeed (2001)；Kozumi 和 Kobayashi (2011)。然而，研究者对 ALD 族的灵活性 Yan 和 Kottas (2017)，以及基于 ALD 似然进行后验推断的有效性 Yang 等人 (2016) 表示担忧。例如，该分布在中位数处始终对称，且其众数由分布的位置参数决定。这些较为严格的性质使人们对 ALD 作为分位数回归中可行似然的灵活性产生了严重怀疑。相较之下，广义非对称拉普拉斯（GAL）分布已成为贝叶斯分位数回归的一种可能似然 Yan 和 Kottas (2017)；Rahman 和 Karnawat (2019)；Kobayashi 等人 (2021)。尽管 GAL 作为贝叶斯分位数回归的工作似然正得到越来越广泛的应用，但它尚未用于函数型协变量存在潜在复杂测量误差的分位数回归情形。

本文旨在填补贝叶斯估计中带测量误差的标量对函数型协变量分位数回归所存在的空白。据我们所知，目前贝叶斯文献中尚无方法能够估计函数型协变量对结果变量分布不同方面的影响，尤其是在这些函数型协变量存在测量误差且误差可能具有复杂异方差性的情况下。为使模型更加灵活，我们的方法采用 GAL 的截断狄利克雷过程混合（tDPM）作为响应变量的似然。随后，我们考虑了两种校正测量误差的方法：一种是基于联合模型的方法，与我们之前的工作类似Zoh 等人（2022）；另一种是受 Cui 等人（2022a）研究启发的快速回归校准方法。本文组织如下：第3 节简要介绍研究动机、模型及估计方法；第4 节讨论模拟设置及结果。随后，在第5 节，我们将所提出的模型应用于 NHANES 2013—2014 年数据，分析设备测量的身体活动对身体质量指数分布的影响，并在第6节给出总结性意见。

2 动机示例

本文提出的模型旨在评估美国国家健康与营养调查数据集中一组美国成年人的身体活动监测器测量数据对身体质量指数各分位数的影响。

我们使用的数据来自公开发布的美国国家健康与营养调查（NHANES）2013—2014年数据。NHANES是一项调查，每年对约5000名具有全国代表性的样本进行检查。该调查包括由疾病控制与预防中心（CDC）开展的人口统计问卷、检查问卷和实验室检测。在分析中，我们从人口统计数据中提取身体质量指数及其他人口统计信息。除人口统计信息外，我们还从检查数据集中获取了身体活动监测器（PAM）数据。NHANES自2011年起开始收集身体活动监测器数据。研究参与者需佩戴监测器至少7天，每天24小时；监测器以80 Hz（每1/80秒一次）的频率测量3个轴（x、y、z）的加速度。研究结束时，参与者归还设备，随后下载数据，并检查设备是否发生故障。之后再下载数据并进行处理，以供分析。

3 模型设定

假设从个体 $i = 1$、$\cdots$、n $\mathcal{D}_i$ =（Y_i、$\boldsymbol{W}_i$）处获得以下数据，其中 $Y_i \in \mathbb{R}$、$\boldsymbol{W}_i \in \mathbb{R}^T$以及 T 是在不同时间点观测 $\boldsymbol{W}_i$的一个数值。类似于 Tekwe等人 （2022）提出的模型，带有测量误差的标量对函数 τ分位数回归可表示为

$$Q_\tau(Y_i|\boldsymbol{X}, \boldsymbol{Z}) = \widetilde{\beta}_{0,\tau} + \boldsymbol{Z}_i^{\mathrm{T}}\widetilde{\boldsymbol{\theta}}_\tau + \int_{\mathcal{T}} \boldsymbol{X}_i(t)\boldsymbol{\beta}_{1,\tau}(t)dt \quad (1)$$

$$W_{ij}(t) = X_i(t) + U_{ij}(t),\ j = 1, 2, \cdots, J \quad (2)$$

$\boldsymbol{Z}_i$ 是长度为 p的无测量误差协变量向量， $\boldsymbol{\theta}$ 是其对应的系数； $\boldsymbol{X}_i$ ={$X_i(t_1)$ $\cdots$, $X_i(t_T)$} 是函数型协变量，$\boldsymbol{\beta}_{1,\tau}(t)$则是其对 Y的第τ个分位数的（函数型）效应。遗憾的是， $\boldsymbol{X}_i$并未被真正观测到，但其（无偏）代理变量 $\boldsymbol{W}_{ij}$ ={$W_{ij}(t_1)$ $\cdots$, $W_{ij}(t_{T_i})$} 可以被观测到。我们假设 $J \geq 2$ 表示对 $\boldsymbol{W}$的重复测量次数。方程 2 描述了测量误差模型，并假设误差项 $U_{ij}(t)$来自均值为零 $E\{U_{ij}(t)\} = 0$ 且协方差函数可能复杂或任意的高斯过程。我们仅要求，对于任意两个不同的时间点 $t_j \neq t_l$, $\mathrm{cov}\{U_{ij_1}(t_j)\ U_{ij_2}(t_l)\} < \infty$。方程 1 和 2 中模型的估计因诸多原因而十分困难，包括参数的维度很高（无限维）（函数型参数$\boldsymbol{\beta}_{1,\tau}(t)$），以及代理变量 $\boldsymbol{W}_{ij}$缺乏重复测量（即 $J = 1$）。对于 $J = 1$（ $\boldsymbol{W}_i$没有重复测量）的情形，Zoh 等人（2022）和 Tekwe 等人（2022）提出了一种利用工具变量（IV）估计方程 1 和 2 中模型参数的方法。类似的方法可以轻松地应用于我们的贝叶斯模型，且不会增加复杂性。我们通过基函数展开来缩减参数空间。具体而言，若记$\boldsymbol{B}_{i,K_n}$ =($\mathrm{b}_{i,1}, \cdots, \mathrm{b}_{i,K_n}$) $\in \mathbb{R}^{T_i \times K_n}$, $\boldsymbol{b}_k$ 为第 k个基函数（$k = 1, \cdots, K_n$）； T_i 为个体 i 的函数型数据 $\boldsymbol{W}_i$ 被观测到的不同时间点数， K_n 为基函数的数量。于是，方程 1 和 2 的降维参数形式

