arXiv:2602.06360v1 [math.ST] 6 Feb 2026

条件异方差时间序列模型中的稳健贝叶斯估计

Jeongho Lee[1] 和 Junmo Song[1]

[1]庆北国立大学统计学系

2026年2月9日

摘要

离群值会使似然函数过度敏感，从而严重扭曲统计推断，损害贝叶斯估计的可靠性。为解决这一问题，我们将密度幂散度（DPD）框架扩展至贝叶斯设定，进而提出一种适用于条件异方差时间序列模型的稳健贝叶斯估计方法。所得的基于DPD的后验分布由调节参数控制，在效率与稳健性之间实现平滑平衡。我们建立了所提估计量的渐近性质；具体而言，证明基于DPD的后验分布满足 Bernstein–von Mises 型定理，并收敛于以最小DPD估计量（MDPDE）为中心的正态分布。此外，相应的贝叶斯估计量，即基于DPD的后验分布下的后验均值（EDPE），在渐近意义上等价于MDPDE。基于GARCH(1,1)模型的蒙特卡罗模拟证实，所提出的EDPE在未污染和污染数据下均表现良好；当普通贝叶斯估计量出现严重偏差时，它仍能保持稳健性。对BTC-USD收益的实证应用进一步展现了所提出的稳健贝叶斯框架在金融时间序列分析中的实际优势。

关键词和短语：稳健性，离群值，伪后验，密度幂散度，条件异方差时间序列模型，GARCH模型。

1 引言

离群值的存在会显著影响统计推断的准确性和可靠性。离群值广泛出现在金融、经济学和环境研究等各类应用中。为减轻其影响，研究人员已在频率学派框架下广泛研究稳健统计方法（例如，见 Huber（1981）；Ronchetti（2021）；Loh（2024））。这类研究的一项公认结论是，基于似然的推断尤其

容易受到离群观测值的影响。由于似然函数在贝叶斯更新中发挥着核心作用，这种敏感性自然也会延伸到贝叶斯推断。因此，尽管贝叶斯推断因其优良性质而得到广泛应用，但基于普通后验分布、采用二次损失的贝叶斯估计量仍极易受到数据污染的影响。

在贝叶斯分析中，实现稳健性的一种传统策略是修改假定的数据生成机制，通常通过为模型误差采用重尾分布来实现（例如，参见 Andrade 和 O'Hagan（2011）；O'Hagan 和 Pericchi（2012））。虽然这类方法可以减轻极端观测的影响，但在污染程度较轻或不存在污染时，可能会造成不可忽略的效率损失。受这一权衡的启发，越来越多的研究尝试通过构造伪后验分布，将稳健频率学派思想引入贝叶斯框架。值得关注的例子包括基于 Hellinger 距离的稳健贝叶斯程序（Hooker 和 Vidyashankar，2014）、基于密度幂散度的稳健贝叶斯程序（Ghosh 和 Basu，2016），以及基于 γ-散度的稳健贝叶斯程序（Nakagawa 和 Hashimoto，2020）。此外，还有一些相关方法在无似然或近似贝叶斯框架中采用稳健估计方程，例如 Ruli 等人（2020）提出的基于 ABC 的稳健后验。这些方法共同构造出的贝叶斯估计量在模型成立时仅造成轻微的效率损失，同时在存在离群值时表现稳定。

尽管取得了这些进展，且时间序列模型已广泛应用于各个领域，但现有的大多数稳健贝叶斯方法都是针对独立同分布数据或回归类模型开发的。相比之下，针对时间序列数据的稳健贝叶斯推断仍未得到充分探索，尽管时间依赖性和动态模型结构可能会显著放大离群值的影响。然而，将上述贝叶斯方法扩展到时间序列领域并非易事。这主要是因为这需要谨慎处理条件似然、依赖结构，以及许多时间序列模型所固有的递归成分。

本研究通过构建一种基于散度的稳健贝叶斯框架，应用于条件异方差时间序列模型，弥补了这一研究空白。具体而言，我们将 Ghosh 和 Basu（2016）的密度幂散度（DPD）方法推广至一大类模型，这些模型的特点是条件方差随时间变化。DPD 最初由 Basu 等人（1998）提出，在 Kullback–Leibler 散度与 L_2 距离之间提供平滑过渡，从而为平衡效率与稳健性提供了一种原则性机制。这种基于散度的方法已成功用于在频率学派框架下构建稳健估计量（例如，参见 Fujisawa和Eguchi（2008）；Lee和Song（2009）；Ghosh 和 Basu（2013）；Song（2017））。Ghosh 和 Basu（2016）还表明，在独立同分布情形下，由 DPD 导出的伪后验继承了这些理想性质。基于这一见解，我们将基于 DPD 的贝叶斯方法应用于时间序列数据，在平稳性和遍历性条件下确立其大样本性质，并通过模拟研究以及对金融时间序列数据的

实证分析证明其有效性。

本文的主要贡献有三点。首先，我们为条件异方差时间序列模型构建了基于 DPD 的伪后验分布，并将相应的贝叶斯估计量定义为后验均值。其次，在适当的正则条件下，我们证明了相应的频率学派估计量和基于 DPD 的后验分布均具有一致性和渐近正态性，并得到了伯恩斯坦–冯·米塞斯类型的结果。第三，通过蒙特卡罗实验以及对比特币收益率数据的应用，我们表明，所提出的方法在无污染设定下仅造成轻微的效率损失，同时显著提升了稳健性。

本文其余部分安排如下。第2节介绍所提出的基于 DPD 的贝叶斯估计程序，并给出主要理论结果。第3节报告评估有限样本表现的蒙特卡罗研究结果。第4节将该方法应用于比特币收益率数据，第5节总结全文。

2 主要结果

考虑参数为 $\theta \in \mathbb{R}^d$ 的以下时间序列模型：

$$X_t = \sigma_t(\theta)\,\epsilon_t, \tag{1}$$

其中，$\sigma_t^2(\theta)$表示给定 $\mathcal{F}_{t-1} = \sigma(X_s \mid s \le t-1)$时的条件方差，$\{\epsilon_t \mid t \in \mathbb{Z}\}$ 是一个独立同分布随机变量序列，其密度函数为 f_ϵ，均值为零、方差为一。我们假设，由上述模型在真实参数 θ_0 下生成的 $\{X_t \mid t \in \mathbb{Z}\}$ 严格平稳且遍历。广泛的一类平稳时间序列模型都属于这一框架，包括标准GARCH模型，以及幂变换和阈值GARCH模型等非线性和非对称扩展。

基于观测值 $\{X_1, \cdots, X_n\}$估计模型时，由于初始值问题，通常无法显式求得条件方差$\sigma_t^2(\theta)$。由于 $\sigma_t^2(\theta)$通常由递推方程定义，因此可以通过递归并选取适当的初始值来近似求得。例如，标准GARCH模型可参见Francq和Zako¨ıan（2004），非线性GARCH模型可参见Hamadeh和Zako¨ıan（2011）。本研究假设，使用这种递归方法可获得 $\sigma_t^2(\theta)$的代理。下文将所得的近似过程记为 $\{\tilde{\sigma}_t^2(\theta) \mid t = 1, \cdots, n\}$。

利用条件方差代理构造的可行似然为

$$\tilde{L}(\theta \mid \mathbf{X}_n) = \prod_{t=1}^{n} \tilde{f}_\theta(X_t \mid \mathcal{F}_{t-1}),$$

其中 $\mathbf{X}_n = (X_1, \cdots, X_n)$ 与 $\tilde{f}_\theta(x|\mathcal{F}_{t-1}) = \frac{1}{\tilde{\sigma}_t(\theta)} f_\epsilon\left(\frac{x}{\tilde{\sigma}_t(\theta)}\right)$。后验分布对应-

于模型 (1) 则由下式给出

$$\pi(\theta \mid \mathbf{X}_n) \propto \tilde{L}(\theta \mid \mathbf{X}_n)\, \pi(\theta), \tag{2}$$

其中， $\pi(\theta)$ 表示一个先验分布。在平方误差损失下，相应的贝叶斯估计量称为期望普通后验估计量（EOPE），其表达式为

$$\hat{\theta}_n^{EOPE} = \int \theta\, \pi(\theta \mid \mathbf{X}_n)\, d\theta. \tag{3}$$

如引言所述，离群值可能严重扭曲似然，导致后验分布不可靠，进而使贝叶斯估计量产生偏差。下面的模拟研究展示了这一影响。为解决这一问题，我们提出一种基于散度的方法来构造稳健贝叶斯估计量。

2.1 基于密度幂散度的贝叶斯估计

本小节回顾 Ghosh 和 Basu（2016）提出的基于 DPD 的稳健贝叶斯估计方法，并将其扩展到时间序列模型（1）。

Basu 等人（1998）提出的 DPD 是一种散度度量，用于量化两个概率密度函数之间的差异。具体而言，对于概率密度函数 f 和 g，它们之间的 DPD 定义如下：

$$d_\gamma(g, f) = \begin{cases} \displaystyle\int \Big\{ f^{1+\gamma}(x) - \Big(1 + \frac{1}{\gamma}\Big)\, g(x)\, f^{\gamma}(x) + \frac{1}{\gamma}\, g^{1+\gamma}(x) \Big\} dx & , \gamma > 0, \\ \displaystyle\int g(x) \big\{ \log g(x) - \log f(x) \big\} dx & , \gamma = 0. \end{cases}$$

考虑一族参数密度 $\{f_\theta \mid \theta \in \Theta\}$，设 $X_1, \cdots, X_n$ 是来自真实密度函数 f_{θ_0} 的随机样本。最小 DPD估计量（MDPDE）定义为散度 $d_\gamma(f_{\theta_0}, f_\theta)$ 的经验形式的最小化者。具体而言，

$$\hat{\theta}_{\gamma,n} = \underset{\theta \in \Theta}{\operatorname{argmax}}\, Q_n^{(\gamma)}(\theta), \tag{4}$$

其中

$$Q_n^{(\gamma)}(\theta) = \begin{cases} \displaystyle\frac{1}{\gamma} \sum_{t=1}^{n} f_\theta^{\gamma}(X_t) - \frac{n}{1+\gamma} \int f_\theta^{1+\gamma}(x) dx & , \gamma > 0, \\ \displaystyle\sum_{t=1}^{n} \log f_\theta(X_t) & , \gamma = 0. \end{cases}$$

调节参数 γ 控制效率与稳健性之间的权衡。具体而言，

众所周知，当 γ 接近 0 时，MDPDE 在保持高效率的同时也具有稳健性。我们还注意到，$\gamma = 0$ 对应的 MDPDE 即为最大似然估计量。

基于上述内容，Ghosh和Basu（2016）通过在贝叶斯公式中用 $Q_n^{(\gamma)}$替换对数似然，引入了所谓的 γ稳健化后验密度。具体而言，其稳健后验分布按如下方式得到：

$$\pi_\gamma(\theta \mid \mathbf{X}_n) \propto \exp\left(Q_n^{(\gamma)}(\theta)\right)\pi(\theta),$$

其中 $\pi(\theta)$是 θ的先验分布。注意，当 $\gamma = 0$时，上述后验分布退化为普通后验分布。

注1. 可以证明，对于每个固定的 $\theta \in \Theta$,

$$Q_n^{(\gamma)}(\theta) = \frac{n}{\gamma} + \sum_{t=1}^{n} \log f_\theta(X_t) - n + o(1) \qquad as\ \gamma \downarrow 0.$$

因此，我们有

$$Q_n^{(\gamma)}(\theta) - \frac{n}{\gamma} \longrightarrow \sum_{t=1}^{n} \log f_\theta(X_t) - n,$$

在 θ上逐点成立。由于后验密度仅在相差一个归一化常数的意义下定义，发散项 n/γ 与加性常数 $-n$ 在归一化意义下不影响后验密度。因此，上述展开在启发式层面上表明，当 $\gamma \downarrow 0$时，基于*DPD*的后验分布平滑地连接到普通贝叶斯后验分布。然而，上述逐点收敛不足以证明基于*DPD*的后验分布与普通后验分布具有相似的分布形状。这种解释需要更强的收敛概念，例如未归一化后验密度的一致收敛，或者更强地，全变差收敛。对于时间序列模型(1)，这些问题在适当条件下由下文的定理*1*加以处理。

这种类型的伪后验也可以通过将适用于独立同分布情形的 DPD 扩展到相关数据，进而在时间序列模型中推导出来。该扩展通过以下密度功效散度的条件版本实现，该版本比较两个条件密度 $f_{\theta_0}(\cdot \mid \mathcal{F}_{t-1})$ 和 $f_\theta(\cdot \mid \mathcal{F}_{t-1})$：

$$d_\gamma\left(f_{\theta_0}(\cdot \mid \mathcal{F}_{t-1}), f_\theta(\cdot \mid \mathcal{F}_{t-1})\right)$$

$$= \begin{cases} \displaystyle\int \left\{ f_\theta^{1+\gamma}(x|\mathcal{F}_{t-1}) - \left(1+\frac{1}{\gamma}\right) f_{\theta_0}(x|\mathcal{F}_{t-1}) f_\theta^{\gamma}(x|\mathcal{F}_{t-1}) + \frac{1}{\gamma} f_{\theta_0}^{1+\gamma}(x|\mathcal{F}_{t-1}) \right\} dx & ,\gamma > 0 \\ \displaystyle\int f_{\theta_0}(x|\mathcal{F}_{t-1}) \big\{ \log f_{\theta_0}(x|\mathcal{F}_{t-1}) - \log f_\theta(x|\mathcal{F}_{t-1}) \big\} dx & ,\gamma = 0. \end{cases}$$

