arXiv:2603.05119v1 [q-fin.ST] 5 Mar 2026

高频 CIR 和 CKLS 模型中扩散与跳跃分量的渐近可分性

Sourojyoti Barick 跨学科统计研究部，印度统计研究所，加尔各答

2026年3月6日

摘要

本文基于最小密度幂散度估计量（MDPDE），针对具有高频渐近性质、离散观测的 CKLS 型跳跃—扩散过程，构建了一个用于跳跃检测的稳健参数框架。该方法利用了扩散增量与跳跃增量之间内在的渐近尺度分离：扩散增量以 $\sqrt{\Delta_n}$ 的速率衰减，而跳跃增量则保持非消失的随机幅度。通过采用基于稳健 MDPDE 的漂移和扩散系数估计量，我们构造了标准化残差，其极值行为为连续与不连续分量之间的统计判别提供了合理依据。我们证明，在扩散区间内，归一化残差的最大值收敛于 Gumbel 极值分布，从而得到明确且渐近有效的检测阈值。在此基础上，我们证明了所提出稳健检测程序的分类一致性：在适当的渐近条件下，正确识别所有跳跃增量和扩散增量的概率趋近于 1。基于 MDPDE 的归一化能够减弱异常增量的影响，并在存在不连续性的情况下稳定检测边界。模拟结果表明，稳健性能够提升有限样本稳定性并减少虚假检测，同时不损害渐近有效性。所提出的方法为高频随机系统中的跳跃识别提供了一种理论严谨且实践稳健的方法。

1 引言

随机微分方程（SDE）构成了一个用于对受随机扰动影响的动态演化系统进行建模的基础数学框架。在金融计量经济学中，此类模型在分析利率动态、资产价格和波动率过程中发挥着核心作用。在众多重要模型设定中，由Cox 等人（1985）提出的 Cox–Ingersoll–Ross（CIR）过程因能够刻画均值回复行为，同时保持状态变量的严格正性而受到广泛关注。CIR模型是更一般的 Chan–Karolyi–Longstaff–Sanders（CKLS）模型的一个特例，该模型由 Chan 等人（1992）提出，提供了一个灵活的参数族，能够容纳实证观察到的特征，例如状态依赖波动率和异质的扩散弹性。

均值回复扩散过程在利率建模中的实证相关性已得到充分确立。短期利率呈现出持续回归随机均衡水平的倾向，反映了宏观经济基本面、货币政策干预和市场预期。这类动态自然地由连续时间扩散模型表示，它们以简约且易于分析的方式描述短期利率演化。这些模型构成现代无套利期限结构理论的结构性基础，并已系统地纳入一般无套利框架；例如，Heath 等人（1992）和Duffie 等人（1993）。由于衍生品估值、对冲和风险管理明确依赖于支配利率动态的漂移和扩散系数，因此，对离散观测扩散进行可靠的统计推断至关重要。这推动了对统计效率估计程序的大量研究，其中最著名的是最大似然估计（MLE）和条件最小二乘法（CLS）；在高频抽样方案下，这些方法可提供一致且渐近正态的估计量。

关于CIR过程统计推断的早期研究包括Overbeck（1998）的工作，他提出了CLS型估计量，并证明了这些估计量在离散和连续观测方案下均具有强一致性。后续研究包括Alaya等人（2012）和Alaya等人（2013），结果表明，与CLS估计量相比，基于似然的方法通常具有更优的有限样本效率和渐近效率，尤其是在高频抽样机制下。这些研究确立了最大似然估计量的一致性、渐近正态性和更优效率性质，尤其是在过程趋近边界区域时。Li等人（2013）将研究扩展至更一般的CKLS框架，在放宽的正则条件和更一般的扩散结构下推导了CLS估计量的渐近性质。Feng等人（2022）则发展了其他推断范式，包括融入先验结构信息的贝叶斯估计程序，从而提升了小样本环境下推断的稳定性。

尽管应用十分广泛，经典扩散模型从根本上依赖于底层随机过程随时间连续演化这一假设。然而，高频金融数据的经验证据明确表明存在不连续性，即跳跃，其源于宏观经济公告、流动性冲击、机构交易以及其他结构性市场事件。这些观测推动了跳跃—扩散模型的发展，该类模型在连续扩散分量之外加入不连续的纯跳过程；例如，可参见Merton(1976) 与Duffie 等人 (2000)的开创性贡献。

从统计学角度看，跳跃的存在给推断带来了深刻挑战。在高频渐近理论下，扩散驱动的增量以速率 $\mathcal{O}(\sqrt{\Delta t})$标度，而跳跃诱导增量仍保持 $\mathcal{O}(1)$阶。渐近标度上的这一根本差异为连续与不连续分量之间的统计判别提供了理论基础，并推动了大量致力于跳跃检测与半鞅动态分解的文献。

Barndorff-Nielsen等人Barndorff-Nielsen等人，2004；2006 率先提出了基于实现变差度量的非参数跳跃检测程序，利用跳跃存在时二次变差与双幂变差不同的渐近行为。这些方法随后由Jacod（2008）和A¨ıt-Sahalia等人（2009）进一步发展；他们在一般半鞅模型下建立了严格的极限定理和用于检测不连续性的相合统计检验。其中一项影响深远的成果是Lee等人（2008）提出的局部归一化跳跃检测统计量：在不存在跳跃时，该统计量弱收敛于标准高斯分布；而在存在不连续性时，则表现出发散行为。这一框架为识别高频观测中的跳跃发生提供了一种具有良好统计可处理性且具有渐近理论依据的机制。

尽管取得了这些进展，现有的大多数跳跃检测方法仍是在完全非参数框架下构建的，并未明确利用参数扩散模型所固有的结构信息。在许多实际应用中，CIR和CKLS等参数模型由于具有可解释性、可处理性，并且与结构分析和预测用途密切相关，仍然不可或缺。一种常见的实证策略是将非参数跳跃检测程序作为初步过滤步骤，然后基于过滤后的样本进行参数估计。然而，这种两阶段程序会引入额外的随机变异性，并且本质上容易产生误分类错误；这些错误可能传递至参数估计阶段，从而降低统计效率。

一个核心推断难题源于基于似然的估计量对跳跃诱导增量污染的内在敏感性。经典似然和拟似然方法在高斯近似下会对大偏差施加二次惩罚，因此对非典型观测高度敏感。因此，即使跳跃数量相对较少，也可能对所得参数估计产生不成比例的影响，导致偏差、低效和推断不稳定。离散观测扩散领域已对这一现象进行了严格分析；例如，参见Yoshida (1988)、Kessler (1997)和Uchida etal. (2012)。

这些局限凸显了稳健统计方法的必要性：这类方法能够减轻污染的影响，同时在正确设定模型下保留渐近效率。基于散度的估计方法提供了一种原则清晰且理论基础扎实的途径，它以经验分布与模型隐含分布之间的差异泛函取代经典似然函数。特别是，Basu et al. (1998) 提出的最小密度幂散度估计量通过一个调节参数，在最大似然估计与高度稳健的方法之间实现连续过渡，从而控制效率与稳健性之间的权衡。