变为

$$Y_i|\boldsymbol{X}_i, \boldsymbol{Z}_i = \beta_0 + \boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\beta}_z + \sum_{k=1}^{K_n}\phi_k X_{i,k} + \epsilon_i \quad (3)$$

$$W_{i,k,j} = X_{i,k} + U_{i,k,j}, \quad (4)$$

其中 $i = 1, \cdots, n$; $j = 1, \cdots, J$, 以及 $k = 1, \cdots, K_n$, 其中 $X_{i,k} = \int \mathrm{b}_{i,k}(t)X_i(t)dt$; $\phi_k = \int \mathrm{b}_{i,k}(t)\beta_1(t)dt$; ϵ_i 具有密度函数 $f_\epsilon(.)$, 使得 $\int_{-\infty}^{0} f_\epsilon(\epsilon) = \tau$; $W_{i,k} = \int \mathrm{b}_{i,k}(t)W_i(t)dt$; $U_{i,k} = \int \mathrm{b}_{i,k}(t)U_i(t)dt$. 在贝叶斯框架下，通过为误差项 ϵ_i指定一个灵活的分布，完成分位数回归模型的设定。我们的完全贝叶斯方法将基于方程 3 和 4 中的模型展开。

3.1 广义非对称拉普拉斯分布（GAL）

鉴于贝叶斯分位数回归中 ALD 存在上述问题，研究者提出了其他误差分布的替代方案。最近，Yan 和 Kottas (2017) 提出将 GAL 分布作为贝叶斯分位数设定下误差项的一种替代方案。GAL 是偏态正态分布与定义在正实轴上的混合截断正态分布所构成的混合分布（Azzalini 和 Regoli, 2012）。也就是说，若 ϵ 服从 GAL 分布，则 ϵ 具有如下混合表示

$$\epsilon = \alpha s + A\nu + u\sqrt{\sigma B\nu}, \qquad \text{where } s \sim \text{Normal}^{+}(0, \sigma^2),\ \nu \sim \text{Exp}(\sigma),\ u \sim \text{Normal}(0, 1), \quad (5)$$

其中 $\alpha \in \mathbb{R}$ 为偏度参数；Normal(a, b) 表示均值为 a、方差为 b的正态分布；Exp(.|a) 表示速率为 a的指数分布；Normal^{+}(.|0, 1) 表示定义在正实轴上的截断标准正态分布。与 ALD 相比，GAL 在分位数建模方面更加灵活。也就是说，若 $\alpha \in \mathbf{R}$、$\mu \in \mathcal{R}$、$\sigma > 0$，则当 ϵ 的密度可表示为下式时，ϵ 服从 GAL 分布：

$$f_{GAL}(\epsilon|\tau, \alpha, \sigma) = \frac{2\tau(1-\tau)}{\sigma}\left([\Phi(\epsilon^*/\alpha - \alpha p_{\alpha-}) - \Phi(-\alpha p_{\alpha-})]\exp\left\{-\epsilon^* p_{\alpha-} + (-\alpha p_{\alpha-})^2/2\right\}\right.$$
$$\left.\times I(\epsilon^*/\alpha > 0) + \Phi\left(-\alpha p_{\alpha+} - \epsilon^*/\alpha I(\epsilon^*/\alpha > 0)\right)\exp\left\{-\epsilon^* p_{\alpha+} + (\alpha p_{\alpha+})^2\right\}\right) \quad (6)$$

有趣的是，对于 $\alpha = 0$，GAL 恰好就是非对称拉普拉斯分布（Yan 和 Kottas（2017））。另外需要注意的是，如果 $\epsilon \sim$ GAL(α, τ, σ)，其中 $\epsilon \in$ R，那么 $\int_{-\infty}^{0}$ GAL($\epsilon|\mu, \alpha, \tau, \sigma)d\epsilon = \tau_0 = h(\gamma)\tau \neq \tau$，其中 $h(\gamma)$ =2$\Phi(-|\gamma|)$ exp$\{0.5\gamma^2\}$，且 $\gamma =$ (I($\alpha > 0$)$-\tau$)$|\alpha|$，GAL 表示 GAL 分布的密度。这意味着我们需要指定 τ_0，随后得到 $\tau =$ I($\gamma < 0$) + $\{[\tau_0 -$ I($\gamma < 0$)$]/h(\gamma)\}$。注意，对于 $\tau_0 \in$ (0, 1) 的选择，γ是有界的，并且 $\gamma \in (\gamma_L, \gamma_U)$，其中 γ_L是 $h(\gamma) - (1 - \tau_0)$ 的负根，而 γ_U是 $h(\gamma) - \tau_0$ 的正根；使用 R 中的任意优化程序（如 optim）即可轻松获得这些根（Yan 和 Kottas，2017）。最后，GAL 具有如下

混合表示（因此 τ^{th} 分位数位于零点）

$$\epsilon = C|\gamma|s + A\nu + u\sqrt{\sigma B\nu}, \qquad \text{where } s \sim \text{Normal}^{+}(0,\sigma^2),\ \nu \sim \text{Exp}(\sigma),\ u \sim \text{Normal}(0,1), \quad (7)$$

GAL 作为一种灵活的分布用于分位数回归的情况日益增多。Rahman 和 Karnawat(2019) 研究了有序数据分位数回归建模中的 GAL；Kobayashi 等人(2021) 在贝叶斯分位数回归中考虑了 GAL 的狄利克雷过程混合分布（将 GAL 狄利克雷混合过程记为 GALDP）。该混合分布是关于参数向量 $(\gamma,\ \sigma) \in [\gamma_L,\gamma_U] \times \mathcal{R}^{+}$的混合分布。这里我们采用类似的截断模型表述。也就是说，我们假设 $\epsilon_i \overset{iid}{\sim} \boldsymbol{G}$，其中

$$\epsilon_i \overset{iid}{\sim} f_\epsilon(\epsilon|\boldsymbol{\Omega}_\epsilon) \quad = \quad \sum_{k=1}^{K\epsilon} \pi_{\epsilon,k}\text{GAL}(\tau_0,\gamma_k,\sigma_k) \qquad (8)$$

其中 $\boldsymbol{\Omega}_\epsilon = \{(\pi_{\epsilon,\ k},\ \gamma_{\epsilon,\ k},\ \sigma_{\epsilon,\ k})\}_{k=1}^{K_\epsilon}$，对于某个选定的 $K_\epsilon < \infty$。利用 Sethuraman (1994) 的截断构造，我们得到 $(\pi_{\epsilon,\ 1},\ \cdots,\ \pi_{\epsilon,\ K_n})^{\text{T}} \sim$Dirichlet分布$(\alpha/K_n,\ \cdots,\ \alpha/K_n)$ 和 $(\gamma_{\epsilon,\ 1},\ \sigma_{\epsilon,\ 1})$, $\cdots$, $(\gamma_{\epsilon,\ K_\epsilon},\ \sigma_{\epsilon,\ K_\epsilon}) \overset{iid}{\sim}$Unif$(\gamma_L,\ \gamma_U)Ga(a_0,\ b_0)$，浓度参数为 $\alpha > 0$。