与式（4）中 MDPDE 的推导方式类似，可以推导出模型（1）的 MDPDE，如下：

所示：

$$\hat{\theta}_{\gamma,n} = \underset{\theta\in\Theta}{\operatorname{argmax}}\ \widetilde{Q}_n^{(\gamma)}(\theta), \tag{5}$$

其中

$$\widetilde{Q}_n^{(\gamma)}(\theta) = \begin{cases} \dfrac{1}{\gamma}\displaystyle\sum_{t=1}^{n} \tilde{f}_\theta^{\gamma}(X_t|\mathcal{F}_{t-1}) - \dfrac{1}{1+\gamma}\sum_{t=1}^{n}\int \tilde{f}_\theta^{1+\gamma}(x|\mathcal{F}_{t-1})dx & ,\gamma > 0, \\ \displaystyle\sum_{t=1}^{n} \log \tilde{f}_\theta(X_t|\mathcal{F}_{t-1}) & ,\gamma = 0 \end{cases} \tag{6}$$

（参见 Song（2021）中的第 3.1 小节）。使用上述定义的目标函数 $\widetilde{Q}_n^{(\gamma)}(\theta)$，时间序列模型（1）的 γ-稳健后验密度为

$$\pi_\gamma(\theta \mid \mathbf{X}_n) \propto \exp\left(\widetilde{Q}_n^{(\gamma)}(\theta)\right)\pi(\theta).$$

相应地，EOPE（3）的 DPD 类比量（称为期望 DPD 后验估计量（EDPE））定义为

$$\hat{\theta}_{\gamma,n}^{EDPE} = \int \theta\, \pi_\gamma(\theta \mid \mathbf{X}_n)d\theta.$$

备注 2. $\widetilde{Q}_n^{(\gamma)}(\theta)$ 中包含的积分可能会带来计算负担。不过，当 f_ϵ 的支撑集为 $\mathbb{R}$ 时，我们可以预先处理该积分。具体而言，注意到

$$\int \tilde{f}_\theta^{1+\gamma}(x|\mathcal{F}_{t-1})dx = \frac{1}{\tilde{\sigma}_t^{\gamma}(\theta)}\int f_\epsilon^{1+\gamma}(x)dx.$$

由于积分 $\int f_\epsilon^{1+\gamma}(x)\,dx$ 不涉及任何参数，因此可以针对每个 γ 预先进行解析或数值计算。例如，若 ϵ_t 服从 $N(0,1)$，该积分的计算结果为$(2\pi)^{-\gamma/2}(1+\gamma)^{-1/2}$。使用此预先计算的值代替积分，可以显著降低计算负担。

如前所述，调节参数 γ 控制 MDPDE 对离群值的稳健性。同理，γ 也决定了基于DPD的后验分布以及由此得到的上述 EDPE 的稳健性。这一特性源于基于DPD的后验分布与普通贝叶斯后验分布之间的联系。为更精确地确立这种联系，我们提出以下正则条件。

C1. 创新密度 f_ϵ 在 $\mathbb{R}$上严格为正。

C2. 对于每个 x 和 t，$f_\theta(x \mid \mathcal{F}_{t-1})$ 关于 θ 连续。

C3. 对于所有充分小的 $\gamma > 0$,

$$\sup_{\theta\in\Theta}\left|\widetilde{Q}_n^{(\gamma)}(\theta) - \sum_{t=1}^{n}\log\tilde{f}_\theta(X_t \mid \mathcal{F}_{t-1})\right| \le V_n\gamma,$$

其中， $V_n > 0$ 是一个不涉及 θ的随机变量。

C4. 先验密度 $\pi(\theta)$在紧参数空间 Θ上连续。

C5. 对于某个 $\theta^* \in \Theta$， $\pi(\theta^*) > 0$。

条件**C1** 是 GARCH 等时间序列模型中的标准条件，其创新密度通常假定在 $\mathbb{R}$ 上具有完整支撑集（例如高斯分布或 Student-t 分布）。该条件˜ ˜确保 $f_\theta(x \mid \mathcal{F}_{t-1})$和 $L(\theta \mid \mathbf{X}_n)$严格为正。条件**C5** 是确保普通后验分布 $\pi(\theta \mid \mathbf{X}_n)$的归一化常数为正所必需的。结合似然 $L(\theta \mid \mathbf{X}_n)$在 Θ 上为正这一条件，可保证极限后验分布的归一化常数严格为正。

定理1. 假设条件**C1– C4** 成立，且参数空间 Θ 是紧集。那么，未归一化后验密度满足如下的一致收敛

$$\sup_{\theta\in\Theta}\left|\exp\big(\widetilde{Q}_n^{(\gamma)}(\theta)\big)\pi(\theta) - \tilde{L}(\theta \mid \mathbf{X}_n)\pi(\theta)\right| \longrightarrow 0 \quad as\ \gamma \downarrow 0.$$

此外，若条件 **C5** 成立，则基于*DPD*的后验分布 $\pi_\gamma(\theta \mid \mathbf{X}_n)$ 在全变差意义下收敛到普通后验分布 $\pi(\theta \mid \mathbf{X}_n)$，即

$$\frac{1}{2}\int_\Theta |\pi_\gamma(\theta \mid \mathbf{X}_n) - \pi(\theta \mid \mathbf{X}_n)|\, d\theta \longrightarrow 0 \quad as\ \gamma \downarrow 0.$$

这些极限结果表明，基于DPD的后验分布构成了对普通贝叶斯后验分布的一种连续稳健化。当 γ 非常接近零时，$\pi_\gamma(\theta \mid \mathbf{X}_n)$ 与普通后验分布 $\pi(\theta \mid \mathbf{X}_n)$ 仍高度相似，因此能够保持较高的效率。同时，任何 $\gamma > 0$ 都会降低极端观测的影响，从而提供类似于MDPDE的稳健性。因此，调节参数 γ 在贝叶斯推断中也能平滑地权衡效率与稳健性。

作为定理1的直接推论，后验分布在全变差意义下的收敛，加上 Θ 的紧性，意味着相应后验期望也会收敛。由此得到以下推论。

推论 1. 在定理*1*的条件下，我们有

$$\hat{\theta}_{\gamma,n}^{EDPE} = \int_\Theta \theta\, \pi_\gamma(\theta \mid \mathbf{X}_n)\, d\theta \ \longrightarrow\ \hat{\theta}_n^{EOPE} = \int_\Theta \theta\, \pi(\theta \mid \mathbf{X}_n)\, d\theta \quad as\ \gamma \downarrow 0.$$

注3. 尽管定理*1*中的条件 **C3** 是针对目标函数$\widetilde{Q}_n^{(\gamma)}(\theta)$表述的，但在实际应用中，使用其平移版本通常更为方便。具体而言，对于任意与 θ无关的函数 $K_n(\gamma)$，定义

$$\widetilde{Q}_{n,K}^{(\gamma)}(\theta) := \widetilde{Q}_n^{(\gamma)}(\theta) - K_n(\gamma).$$

那么， $\widetilde{Q}_{n\,K}^{(\gamma)}(\theta)$ 和 $\widetilde{Q}_n^{(\gamma)}(\theta)$ 会导出相同的后验分布。因此，我们有

$$\pi_\gamma(\theta \mid \mathbf{X}_n) \propto \exp\big(\widetilde{Q}_n^{(\gamma)}(\theta)\big)\pi(\theta) \propto \exp\big(\widetilde{Q}_{n,K}^{(\gamma)}(\theta)\big)\pi(\theta),$$

因此，添加或减去一个与 θ 无关的项，不会改变所得的后验分布。所以，在验证给定模型的条件 ***C3*** 时，可以等价地对$\widetilde{Q}_{n\,K}^{(\gamma)}(\theta)$ 验证该条件，而不是直接对 $\widetilde{Q}_n^{(\gamma)}(\theta)$ 进行验证。

注释4. 调节参数 γ 所起的作用与模型参数 θ有着根本区别。 θ 的各个分量与数据生成动态相关，并具有真实的潜在取值，而参数 γ 并不存在真实取值。相反，对于给定的数据集，只有相对于特定的推断或预测目标，才能定义 γ 的最优选择；这反映了程序层面在稳健性与效率之间的权衡。

因此，从贝叶斯视角来看，可以将 γ 视为决策参数或设计参数，而非真正的模型参数。尽管理论上可以为 γ 指定先验分布，并与 θ一起进行估计，但这样做会改变问题本身的性质，因为由此得到的后验分布将不再对应于对数据生成过程参数的推断。此外，下一小节中确立的伯恩斯坦–冯·米塞斯类型结果，关键依赖于将 γ 视为固定调优常数时 *MDPDE* 的频率学派渐近正态性。出于这些原因，并为了保持渐近分析清晰的理论解释，在整个理论推导过程中，我们将 γ 视为固定值。

备注5。 在实证应用中，选择合适的 γ 值是一项重要的实际问题。在关于最小距离概率密度估计的频率学派文献中， γ 通常被视为一个用于衡量稳健性的干扰参数，其取值需在效率与稳健性之间权衡；例如，参见*Warwick*和*Jones*（*2005*）以及*Fujisawa*和*Eguchi*（*2008*）。实践中通常建议采用相对较小的 γ 值，因为当污染程度有限时，过大的取值可能会导致效率大幅下降。更重要的是，这类γ 值通常足以在稳健性与效率之间实现良好的平衡；例如，参见*Lee*和*Song*（*2009*）以及*Song*（*2017*）。

从贝叶斯决策理论视角来看，对于给定的数据集，也可以通过最小化样本外预测损失来定义 γ 的最优值。具体而言，基于一步前瞻波动率预测的均方根误差来选择 γ，提供了一种简单而有效的数据驱动规则，它直接反映了预测性能方面的稳健性与效率之间的权衡。在实际

第*4*节的数据分析中，我们采用这一方法，并通过最小化样本外预测均方根误差来选择 γ ，由此得到的 γ 取值位于 $[0.2, 0.3]$范围内，与现有文献中的实证发现一致。

2.2 渐近性质

为建立 $\hat{\theta}_{\gamma n}^{EDPE}$ 的渐近性质，我们施加正则条件。为此，令 $f_\theta(x|\mathcal{F}_{t-1})$ 表示将 $f_\theta(x|\mathcal{F}_{t-1})$ 中的 $\tilde{\sigma}_t(\theta)$ 替换为 $\sigma_t(\theta)$ 后得到的不可行条件密度。尽管在独立同分布情形下的式（4）中使用了相同的记号 $Q_n^{(\gamma)}(\theta)$，但现在我们用$Q_n^{(\gamma)}(\theta)$表示式（5）中 $\widetilde{Q}_n^{(\gamma)}$的对应形式，即将 $\tilde{f}_\theta(x|\mathcal{F}_{t-1})$ 替换为 $f_\theta(x|\mathcal{F}_{t-1})$。此外，令$\tilde{q}_t^{(\gamma)}(\theta)$ 为满足 $\widetilde{Q}_n^{(\gamma)}(\theta) = \sum_{t=1}^n \tilde{q}_t^{(\gamma)}(\theta)$ 的函数，并以相同方式定义 $q_t^{(\gamma)}(\theta)$。下文中，$\|\cdot\|$ 表示矩阵和向量的 l_1 范数。为方便记号，我们分别用∂_θ 和 $\partial^2_{\theta\theta'}$ 表示 $\frac{\partial}{\partial\theta}$ 和 $\frac{\partial^2}{\partial\theta\partial\theta'}$。现在介绍以下条件：

A1. $q_t^{(\gamma)}(\theta)$ 是连续的关于 θ，且对于每个 $\theta \in \Theta$，$\{q_t^{(\gamma)}(\theta)|t \in \mathbb{Z}\}$ 严格平稳且遍历

A2. $\mathrm{E}\sup_{\theta\in\Theta}|q_t^{(\gamma)}(\theta)| < \infty$ 和 $\frac{1}{n}\sum_{t=1}^n \sup_{\theta\in\Theta}|q_t^{(\gamma)}(\theta) - \tilde{q}_t^{(\gamma)}(\theta)| = o(1)\ a.s.$

A3. $\{\partial_\theta q_t^{(\gamma)}(\theta_0)\ \mathcal{F}_t \mid t \in \mathbb{Z}\}$ 是鞅差序列。

A4. $\frac{1}{\sqrt{n}}\sum_{t=1}^n \left\|\partial_\theta\, q_t^{(\gamma)}(\theta_0) - \partial_\theta\, \tilde{q}_t^{(\gamma)}(\theta_0)\right\| = o(1)\ a.s.$

A5. 对于 θ_0 的某个邻域 $N_1(\theta_0)$,

$$\frac{1}{n}\sum_{t=1}^n \sup_{\theta\in N_1(\theta_0)} \left\|\partial^2_{\theta\theta'} q_t^{(\gamma)}(\theta) - \partial^2_{\theta\theta'}\tilde{q}_t^{(\gamma)}(\theta)\right\| = o(1) \quad a.s.$$