密度幂散度估计量具有若干理想的理论性质，包括一致性、渐近正态性和有界影响函数，因此尤其适用于存在污染或模型设定错误时的推断。其统计性质已在独立和相关数据情形下得到广泛研究；例如，Ghosh 等人（2013）和Lee 等人（2013）。对于离散观测扩散过程，基于散度的方法提供了一种自然且有理论依据的机制，可减弱跳跃诱导增量的影响；这些增量表现为对连续扩散动力学的局部偏离。

受上述考虑启发，本文为 CKLS 型扩散模型中的稳健跳跃识别与参数估计构建了一个统一的参数框架。所提出的方法将基于散度的稳健估计与一种具有统计学依据的跳跃识别机制相结合，该机制源自局部标准化残差的渐近行为。该方法在完全参数化的设定下，提供了一个连贯的推断框架，可同时实现稳健参数估计和一致的跳跃检测。

本文其余部分安排如下。第 2 节回顾参数化 CKLS 模型，并总结经典估计量在一致性和渐近正态性方面的现有结果。第 2.1节介绍所提出的稳健估计框架，并确立其渐近性质。第 3节构建具有严格统计学依据的跳跃检测程序，并推导相应的渐近临界阈值。第 4节通过模拟研究展示所提出方法的有限样本表现。结论见第5节。所有技术性证明和辅助结果均置于附录 A，更多模拟结果见补充材料。

2 理论背景与说明

本节概述概率框架及经典渐近结果，这些内容构成了后文鲁棒方法的基准。所有论述均为标准结论，仅用于明确记号并阐明参考模型。详细推导可参见所引文献。

我们考虑由Chan 等人（1992）提出的 CKLS 扩散过程，

$$dX_t = (\beta_1 - \beta_2 X_t)\,dt + \sigma X_t^{\gamma}\,dW_t, \tag{1}$$

其中 $\gamma \in [1/2, 1]$，而 (W_t) 是标准布朗运动。本文始终采用以下条件。

假设 1。 参数满足 $\beta_1 > 0$、 $\beta_2 > 0$ 和 $\sigma > 0$，且初始条件满足$X_0 > 0$。

(1) 的遍历性和边界性质已得到充分研究；参见 Chan 等人（1992）、Karlin 等人（1981）以及 Andersen 等人（2007）。在假设 1 下，该过程存在唯一的平稳分布。对于 $\gamma \in (1/2,\ 1]$，零边界不可达；而在边界情形 $\gamma = 1/2$ 下，只要满足 $2\beta_1 > \sigma^2$，该边界仍不可达。平稳密度的形式为

$$f(r) = C(\gamma)\, r^{-2\gamma} \exp\{Q(r;\gamma)\}, \tag{2}$$

其中 $Q(;\gamma)$ 由尺度测度与速度测度确定。

为说明估计框架，先考虑 CIR 设定（$\gamma = 1/2$），

$$dX_t = (\beta_1 - \beta_2 X_t)\,dt + \sigma\sqrt{X_t}\,dW_t, \tag{3}$$

由Cox 等人（1985年）提出。令 $\{X_{t_i}\}_{i=0}^{n}$ 表示在网格间距为 Δ_n的等距网格上取得的观测。欧拉–丸山近似（Kloeden 等人，1992）给出

$$X_{t+\Delta_n} = X_t + (\beta_1 - \beta_2 X_t)\Delta_n + \sigma\sqrt{X_t}\Delta W_t.$$

定义

$$y_t = \frac{X_{t+\Delta_n} - X_t}{\sqrt{X_t \Delta_n}}, \qquad z_{1t} = \frac{\sqrt{\Delta_n}}{\sqrt{X_t}}, \qquad z_{2t} = -\sqrt{X_t \Delta_n}.$$

那么，离散化模型可表示为线性形式

$$y_t = \beta_1 z_{1t} + \beta_2 z_{2t} + \varepsilon_t, \qquad \varepsilon_t = \frac{\sigma}{\sqrt{\Delta_n}} \Delta W_t, \tag{4}$$

这是离散观测扩散经典推断的基础；参见 Kessler (1997) 和 Dehtiar 等人（2022）。令 $\hat{\boldsymbol{\beta}}_n$ 表示与（4）相对应的普通最小二乘估计量，并令 $\widehat{\sigma}_n^2$ 表示相应的残差方差估计量。

我们考虑高频渐近情形

$$\Delta_n \to 0, \qquad n\Delta_n \to \infty, \tag{5}$$

在此情形下，随着样本量增大，观测网格变得越来越密。这一框架对应于一种填充式 渐近方案：时间离散化逐步细化，同时观测时域不断延长。

在（5）条件下，标准 鞅论证表明，这些估计量具有一致性和渐近正态性 .

结果 1（一致性）。$\hat{\boldsymbol{\beta}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\beta}, \qquad \widehat{\sigma}_n^2 \xrightarrow{\mathbb{P}} \sigma^2.$

结果 2（渐近正态性）。$\frac{\sqrt{n\Delta_n}}{\widehat{\sigma}_n} \left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} \right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \ \Sigma^{-1}\right)$，其中 Σ 取决于平稳分布的矩；参见 *Kessler (1997)*。

对于 CIR 情形,

$$\Sigma = \begin{pmatrix} \frac{\beta_2}{\beta_1 - \sigma^2/2} & -1 \\ -1 & \frac{\beta_1}{\beta_2} \end{pmatrix},$$

而对于一般的 CKLS模型,

$$\Sigma = \begin{pmatrix} \int_0^\infty r^{-2\gamma} f(r)\, dr & -\int_0^\infty r^{1-2\gamma} f(r)\, dr \\ -\int_0^\infty r^{1-2\gamma} f(r)\, dr & \int_0^\infty r^{2-2\gamma} f(r)\, dr \end{pmatrix}.$$

在填充渐近条件下，普通最小二乘估计量和高斯准最大似然估计量均对同一个真实参数向量具有一致性；参见Dehtiar 等人（2022）和Mishura 等人（2022）。由于这两个估计量都依概率收敛到相同的确定性极限，它们之间的差异必然渐近趋于零。因此，前述结果同样适用于基于似然的估计量。

前述渐近分析是在纯扩散设定所隐含的高斯增量结构下推导得出的。这种结构确保了似然的局部二次近似，并导出标准的准似然估计方程。然而，在实证应用中，高频金融序列往往表现出超额峰度和与高斯基准不相容的剧烈波动。由于最小二乘法和高斯准似然估计量对所有增量赋予相同权重，它们在有限样本中的表现可能会受到异常大增量的不利影响。这一考虑促使我们采用一种备择假设下的估计框架，该框架在扩散模型下保持一致性，同时降低极端观测值的影响。