3.2 先验设定

现在讨论所有模型参数的先验选择。对于 $\boldsymbol{\beta}_z \sim$ MVN$(0,\Sigma_z,0)$；对于参数向量 $\boldsymbol{\phi} = (\phi_1,\cdots,\phi_{K_n})^{\text{T}}$，我们采用 Lang 和 Brezger (2004) 提出的贝叶斯P样条先验，并假设 $\boldsymbol{\phi} \sim$MVN$(0,\theta^2\boldsymbol{P}^{-1})$, 其中 P是二阶差分矩阵。此外，对于 θ^2的先验，我们采用 Klein 和 Kneib (2016) 提出的方法。对于测量误差向量 $\boldsymbol{U}_i \overset{iid}{\sim} \sum_{k=1}^{K_u} \pi_{u,\ k}MVN(\cdot|\boldsymbol{\mu}_{u,k},\Sigma_{k,u})$，其中$\boldsymbol{\mu}_{u,\ k} \sim$MVN$(\cdot|\boldsymbol{\mu}_{u,}0,\Sigma_{u,}0)$，并约束为 $\sum_{k=1}^{K_\epsilon} \pi_{u,\ k}\boldsymbol{\mu}_{u,\ k} = \boldsymbol{0}$；$\Sigma_{k,\ u} \sim$Inv-W$(\cdot|\nu_u,0,\Psi_{0,u})$；且$\boldsymbol{\pi}_u = (\pi_1,\cdots,\pi_{K_u})^{\text{T}} \sim$Dirichlet分布$(\alpha_u/K_u,\cdots,\alpha_u/K_u)$。

对于 IV 的误差模型，$\boldsymbol{\omega}_i^* \overset{iid}{\sim} \sum_{k=1}^{K_\omega} \pi_{\omega,\ k}MVN(\cdot|\boldsymbol{\mu}_{\omega,\ k},\ \Sigma_{\omega,\ k})$，$\boldsymbol{\mu}_{\omega,\ k} \sim$MVN$(\cdot|\boldsymbol{\mu}_{\omega,\ 0},\ \Sigma_{\omega,\ 0})$, 以及 $\Sigma_{\omega,\ k} \sim$Inv-W$(\cdot|\nu_{0,\ \omega})$，并限制为 $\sum_{k=1}^{K_\omega} \pi_{\omega,\ k}\boldsymbol{\mu}_{\omega,\ k} = \boldsymbol{0}$。

对于聚类概率，我们假设 $\boldsymbol{\pi}_\omega = (\pi_1,\cdots,\pi_{K_\omega})^{\text{T}} \sim$Dirichlet分布$(\alpha_u/K_\omega,\cdots,\alpha_u/K_\omega)$。此外，我们假设 $\boldsymbol{X}_i \sim \sum_{k=1}^{K_x} \pi_{x,k}MVN(\cdot|\boldsymbol{\mu}_{x,k},\Sigma_{x,k})$，其中 $\boldsymbol{\mu}_k \sim$MVN$(\cdot|\boldsymbol{\mu}_0,\Sigma_0)$，$\Sigma_{x,k} \sim$Inv-W$(\cdot|\nu_x,0,\Psi_0)$, 并且 $\boldsymbol{\pi}_x = (\pi_1,\cdots,\pi_{K_x})^{\text{T}} \sim$Dirichlet分布$(\alpha_x/K_x,\cdots,\alpha_x/K_x)$。

3.3 联合后验分布

我们首先定义以下参数。令$\boldsymbol{\Omega}_u = \{(\pi_{u,k},\boldsymbol{\mu}_{u,k},\Sigma_{u,k})\}_{k=1}^{K_u}$、$\boldsymbol{\Omega}_\omega = \{(\pi_{\omega,k},\sigma_{\omega,k})\}_{k=1}^{K_\omega}$、$\boldsymbol{\Omega}_x = \{(\pi_{\omega,k},\sigma_{x,k})\}_{k=1}^{K_x}$，以及$\boldsymbol{\Omega}_\epsilon$，如上所定义。给定所有参数后，数据的联合似然

参数与.....成正比

$$\ell(\boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{M}_i|\tau_0, \nu_i, s_i, \boldsymbol{X}_i, \boldsymbol{\Omega}_\epsilon, \boldsymbol{\Omega}_u, \boldsymbol{\Omega}_\omega, \boldsymbol{\Omega}_x) \propto f_{GAL}(Y_i|\tau_0, \nu_i, s_i, \boldsymbol{X}_i, \boldsymbol{\Omega}_\epsilon) f_{u_i}(\boldsymbol{W}_i|\boldsymbol{X}_i, \boldsymbol{\Omega}_u) f_{\omega_i}(\boldsymbol{M}_i|\boldsymbol{X}_i, \boldsymbol{\Omega}_\omega).$$

基于该似然，联合后验分布正比于

$$\begin{aligned} Pr(\boldsymbol{\nu}, \boldsymbol{s}, \boldsymbol{X}, \boldsymbol{\Omega}_\epsilon, \boldsymbol{\Omega}_u, \boldsymbol{\Omega}_\omega, \boldsymbol{\Omega}_x|Data) \propto & \prod_{i=1}^{n} \{\ell(\boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{M}_i|\tau_0, \nu_i, s_i, \boldsymbol{X}_i, \boldsymbol{\Omega}_\epsilon, \boldsymbol{\Omega}_u, \boldsymbol{\Omega}_\omega, \boldsymbol{\Omega}_x) \\ & \times Pr(\boldsymbol{X}_i|\boldsymbol{\Omega}_x) Pr(\nu_i, s_i)\} Pr(\beta_0, \boldsymbol{\beta}_z) \\ & \times Pr(\boldsymbol{\Omega}_\epsilon) Pr(\boldsymbol{\Omega}_u) Pr(\boldsymbol{\Omega}_\omega) Pr(\boldsymbol{\Omega}_x) \end{aligned} \tag{9}$$

直接从联合后验分布中抽样较为复杂，因此我们采用一系列吉布斯步骤从联合后验分布中抽样。我们采用 Rahman 和 Karnawat (2019提出的结合 Metropolis-Hasting算法 (MH) 与部分折叠吉布斯抽样的方法来更新参数 (γ_k, σ_k)。