A6. 对于 θ_0 的某个邻域 $N_2(\theta_0)$,

$$\mathrm{E}\sup_{\theta\in N_2(\theta_0)} \left\|\partial^2_{\theta\theta'} q_t^{(\gamma)}(\theta)\right\| < \infty.$$

A7. 矩阵 $\mathcal{J}_\gamma = -\mathrm{E}\left[\partial^2_{\theta\theta'} q_t^{(\gamma)}(\theta_0)\right]$ 和 $\mathcal{I}_\gamma = \mathrm{E}\left[\partial_\theta q_t^{(\gamma)}(\theta_0)\partial_{\theta'} q_t^{(\gamma)}(\theta_0)\right]$ 存在，且 $\mathcal{J}_\gamma$ 为正定矩阵。

假设 **A1**–**A2** 确保目标函数满足一致大数定律，并控制基于 $f_\theta(x|\mathcal{F}_{t-1})$的不可行准则与基于 $f_\theta(x|\mathcal{F}_{t-1})$的可行准则之间的差异。假设 **A3**–**A4** 提供了可导出中心极限定理的鞅差结构。假设 **A5**–**A6** 是建立二阶导数 $\partial^2_{\theta\theta'}\widetilde{Q}_n^{(\gamma)}(\theta)$一致收敛的技术条件。

这些假设与用于建立相依时间序列模型中M估计量的一致性和渐近正态性的标准正则条件十分接近；例如参见 Ling and McAleer（2010）。对于时间序列情境下基于DPD的M估计，Song（2021）也采用了类似条件。重要的是，以下所述伯恩斯坦–冯·米塞斯类型结果的证明同样使用了这组假设。

ˆ 我们首先从频率学派的角度证明 MDPDE 的一致性和渐近正态性$\theta_{\gamma,\ n}$ 。随后转向贝叶斯对应方法，研究基于DPD的后验分布的渐近行为。

ˆ**定理2**。假设条件**A1** 和**A2** 成立。那么，$\theta_{\gamma,\ n}$ 几乎必然收敛到θ_0。此外，若条件**A3–A7** 也成立，则可得

$$\sqrt{n}(\hat{\theta}_{\gamma,n} - \theta_0) \ \xrightarrow{d}\ N_d(0, \mathcal{J}_\gamma^{-1}\mathcal{I}_\gamma\mathcal{J}_\gamma^{-1}).$$

定理3。假设条件**A1**、**A2**和**A7** 成立，并令 $\pi(\theta)$为在 θ_0处取正值且连续的任意先验密度。则有

$$\int_{\mathbb{R}^d} \left|\pi_\gamma(t|\mathbf{X}_n) - \frac{|\mathcal{J}_\gamma|^{1/2}}{(2\pi)^{d/2}} e^{-\frac{1}{2}t'\mathcal{J}_\gamma t}\right| dt \ \xrightarrow{a.s.}\ 0,$$

ˆ其中，$\pi_\gamma(t|\mathbf{X}n)$ 是给定 $\mathbf{X}_n$时 $t = \sqrt{n}(\theta - \theta_{\gamma,\ \bar{}n})$ 的 γ-后验密度。

定理3表明，基于DPD的后验分布满足 Bernstein–von Mises型定理。特别是，该结果意味着 EDPE 与 MDPDE 渐近等价，并具有相同的渐近分布，下一定理将对此作出形式化表述。

定理4。 除定理*3*的假设外，假设 $\int \|\theta\|\pi\ (\theta)\ d\theta < \infty$。则有

$$\sqrt{n}(\hat{\theta}_{\gamma,n}^{EDPE} - \hat{\theta}_{\gamma,n}) \ \xrightarrow{a.s.}\ 0.$$

此外，如果假设 **A3–A6** 也满足，则可得

$$\sqrt{n}(\hat{\theta}_{\gamma,n}^{EDPE} - \theta_0) \ \xrightarrow{d}\ N_d(0, \mathcal{J}_\gamma^{-1}\mathcal{I}_\gamma\mathcal{J}_\gamma^{-1}).$$

2.3 GARCH模型的应用

在本小节中，我们验证在标准正则条件下，上一小节提出的条件和假设对 GARCH(p, q)模型均成立。考虑以下 GARCH(p, q)模型：

$$\begin{aligned} X_t &= \sigma_t(\theta)\,\epsilon_t, \quad \epsilon_t \sim\ i.i.d.\ N(0,1), \\ \sigma_t^2(\theta) &= \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2(\theta), \end{aligned} \tag{7}$$

其中 $\theta = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)'$，其中 $\omega > 0$ 且 $\alpha_i, \beta_j \geq 0$。我们假设由 (7) 在真实参数 θ_0 下生成的过程$\{X_t \mid t \in \mathbb{Z}\}$ 是严格平稳且遍历的。众所周知，对于 GARCH模型，严格平稳性和遍历性成立，当且仅当最大李雅普诺夫指数严格为负；例如，参见 Francq 和 Zako¨ıan (2004)。我们进一步假设参数空间 Θ 是 $(0, \infty) \times [0, \infty)^{p+q}$的一个紧子集。

令 $\{\tilde{\sigma}_t^2(\theta)\}_{t=1}^n$ 为条件方差过程 $\{\sigma_t^2(\theta)\}_{t=1}^n$ 的代理过程，递归定义如下

$$\tilde{\sigma}_t^2(\theta) = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \tilde{\sigma}_{t-j}^2(\theta), \tag{8}$$

其中为 $X_0^2 = \cdots = X_{1-p}^2$ 和 $\tilde{\sigma}_0^2 = \cdots = \tilde{\sigma}_{1-q}^2$ 选择适当的初始值。由于

$$\tilde{f}_\theta(x|\mathcal{F}_{t-1}) = \frac{1}{\sqrt{2\pi}\tilde{\sigma}_t(\theta)} \exp\Big(- \frac{x^2}{2\tilde{\sigma}_t^2(\theta)}\Big),$$

模型 (7) 的目标函数 (6) 为

$$\begin{aligned} \widetilde{Q}_n^{(\gamma)}(\theta) &= \begin{cases} \displaystyle\sum_{t=1}^{n} \left(\frac{1}{\sqrt{2\pi}\,\tilde{\sigma}_t(\theta)}\right)^{\gamma} \left\{ \frac{1}{\gamma} \exp\!\left(-\frac{\gamma X_t^2}{2\tilde{\sigma}_t^2(\theta)}\right) - \left(\frac{1}{1+\gamma}\right)^{\frac{3}{2}} \right\} & , \gamma > 0, \\ \displaystyle\sum_{t=1}^{n} \Big(- \log\sqrt{2\pi}\tilde{\sigma}_t(\theta) - \frac{X_t^2}{2\tilde{\sigma}_t^2(\theta)} \Big) & , \gamma = 0. \end{cases} \\ &:= \sum_{t=1}^{n} \tilde{q}_t^{(\gamma)}(\theta). \end{aligned} \tag{9}$$

相应的基于DPD的后验分布为

$$\pi_\gamma(\theta \mid \mathbf{X}_n) \propto \exp\big(\widetilde{Q}_n^{(\gamma)}(\theta)\big)\pi(\theta).$$

为了使定理1适用于上述GARCH模型，我们现在检查与模型有关的条件是否满足。条件 **C1** 和**C2** 显然满足；根据注3，引理 3 证明了条件**C3** 的一个平移版本成立。因此，我们得到以下结果。

命题1. 如果先验分布 $\pi(\theta)$ 满足条件**C4** 和**C5**，那么定理*1*的结论对(7)中定义的*GARCH*模型成立。

上述 GARCH模型通常采用的标准条件如下：

G1. $\sup_{\theta\in\Theta} \sum_{j=1}^{q} \beta_j < 1.$

G2. 若 $q > 0$、 $\mathcal{A}_{\theta_0}(z)$ 和 $\mathcal{B}_{\theta_0}(z)$ 没有公共根， $\mathcal{A}_{\theta_0}(1) \neq 1$，且 $\alpha_{0p} + \beta_{0q} \neq 0$，其中$\mathcal{A}_{\theta}(z) = \sum_{i=1}^{p} \alpha_i z^i$，而 $\mathcal{B}_{\theta}(z) = 1 - \sum_{j=1}^{q} \beta_j z^j$。（按照惯例，若 $p = 0$，则 $\mathcal{A}_{\theta}(z) = 0$；若 $q = 0$，则 $\mathcal{B}_{\theta}(z) = 1$。）

G3. θ_0 位于 Θ的内部。

上述 GARCH模型的定理2已由Lee和Song（2009）证明。因此，为了应用定理3和定理4，仍需验证在条件**G1**–**G3** 下，第2.2小节中的假设**A1**–**A7** 均成立。

与 $\tilde{q}_t^{(\gamma)}(\theta)$对应的函数 $q_t^{(\gamma)}(\theta)$定义为：在式（7）中用 $\sigma_t(\theta)$替换 $\tilde{\sigma}_t(\theta)$。显然， $q_t^{(\gamma)}(\theta)$在 θ 中连续。条件**G1**保证过程 $\{\sigma_t(\theta) \mid t \in \mathbb{Z}\}$的平稳性和遍历性，这意味着假设**A1**的第二部分成立。此外，注意到

$$|q_t^{(\gamma)}(\theta)| \leq \left(\frac{1}{\sqrt{2\pi}}\right)^{\gamma} \left\{\frac{1}{\gamma} + \left(\frac{1}{1+\gamma}\right)^{3/2}\right\} \sup_{\theta \in \Theta} \left(\frac{1}{\omega}\right)^{\gamma/2}$$

并且由于假设 $\Theta \subset (0, \infty) \times [0, \infty)^{p+q}$具有紧性， ω 有正下界，因此假设**A2** 的第一部分成立。第二部分已由Lee和Song（2009）在条件**G1**和**G2**下证明。在附加条件**G3**下，Lee和Song（2009）以及Song和Kang（2021）分别证明了假设**A3**–**A5** 和**A6** 成立。最后，Song（2021）的引理12证明了假设**A7** 成立。因此，我们得到以下命题。

命题2。 假设 *GARCH*(p, q) 模型(7) 满足条件**G1**–**G3**。若先验分布$\pi(\theta)$ 在 θ_0处为正且连续，则定理*3*和定理*4*中所述的渐近结果对 *GARCH*(p, q) 模型成立。

基于 DPD 的目标函数所诱导的后验分布在解析上难以处理，因为感兴趣的后验量无法以闭式求得。因此，后验推断通过马尔可夫链蒙特卡洛（MCMC）方法进行。具体而言，我们建议使用哈密顿蒙特卡洛（HMC）算法及其自适应变体无U型转弯采样器，从基于DPD的后验分布中生成样本。

HMC 利用对数后验 $\nabla_\theta \log \pi_\gamma(\theta \mid \mathbf{X}_n)$的梯度信息，通过模拟哈密顿动力学，高效探索目标分布（参见 Hoffman等人（2014））。这种方法尤其适用于 GARCH类模型，其中条件方差递推（8）会导致参数之间产生复杂的非线性依赖，这可能使标准 MCMC 方案（如随机游走 Metropolis–Hastings）的混合速度变慢。

令 $\{\theta_\gamma^{(1)}, \ldots, \theta_\gamma^{(m)}\}$ 表示由 HMC 算法生成的后验样本。这些样本用于近似感兴趣的后验量，包括贝叶斯估计量，例如

表 1. 无污染设定下，$\theta_1 = (1，0。2，0。4)$ 的 EOPE 和 EDPE 样本均值（括号内为总缩放均方根误差）

			EDPE（γ）						
n	θ_1	EOPE	0.05	0.10	0.20	0.30	0.50	0.75	1.00
500	ω	1.187	1.194	1.201	1.215	1.224	1.234	1.245	1.265
	α	0.217	0.220	0.223	0.231	0.240	0.265	0.298	0.326
	β	0.318	0.314	0.311	0.304	0.299	0.291	0.286	0.285
		(0.975)	(0.956)	(0.941)	(0.932)	(0.939)	(0.991)	(1.109)	(1.228)
1000	ω	1.127	1.141	1.154	1.178	1.195	1.224	1.236	1.248
	α	0.208	0.210	0.212	0.215	0.220	0.233	0.259	0.290
	β	0.346	0.340	0.334	0.323	0.315	0.302	0.294	0.288
		(0.747)	(0.738)	(0.736)	(0.743)	(0.757)	(0.809)	(0.900)	(1.034)
2000	ω	1.080	1.090	1.103	1.127	1.152	1.191	1.219	1.233
	α	0.205	0.206	0.207	0.210	0.213	0.220	0.234	0.257
	β	0.365	0.361	0.355	0.344	0.332	0.314	0.301	0.293
		(0.569)	(0.574)	(0.581)	(0.602)	(0.629)	(0.683)	(0.759)	(0.860)

后验均值

$$\hat{\theta}_{\gamma,n}^{EDPE} = \int \theta \pi_\gamma(\theta \mid \mathbf{X}_n)\, d\theta \approx \frac{1}{m} \sum_{i=1}^{m} \theta_\gamma^{(i)}.$$