2.1 通过密度幂散度进行稳健估计

为形式化稳健性，我们按照Basu etal. (1998)的思路，以密度幂散度目标函数取代高斯准似然准则。令 $f_{\boldsymbol{\theta}}$ 表示工作高斯增量密度。对于$\alpha \geq 0$，经验分布与 $f_{\boldsymbol{\theta}}$ 之间的密度幂散度给出如下估计准则

$$\mathcal{L}_n^{(\alpha)}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \int f_{\boldsymbol{\theta}}^{1+\alpha}(x)\, dx - \left(1 + \frac{1}{\alpha}\right) f_{\boldsymbol{\theta}}^{\alpha}(X_i) + \frac{1}{\alpha} \right\},$$

通常的似然在 $\alpha = 0$处得以恢复。

由此得到的最小密度幂散度估计量（MDPDE）会降低模型隐含密度较小的增量的权重，从而限制极端观测值的影响，同时在正确设定下保持一致性。

令 $\theta = (\beta_1,\ \beta_2,\ \sigma)^\top$ 表示未知参数向量，并考虑与扩散过程的欧拉离散化相关联的高斯条件密度，

$$f_\theta(y_t \mid X_t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_t - \beta_1 z_{1t} - \beta_2 z_{2t})^2}{2\sigma^2}\right),$$

其中，X_t 表示时刻 t 可用的条件信息，而 $(z_{1t},\ z_{2t})$ 是相应的回归变量。该高斯近似作为增量条件分布的工作模型，是经典推断和稳健推断的基础。

对于稳健性　　调节参数 $\alpha > 0$，经验密度幂散度目标函数 i　　s
定义为

$$H_n^{(\alpha)}(\theta) = \sum_{t=1}^{n} \left[\int f_\theta^{1+\alpha}(y \mid X_t)\, dy - \frac{1+\alpha}{\alpha} f_\theta^{\alpha}(y_t \mid X_t) + \frac{1}{\alpha} \right].$$

第一项是仅取决于模型的归一化项，而第二项则降低在假定的条件密度下不太可能出现的观测的权重。当 $\alpha \downarrow 0$ 时，准则$H_n^{(\alpha)}(\theta)$ 收敛于负对数似然，相应的估计量则退化为经典最大似然估计量。

MDPDE 定义如下：

$$\hat{\theta}_n^{(\alpha)} = \arg\min_{\theta \in \Theta} H_n^{(\alpha)}(\theta),$$

其中，Θ 表示参数空间。调节参数 α 控制稳健性与效率之间的权衡；α 的值越大，在正确设定的高斯模型下，对异常值的抵抗力越强，但效率会有所损失。

结果3（MDPDE 的渐近性质）。假设 *Lee 等人（2013）* 的正则条件成立，且抽样方案满足填充渐近机制（5）。若此外，$n(\Delta_n)^q \to 0$ 对某个 $q > 1$，均成立，则最小密度幂散度估计量满足

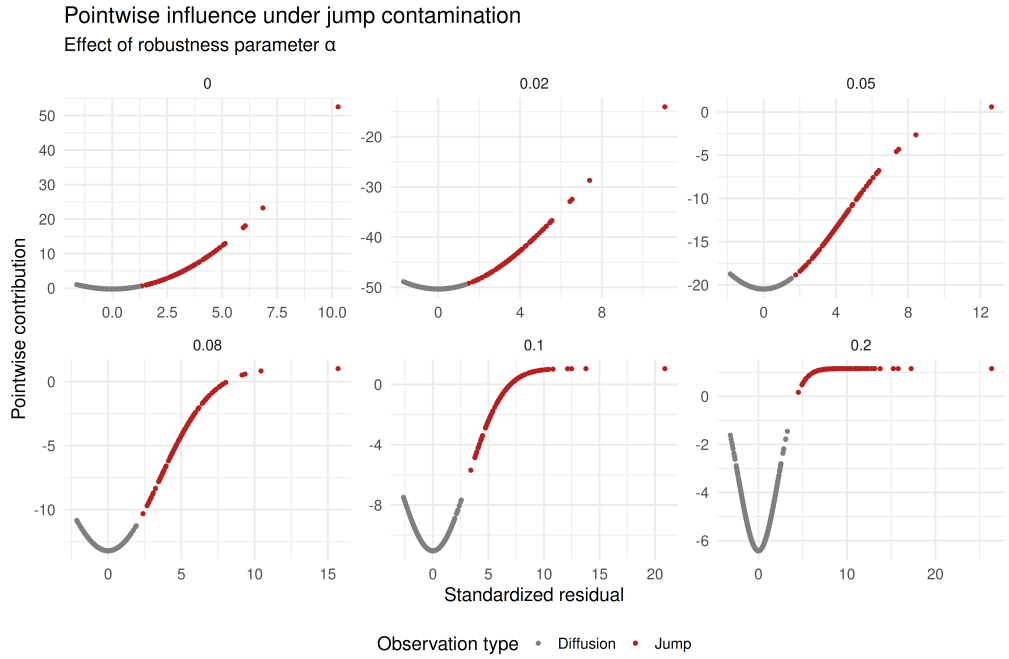
$$\hat{\theta}_n^{(\alpha)} \xrightarrow{\mathbb{P}} \theta_0,$$

其中，θ_0 表示真实参数值。此外，若 $n(\Delta_n)^2 \to 0$，则该估计量渐近正态，其协方差矩阵的显式表达式见 *Lee 等人（2013）*。

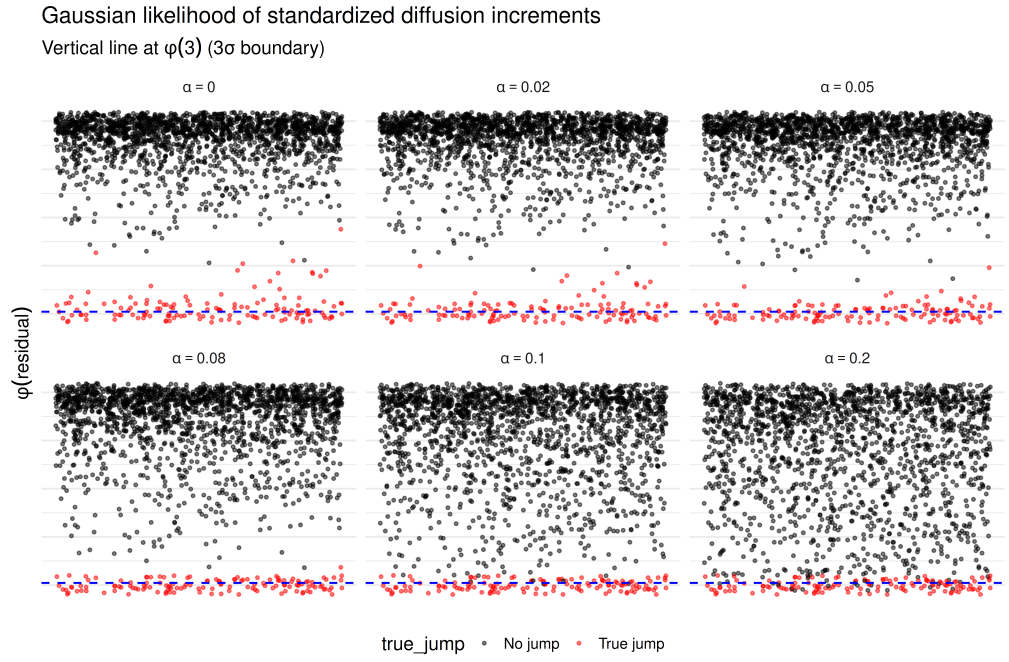
注释1。最小密度幂散度估计量的一项基本性质是其影响函数有界，这确保了在出现异常大的观测值时仍具有稳定性。在离散观测扩散模型中，异常大的增量可能源于未建模的跳跃分量，或与纯扩散设定存在轻微偏离。与影响函数无界的经典似然估计量不同，*MDPDE* 会自动降低此类增量的权重，避免它们在估计方程中占据主导地位。因此，经典估计量与稳健估计量之间的系统性差异可作为结构性设定错误的自然诊断信号，包括跳跃的存在。