4 模拟与结果

4.1 模拟设置

对于每个 $i = 1$、$\cdots$、n，我们分别从具有 $E\{X_i(t)\}$ ={sin(2πt) + 1.25}/2、方差 $\{X_i(t)\} = \sigma_x^2$ 和协方差 cov$\{X_i(t_l)\ X_i(t_j)\} = \rho_x \sigma_x^2$ 的高斯过程中模拟 $X_i(t)$。我们模拟了 $W_{ij}(t)$ =$X_i(t)$ + $U_{ij}(t)$, 其中 $j = 1$、$\cdots$、$J = 5$, 且 $U(t)$ 是具有 $E\{U(t)\} = 0$、方差 $\{U_i(t)\} = \sigma_u^2$ 和协方差 cov$\{U_i(t_l)\ U(t_j)\} = \rho_u \sigma_u^2$ 的高斯过程；最后，我们将每个个体的响应变量模拟为 $Y_i = \boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{\beta}_z + \int_0^1 \boldsymbol{\beta}(t) X_i(t) dt + \epsilon_i$，其中 $\epsilon_i \sim \boldsymbol{G}$，服从某个分布 $\boldsymbol{G}$。在每次模拟中，我们使用平均偏差平方 (ABias^2)、平均样本方差 (渐近方差) 和积分均方误差 (MISE)，评估各估计量在 $\boldsymbol{\beta}_\tau(t)$ ($\tau = 0.25$、0.50、0.90) 下的表现:

$$\begin{aligned} ABias^2(\widehat{\beta}_\tau) &= \frac{1}{n_{grid}} \sum_{l=1}^{n_{grid}} \left\{\bar{\beta}_\tau(t_l) - \beta_\tau(t_l)\right\}^2, \\ Avar(\widehat{\beta}_\tau) &= \frac{1}{n_s n_{grid}} \sum_{i=1}^{n_r} \sum_{j=1}^{n_{grid}} \left\{\widehat{\beta}_\tau(t_j) - \bar{\beta}_\tau(t_j)\right\}^2, and \\ MISE &= ABias^2 + Avar, \end{aligned}$$

其中，n_{grid} 表示在 0 和 1 之间等距选取的网格点数量，$\bar{\beta}_\tau(t_j) = \frac{1}{n_r} \sum_{j=1}^{n_r} \widehat{\beta}_\tau^r(t_j)$ 表示在特定时间点 t_j处，$\beta_\tau(t_j)$ 在 n_r 个重复样本上的逐点平均值。我们将本方法得到的 $\beta_\tau(t)$ 估计值记为 $\boldsymbol{\beta}_{\tau,\ fBQ}(t)$（完全贝叶斯方法）。我们还考虑了朴素方法，该方法将 $\boldsymbol{W}$ 视为精确观测到的协变量，并使用 $\boldsymbol{W}$ 作为

^真实协变量 $\boldsymbol{X}$。朴素方法得到的估计值记为 $\boldsymbol{\beta}_{\tau,\ naive}(t)$。完全贝叶斯估计方法的一个潜在缺点是速度较慢，尤其是在数据集较大的情况下。我们还考虑了另一种方法，即两阶段方法。在第一阶段，我们基于观测到的代理数据 $\boldsymbol{X}$，并使用混合效应模型来近似函数型协变量 $\boldsymbol{W}$ 的值。然后在第二阶段，仅使用模型 3 进行贝叶斯估计。有关该方法的更多详情，请参见 Cui 等人 (2022a)。我们将该方法得到的参数估计值记为$\widehat{\boldsymbol{\beta}}_{\tau,\ fast}(t)$。在所有模拟中，我们均使用三个 GAL 分量的混合。

4.2 模拟结果

我们的模拟研究考察了所提出方法的各个方面，并分为不同情形。对于所有情形，除非另有说明，我们假设 $\sigma_x = 4$, $\rho_x = 0.5$, $\sigma_u = 4$,$\rho_u = 0.5$。在模拟**情形1**中，我们考虑一个简单情形，并假设 $\boldsymbol{G}$ （响应变量 Y 的误差分布）服从均值为零、标准差为 $\sigma_e = 1$的正态分布。我们考虑以下样本量 n $= 200$、500 和 1000。我们在表1中报告所考察方法的表现。在所考察的各个分位数下，完全贝叶斯方法、$\widetilde{\boldsymbol{\beta}}_{\tau、fBQ}(t)$ 以及回归校准方法$\boldsymbol{\beta}_{\tau、fast}(t)$ 的表现非常相似；不过，完全贝叶斯估计量在样本量较小时往往具有较高的均方误差（积分均方误差），而在样本量较大时其^积分均方误差较大。基于朴素方法的估计量$\boldsymbol{\beta}_{\tau、naive}(t)$ 在样本量较小时表现最佳，但在样本量较大时具有最高的积分均方误差。这表明，与测量误差校正方法相比，使用朴素方法会导致效率损失。正如预期，朴素方法的偏差趋于较高，并且即使在样本量较大时仍保持较高水平。

在**情形2**的模拟中，我们考虑误差服从偏态分布的情况。首先假设误差项服从位置参数为 $\xi = 0$、自由度为 $\nu = 5$、偏斜参数为 $\alpha = 2.0$的偏斜 t 分布。使用Azzalini (2022) 软件包即可轻松在 R 中模拟偏斜 t 分布。再次假设与**情形1**相同的设定，样本量 n 为 $= 500$。**情形2**各场景的结果见表2。总的来说，完全贝叶斯方法和回归校准（RC）方法的表现非常相似，不过完全贝叶斯方法表现最佳。朴素方法的表现仍然最差，且偏差较大。

在**案例 3**中，我们研究了中等到较大的测量误差方差 σ_u 对函数型参数估计 $\beta_\tau(t)$的影响。我们在假设 $\sigma_u = 1$、4和16这些取值的情况下运行模拟。我们在表3中报告了结果。随着测量误差增大（MSIE 增加），朴素估计量的性能变差，这表明有必要对测量误差进行校正。我们注意到，与朴素方法相比，尽管我们的快速测量误差校正方法在测量误差方差较低的情况下（$\sigma_u = 1$）往往具有较高的 MSE，但完整的贝叶斯