由此得到的后验样本及其蒙特卡罗汇总随后用于以下模拟研究中的后验推断、收敛诊断和性能评估。

3 模拟研究

本节将考察所提出的稳健贝叶斯程序与传统贝叶斯方法相比的性能。具体而言，我们研究以下 GARCH(1，1) 模型下 EDPE 和 EOPE 的有限样本行为：

$$\begin{aligned} X_t &= \sigma_t(\theta)\, \epsilon_t, \qquad \epsilon_t \sim i.i.d.\ N(0,1), \\ \sigma_t^2(\theta) &= \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2(\theta), \end{aligned} \tag{10}$$

其中 $\theta = (\omega，\ \alpha，\ \beta)$，且 $\omega > 0$ 和 $\alpha，\ \beta \geq 0$。

我们为 θ 指定一个先验分布，使得联合先验分布可分解为

$$\pi(\theta) = \pi(\omega)\, \pi(\alpha, \beta).$$

尽管为理论发展假设了参数空间的紧性，在前-

表 2. 无污染设定下，$\theta_2 = (1, 0.15, 0.8)$ 的 EOPE 和 EDPE 样本均值，括号内为总缩放均方根误差

			EDPE（γ）						
n	θ_2	EOPE	0.05	0.10	0.20	0.30	0.50	0.75	1.00
500	ω	1.469	1.504	1.539	1.602	1.656	1.721	1.750	1.736
	α	0.171	0.175	0.179	0.188	0.198	0.222	0.256	0.291
	β	0.753	0.748	0.742	0.730	0.717	0.688	0.640	0.583
		(0.984)	(1.003)	(1.026)	(1.100)	(1.193)	(1.398)	(1.680)	(1.957)
1000	ω	1.262	1.297	1.335	1.414	1.495	1.623	1.719	1.749
	α	0.162	0.165	0.168	0.174	0.182	0.201	0.226	0.253
	β	0.774	0.770	0.765	0.755	0.744	0.719	0.685	0.644
		(0.644)	(0.663)	(0.692)	(0.774)	(0.883)	(1.117)	(1.396)	(1.645)
2000	ω	1.140	1.162	1.189	1.253	1.325	1.479	1.630	1.719
	α	0.156	0.158	0.160	0.164	0.170	0.186	0.208	0.229
	β	0.786	0.784	0.781	0.773	0.764	0.742	0.713	0.683
		(0.423)	(0.438)	(0.460)	(0.530)	(0.620)	(0.850)	(1.154)	(1.405)

述一节中，我们在本模拟研究中对 ω 采用截断正态先验：

$$\omega \sim N_{(0,\infty)}(\mu_w, \sigma_w^2),$$

其中，μ_w 和 σ_w^2 表示固定的超参数。该先验分布可视为具有有效紧支撑集的弱信息先验分布，因为高斯尾部衰减迅速，对极大值赋予的先验质量可忽略不计。

对于 ARCH 和 GARCH 系数 α 和 β，我们不依赖基于 Lyapunov 指数为负的一般平稳性与遍历性条件，而是为简便起见施加充分条件 $\alpha + \beta < 1$，该条件可保证 GARCH(1,1) 过程同时满足严格平稳性和遍历性。因此，我们在平稳区域上赋予均匀先验，

$$\pi(\alpha, \beta) \propto I\big(\alpha \geq 0,\ \beta \geq 0,\ \alpha + \beta < 1\big),$$

其中，$I(\cdot)$ 表示示性函数。结合这些选择，联合先验分布为

$$\pi(\theta) \propto \exp\left(-\frac{(\omega - \mu_w)^2}{2\sigma_w^2} \right) I(\omega > 0)\, I(\alpha \geq 0, \beta \geq 0, \alpha + \beta < 1).$$

给定观 由(10)生成的观测$\mathbf{X}_n = (X_1, \cdots, X_n)$，普通后验分布所依据

上述先验分布由下式给出

$$\pi(\theta \mid \mathbf{X}_n) \propto \prod_{t=1}^{n} \frac{1}{\tilde{\sigma}_t(\theta)} \exp\Big(- \frac{X_t^2}{2\tilde{\sigma}_t^2(\theta)} \Big) \pi(\theta),$$

基于DPD的后验分布如下:

$$\pi_\gamma(\theta \mid \mathbf{X}_n) \propto \prod_{t=1}^{n} \exp\left[\left(\frac{1}{\sqrt{2\pi}\tilde{\sigma}_t(\theta)} \right)^{\gamma} \left\{ \frac{1}{\gamma} \exp\left(-\frac{\gamma X_t^2}{2\tilde{\sigma}_t^2(\theta)} \right) - \left(\frac{1}{1+\gamma} \right)^{\frac{3}{2}} \right\} \right] \pi(\theta).$$

代理过程 $\{\tilde{\sigma}_t^2(\theta)\}_{t=1}^n$ 的初始值设为 $X_0^2 = \tilde{\sigma}_t^2(\theta) = X_1^2$。

在本次模拟中，我们考虑以下两种参数配置:

$$\theta_1 = (1, 0.2, 0.4) \quad \text{and} \quad \theta_2 = (1, 0.15, 0.8),$$

其中， $\alpha + \beta$ in θ_1 对应于一个中等持久性的 GARCH 过程，而 θ_2 则表示一个高度持久的过程，其 $\alpha + \beta$ 接近于 1；这一特征在实证应用中十分常见。样本量 n 设为 500、1000 和 2000。我们生成长度为 $n + 1000$ 的样本路径，并丢弃前 1000 个观测值，以减轻初始化效应。

回顾一下，EOPE 和 EDPE 分别定义为普通后验分布和基于 DPD 的后验分布下的后验均值。由于这些后验期望没有闭式表达式，因此使用相应 MCMC 样本的经验均值进行近似。基于上一小节所述原因，后验推断采用结合自适应无转向采样器的 HMC。对于每条样本路径，运行 4 条并行链，每条链进行 500 次预热迭代和 1000 次采样迭代以获得后验样本，并据此计算 EOPE 和 EDPE。所有计算均通过 R 中的 cmdstanr 接口，使用 Stan 概率编程语言完成。通过潜在尺度缩减因子 R 和有效样本量评估收敛情况；在所考察的所有设置下，各链均表现出令人满意的收敛性。

对于每种模拟设置，该过程重复 $R = 200$ 次。所得 EOPE 和 EDPE 的平均值列于表1和表2。为了比较各估计量的表现，我们采用总尺度化均方根误差，其定义与 Bai等（2014）中的定义相同,

$$\text{Total scaled RMSE} = \sum_{j=1}^{3} \frac{1}{|\theta_{0j}|} \sqrt{\frac{1}{R} \sum_{r=1}^{R} \left(\hat{\theta}_j^{(r)} - \theta_{0j} \right)^2},$$

其中， θ_{0j} 表示真实参数向量 θ_0的第 j个分量，$\hat{}\theta_j^{(r)}$ 表示基于第 r次重复计算得到的第 j个参数后验均值的蒙特卡洛估计。正如Bai等（2014）所倡导的，这一准则提供了一个单一的无尺度汇总指标，用于衡量总体估计精度，因此在比较涉及

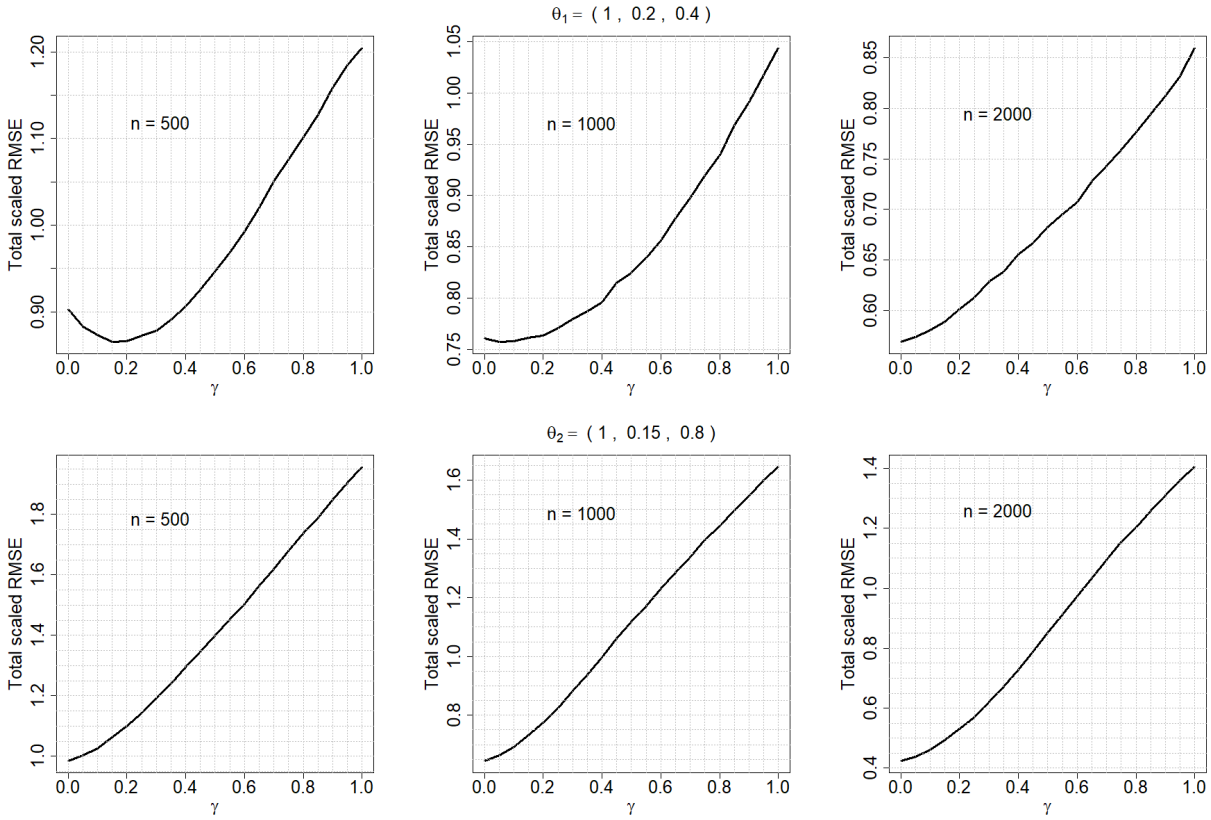


图1. 无污染设定下后验均值关于 γ 的总尺度化均方根误差

数量级差异显著的参数的模型中的估计量时尤其有用。

表1和表2分别报告了无污染设定下关于 θ_1 和 θ_2的模拟结果。对于 θ_2，在所有考察的样本量下，EOPE的总尺度化均方根误差都最小，这与模型设定正确时标准贝叶斯推断的效率一致。然而，对于 θ_1，有限样本下的最小总尺度化均方根误差出现在一个较小的正调节参数处：当 $n = 500$ 时，该值在 $\gamma = 0.2$ 处取得；当 $n = 1000$时，则在 $\gamma = 0.1$ 处取得，而当 $n = 2000$时，EOPE表现最佳。总体而言，总尺度化均方根误差往往随γ增加，但 θ_1 结果表明，在较小的 n下，采用较小 γ 的EDPE表现可与EOPE相当（或略优），这或许反映的是有限样本和蒙特卡洛变异，而非真正的效率提升。随着 n 从500增加到2000，所有估计量的均方根误差值均有所下降，且优势转向EOPE，这支持了真实模型下预期的渐近效率。图1直观地印证了这些模式。

为考察所提方法抵御数据污染的稳健性，我们将创新值异常值纳入模型。具体而言，我们不采用标准正态创新值，而是按照以下方案生成受污染的创新项：

$$\epsilon_t = \epsilon_t^{(0)} + s \cdot p_t \cdot \text{sign}(\epsilon_t^{(0)}),$$

表 3. 1% 污染设定下，$\theta_1 = (1, 0.2, 0.4)$ 的 EOPE 和 EDPE 样本均值，括号内为总缩放均方根误差

		EOPE（f_ϵ）			EDPE（γ）						
n	θ_1	$N(0,1)$	$t(5)$	$t(7)$	0.05	0.10	0.20	0.30	0.50	0.75	1.00
500	ω	1.431	1.441	1.339	1.316	1.247	1.204	1.211	1.235	1.261	1.286
	α	0.301	0.295	0.272	0.271	0.253	0.244	0.249	0.273	0.306	0.332
	β	0.356	0.351	0.346	0.344	0.334	0.320	0.309	0.295	0.287	0.285
		(1.894)	(1.430)	(1.265)	(1.421)	(1.120)	(0.944)	(0.940)	(1.022)	(1.162)	(1.279)
1000	ω	1.410	1.385	1.281	1.260	1.180	1.139	1.150	1.192	1.232	1.259
	α	0.287	0.280	0.257	0.255	0.236	0.226	0.229	0.244	0.271	0.302
	β	0.367	0.372	0.369	0.366	0.361	0.349	0.337	0.316	0.299	0.290
		(1.597)	(1.156)	(0.979)	(1.105)	(0.828)	(0.693)	(0.698)	(0.783)	(0.938)	(1.095)
2000	ω	1.391	1.370	1.262	1.239	1.152	1.103	1.110	1.156	1.208	1.239
	α	0.272	0.271	0.248	0.244	0.227	0.216	0.217	0.227	0.244	0.269
	β	0.382	0.379	0.378	0.378	0.374	0.365	0.355	0.333	0.310	0.297
		(1.333)	(1.028)	(0.838)	(0.916)	(0.680)	(0.551)	(0.554)	(0.637)	(0.772)	(0.919)