这种区别在高频情形下尤为重要。扩散样本路径几乎必然连续，其增量以 $\sqrt{\Delta_n}$ 的速率趋于零，而 $\Delta_n \to 0$，而跳跃诱导增量则保持有限幅度。因此，相对于连续扩散动力学，跳跃增量表现为局部异常值，并可能对基于似然的方法产生不成比例的影响，往往导致估计值出现偏差或不稳定。基于稳健散度的估计量通过限制极端观测值的贡献来减轻这种影响，从而保持拟合模型的稳定性和可解释性。

在受约束参数模型中，这还带来一项额外优势：极端增量可能会使基于似然的估计值趋向参数空间边界，从而可能违反正性或遍历性等结构条件。通过减弱非典型观测的影响，稳健估计能够维持数值稳定性，并确保估计参数处于具有经济和统计意义的范围内。基于这些考虑，我们考虑 CKLS 扩散



(a) *Pointwise contribution (influence) of standardized increments under jump contamination for different values of the robustness parameter α. Diffusion-driven increments are shown in gray, while jump-induced increments are shown in red.*



(***b***) *Gaussian likelihood values φ(Residuals) of standardized increments with a horizontal reference line at φ(3), corresponding to the classical 3σ boundary, shown for different values of the robustness parameter α.*

图 1. *Effect of robustness on pointwise influence and likelihood contributions under jump contamination. Increasing α progressively bounds the influence of large, jump-induced increments while preserving the contribution of diffusion-driven observations.*

加入跳跃分量。具体而言，对于已知的弹性参数 γ，我们研究跳扩散模型

$$dX_t = (\beta_1 - \beta_2 X_t)\,dt + \sigma X_t^{\gamma}\,dW_t + dJ_t, \tag{6}$$

其中，$\{J_t\}_{t\geq 0}$ 是一个与 W_t 独立的纯跳过程。这类模型广泛应用于利率和波动率建模，用以捕捉市场的突然波动。经典的 Cox–Ingersoll–Ross 模型是其特例，对应于 $\gamma = 1/2$。

图 1a 和 1b展示了稳健性对标准化增量逐点贡献的影响。当 $\alpha = 0$ 时，对应于高斯似然估计，极端增量的影响无界，跳跃观测可能会主导估计方程。随着稳健性参数α 增大，影响函数变得有界，跳跃诱导增量的贡献逐渐减弱，而扩散驱动的增量基本不受影响。这一行为体现了 MDPDE 的内在稳健性，以及其从跳跃污染中分离出连续扩散结构的能力。

综上所述，这些考量凸显了基于散度的估计在高频扩散模型中的一项根本优势。通过利用扩散增量与跳跃增量不同的缩放特性，MDPDE 能够提供稳定的参数估计值，并为检测偏离纯扩散假设的情况提供有原则的依据。这种诊断能力将在后续章节所介绍的稳健参数化跳跃检测方法的发展中发挥核心作用。

3 跳跃检验方法

设 $\{X_{t_i}\}_{i=0}^{n}$ 表示上文方程（6）中引入的过程的离散观测，记录于等距时刻 $t_i = i\Delta_n$，覆盖固定时间范围，其中 $\Delta_n \to 0$。将增量记为

$$\Delta_i^n X := X_{t_i} - X_{t_{i-1}}$$

并将跳跃增量定义为

$$\Delta_i^n J = J_{t_i} - J_{t_{i-1}}, \tag{7}$$

^ ^ ^其中指标 i 处存在跳跃对应于事件 $\Delta_i J \neq 0$。设 $\boldsymbol{\theta}_n := (\beta_{1n},\ \beta_{2n},\ \hat{}\sigma_n)$表示模型参数的相合估计量，且满足

$$\hat{\beta}_{1n} \xrightarrow{P} \beta_1, \qquad \hat{\beta}_{2n} \xrightarrow{P} \beta_2, \qquad \hat{\sigma}_n \xrightarrow{P} \sigma, \qquad \Pr(\hat{\sigma}_n > 0) \to 1.$$

我们定义标准化增量统计量

$$Z_{i,n} = \frac{\Delta_i^n X - (\hat{\beta}_{1n} - \hat{\beta}_{2n} X_{t_{i-1}})\Delta_n}{\hat{\sigma}_n X_{t_{i-1}}^{\gamma}\sqrt{\Delta_n}}. \tag{8}$$

该归一化去除估计漂移，并按估计的局部扩散幅度对增量进行重标度。因此，由连续鞅分量产生的增量被调整到单位高斯尺度，而跳跃诱导增量由于其更大的阶而保持渐近可分离。这种分离构成了下文所发展的参数化跳跃检测程序的基础。

下述定理建立了标准化增量统计量向其连续鞅分量和跳跃分量的渐近分解。

Theo注 1. 假设参数估计量是一致的。那么，对于每个固定的指标 i,

$$Z_{i,n} = \frac{\Delta_i^n W}{\sqrt{\Delta_n}} + \frac{\Delta_i^n J}{\sigma X_{t_{i-1}}^{\gamma}\sqrt{\Delta_n}} + o_p(1). \tag{9}$$

因此，若 $\Delta_i^n J = 0$，则

$$Z_{i,n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),$$

而如果 $\Delta_i^n J \neq 0$ 且跳跃幅度非退化，

$$|Z_{i,n}| \xrightarrow{P} \infty.$$

定理1 对连续与不连续分量之间的逐点渐近分离进行了形式化。归一化后，扩散驱动的增量在随机意义下保持有界，而跳跃引起的增量则以 $\Delta_n^{-1/2}$的速率发散。