	方法	n	偏差^2	AVar	MISE
$\tau = 0.25$	$\beta_{\tau,\ fast}(t)$	200	0.691	0.023	0.714
		500	0.048	0.003	0.051
		1000	0.057	0.001	0.059
	$\beta_{\tau,\ naive}(t)$	200	0.688	0.012	0.700
		500	0.134	0.002	0.136
		1000	0.127	0.001	0.128
	$\beta_{\tau,\ FBQ}(t)$	200	0.969	0.042	1.011
		500	0.036	0.008	0.044
		1000	0.032	0.003	0.035
$\tau = 0.50$	$\beta_{\tau,\ fast}\ (t)$	200	0.693	0.018	0.711
		500	0.071	0.002	0.074
		1000	0.078	0.001	0.079
	$\beta_{\tau,\ naive}\ (t)$	200	0.708	0.010	0.718
		500	0.176	0.002	0.178
		1000	0.167	0.001	0.168
	$\beta_{\tau,\ FBQ}\ (t)$	200	0.869	0.036	0.905
		500	0.044	0.007	0.050
		1000	0.038	0.003	0.041
$\tau = 0.90$	$\beta_{\tau,\ fast}\ (t)$	200	0.600	0.036	0.636
		500	0.039	0.005	0.044
		1000	0.043	0.003	0.046
	$\beta_{\tau,\ naive}(t)$	200	0.663	0.019	0.682
		500	0.104	0.003	0.107
		1000	0.091	0.002	0.093
	$\beta_{\tau,\ FBQ}(t)$	200	1.391	0.083	1.474
		500	0.055	0.015	0.070
		1000	0.070	0.006	0.076

表 1：不同样本量下情形1的模拟汇总。$\beta_{\tau,\ FBQ}(t)$表示我们的完全贝叶斯方法；$\beta_{\tau,\ fast}(t)$表示快速回归校准方法；$\beta_{\tau,\ naive}(t)$表示忽略测量误差的朴素方法。

方法在测量误差较小时与朴素方法的表现非常相似，但在 ME 方差较大时优于朴素方法。对于较大的 ME 方差 σ_u，朴素方法往往表现不佳（ $MISE$ 较大）

最后，在**情形4**中，我们研究重复样本数量对所考察估计量性能的影响，因为两种经测量误差校正的方法（完全贝叶斯方法和快速单变量方法）都至少需要两个重复样本。我们采用与**情形1**类似的方式进行模拟， $n = 500$ 并允许每个个体的重复样本数量从 $J = 2$、3、4变化。在表 4中，总体而言，与较少重复样本数量相比，当重复样本数量较大时，两种经测量误差校正的方法表现非常相似。在几乎所有设置下，两种经测量误差校正的方法再次优于朴素方法。

5 应用

我们重新审视了这项案例研究。我们的目标是利用 NHANES 数据集，估计美国 50 岁以上成年人平均身体活动模式（行为）对不同 BMI 分位数的影响。

		偏差^2	Var	MISE
$\tau = 0.25$	$\beta_{\tau,\ fast}(t)$	0.049	0.001	0.050
	$\beta_{\tau,\ naive}(t)$	0.130	0.001	0.131
	$\beta_{\tau,\ FBQ}(t)$	0.030	0.003	0.033
$\tau = 0.50$	$\beta_{\tau,\ fast}(t)$	0.054	0.001	0.055
	$\beta_{\tau,\ naive}(t)$	0.146	0.001	0.147
	$\beta_{\tau,\ FBQ}(t)$	0.034	0.002	0.036
$\tau = 0.90$	$\beta_{\tau,\ fast}(t)$	0.043	0.002	0.044
	$\beta_{\tau,\ naive}(t)$	0.110	0.001	0.111
	$\beta_{\tau,\ FBQ}(t)$	0.031	0.003	0.034

表2：案例 2 的模拟汇总。数据是在位置参数为 $\xi = 0$、自由度为 $\nu = 5$、偏斜参数为 $\alpha = 2.0$ 的偏斜 t 分布下进行模拟的。模拟是在假设 $n = 500$ 的情况下完成的。

自 2012 年以来，NHANES 除了收集自我报告的身体活动行为数据外，也开始纳入更客观的身体活动数据测量（基于设备）。研究参与者需在至少一周内，将身体活动监测设备佩戴于非惯用手腕上，每次佩戴 24 小时。随后，参与者需归还身体活动监测设备。之后提取数据，并按照 NHANES 数据质量检查进行处理nha。在我们的分析中，我们使用了监测器独立运动摘要（MIMS）三轴汇总数据，其中记录了每位个体每分钟的身体活动情况。对于完整数据，每位个体每天有 1440 条观测。NHANES 还报告了数据质量标记，以及对每分钟身体活动状态（清醒／睡眠／未佩戴）的预测。按照 Karas 等人 (2019) 概述的数据处理步骤，我们仅纳入没有质量标记、且至少有 3 个有效工作日（周一至周五）的个体数据。有效日定义为无效分钟不超过 10%（144 分钟）的一天。若某分钟未触发任何质量标记，且被分类为佩戴状态，则该分钟视为有效。为消除极端值对分析的影响，我们对数据进行缩尾处理，并将高于数据 99.9 分位数的 MIMS 值直接替换为 99.9 分位数值。最后，将标记为未佩戴或值为−0.01 的分钟记为缺失值。与 Karas 等人(2019)；Cui 等人 (2022b) 类似，我们使用 R 中 R 软件包 Refund 的函数 *fpca.ssvd* ，借助 R CoreTeam (2022)，获得每位个体预测的身体活动数据。随后，使用 *fpca.sc* 得到的预测值替代缺失值。

我们拟合贝叶斯分位数回归模型，以估计BMI在以下无测量误差协变量条件下的0.1、0.5和0.9分位数：性别（男性 =参考组/女性 = 0）；种族（黑人 =参考组/西班牙裔/白人/混血/其他）；自我报告的健康状况（优秀 =参考组/良好/一般和较差）；年龄（以年计）。我们还希望校正通常工作日身体活动模式对BMI的影响，但该模式不可观测，只能通过工作日身体活动模式进行近似。我们首先对观测到的工作日身体活动模式中的测量误差进行校正 $\boldsymbol{W}_{ij}(t)$，采用Cui 等人$\widehat{\ }$(2022a)提出的方法的快速版本。随后，我们得到 $X_i(t)$ 的估计值，并将其作为函数型协变量输入