表4. 在1% 污染设定下，$\theta_2 = (1,\ 0.15,\ 0.8)$的 EOPE 和 EDPE 样本均值，括号内为总缩放均方根误差

		EOPE			EDPE (γ)						
n	θ_2	$N(0,\ 1)$	$t(5)$	$t(7)$	0.05	0.10	0.20	0.30	0.50	0.75	1.00
500	ω	1.918	1.857	1.718	1.662	1.544	1.497	1.533	1.619	1.668	1.668
	α	0.189	0.196	0.195	0.186	0.184	0.186	0.195	0.220	0.256	0.295
	β	0.774	0.776	0.769	0.769	0.763	0.751	0.738	0.706	0.653	0.588
		(1.698)	(1.361)	(1.261)	(1.279)	(1.064)	(0.955)	(1.015)	(1.250)	(1.585)	(1.911)
1000	ω	1.798	1.699	1.532	1.506	1.376	1.328	1.374	1.514	1.631	1.679
	α	0.188	0.195	0.190	0.182	0.176	0.175	0.181	0.200	0.226	0.255
	β	0.783	0.784	0.781	0.781	0.777	0.770	0.760	0.735	0.699	0.653
		(1.432)	(1.144)	(0.977)	(1.004)	(0.782)	(0.673)	(0.731)	(0.976)	(1.287)	(1.576)
2000	ω	1.620	1.538	1.367	1.346	1.234	1.191	1.226	1.364	1.537	1.645
	α	0.184	0.192	0.184	0.177	0.169	0.165	0.168	0.183	0.206	0.229
	β	0.796	0.793	0.792	0.793	0.790	0.784	0.778	0.759	0.729	0.696
		(1.132)	(0.939)	(0.751)	(0.765)	(0.572)	(0.475)	(0.510)	(0.710)	(1.033)	(1.315)

其中，$\epsilon_t^{(0)}$ 和 p_t 是分别从 $N(0,\ 1)$ 和《伯努利》(0.01)中抽取的独立随机变量序列。缩放因子设为 $s = 5$，用于确定污染的幅度。因此，以 1% 的概率，创新项会沿其符号方向偏移 $+5$ 或 -5，以模拟系统受到的极端冲击。

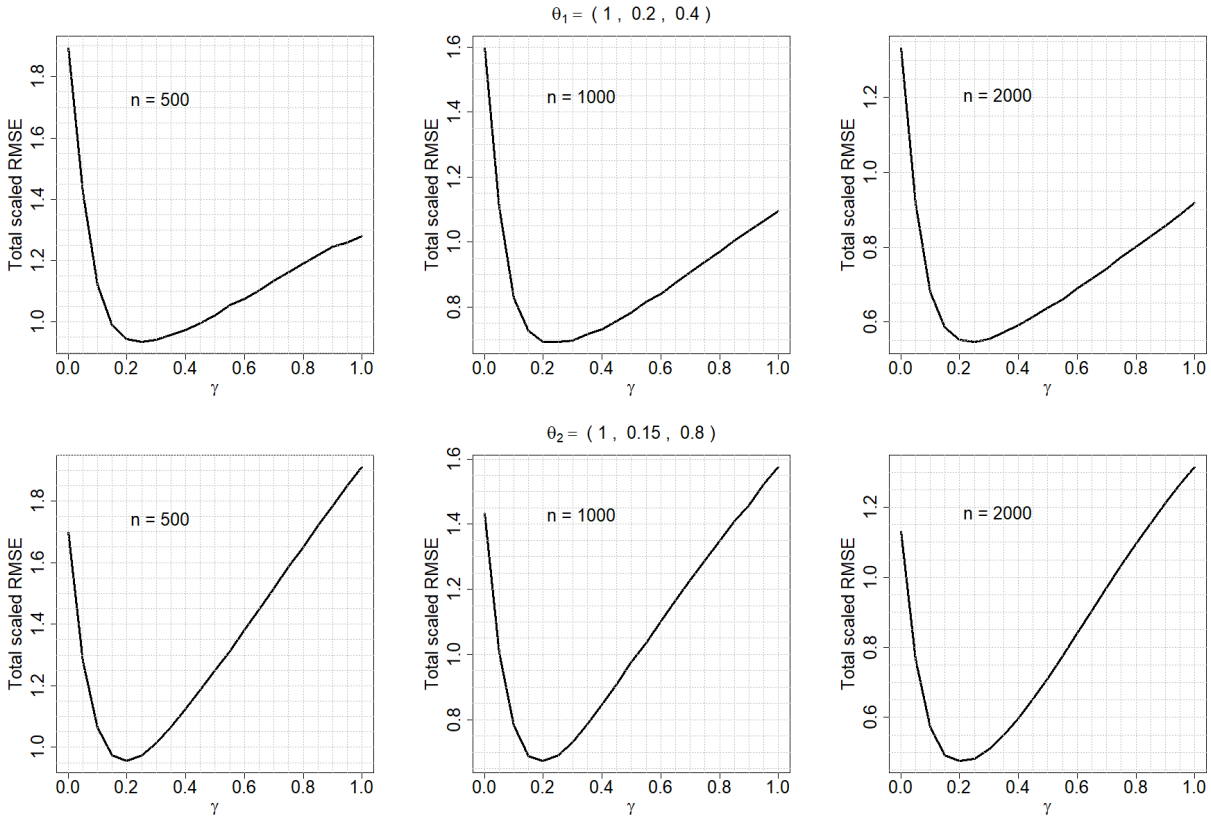


图2. 1% 污染设定下后验均值相对于 γ 的总尺度化均方根误差

面对这类重尾数据或离群值，在贝叶斯推断中，通常会采用基于重尾分布的似然函数，例如 Student'st分布，而非高斯分布。为进行全面比较，我们除了计算标准高斯 EOPE 外，还使用自由度为 5 和 7 的 Student's t 似然计算 EOPE。

表3和表4总结了在1%污染下 θ_1 和 θ_2 的模拟结果。可以得出几点重要观察。首先，与未污染情形相比，标准高斯 EOPE（$N(0,\ 1)$）的表现显著下降，在所有设置下的总缩放均方根误差均为最高。其次，基于 Student t分布的 EOPE 虽然通过适应重尾特性，表现明显优于高斯估计量，但仍不及所提出的基于 DPD 的估计量。具体而言，调节参数约为 $\gamma = 0.2$ 至 $\gamma = 0.3$ 的 EDPE 始终取得最低的均方根误差值（以红色突出显示）。这表明，与仅切换为 t分布这一传统策略相比，所提出的方法对离群值具有更强的稳健性。图2进一步展现了这一优势；与未污染情形中观察到的单调上升不同，此时均方根误差曲线呈现凸形（U 形）模式，这表明较小但为正的 γ 能够最好地平衡

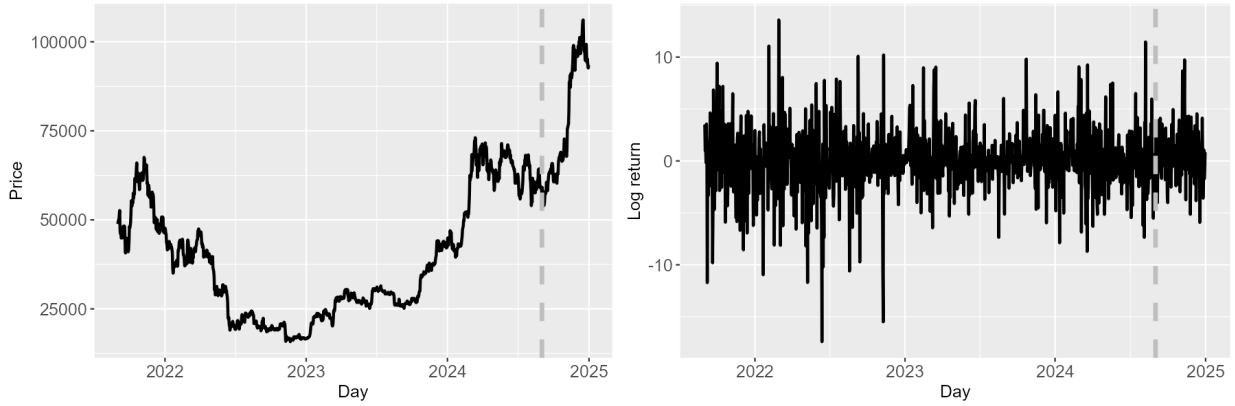


图3. 2021年9月至2024年12月 BTC-USD 的每日收盘价（L）和对数收益率（R）；垂直虚线标示样本外评估期的开始

污染下的稳健性和效率。

总之，模拟研究凸显了所提出的稳健贝叶斯方法在实践中的实用价值。在无污染设定下，调节参数 γ 较小的基于DPD的估计量表现与高效的 EOPE 相当，尤其是随着样本量增加时。然而，存在离群值时，标准 EOPE 无法提供可靠的估计；即使采用重尾似然（例如 Student's t）这一策略，与所提出的方法相比，结果仍不理想。通过恰当选择调节参数（例如 $\gamma \approx 0.2$），EDPE 能够成功地兼顾稳健性与效率，并通过有效降低离群值的影响，大幅降低均方根误差。这些发现证实，所提出的方法为 GARCH 建模提供了一种稳妥且稳健的替代方案，尤其适用于数据似乎含有离群值或异常的情况。

4 实证数据分析

本节分析比特币价格数据，以说明所提出的方法。令 $\{X_t\}$表示每日收盘价，$\{r_t\}$ 表示相应的对数收益率序列，定义为$r_t = 100\log(X_t/X_{t-1})$。数据集涵盖2021年9月至2024年12月（共1,218个观测值），如图3所示。我们使用2021年9月至2024年8月的收益率进行模型估计，并将其余观测值留作样本外评估。

使用 Ljung-Box检验进行初步分析表明，对于样本内对数收益率，在滞后阶数 1 至 30 中，最小的 p-值为 0。1241（滞后 25 阶），说明不存在显著的线性序列相关。相比之下，对于平方对数收益率，最大的 p-值为 0。0002（滞后 6 阶），有力地表明存在条件异方差。我们还观察到收益率序列中典型的波动聚集现象。因此，我们使用（10）中定义的 GARCH(1，1) 模型，并

表5. BTC-USD 数据的 GARCH 参数估计及一步预测表现

		EDPE（γ）						
	EOPE	0.05	0.10	0.20	0.30	0.50	0.75	1.00
$\hat{\omega}$	1.859	1.403	1.078	0.726	0.738	0.990	1.194	1.306
$\hat{\alpha}$	0.167	0.134	0.113	0.092	0.094	0.125	0.179	0.251
$\hat{\beta}$	0.637	0.694	0.736	0.776	0.746	0.617	0.484	0.396
RMSE	3.553	3.531	3.519	3.512	3.517	3.534	3.553	3.577
MAE	7.908	7.563	7.277	6.874	6.666	6.515	6.593	6.832
风险价值违规率	0.016	0.016	0.016	0.049	0.049	0.057	0.066	0.074

将参数 $\theta = (\omega,\ \alpha,\ \beta)$ 拟合到收益率序列 $\{r_t\}$，这是实证波动率建模中常用的一种标准且简约的设定。

如图3右侧面板所示，该序列还包含若干可能对常规贝叶斯推断产生不利影响的离群观测值。为解决这一问题，我们使用本研究提出的 EDPE 估计模型。作为比较，我们也考察了传统的 EOPE。我们通过比较这两种方法在样本外期间对条件方差及其对应的 95% 风险价值（VaR）的一步预测来评估它们。具体而言，我们使用 2024年9月至2024年12月的收益率（122 天）进行样本外分析。

对于每 个时间点 t，时刻 $t+1$ 的一步条件方差由 y

$$\hat{\sigma}^2_{t+1|t} = \hat{\omega}_t + \hat{\alpha}_t r_t^2 + \hat{\beta}_t \tilde{\sigma}_t^2,$$

ˆ 其中 ˆω_t、ˆα_t 和 β_t 是利用截至时刻 t 的收益率序列，由 EOPE 或 EDPE 得到的参数估计。项 $\tilde{\sigma}_t^2$ 依据 (8) 使用相应的代入估计递归计算，并采用初始值 $r_0^2 = \tilde{\sigma}_0^2 = r_1^2$。预测表现使用均方根误差（RMSE）和平均绝对误差（MAE）进行评估。由于真实条件方差 σ^2_{t+1} 不可观测，我们使用 r^2_{t+1} 作为其代理，并注意到在 GARCH模型下 E($r^2_{t+1}|\mathcal{F}_t$) $= \sigma^2_{t+1}$。因此，在 122 个样本外天数上，RMSE 和 MAE 的计算如下

$$\text{RMSE} = \sqrt{\frac{1}{122}\sum_{t=1096}^{1217}(\hat{\sigma}^2_{t+1|t} - r^2_{t+1})^2} \quad \text{and} \quad \text{MAE} = \frac{1}{122}\sum_{t=1096}^{1217}\left|\hat{\sigma}^2_{t+1|t} - r^2_{t+1}\right|,$$