接下来，我们刻画无跳跃区间上统计量的最大波动，这决定了自然检测边界。

定理 2. 设

$$\mathcal{C}_n = \{i \in \{1, \dots, n\} : \Delta_i^n J = 0\}.$$

在参数估计量一致的条件下，

$$\max_{i \in \mathcal{C}_n} |Z_{i,n}| = \sqrt{2\log(n)} + O_p(1),$$

更确切地说，

$$\frac{\max_{i \in \mathcal{C}_n} |Z_{i,n}| - a_n}{b_n} \Rightarrow \Lambda,$$

其中

$$a_n = \sqrt{2\log n} - \frac{\log\log n + \log(\pi)}{2\sqrt{2\log n}}, \qquad b_n = \frac{1}{\sqrt{2\log n}},$$

且 Λ 表示标准 *Gumbel* 分布。

这一极值刻画确定了连续路径最大值的内在增长率，并为阈值选择提供了有据可依的依据。

推论 1. 在相同假设下，

$$\Pr\left(\max_{i \in \mathcal{C}_n} |Z_{i,n}| < \sqrt{2\log(n)} < \min_{i:\Delta_i^n J \neq 0} |Z_{i,n}|\right) \to 1.$$

推论 1确立了连续分量与跳跃分量之间的一致渐近分离，从而确保量级为 $\sqrt{2\log(1/\Delta_n)}$ 的阈值能够在渐近意义下区分这两种情形。

定理 3 （参数化跳跃检测的一致性）。 假设定理 *1*、*2*中的假设成立，且估计量具有一致性。那么，

$$\Pr\left(\max_{i \in \mathcal{C}_n} |Z_{i,n}| > \xi_n\right) \to 0,$$

并且，对于任意满足 $\Delta_i^n J \neq 0$ 的指标 i，

$$\Pr(|Z_{i,n}| \leq \xi_n) \to 0.$$

因此，

$$\Pr(\textit{correct classification of all increments}) \to 1.$$

可靠的分离需要对扩散尺度参数进行稳定估计。经典似然估计量可能会受到跳跃诱导的污染而严重扭曲，从而影响归一化。为解决这一问题，我们采用基于最小密度幂散度的稳健估计量。

令 $\hat{\boldsymbol{\theta}}_n^\alpha := (\beta_{1n}^\alpha,\ \beta_{2n}^\alpha,\ \hat{\sigma}_n^\alpha)$ 表示稳健估计量，并定义相应的归一化统计量

$$Z_{i,n}^\alpha = \frac{\Delta_i^n X - (\hat{\beta}_{1n}^\alpha - \hat{\beta}_{2n}^\alpha X_{t_{i-1}})\Delta_n}{\hat{\sigma}_n^\alpha X_{t_{i-1}}^\gamma \sqrt{\Delta_n}}. \tag{10}$$

根据稳健估计量的一致性，渐近展开（9）以及定理 2 的极值结果对 $Z_{i,\ n}^\alpha$ 仍然成立，同时增强了污染条件下的稳定性。

我们将指数为 i 的跳跃识别表述为检验问题

$$H_{0,i} : \Delta_i^n J = 0, \qquad H_{1,i} : \Delta_i^n J \neq 0.$$

在 $H_{0,\ i}$ 下，

$$Z_{i,n}^{\alpha} \Rightarrow \mathcal{N}(0,1),$$

而在 $H_{1,\ i}$ 跳跃幅度非零的情况下，

$$|Z_{i,n}^{\alpha}| \xrightarrow{P} \infty.$$

因此，连续与不连续状态可以通过标准化统计量的大小渐近地区分。受定理 2启发，我们引入检测阈值

$$\xi_n = \sqrt{2\log n} + c_n, \qquad c_n \to \infty,$$

并在满足以下条件时判定时刻 t_i 发生跳跃

$$|Z_{i,n}^{\alpha}| > \xi_n. \tag{11}$$

将真实跳跃索引集和估计跳跃索引集定义为

$$\mathcal{J}_n = \{i \in \{1,\dots,n\} : \Delta_i^n J \neq 0\}, \qquad \hat{\mathcal{J}}_n = \left\{i \in \{1,\dots,n\} : |Z_{i,n}^{\alpha}| > \xi_n\right\}. \tag{12}$$

结合稳健参数估计量的一致性与定理 3 可得

$$\Pr\left(\hat{\mathcal{J}}_n = \mathcal{J}_n\right) \to 1.$$

因此，在跳跃索引被一致识别的条件下，检测时刻的跳跃幅度估计为

$$\Delta_i^n \hat{J} = \Delta_i^n X - \left(\hat{\beta}_{1n}^{\alpha} - \hat{\beta}_{2n}^{\alpha} X_{t_{i-1}}\right)\Delta_n, \qquad i \in \hat{\mathcal{J}}_n.$$

前述结果确立了连续扩散分量与跳跃分量之间的渐近分离。具体而言，最大标准化扩散增量的极值增长决定了适当的检测边界，而存在跳跃时统计量的发散则确保了对不连续性的相合识别。稳健参数归一化进一步稳定了尺度估计，从而提高了在跳跃污染和轻微模型设定错误情况下的可靠性。

4 模拟研究

我们开展模拟研究，以考察所提出的稳健参数估计与跳跃检测程序的有限样本表现。数据生成过程由（6）给出。跳跃分量（J_t）设定为复合泊松过程

$$J_t = \sum_{k=1}^{N_t} Y_k, \tag{13}$$

其中，(N_t) 是强度为 λ 的泊松过程，跳跃大小 (Y_k) 独立同分布，且

$$Y_k \sim \mathcal{N}(\mu_J, \sigma_J^2).$$

该过程在等距时间点上观测，采样间隔为 $\Delta_n = n^{-0.55}$，对应于满足 $\Delta_n \to 0$ （当 $n \to \infty$ 时）的高频渐近情形。底层扩散参数设为 $\beta_1 = 1$、 $\beta_2 = 0.8$、 $\sigma = 0.3$ 和 $\gamma = 0.7$。

模拟研究考虑了样本量和跳跃分量参数的以下配置：

$$n \in \{200, 500, 1000, 1500, 2000\}, \quad \lambda \in \{1,2,3,5\}, \quad \mu_J \in \{1,2,3,4,5\}, \quad \sigma_J = 0.1, \tag{14}$$

以评估所提出的用于跳跃隔离和检测的稳健程序的性能。

我们首先通过对真实跳跃位置与检测到的跳跃位置进行直观比较来考察性能。为便于说明，在配置（n、 λ、 μ_J）=(1000、 5、3)下生成了一条具有代表性的轨迹。基于这条固定的样本路径，对以下稳健性参数进行跳跃分类

$$\alpha \in \{0, 0.05, 0.1, 0.15, 0.2, 0.25\},$$

其中，参数估计通过稳健程序进行，并使用所得估计值将增量分类为由跳跃或扩散驱动。情形 $\alpha = 0$ 对应于经典最小二乘估计量，可作为非稳健基准。

图 2 展示了上述稳健性参数 α 取值下的跳跃检测结果。我们没有展示样本路径本身，而是绘制了一阶增量（$\Delta_i^n X$），以直接突出不连续性。根据定理 2 得到的检测阈值为 $\xi_n = 3.512$。真实跳跃时刻以蓝色标记表示，而基于阈值化标准化统计量检测到的跳跃则以红色表示。