			$Bias^2$	Var	MISE
$\sigma_u = 1.0$	$\tau = 0.25$	$\beta_{\tau,\ fast}(t)$	0.065	0.002	0.068
		$\beta_{\tau,\ naive}(t)$	0.034	0.002	0.036
		$\beta_{\tau,\ FBQ}(t)$	0.050	0.005	0.056
	$\tau = 0.5$	$\beta_{\tau,\ fast}(t)$	0.072	0.002	0.074
		$\beta_{\tau,\ naive}(t)$	0.036	0.001	0.038
		$\beta_{\tau,\ FBQ}(t)$	0.039	0.004	0.044
	$\tau = 0.9$	$\beta_{\tau,\ fast}(t)$	0.045	0.004	0.049
		$\beta_{\tau,\ naive}(t)$	0.035	0.004	0.039
		$\beta_{\tau,\ FBQ}(t)$	0.095	0.009	0.104
$\sigma_u = 4.0$	$\tau = 0.25$	$\beta_{\tau,\ fast}(t)$	0.048	0.003	0.051
		$\beta_{\tau,\ naive}(t)$	0.134	0.002	0.136
		$\beta_{\tau,\ FBQ}(t)$	0.036	0.008	0.044
	$\tau = 0.50$	$\beta_{\tau,\ fast}\ (t)$	0.071	0.002	0.074
		$\beta_{\tau,\ naive}\ (t)$	0.176	0.002	0.178
		$\beta_{\tau,\ FBQ}\ (t)$	0.044	0.007	0.050
	$\tau = 0.90$	$\beta_{\tau,\ fast}\ (t)$	0.039	0.005	0.044
		$\beta_{\tau,\ naive}\ (t)$	0.104	0.003	0.107
		$\beta_{\tau,\ FBQ}\ (t)$	0.055	0.015	0.070
$\sigma_u = 16.0$	$\tau = 0.25$	$\beta_{\tau,\ fast}(t)$	0.480	0.009	0.489
		$\beta_{\tau,\ naive}(t)$	0.830	0.001	0.831
		$\beta_{\tau,\ FBQ}(t)$	0.733	0.013	0.747
	$\tau = 0.50$	$\beta_{\tau,\ fast}(t)$	0.703	0.006	0.709
		$\beta_{\tau,\ naive}(t)$	0.900	0.000	0.901
		$\beta_{\tau,\ FBQ}(t)$	0.867	0.007	0.874
	$\tau = 0.90$	$\beta_{\tau,\ fast}(t)$	0.270	0.020	0.290
		$\beta_{\tau,\ naive}(t)$	0.747	0.001	0.748
		$\beta_{\tau,\ FBQ}(t)$	0.436	0.052	0.488

表 3：案例 3 模拟结果汇总。数据是在测量误差方差水平不同的条件下模拟得到的（$\sigma_u = 1$，4，16）。

我们的贝叶斯分位数回归模型。作为第二种方案，我们拟合了类似的贝叶斯分位数回归模型，简单地对每个人在5个工作日内的身体活动模式进行平均¯。随后，这些 W_i（t）被用作朴素贝叶斯分位数回归模型的输入。我们拟合了多个具有不同GAL混合数量的贝叶斯模型，其中包括仅含一个GAL分量的模型。根据广泛适用信息准则，仅含一个GAL分量的模型往往对这些数据拟合得最好。我们按照贝叶斯工作流程中的建议，使用后验预测检验进行模型检查（Gelman 等人，2020）。模型检查未发现任何问题。为便于实现我们的方法，我们编写了R代码，在《贝叶斯分析》软件包Brms（Bürkner，2017）和《贝叶斯分析》软件包Rstan（Carpenter 等人，2017）中广泛使用的brm函数内实现了该方法。

我们报告了图 $\boldsymbol{\beta}$（t）中有测量误差校正和无测量误差校正时估计值的绘图，见图1。我们观察到，对于所有分位数，估计权重 $\boldsymbol{\beta}$（t）在大多数情况下均为负。然而，我们注意到，对于所有分位数，经过测量误差校正的估计值都低于所有朴素方法的估计值。事实上，对于所有分位数，朴素方法得到的权重似乎都向零衰减

		方法	偏差²	AVar	MISE
		$\beta_{\tau,\ fast}(t)$	0.073	0.004	0.077
	$\tau = 0.25$	$\beta_{\tau,\ naive}(t)$	0.307	0.002	0.309
		$\beta_{\tau,\ FBQ}(t)$	0.162	0.036	0.197
		$\beta_{\tau,\ fast}(t)$	0.145	0.003	0.148
$J = 2$	$\tau = 0.25$	$\beta_{\tau,\ naive}(t)$	0.380	0.001	0.382
		$\beta_{\tau,\ FBQ}(t)$	0.230	0.029	0.259
		$\beta_{\tau,\ fast}(t)$	0.039	0.006	0.045
	$\tau = 0.90$	$\beta_{\tau,\ naive}(t)$	0.237	0.003	0.240
		$\beta_{\tau,\ FBQ}(t)$	0.282	0.063	0.344
		$\beta_{\tau,\ fast}(t)$	0.053	0.003	0.057
	$\tau = 0.90$	$\beta_{\tau,\ naive}(t)$	0.214	0.002	0.216
		$\beta_{\tau,\ FBQ}(t)$	0.060	0.023	0.083
		$\beta_{\tau,\ fast}(t)$	0.097	0.003	0.100
$J = 3$	$\tau = 0.50$	$\beta_{\tau,\ naive}(t)$	0.272	0.001	0.273
		$\beta_{\tau,\ FBQ}(t)$	0.087	0.020	0.107
		$\beta_{\tau,\ fast}(t)$	0.036	0.006	0.042
	$\tau = 0.90$	$\beta_{\tau,\ naive}(t)$	0.165	0.003	0.168
		$\beta_{\tau,\ FBQ}(t)$	0.122	0.044	0.166
		$\beta_{\tau,\ fast}(t)$	0.048	0.003	0.051
	$\tau = 0.25$	$\beta_{\tau,\ naive}(t)$	0.161	0.002	0.162
		$\beta_{\tau,\ FBQ}(t)$	0.041	0.012	0.053
		$\beta_{\tau,\ fast}(t)$	0.078	0.002	0.080
$J = 4$	$\tau = 0.50$	$\beta_{\tau,\ naive}(t)$	0.212	0.002	0.213
		$\beta_{\tau,FBQ}(t)$	0.051	0.011	0.062
		$\beta_{\tau,\ fast}(t)$	0.034	0.005	0.040
	$\tau = 0.90$	$\beta_{\tau,\ naive}(t)$	0.123	0.003	0.126
		$\beta_{\tau,\ FBQ}(t)$	0.076	0.024	0.100

表4：**情形4**的模拟总结。数据按照**情形1** 下的设定进行模拟，采用 $n = 500$，并改变重复样本数量 J($J = 2$，3，4)。

与测量误差校正方法相比时所考虑的情况。这是误差测量的协变量中所观察到的衰减效应的特征。这可能表明，在将这些设备测量的身体活动作为函数型协变量报告时，对测量误差进行校正的重要性。此外，我们还报告了无测量误差协变量的后验推断（见表5）。