其中预测期从2024年9月1日持续至2024年12月31日。

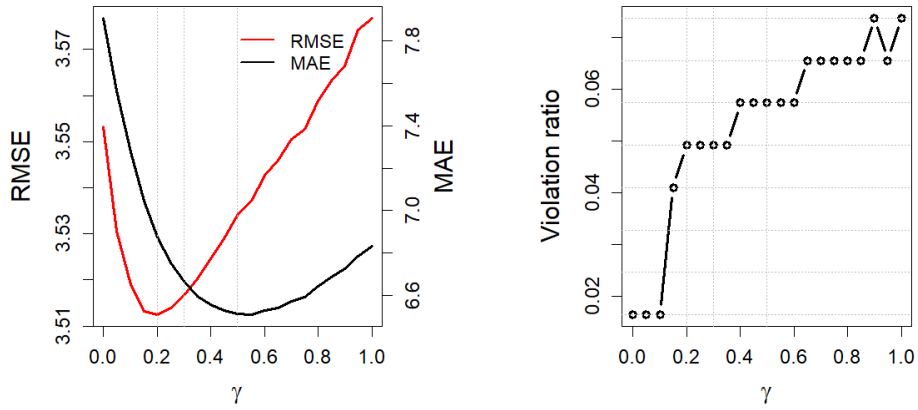


图 4. 条件方差预测的均方根误差和平均绝对误差（左）以及经验 95% 风险价值违规比率（右）

时刻 $t+1$ 的一步超前95%风险价值预测定义为

$$VaR_{t+1|t} = F_{\epsilon}^{-1}(0.05)\,\hat{\sigma}_{t+1|t}$$

其中， F_{ϵ} 是 GARCH 模型中创新项 ϵ_t 的累积分布函数（CDF）。由于在本分析中假设 ϵ_t 服从标准正态分布，因此我们通过计算经验违约率来回测风险价值预测

$$\frac{1}{122}\sum_{t=1096}^{1217} I\left(r_{t+1} < \Phi^{-1}(0.05)\,\hat{\sigma}_{t+1|t}\right),$$

其中， $I(\cdot)$ 表示示性函数。在模型设定正确的情况下，r_{t+1} 低于 $VaR_{t+1|t}$ 的概率等于 0.05，因此，接近 0.05 的经验违约率表明拟合模型校准良好，并能合理捕捉收益分布的尾部特征。

表5的前三行报告了针对不同调节参数值 γ，使用 EOPE 和 EDPE 得到的 GARCH(1,1) 参数后验均值估计。可以看到，EOPE 与 EDPE 的估计值存在明显差异，尤其是参数 ω。具体而言，在 γ 的整个取值范围内，EOPE 得到的值都显著高于 EDPE。回想模拟研究的结果，在不存在离群值时，对于较小的 γ，EOPE 和 EDPE 得到的估计值相近。相比之下，在存在离群值时，对离群值敏感的 EOPE 往往会高估 ω，导致其估计值与 EDPE 的估计值存在显著差异，即使 γ 较小也是如此。鉴于这些发现，EOPE 得到的 ω 膨胀估计值强烈表明，收益率序列中存在表现得类似离群值、并对 EOPE 产生不可忽略影响的观测值。

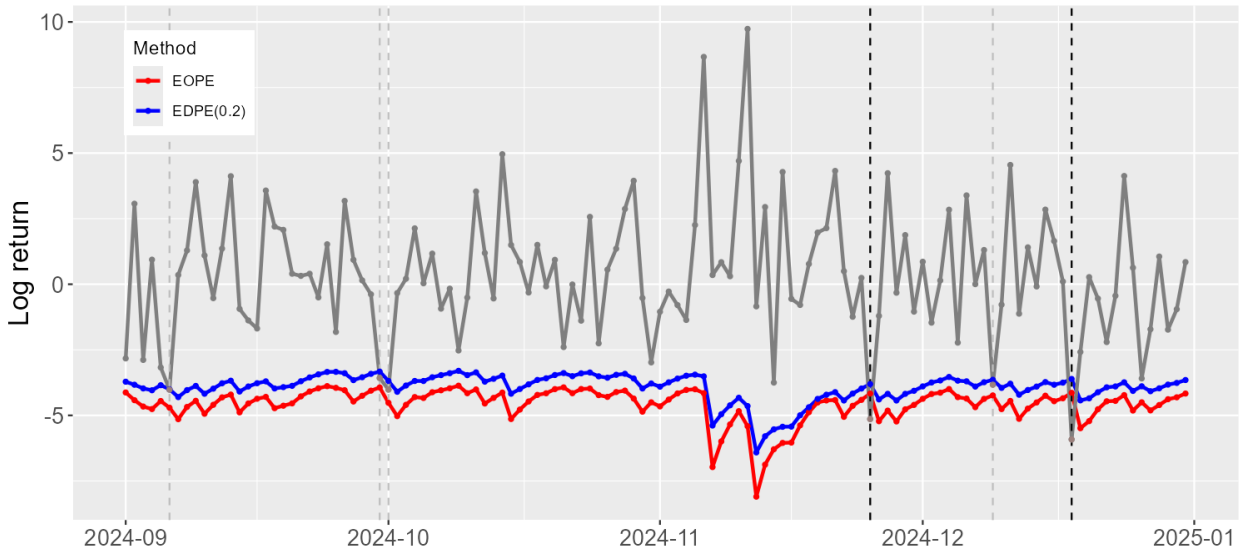


图5. 2024年9月至12月EOPE与EDPE下的每日对数收益率及一步向前95%风险价值预测

表5的末几行报告了单步预测表现，包括均方根误差、平均绝对误差以及 95%风险价值的经验违约率；图4则展示了这些指标如何随调节参数 γ 在范围 $[0,1]$ 内变化。正如预期，对于大多数 γ 值，EDPE 在均方根误差和平均绝对误差方面均优于 EOPE，这表明当数据中可能含有具有类似异常值影响的观测值时，所提出的贝叶斯估计量可作为一种实用的替代方法。

我们还观察到，均方根误差和平均绝对误差分别在 $\gamma = 0.2$和 γ =0.5处达到最小值。尽管报告了这两项指标以评估预测准确性，但我们更重视均方根误差，因为它对较大预测误差的惩罚重于平均绝对误差，因而对波动率建模和风险管理更具参考价值，而在这些领域捕捉极端市场波动至关重要。这一考量也促使我们采用备注5中讨论的针对 γ的数据驱动的选择规则，该规则基于最小化样本外均方根误差。根据这一准则，我们选择 $\gamma = 0.2$ 作为给定数据集的最优调节值。

值得注意的是，这一 γ 的选择也与接近最优的VaR校准相关联。如表中最后一行所示，EOPE的违背率为0.016（122次中有2次违背），远低于名义水平0.05，表明VaR估计过于保守。相比之下，当 γ 介于0.20与0.35之间时，EDPE的违背率为0.049（122次中有6次违背），非常接近目标水平0.05。当 γ取值更大时，违背率超过0.05，意味着VaR预测变得过于激进，从而低估了下行风险。总体而言，适中的 γ 取值（大致介于0.2与0.5之间）似乎能在预测准确性与VaR校准之间提供有利的权衡。

图 5 显示　　显示对数收益率以及使用 EOPE 和 EDPE 获得的一步前瞻 95%风险价值预测

，其中 $\gamma = 0.2$。灰线表示实际收益率序列，红线和蓝线分别表示由 EOPE 和 EDPE 计算得出的相应风险价值预测。黑色竖直虚线表示两种估计量均出现风险价值突破的日期，而灰色竖直虚线则表示仅 EDPE 预测出现突破的日期。我们可以清楚地看到，EOPE 得出的风险价值预测区间相对较宽，这似乎与 ω 的估计值相对较大有关；正如前文所述，这可能受到类似异常值的观测的影响。相比之下，EDPE 得出的风险价值预测既不过于保守，也不过于激进，而是与名义覆盖水平相当吻合。

总体而言，BTC-USD 的真实数据分析证实了所提出的稳健贝叶斯估计量在实践中的有效性。与传统的 EOPE 相比，EDPE 在整个评估期内对条件方差的样本外预测有所改善。此外，风险价值回测结果表明，EOPE 给出的风险预测过于保守，而采用适中调节参数的 EDPE 所得经验违约率接近名义水平，因此能够更好地校准尾部风险。综合来看，这些结果表明，对于偶尔出现极端冲击的金融收益序列，所提出的 EDPE 在预测准确性与风险校准之间实现了较好的权衡，因此可作为 GARCH 类模型标准贝叶斯推断的一种稳健替代方法。

5 结论

离群值和突发冲击可能严重影响对条件异方差时间序列的贝叶斯推断，因为它们会主导似然，进而左右后验分布。为降低这种敏感性，我们提出了一种稳健贝叶斯框架，该框架基于由调节参数 γ 索引的 DPD 驱动伪后验分布。该估计量定义为基于 DPD 的后验分布下的后验均值；当 $\gamma \downarrow 0$ 时，它平滑地过渡到普通贝叶斯估计量，并在 $\gamma > 0$ 时通过降低极端观测的权重而变得稳健。

我们在平稳性和遍历性条件下建立了渐近理论，其中包括基于DPD的后验分布的伯恩斯坦–冯·米塞斯型近似，以及EDPE与MDPDE之间的渐近等价。我们还表明，在GARCH文献中通常采用的标准正则条件下，这些渐近结果所需的假设对GARCH(p，q)模型均成立。针对GARCH(1，1)的数值研究表明，γ取适中值时，能够显著提高模型在污染条件下的准确性，且优于基于重尾似然的替代方法；同时，在未污染样本中仍具有竞争力。在BTC-USD收益的应用中，EDPE改善了一步前瞻波动率预测，并提供校准更准确的95%风险价值，使经验违约率接近名义水平。

总之，所提出的方法为存在偶发极端冲击的情形下进行波动率建模，提供了一种实用且有理论依据的稳健贝叶斯工具。重要的扩展方向

包括将该方法应用于整数值时间序列模型（其中极端计数和离群值往往会产生重大影响），以及将其拓展至多元贝叶斯设定，以进一步扩大其适用范围。

6 附录

在本附录中，我们给出主要定理及相关引理的证明。

定理1的证明
设 $\Delta_\gamma(\theta) := \widetilde{Q}_n^{(\gamma)}(\theta) - \sum_{t=1}^{n} \log \tilde{f}_\theta(X_t|\mathcal{F}_{t-1})$。则由条件**C3** 可知

$$e^{-V_n\gamma} \le e^{\Delta_\gamma(\theta)} \le e^{V_n\gamma} \qquad \text{for all } \theta \in \Theta$$

从而，我们有

$$\sup_{\theta\in\Theta} \left|e^{\Delta_\gamma(\theta)} - 1\right| \le \max\{e^{V_n\gamma} - 1,\, 1 - e^{-V_n\gamma}\} \longrightarrow 0 \quad as\ \gamma \downarrow 0,$$

这意味着

$$\sup_{\theta\in\Theta} \left| \frac{\exp\big(\widetilde{Q}_n^{(\gamma)}(\theta)\big)}{\tilde{L}(\theta \mid \mathbf{X}_n)} - 1 \right| = \sup_{\theta\in\Theta} \left|e^{\Delta_\gamma(\theta)} - 1\right| \longrightarrow 0 \quad as\ \gamma \downarrow 0.$$

由于 Θ 是紧的，且根据条件 **C2**，$\theta \mapsto L(\theta \mid \mathbf{X}_n)$ 在 Θ 上连续，因此函数 $L(\theta \mid \mathbf{X}_n)$ 在 Θ 上有界。此外，根据条件 **C4**，$\pi(\theta)$ 在 Θ 上连续，因此也有界。因此，我们得到

$$\begin{aligned} \sup_{\theta\in\Theta} \left|\exp(\widetilde{Q}_n^{(\gamma)}(\theta))\pi(\theta) - \tilde{L}(\theta \mid \mathbf{X}_n)\pi(\theta)\right| \;&\le\; \sup_{\theta\in\Theta} \left|\tilde{L}(\theta \mid \mathbf{X}_n)\pi(\theta)\right| \sup_{\theta\in\Theta} \left| \frac{\exp(\widetilde{Q}_n^{(\gamma)}(\theta))}{\tilde{L}(\theta \mid \mathbf{X}_n)} - 1 \right| \\ &\longrightarrow\; 0 \quad \text{as } \gamma \downarrow 0. \end{aligned} \tag{11}$$

接下来，令 $u_\gamma(\theta) := \exp(\widetilde{Q}_n^{(\gamma)}(\theta))\pi(\theta)$，且 $u_0(\theta) := \tilde{L}(\theta \mid \mathbf{X}_n)\pi(\theta)$，其归一化常数分别为$Z_\gamma := \int_\Theta u_\gamma(\theta)\, d\theta$ 和 $Z_0 := \int_\Theta u_0(\theta)\, d\theta$。由 (11)，我们有 $\sup_{\theta\in\Theta} |u_\gamma(\theta) - u_0(\theta)| \to 0$。因此，由于 Θ 的紧性，可得

$$|Z_\gamma - Z_0| \le \int_\Theta |u_\gamma(\theta) - u_0(\theta)|\, d\theta \le \sup_{\theta\in\Theta} |u_\gamma(\theta) - u_0(\theta)| \int_\Theta 1\, d\theta \longrightarrow 0 \quad \text{as } \gamma \downarrow 0. \tag{12}$$