当 $\alpha = 0$ 时，对应于经典普通最小二乘估计量，该程序未能识别出若干明显分离的跳跃。随着 α 增大，检测效果显著改善，这体现了稳健归一化的稳定作用。对于较小的正值， α 的检测率迅速提高，并在 $\alpha = 0.15$附近趋于稳定；超过该值后，所有真实跳跃都能被持续识别。

这些结果说明，稳健性参数在控制灵敏度与稳定性之间的权衡方面发挥着作用。 α 的取值较小时，归一化容易受到扩散驱动的极端增量影响，导致漏检或不稳定检测；而取值适中时，则能减弱这种影响，减少伪分类，同时保持检测能力。

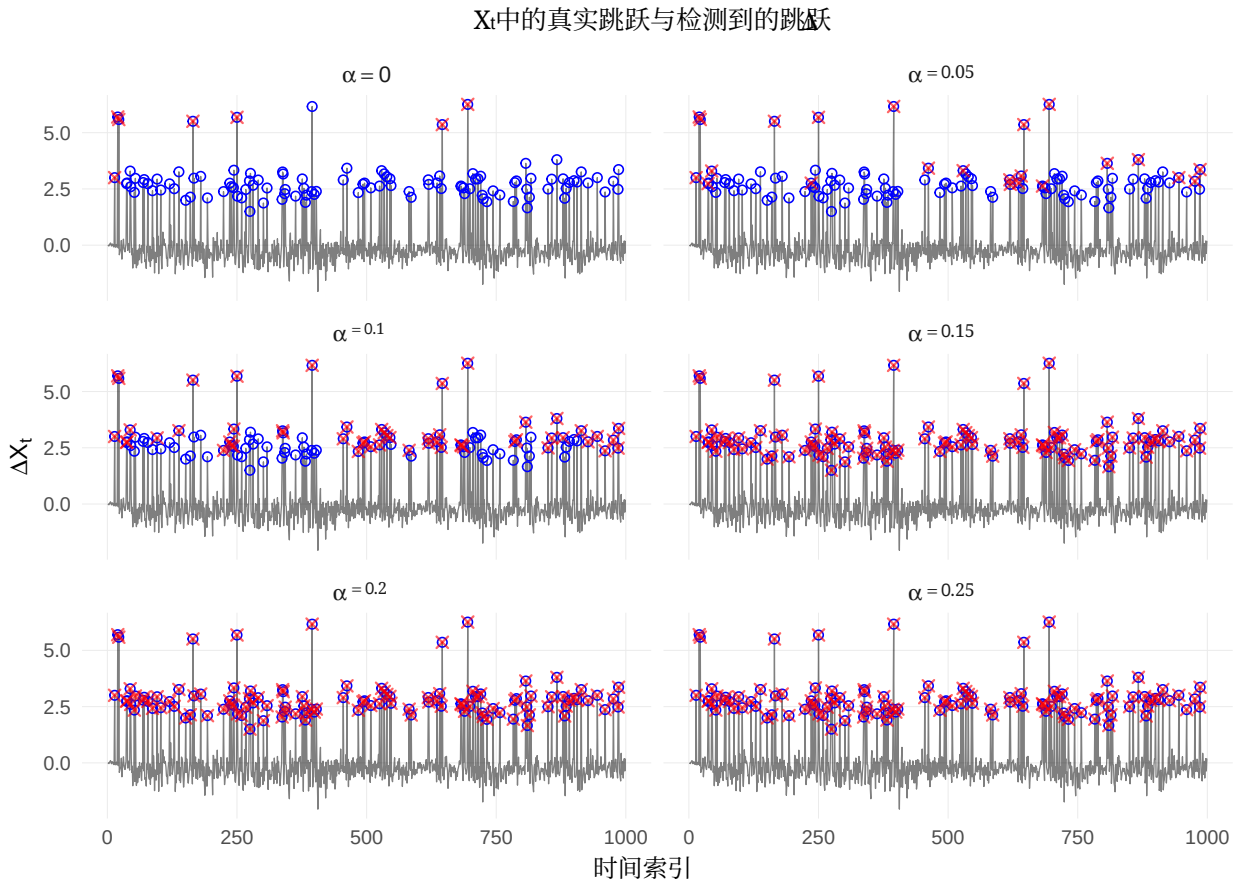


图 2 *Visual comparison of true and detected jumps in the increment process $\Delta_i^n X$ for different values of the robustness parameter α. ○ denotes true jumps whereas ✕ denotes jumps detected by the proposed procedure.*

为更全面地评估所提出的方法，我们汇总了在（14）中针对不同稳健性水平 $\alpha \in \{0, 0.05, 0.1, \ldots, 0.5\}$ 所指定的参数配置下的表现。我们使用分类准确率和参数恢复指标评估性能，具体包括 F1分数和尺度化估计误差。

令 $\mathcal{J}_n$ 和 $\hat{\mathcal{J}}_n$ 分别表示式（12）中定义的真实跳跃集和估计跳跃集。我们定义分类计数

$$\mathrm{TP} = |\mathcal{J}_n \cap \hat{\mathcal{J}}_n|, \qquad \mathrm{FN} = |\mathcal{J}_n \setminus \hat{\mathcal{J}}_n|, \qquad \mathrm{FP} = |\hat{\mathcal{J}}_n \setminus \mathcal{J}_n|.$$

相应的性能指标为

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \qquad \mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}},$$
$$\text{F1-score} = 2\,\frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}.$$

尽管模拟中跳跃强度参数 λ 固定不变，但每条轨迹中已实现的跳跃次数各不相同。因此，在有限样本中依据理论参数进行评估可能会产生误导。为此，我们将估计量与已实现的跳跃特征进行比较。将已实现跳跃均值和跳跃强度定义为

$$\tilde{\mu}_J = \frac{1}{|\mathcal{J}_n|} \sum_{i \in \mathcal{J}_n} \Delta_i^n J, \qquad \tilde{\lambda} = \frac{|\mathcal{J}_n|}{n}. \tag{15}$$

基于检测到的跳跃得到的相应估计量为

$$\hat{\mu}_J = \frac{1}{|\hat{\mathcal{J}}_n|} \sum_{i \in \hat{\mathcal{J}}_n} \Delta_i^n \hat{J}, \qquad \hat{\lambda} = \frac{|\hat{\mathcal{J}}_n|}{n}. \tag{16}$$

当抽样规模 n 和跳跃强度均处于中等水平时，已实现的跳跃次数可能较少，难以可靠地收敛至理论参数。因此，估计准确率是相对于已实现量而非总体值进行评估的。我们通过尺度化误差度量来衡量差异

$$d_M := \sqrt{\left(\frac{\tilde{\mu}_J}{\hat{\mu}_J} - 1\right)^2 + \left(\frac{\tilde{\lambda}}{\hat{\lambda}} - 1\right)^2},$$