6 结论

我们针对函数型协变量存在测量误差的情形，在分位数回归框架下提出了两种贝叶斯测量误差校正方法：一种基于完整模型，另一种基于回归校准方法。通过一系列模拟设置，我们说明了校正函数型协变量测量误差的重要性。对重复测得的函数取平均（朴素方法）完全无法消除测量误差，并会导致函数型系数的估计产生偏差。在我们的应用中，我们表明，与

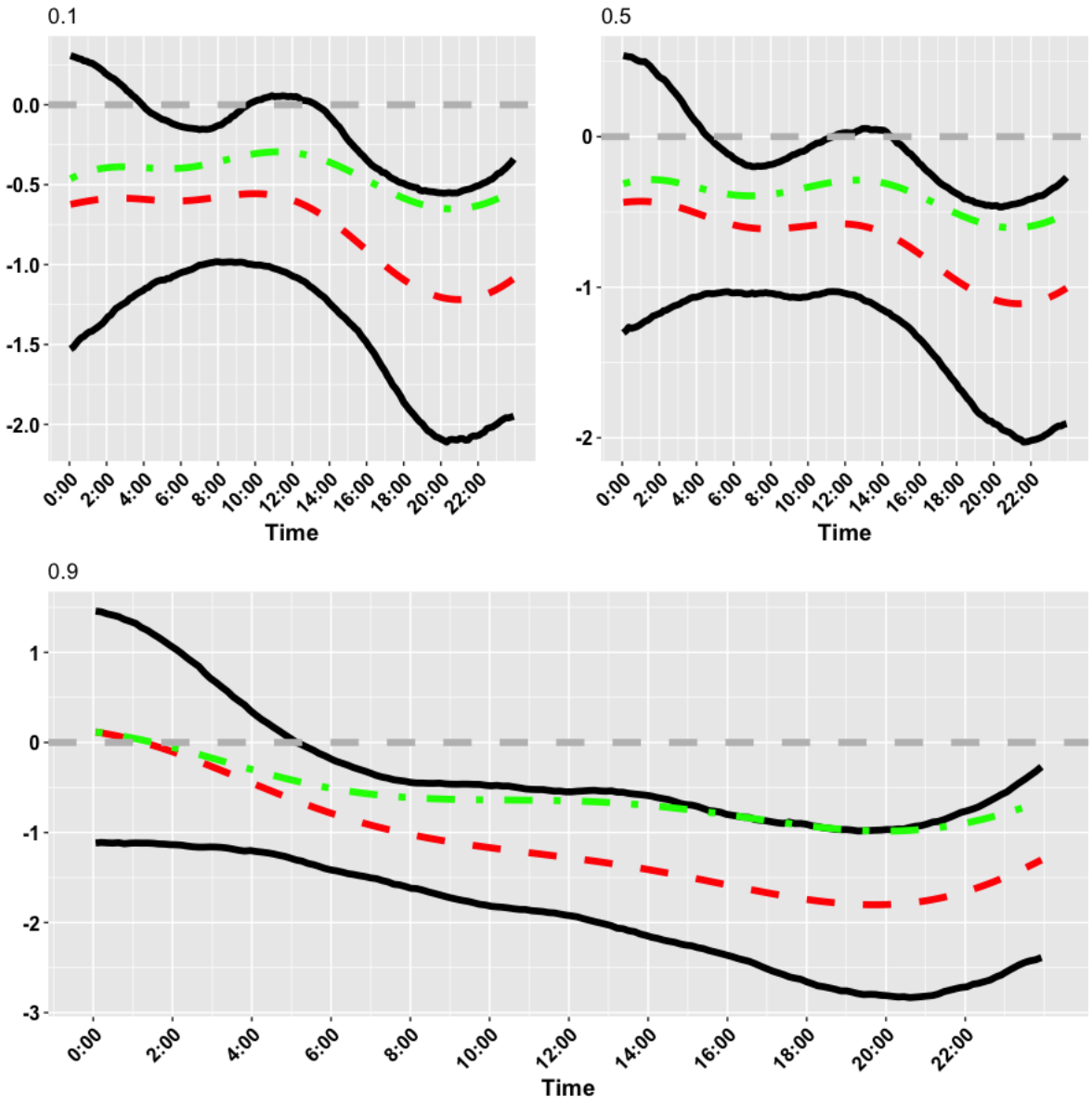


图1： $\beta(t)$的后验估计图。虚线（红色）表示经过测量校正后获得的估计值（RC - ）。点线（绿色）表示采用朴素方法获得的估计值（未进行测量误差调整）。实线（黑人）表示逐点95%后验可信区间的界限。

使用朴素方法相比，采用我们的测量误差校正方法会得到衰减程度较小的函数型效应。尽管我们的完整贝叶斯测量误差校正方法似乎非常适用，因为它能够反映模型中的所有不确定性，但它无法适应样本量的增加。因此，我们建议使用快速单变量方法。为便于实施我们的方法，我们将在作者的 GitHub 上分享分析所用的 R 代码。

	$\tau = 0.1$		$\tau = 0.5$		$\tau = 0.9$	
	mean	（2.5% - 97.5%）	mean	(2.5%-97.5%)	mean	(2.5%-97.5%)
截距	34.30	(31.31 , 37.19)	39.83	(36.92 , 42.67)	58.01	(54.28, 61.62)
性别（男性）	-0.50	(-1.07 , 0.06)	-0.71	(-1.28, -0.13)	-2.25	(-2.97 , -1.53)
种族 - 基线（黑人）	-	-	-	-	-	-
种族（西班牙裔）	0.05	(-0.86 , 0.95)	-0.26	(-1.18, 0.64)	-1.99	(-3.16 , -0.82)
种族（混合/其他）	-3.01	(-4.09 , -1.93)	-3.27	(-4.34 , -2.26)	-5.79	(-7.06 , -4.52)
种族（白人）	-0.58	(-1.32 , 0.17)	-0.80	(-1.58 , -0.03)	-1.59	(-2.58 , -0.62)
健康状况 - 基线（优秀）	-	-	-	-	-	-
健康状况（一般或较差）	0.83	(0.05 , 1.64)	0.78	(0.01, 1.57)	2.39	(1.41 , 3.38)
健康状况（良好）	0.78	(0.13 , 1.43)	0.69	(0.06 , 1.34)	0.81	(0.01 , 1.63)
年龄（岁）	-0.11	(-0.14 , -0.075)	-0.11	(-0.14 , -0.075)	-0.21	(-0.25, -0.17)