此外，结合条件**C4** 和**C5**以及 $L(\theta \mid \mathbf{X}_n)$ 的正性，我们可以证明 $Z_0 > 0$，从而对所有足够小的 γ，有 $Z_\gamma \ge Z_0/2$ 。将归一化密度写为

$p_\gamma(\theta) := u_\gamma(\theta)/Z_\gamma$ 和 $p_0(\theta) := u_0(\theta)/Z_0$，我们得到

$$\int_\Theta |p_\gamma(\theta) - p_0(\theta)|\, d\theta \le \frac{1}{Z_\gamma}\int_\Theta |u_\gamma(\theta) - u_0(\theta)|\, d\theta + \left|\frac{1}{Z_\gamma} - \frac{1}{Z_0}\right| \int_\Theta u_0(\theta)\, d\theta \longrightarrow 0 \quad \text{as } \gamma \downarrow 0,$$

根据 (12)，由于 $Z_\gamma \to Z_0$ 且 $\int_\Theta |u_\gamma - u_0| \to 0$。这意味着全变差收敛，并完成了证明。□

引理1. 假设条件**A1** 和**A2** 成立。那么，对于每个 $\gamma > 0$，我们有

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\widetilde{Q}_n^{(\gamma)}(\theta) - Q^{(\gamma)}(\theta)\right| = o(1) \quad a.s.,$$

其中 $Q^{(\gamma)}(\theta) = \mathrm{E} q_t^{(\gamma)}(\theta)$。

证明。由 $\{q_t^{(\gamma)}(\theta)\}$ 的遍历性可知，对于每个 $\theta \in \Theta$，$\frac{1}{n}Q_n^{(\gamma)}(\theta)$ 几乎必然收敛到 $Q^{(\gamma)}(\theta)$。因此，结合条件 $\mathrm{E}\sup_{\theta\in\Theta}\left|q_t^{(\gamma)}(\theta)\right| < \infty$，应用 Straumann 和 Mikosch (2006) 的定理2.7，我们得到

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}Q_n^{(\gamma)}(\theta) - Q^{(\gamma)}(\theta)\right| = o(1) \quad a.s.$$

因此，由条件**A2**的第二部分可得该引理。 □

定理2的证明

注意，$d_\gamma\left(f_{\theta_0}(\cdot \mid \mathcal{F}_{t-1}),\ f_\theta(\cdot \mid \mathcal{F}_{t-1})\right)$ 几乎必然非负，且在$\theta = \theta_0$处取得唯一最小值。因此，其期望 $\mathrm{E}\, d_\alpha\left(f_{\theta_0}(\cdot \mid \mathcal{F}_{t-1}),\ f_\theta(\cdot \mid \mathcal{F}_{t-1})\right)$ 也在 θ_0 处达到最小值。由于

$$Q^{(\gamma)}(\theta) = -\frac{1}{1+\gamma}\mathrm{E}\, d_\gamma\left(f_{\theta_0}(\cdot \mid \mathcal{F}_{t-1}), f_\theta(\cdot \mid \mathcal{F}_{t-1})\right) + \frac{1}{\gamma(1+\gamma)}\int f_{\theta_0}^{1+\gamma}(x \mid \mathcal{F}_{t-1})dx,$$

由此可知，$Q^{(\gamma)}(\theta)$ 在 θ_0 处取得唯一最大值。因此，根据引理1，利用标准论证可证明估计量 $\theta_{\gamma,n}$ 的强ˆ一致性。

接下来，我们证明渐近正态性。根据假设 **A3** 以及鞅差的中心极限定理，我们有

$$\frac{1}{\sqrt{n}}\partial_\theta Q_n^{(\gamma)}(\theta_0) \xrightarrow{d} N_d(0, \mathcal{I}_\gamma). \tag{13}$$

因此，由假设 **A4** 可得 $\frac{1}{\sqrt{n}}\partial_\theta \widetilde{Q}_n^{(\gamma)}(\theta_0) = O_P(1)$。此外，在假设 **A5** 和 **A6**下，可以证明，对于任意几乎必然收敛到 θ_0 的随机向量序列 $\{\theta_n^*\}$，

$$\frac{1}{n}\partial^2_{\theta\theta'}\widetilde{Q}_n^{(\gamma)}(\theta_n^*) \longrightarrow -\mathcal{J}_\gamma \quad a.s. \tag{14}$$

由于 $\partial_\theta \widetilde{Q}_n^{(\gamma)}(\hat{\theta}_{\gamma,n}) = 0$，由中值定理可得 $\partial_\theta \widetilde{Q}_n^{(\gamma)}(\theta_0) = -\partial^2_{\theta\theta'} \widetilde{Q}_n^{(\gamma)}(\bar{\theta}_n)(\hat{\theta}_{\gamma,n} - \theta_0)$，其中 $\bar{\theta}_n$ 是 $\theta_{\gamma,n}$ 与 θ_0 之间线段上的一点。因此，我们可以看到

$$\begin{aligned}\sqrt{n}(\hat{\theta}_{\gamma,n} - \theta_0) &= \mathcal{J}_\gamma^{-1} \frac{1}{\sqrt{n}} \partial_\theta \widetilde{Q}_n^{(\gamma)}(\theta_0) + \mathcal{J}_\gamma^{-1} \Big(\frac{1}{n} \partial^2_{\theta\theta'} \widetilde{Q}_n^{(\gamma)}(\bar{\theta}_n) + \mathcal{J}_\gamma \Big) \sqrt{n}(\hat{\theta}_{\gamma,n} - \theta_0) \\ &= O_P(1) + o_P(1)\sqrt{n}(\hat{\theta}_{\gamma,n} - \theta_0),\end{aligned}$$

这意味着 $\sqrt{n}(\hat{\theta}_{\gamma,n} - \theta_0) = O_P(1)$。因此，由假设 **A4** 可得

$$\sqrt{n}(\hat{\theta}_{\gamma,n} - \theta_0) = \mathcal{J}_\gamma^{-1} \frac{1}{\sqrt{n}} \partial_\theta Q_n^{(\gamma)}(\theta_0) + o_P(1),$$

结合式（13），可得 $\sqrt{n}(\hat{\theta}_{\gamma,n} - \theta_0)$ 的渐近正态性。证明完毕。□

引理2。假设 **A1** 和 **A2** 成立。对于任意 $\delta > 0$，存在一个 $\epsilon > 0$，使得对于所有充分大的 n，

$$\sup_{\|\theta - \theta_0\| > \delta} \frac{1}{n} \big(\widetilde{Q}_n^{(\gamma)}(\theta) - \tilde{Q}_n^{(\gamma)}(\theta_0) \big) < -\epsilon \quad a.s.$$

证明。回顾一下，$Q^{(\gamma)}(\theta)$ 在 θ_0 处取得唯一最大值。因此，由于 $Q^{(\gamma)}(\theta)$ 的连续性，对于任意 $\delta > 0$，可以选取 $\epsilon > 0$，使得

$$\sup_{\|\theta - \theta_0\| > \delta} Q^{(\gamma)}(\theta) \le Q^{(\gamma)}(\theta_0) - 3\epsilon.$$

利用引理1和上述结果可得，对于任意满足 $\|\theta - \theta_0\| > \delta$ 的 θ，以及所有充分大的 n，

$$\frac{1}{n} \widetilde{Q}_n^{(\gamma)}(\theta) \le Q^{(\gamma)}(\theta) + \epsilon \le Q^{(\gamma)}(\theta_0) - 2\epsilon \quad a.s. \tag{15}$$

此外，对于所有充分大的 n，由于根据引理1有 $\frac{1}{n} \widetilde{Q}_n^{(\gamma)}(\theta_0) \ge Q^{(\gamma)}(\theta_0) - \epsilon \ a.s.$，由式（15）可得

$$\frac{1}{n} \widetilde{Q}_n^{(\gamma)}(\theta) - \frac{1}{n} \widetilde{Q}_n^{(\gamma)}(\theta_0) \le -\epsilon \quad a.s.$$

对所有满足 $\|\theta - \theta_0\| > \delta$ 的 θ，以及所有充分大的 n。至此证明完毕。 □

定理3的证明

令

$$C_n := \int_{\mathbb{R}^d} \exp\Big(\widetilde{Q}_n^{(\gamma)}\Big(\hat{\theta}_{\gamma,n} + \frac{t}{\sqrt{n}} \Big) - \widetilde{Q}_n^{(\gamma)}\big(\hat{\theta}_{\gamma,n} \big) \Big) \pi\Big(\hat{\theta}_{\gamma,n} + \frac{t}{\sqrt{n}} \Big) dt,$$

ˆ γ-后验密度函数 $t = \sqrt{n}(\theta - \theta_{\gamma,\ n})$ 可表示为

$$\pi_\gamma(t|\mathbf{X}_n) = \frac{1}{C_n}\exp\Big(\widetilde{Q}_n^{(\gamma)}\Big(\hat{\theta}_{\gamma,n} + \frac{t}{\sqrt{n}}\Big) - \widetilde{Q}_n^{(\gamma)}\big(\hat{\theta}_{\gamma,n}\big)\Big)\,\pi\Big(\hat{\theta}_{\gamma,n} + \frac{t}{\sqrt{n}}\Big).$$

定义

$$g_n(t) := \pi\Big(\hat{\theta}_{\gamma,n} + \frac{t}{\sqrt{n}}\Big)\exp\Big(\widetilde{Q}_n^{(\gamma)}\Big(\hat{\theta}_{\gamma,n} + \frac{t}{\sqrt{n}}\Big) - \widetilde{Q}_n^{(\gamma)}\big(\hat{\theta}_{\gamma,n}\big)\Big) - \pi(\theta_0)e^{-\frac{1}{2}t'\mathcal{J}_\gamma t}.$$

那么，只需证明以下结论即可得到该定理:

$$\int_{\mathbb{R}^d}\big|g_n(t)\big|dt = o_P(1). \tag{16}$$

为证明上述结论，我们将 $\mathbb{R}^d$ 分为两个区域： $\mathcal{R}_1 = \{t \in \mathbb{R}^d \mid \|t\| \le \delta_0\sqrt{n}\}$ 和 $\mathcal{R}_2 = \{t \in \mathbb{R}^d \mid \|t\| > \delta_0\sqrt{n}\}$ ，其中 $\delta_0 > 0$。

首先，处理 $\mathcal{R}_1$上的积分。由泰勒定理以及 $\partial_\theta \widetilde{Q}_n^{(\gamma)}(\hat{\theta}_{\gamma,n}) = 0$，可得

$$\widetilde{Q}_n^{(\gamma)}\Big(\hat{\theta}_{\gamma,n} + \frac{t}{\sqrt{n}}\Big) - \widetilde{Q}_n^{(\gamma)}\big(\hat{\theta}_{\gamma,n}\big) = \frac{1}{2n}t'\partial^2_{\theta\theta'}\widetilde{Q}_n^{(\gamma)}(\bar{\theta}_{\gamma,n})t, \tag{17}$$

其中 $\bar{\theta}_{\gamma,n}$ 介于 $\hat{\theta}_{\gamma,n} + \frac{t}{\sqrt{n}}$ 和 $\hat{\theta}_{\gamma,n}$之间。对于任意固定的 t，由于 $\bar{\theta}_{\gamma,n}$ 几乎必然收敛到 θ_0，由式 (14) 可知， $\frac{1}{n}\partial^2_{\theta\theta'}\widetilde{Q}_n^{(\gamma)}(\bar{\theta}_{\gamma,n})$ 几乎必然收敛到 $-\mathcal{J}_\gamma$，因此 $g_n(t)$ 几乎必然收敛到 0。

注意，根据假设**A7**， $\mathcal{J}_\gamma$ 是正定的。因此，我们可以看到，对于所有充分大的 n 和任意固定的 t,

$$\frac{1}{n}t'\partial^2_{\theta\theta'}\widetilde{Q}_n^{(\gamma)}(\bar{\theta}_{\gamma,n})t \le -\frac{1}{2}t'\mathcal{J}_\gamma t \quad a.s. \tag{18}$$

ˆ ˆ 此外，对于任意 $t \in \mathcal{R}_1$，我们有$\|\theta - \theta_0\| \le \|\theta_{\gamma,\ n} - \theta_0\| + \|t\|/\sqrt{n} \le \|\theta_{\gamma,\ n} - \theta_0\| + \delta_0$。由于 $\pi(\theta)$ ˆ在 θ_0 处连续，且根据定理2，$\|\theta_{\gamma,\ n} - \theta_0\|$ 几乎必然收敛于零，因此对于任意 $c > 0$，我们可以选择足够小的 $\delta_0 > 0$ ，使得对于所有充分大的 n,

$$\sup_{t\in\mathcal{R}_1}\Big|\pi\Big(\hat{\theta}_{\gamma,n} + \frac{t}{\sqrt{n}}\Big) - \pi(\theta_0)\Big| = \sup_{\|\theta - \hat{\theta}_{\gamma,n}\| \le \delta_0}\Big|\pi(\theta) - \pi(\theta_0)\Big| \le c. \tag{19}$$