该指标综合评估跳跃幅度和频率的估计准确性。

因此，通过上述定义的 F1分数和尺度化误差度量，综合评估检测性能和估计准确率。F1分数的取值范围为 $[0, 1]$, ，数值越大表示分类性能越好；具体而言，数值越接近 1，跳跃索引的识别就越准确。F1分数通过正确分类计数评估检测准确率，而度量 d_M 则量化估计的已实现跳跃特征与实际已实现跳跃特征之间的差异。 d_M 不受尺度影响，且当估计值与其实际对应值完全一致时，当且仅当此时其值为零；也就是说，$\hat{\ }$ $\tilde{\ }$ 即 $\hat{\mu}_J = \tilde{\mu}_J$ 和 $\lambda = \lambda_\circ$ 因此， d_M 的值越小，估计准确率越高；而值越大，则表明其与实际跳跃结构的偏离程度越高，参数估计和检测的可靠性也随之降低。从这个意义上说，F1分数反映分类性能，而 d_M 衡量参数偏离，从而为评估所提出方法的有效性提供互补视角。

图 3 报告了参数配置（见（14））下的 F1分数。结果表明，随着跳跃均值增大，检测性能显著提升。从图中左向右看，随着 μ_J取值增大，F1分数单调上升并趋近于 1，表明跳跃索引几乎可以被完美识别。样本量增大时（从上到下的面板）也能观察到类似的提升。随着稳健性参数 α 增大，样本量越大，F1分数越快收敛至 1。

对于样本量较小且跳跃强度较高（ λ较大）的情形，与可用观测相比，跳跃数量相对较多，因此难以准确识别所有真实跳跃。尽管如此，与 $\alpha = 0$ （对应普通最小二乘估计量）情形相比， $\alpha > 0$下的检测准确率显著提高，体现了所提出的稳健归一化方法的优势。这些发现与定理 3的理论结果一致，该结果意味着扩散增量保持随机阶 $\sqrt{\Delta_n}$ ，而跳跃诱导增量则发散，因此随着$\Delta_n \to 0$，两者的分离程度不断增加。

图4所示的跳跃参数估计结果呈现出类似的模式。随着扩散增量与已实现跳跃均值之间的差异（$\tilde{\mu}_J$）增大（同样是从左向右），所提出的指标更快地趋近于零，表明估计准确率有所提高。当跳跃强度非常小时（$\lambda = 1$），普通最小二乘估计量的表现与稳健估计量在 $\alpha > 0$时几乎无法区分。相比之下，在跳跃强度较高且样本量较小时，估计参数与实际参数之间的差异会更加明显。增加样本量可大幅缩小这一差异，并使该指标更陡峭地趋近于零，反映出估计精度有所提高。

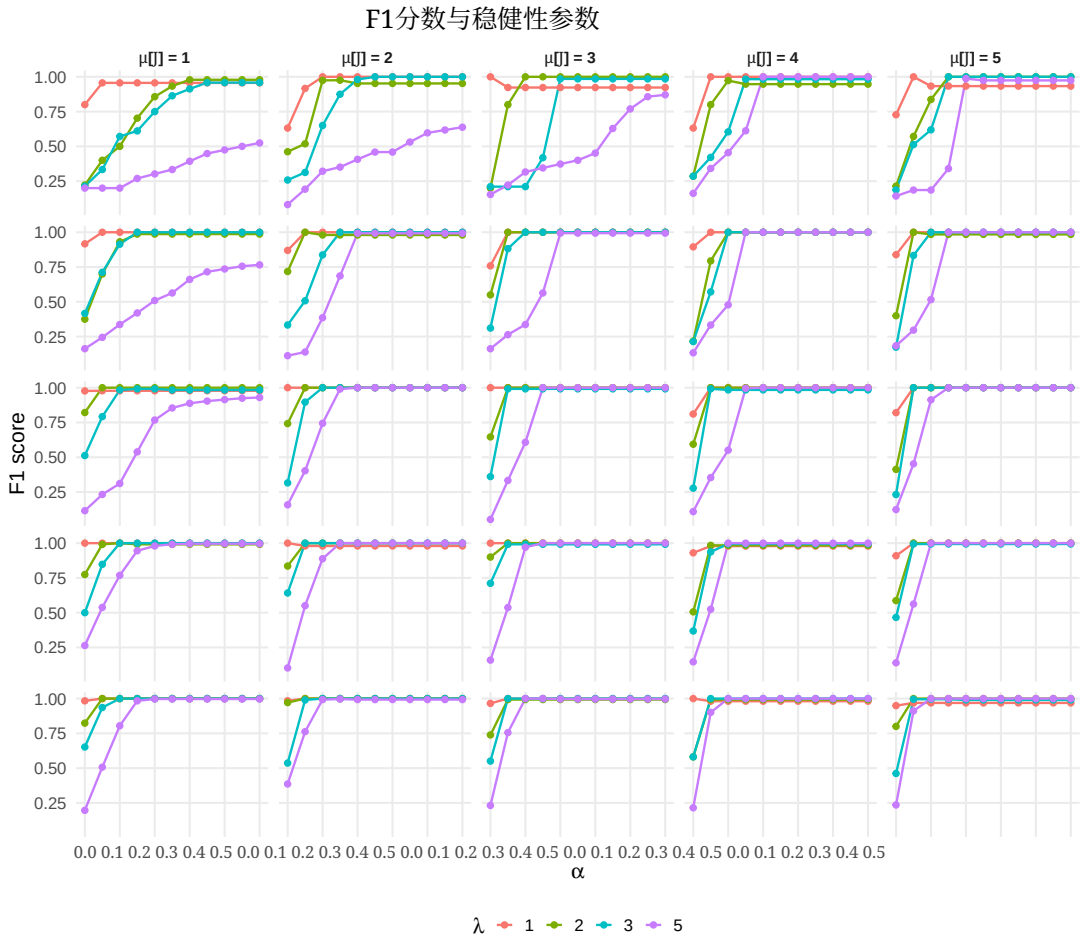


图3. *F1 score as a function of the robustness parameter α for different sample sizes (n), jump intensities (λ), and jump mean values (μ_J). Panels are arranged by sample size (rows) and jump mean (columns), while curves correspond to different jump intensities. The figure illustrates how the robustness parameter affects jump detection accuracy and how this effect varies with jump magnitude, jump frequency, and sampling resolution.*