表5：基于快速单变量RC测量误差校正方法的身体质量指数各分位数$\tau = 0.1$、0。5，0。9的后验均值及相应的95%后验可信区间。

参考文献

美国国家健康与营养调查2013—2014年数据文档、代码本和频数。
https://wwwn.cdc.gov/Nchs/Nhanes/2013-2014/PAXMIN_H.htm。访问日期：2022- 10-22。

A. Azzalini。 *R* 软件包 *sn*：偏正态分布及相关分布，如偏-t 和 *SUN*（版本 *2.1.0*）。 意大利帕多瓦大学，2022。网址 https://cran.r-project.org/package=sn。主页：http://azzalini.stat.unipd.it/SN/。

Adelchi Azzalini 和 Giuliana Regoli。偏对称分布的一些性质。《统计数学研究所年刊》，64(4):857–879，2012年。

Paul-Christian Bürkner。brms：一个使用 Stan 的贝叶斯多层模型 R 软件包。《统计软件期刊》，80:1–28，2017年。

Bob Carpenter、Andrew Gelman、Matthew D Hoffman、Daniel Lee、Ben Goodrich、Michael Be- tancourt、Marcus Brubaker、Jiqiang Guo、Peter Li 和 Allen Riddell。Stan：一种概率编程语言。《统计软件杂志》，76(1)，2017 年。

Erjia Cui、Andrew Leroux、Ekaterina Smirnova 和 Ciprian M Crainiceanu。纵向函数模型的快速单变量推断。《计算与图形统计学杂志》，31 (1):219–230，2022a。

Erjia Cui、E Christi Thompson、Raymond J Carroll 和 David Ruppert。身体活动的半参数风险评分。《医学统计学》，41(7):1191–1204，2022b。

Bernd Fitzenberger、Roger Koenker 和 José AF Machado。分位数回归的经济学应用。施普林格科学与商业媒体，2001 年。

Alan E Gelfand 和 Athanasios Kottas。使用狄利克雷过程混合模型进行完整非参数贝叶斯推断的计算方法。计算与图形统计学杂志，11(2):289–305，2002年。

Andrew Gelman、Aki Vehtari、Daniel Simpson、Charles C Margossian、Bob Carpenter、Yuling Yao、Lauren Kennedy、JonahGabry、Paul-Christian Bürkner 和 Martin Modrák。贝叶斯工作流程。*arXiv* 预印本 *arXiv:2011.01808*，2020年。

Hyokyoung G Hong、David C Christiani 和 Yi Li。用于现代癌症研究的生存数据分位数回归：拓展精准医学的统计工具。《精准临床医学》，2（2）：90–99，2019。

Marta Karas、Jiawei Bai、Marcin Strączkiewicz、Jaroslaw Harezlak、Nancy W Glynn、Tamara Harris、Vadim Zipunnikov、Ciprian Crainiceanu 和 Jacek K Urbanek。健康研究中的加速度测量数据：挑战与机遇。《生物科学统计》，11（2）：210–237，2019。Nadja Klein 和 Thomas Kneib。结构化加性分布回归中方差参数的尺度依赖先验。《贝叶斯分析》，11（4）：1071–1106，2016。Genya Kobayashi、Taeyoung Roh、Jangwon Lee 和 Taeryon Choi。在广义非对称拉普拉斯分布的狄利克雷过程混合下，结合形状约束的灵活贝叶斯分位数曲线拟合。《加拿大统计学杂志》，49（3）：698–730，2021。Roger Koenker。分位数回归：四十年回顾。《经济学年度评论》，9：155–176，2017。Roger Koenker、Victor Chernozhukov、Xu ming He 和 Limin Peng。《分位数回归手册》。CRC出版社，2017。Hideo Kozumi 和 Genya Kobayashi。贝叶斯分位数回归的吉布斯抽样方法。《统计计算与模拟杂志》，81（11）：1565–1578，2011。Tony Lancaster 和 Sung Jae Jun。贝叶斯分位数回归方法。《应用计量经济学杂志》，25（2）：287–307，2010。Stefan Lang 和 Andreas Brezger。贝叶斯P样条。《计算与图形统计学杂志》，13（1）：183–212，2004。Pedro S Martins 和 Pedro T Pereira。教育能否减少工资不平等？来自16个国家的分位数回归证据。《劳动经济学》，11（3）：355–371，2004。R Core Team。*R*：一种用于统计计算的语言和环境。R统计计算基金会，奥地利维也纳，2022年。网址：https://www.R-project.org/。

Mohammad Arshad Rahman 和 Shubham Karnawat。《有序模型中的灵活贝叶斯分位数回归》。载于识别、有限因变量、部分可观测性、实验与灵活建模专题：*B* 部分。Emerald出版有限公司，2019年。

Brian J Reich 和 Luke B Smith。删失数据的贝叶斯分位数回归。《生物计量学》，69 (3):651–660，2013。Jayaram Sethuraman。狄利克雷先验的构造性定义。《中华统计学志》，第 639–650 页，1994。Carmen D Tekwe、Mengli Zhang、Raymond J Carroll、Yuanyuan Luan、Lan Xue、Roger S Zoh、Stephen J Carter、David B Allison 和 Marco Geraci。存在测量误差时稀疏函数型分位数回归的估计：一种 SIMEX方法。《生物统计学》，23(4):1218–1241，05

2022。国际标准连续出版物号 1465-4644。doi: 10.1093/biostatistics/kxac017。URLhttps://doi.org/10.1093/biostatistics/kxac017。

Ying Wei、Anneli Pere、Roger Koenker 和 Xuming He。用于参考生长图表的分位数回归方法。《医学统计学》，25(8):1369–1382，2006。

Yifei Yan 和 Athanasios Kottas。贝叶斯分位数回归的一类新误差分布。*arXiv* 预印本 *arXiv:1701.05666*，2017。

Yunwen Yang、Huixia Judy Wang 和 Xuming He。采用非对称拉普拉斯似然的贝叶斯分位数回归后验推断。《国际统计评论》，84(3):327–344，2016。

Keming Yu 和 Rana A Moyeed。贝叶斯分位数回归。《统计与概率快报》，54 (4):437–447，2001。

Roger S Zoh、Yuanyuan Luan 和 Carmen Tekwe。采用工具变量的含测量误差全贝叶斯半参数标量对函数回归（sofr）。*arXiv* 预印本 *arXiv:2202.00711*，2022。