因此，我们有由 (17)、(18) 和 (19) 可知，对于所有充分大的 n 以及任意固定的 $t \in \mathcal{R}_1$,

$$\begin{aligned}\big|g_n(t)\big| &\le \pi\Big(\hat{\theta}_{\gamma,n} + \frac{t}{\sqrt{n}}\Big)\exp\Big(\frac{1}{2n}t'\partial^2_{\theta\theta'}\widetilde{Q}_n^{(\gamma)}(\bar{\theta}_{\gamma,n})t\Big) + \pi(\theta_0)e^{-\frac{1}{2}t'\mathcal{J}_\gamma t}\\ &\le \big(\pi(\theta_0) + c\big)e^{-\frac{1}{4}t'\mathcal{J}_\gamma t} + \pi(\theta_0)e^{-\frac{1}{2}t'\mathcal{J}_\gamma t} \quad a.s.,\end{aligned}$$

它在 $\mathcal{R}_1$ 上可积。因此，由控制收敛定理可得

$$\int_{\mathcal{R}_1} \big|g_n(t)\big|dt = o(1) \quad a.s. \tag{20}$$

接下来，利用引理2，我们可以取一个常数 $\epsilon > 0$，使得

$$\sup_{\|\theta-\theta_0\|>\delta_0} \frac{1}{n}\big(\widetilde{Q}_n^{(\gamma)}(\theta) - \tilde{Q}_n^{(\gamma)}(\theta_0)\big) < -\epsilon \quad a.s.$$

然后，对于每个 $t \in$ ˆ$\mathcal{R}_2$，根据（14）以及$\theta_{\gamma,\ n}$ 的一致性，我们有对于所有充分大的 n ，

$$\begin{aligned}
\frac{1}{n}\Big(\widetilde{Q}_n^{(\gamma)}\Big(\hat{\theta}_{\gamma,n} + \frac{t}{\sqrt{n}}\Big) - \widetilde{Q}_n^{(\gamma)}\big(\hat{\theta}_{\gamma,n}\big)\Big) &= \frac{1}{n}\Big(\widetilde{Q}_n^{(\gamma)}(\theta) - \widetilde{Q}_n^{(\gamma)}(\theta_0)\Big) + \frac{1}{n}\Big(\widetilde{Q}_n^{(\gamma)}(\theta_0) - \widetilde{Q}_n^{(\gamma)}\big(\hat{\theta}_{\gamma,n}\big)\Big) \\
&\le -\epsilon + \frac{1}{2n}(\theta_0 - \hat{\theta}_{\gamma,n})'\partial^2_{\theta\theta'}\widetilde{Q}_n^{(\gamma)}\big(\bar{\theta}_{\gamma,n}\big)(\theta_0 - \hat{\theta}_{\gamma,n}) \quad a.s. \\
&\le -\frac{\epsilon}{2} \quad a.s.,
\end{aligned}$$

– ˆ，其中 $\theta_{\gamma,\ n}$ 介于 θ_0 和 $\theta_{\gamma,\ n}$ 之间。因此，对于所有充分大的 n，我们有

$$\begin{aligned}
\int_{\mathcal{R}_2} \big|g_n(t)\big|dt &\le e^{-\frac{n\epsilon}{2}} \int_{\mathcal{R}_2} \pi\Big(\hat{\theta}_{\gamma,n} + \frac{t}{\sqrt{n}}\Big)dt + \int_{\mathcal{R}_2} \pi(\theta_0)e^{-\frac{1}{2}t'\mathcal{J}_\gamma t}dt \quad a.s. \\
&\le n^{d/2}e^{-\frac{n\epsilon}{2}} + \pi(\theta_0)\int_{\mathcal{R}_2} e^{-\frac{1}{2}t'\mathcal{J}_\gamma t}dt = o(1),
\end{aligned}$$

结合（20）可得（16）。证明完毕。 □

定理4的证明
按照与定理3的证明相同的步骤，并利用 $\int \|\theta\|\pi(\theta)d\theta$的有限性，我们可以得到

$$\int_{\mathbb{R}^d} \|t\|\Big|\pi_\gamma(t \mid \mathbf{X}_n) - \frac{|\mathcal{J}_\gamma|^{1/2}}{(2\pi)^{d/2}}e^{-\frac{1}{2}t'\mathcal{J}_\gamma t}\Big|dt = o(1) \quad a.s.,$$

这意味着

$$\int_{\mathbb{R}^d} t\pi_\gamma(t \mid \mathbf{X}_n)dt \xrightarrow{a.s.} \int_{\mathbb{R}^d} t\frac{|\mathcal{J}_\gamma|^{1/2}}{(2\pi)^{d/2}}e^{-\frac{1}{2}t'\mathcal{J}_\gamma t}dt = 0.$$

因此，由于ˆ$\theta_\gamma^{EDPE}{}_{,n}$ = E(ˆ$\theta_{\gamma,n} + t/\sqrt{n}$†$\mathbf{X}_n$)，我们有

$$\sqrt{n}(\hat{\theta}_{\gamma,n}^{EDPE} - \hat{\theta}_{\gamma,n}) = \int_{\mathbb{R}^d} t\pi_\gamma(t \mid \mathbf{X}_n)dt \longrightarrow 0 \quad a.s.$$

第二个结果直接由定理2和斯卢茨基定理得出。证明至此完成。□

引理 3. 对于(9)中给出的目标函数 $\widetilde{Q}_n^{(\gamma)}(\theta)$，存在一个正随机变量$V_n$，它与 θ 独立，使得对所有充分小的 $\gamma > 0$,

$$\sup_{\theta\in\Theta}\left|\widetilde{Q}_n^{(\gamma)}(\theta) - K_n(\gamma) - \sum_{t=1}^{n}\log\tilde{f}_\theta(X_t \mid \mathcal{F}_{t-1})\right| \le V_n\gamma,$$

其中 $K_n(\gamma) = n(\gamma^{-1}-1)$。

证明。令 $w_t(\theta) := (2\pi\tilde{\sigma}_t^2(\theta))^{-1/2}$ 和 $B_t(\theta) := X_t^2/(2\tilde{\sigma}_t^2(\theta))$。基于 DPD 的目标函数为

$$\widetilde{Q}_n^{(\gamma)}(\theta) = \sum_{t=1}^{n} w_t^\gamma(\theta)\left\{\frac{1}{\gamma}\exp(-\gamma B_t(\theta)) - \left(\frac{1}{1+\gamma}\right)^{\frac{3}{2}}\right\}.$$

对于足够小的 $\gamma > 0$，我们可以将求和中的各项展开如下 s:

$$\begin{aligned}
w_t^\gamma(\theta) &= 1 + \gamma\log w_t(\theta) + \frac{\gamma^2}{2}(\log w_t(\theta))^2 + O(\gamma^3),\\
\frac{1}{\gamma}\exp(-\gamma B_t(\theta)) &= \frac{1}{\gamma} - B_t(\theta) + \frac{\gamma}{2}B_t^2(\theta) + O(\gamma^2),\\
(1+\gamma)^{-3/2} &= 1 - \frac{3}{2}\gamma + O(\gamma^2).
\end{aligned}$$

回想一下，参数空间 Θ 是 $(0,\infty)\times[0,\infty)^{p+q}$ 的紧子集。由于 $\tilde{\sigma}_t^2(\theta)$ 关于 θ 连续，对于任意固定样本 $\{X_t\}_{t=1}^n$，我们可以看出 $\inf_{\theta\in\Theta}\tilde{\sigma}_t^2(\theta) > 0$ 且 $\sup_{\theta\in\Theta}\tilde{\sigma}_t^2(\theta) < \infty$。因此，$B_t(\theta)$ 和 $\log w_t(\theta)$ 在 Θ 上一致有界，从而保证展开中的余项在 $\theta\in\Theta$ 中也一致有界。

将这些展开式代入求和项 $\tilde{q}_t^{(\gamma)}(\theta)$，我们得到

$$\begin{aligned}
\tilde{q}_t^{(\gamma)}(\theta) &= \left(1 + \gamma\log w_t(\theta) + \frac{\gamma^2}{2}(\log w_t(\theta))^2 + O(\gamma^3)\right)\left\{\left(\frac{1}{\gamma}-1\right) - B_t(\theta) + \gamma\left(\frac{B_t^2(\theta)+3}{2}\right) + O(\gamma^2)\right\}\\
&= \left(\frac{1}{\gamma}-1\right) + (\log w_t(\theta) - B_t(\theta)) + \gamma R_t(\theta) + O(\gamma^2),
\end{aligned}$$

其中 $R_t(\theta) := \frac{1}{2}(B_t^2(\theta) + 3) - (1 + B_t(\theta))\log w_t(\theta) + \frac{1}{2}(\log w_t(\theta))^2$。对 $t = 1,\ldots,n$求和，得到

$$\widetilde{Q}_n^{(\gamma)}(\theta) = K_n(\gamma) + \widetilde{Q}_n^{(0)}(\theta) + \gamma\sum_{t=1}^{n} R_t(\theta) + O(\gamma^2),$$

由于 $R_t(\theta)$ 在紧集 Θ 上连续，$\sup_{\theta\in\Theta}|\sum_{t=1}^n R_t(\theta)| < \infty$。因此，我们可以取一个仅依赖于样本 $\{X_t\}_{t=1}^n$ 的正随机变量 V_n，使得当γ充分小时，

$$\sup_{\theta\in\Theta}\left|\tilde{Q}_n^{(\gamma)}(\theta) - K_n(\gamma) - \tilde{Q}_n^{(0)}(\theta)\right| \le \gamma\left(\sup_{\theta\in\Theta}\left|\sum_{t=1}^{n} R_t(\theta)\right| + O(\gamma)\right) \le V_n\gamma,$$

从而完成证明。 □

参考文献

Andrade, J. A. A. 和 O'Hagan, A. (2011), 《位置和尺度参数的贝叶斯稳健性建模》, 斯堪的纳维亚统计学杂志 **38**(4), 691–711。Bai, Y.、Kang, J. 和 Song, P. X.-K. (2014), 《空间聚类数据的高效成对复合似然估计》, 《生物统计学》**70**(3), 661–670。Basu, A.、Harris, I. R.、Hjort, N. L. 和 Jones, M. (1998), 《通过最小化密度幂散度实现稳健且高效的估计》, 《生物计量学》**85**(3), 549–559。Francq, C. 和 Zako¨ıan, J.-M. (2004), 《纯GARCH和ARMA-GARCH过程的最大似然估计》, 《伯努利》**10**, 605–637。Fujisawa, H. 和 Eguchi, S. (2008), 《在重度污染下具有小偏差的稳健参数估计》, 《多元分析杂志》**99**(9), 2053–2081。Ghosh, A. 和 Basu, A. (2013), 《使用密度幂散度对独立非齐次观测进行稳健估计及其在线性回归中的应用》, *Electronic Journal ofStatistics***7**, 2420–2456。Ghosh, A. 和 Basu, A. (2016), 《使用密度幂散度进行稳健贝叶斯估计》, 《统计数学研究所年刊》**68**(2), 413–437。Hamadeh, T. 和 Zako¨ıan, J.-M. (2011), 《一类非线性GARCH过程的ls和准最大似然估计量的渐近性质》, *Journal of Statistical Planning and Inference***141**(1), 488–507。Hoffman, M. D.、Gelman, A. 等 (2014), 《无U型转弯采样器：在哈密顿蒙特卡洛中自适应设定路径长度。》, *J. Mach. Learn. Res.***15**(1), 1593–1623。Hooker, G. 和 Vidyashankar, A. N. (2014), 《通过差异度实现贝叶斯模型稳健性》, 《*Test*》**23**(3), 556–584。Huber, P. J. (1981), 《稳健统计》, Wiley, 纽约。Lee, S. 和 Song, J. (2009), 《GARCH模型的最小密度幂散度估计量》, 《*Test*》**18**, 316–341。Ling, S. 和 McAleer, M. (2010), 《时间序列模型的一般渐近理论》, 《荷兰统计学》**64**(1), 97–111。Loh, P.-L. (2024), 《现代稳健统计学的理论综述》, 《统计学及其应用年度评论》**11**, 1–28。Nakagawa, T. 和 Hashimoto, S. (2020), 《通过 γ-散度实现稳健贝叶斯推断》, *Communi-cations in Statistics-Theory and Methods* **49**(2), 343–360。

O’Hagan, A. 和 Pericchi, L. (2012)，‘贝叶斯统计与先验–数据冲突问题’，*Statistical Science* **27**(1)，1–31。Ronchetti, E. (2021)，‘稳健统计对统计科学的主要贡献及一项新挑战’，*METRON***79**(2)，127–135。Ruli, E., Sartori, N. 和 Ventura, L. (2020)，‘稳健近似贝叶斯推断’，*Journal ofStatistical Planning and Inference***202**，1–16。Song, J. (2017)，‘离散观测扩散过程的离散参数稳健估计’，*Statistica Sinica***27**(1)，373–388。Song, J. (2021)，‘存在离群值时的序贯变点检验：基于密度幂散度的方法’，*Electronic Journal of Statistics* **15**(1)，3504–3550。Song, J. 和 Kang, J. (2021)，‘存在离群值时的参数变更检验：基于密度幂散度的方法’，*Journal of Statistical Computation and Simulation***91**(5)，1016–1039。Straumann, D. 和 Mikosch, T. (2006)，‘条件异方差时间序列中的拟极大似然估计：一种随机递推方程方法’。Warwick, J. 和 Jones, M. (2005)，‘如何选择稳健性调节参数’，*Journal of StatisticalComputation and Simulation* **75**(7)，581–588。