总体而言，模拟结果为所提出检测程序背后的理论分离机制提供了实证支持。稳健参数归一化在稳定扩散尺度估计的同时，保留了跳跃诱导增量的渐近

发散性。因此，该程序能够在广泛的跳跃强度和幅度范围内可靠地识别不连续性，即使存在由跳跃造成的中度污染，也能保持稳定的性能。

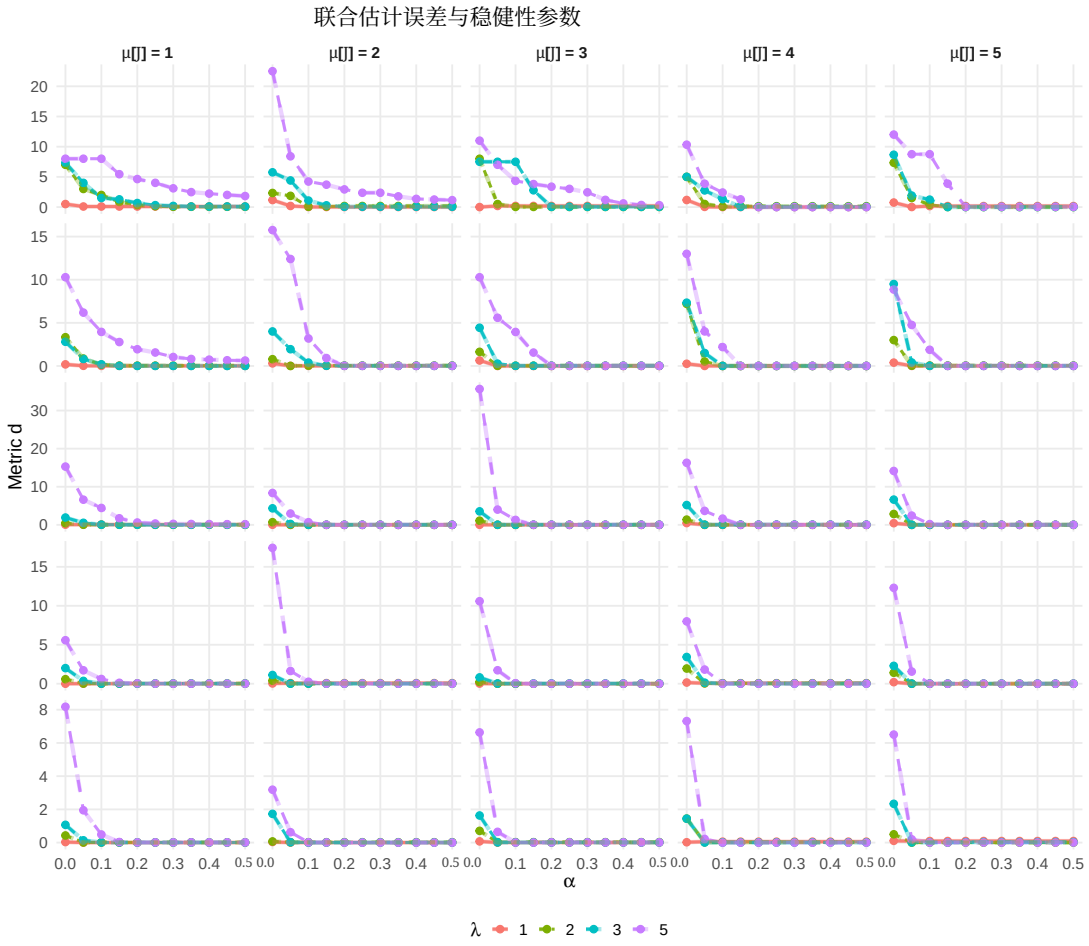


图 4 *Error metric (d_M) as a function of the robustness parameter α under varying sample sizes (n) jump. ,,, intensities (λ) and jump mean values (μ_J). Panels are arranged by sample size (rows) and jump mean (columns) while curves correspond to different jump intensities. The figure illustrates how the robustness tuning parameter influences the accuracy of jump parameter estimation and how this effect interacts with jump magnitude, jump frequency, and sampling resolution.*

5 结论

本文针对高频渐近条件下离散观测的 CKLS 跳扩散过程，构建了一个稳健估计与跳跃检测框架。该方法将基于密度幂散度的漂移参数估计量与一种由标准化残差推导出的极值理论跳跃检测规则相结合。所提出的方法旨在保持扩散基准下渐近有效性的同时，减轻跳跃或其他偏离高斯性的情形所产生的异常增量的影响。

在理论方面，我们证明了稳健估计量在填充渐近下的一致性和渐近正态性，并表明相应的残差标准化可产生一种检测统计量，其极值行为能够渐近有效地分离连续与不连续分量。

分析表明，稳健性能够使标准化过程更加规则，同时不改变渐近分类边界，从而保留跳跃分量的可辨识性。

有限样本实验验证了理论结果。结果表明，适度的稳健性能够提高参数估计和跳跃识别的稳定性，尤其是在跳跃幅度较小或样本量有限的状态下。随着采样频率提高，检测性能逐渐接近渐近状态，证实了扩散增量与跳跃增量之间的理论尺度分离。

总体而言，所提出的框架为跳跃—扩散模型中的稳健推断提供了一种统一且具有理论依据的方法。该方法在广泛的状态下实现了可靠的跳跃检测和稳定的参数估计，同时在纯扩散设定下与经典似然推断保持一致。这些结果表明，稳健性是一种有理论依据的机制，可在不损害渐近效率的前提下改善有限样本表现。

致谢

作者谨向 Prof. Diganta Mukherjee 致以特别感谢，[1] 感谢他在本研究过程中给予的宝贵指导和深刻建议。作者也衷心感谢 Prof. Indranil Sengupta[2] 在本研究开展过程中提供的帮助与支持。他们的贡献对本文最终成果的形成起到了至关重要的作用。

参考文献

A¨ıt-Sahalia, Yacine and Jean Jacod（2009）。“离散观测过程中的跳跃检验”。载于：《统计年鉴》37.1，页184–222。 DOI: 10.1214/07-AOS568（引于第2页）。Alaya, Mohamed Ben and Ahmed Kebaier（2012）。“平方根扩散的参数估计：遍历与非遍历情形”。载于：《随机模型》28.4，页609–634。 DOI: 10.1080/15326349.2012. 726042（引于第2页）。Alaya, Mohamed Ben and Ahmed Kebaier（2013）。“遍历与非遍历平方根扩散的最大似然估计量的渐近行为”。载于：《随机分析与应用》31.4，页552–573。 DOI: 10.1080/07362994.2013.798175（引于第2页）。Andersen, Leif and Vladimir Piterbarg（2007）。“随机波动率模型中的矩爆炸”。载于：《金融与随机过程》11，页29–50。 DOI: 10.1007/s00780-006-0011-7（引于第3页）。Barndorff-Nielsen, Ole E. and Neil Shephard（2004）。“幂变差与双幂变差”。载于：《金融计量经济学杂志》（引于第2页）。Barndorff-Nielsen, Ole E. and Neil Shephard（2006）。“利用双幂变差检验金融经济学中的跳跃的计量经济学方法”。载于：《金融计量经济学杂志》4.1，页1–30。 DOI: 10.1093/ jjfinec/nbi022（引于第2页）。Basu, Ayanendranath, Ian R. Harris, Nils L. Hjort, et al.（1998）。“通过最小化密度幂散度实现稳健且有效的估计”。载于：《生物计量学》85.3，页549–559。 URL: http://www.jstor. org/stable/2337385（引于第2、4页）。Chan, K. C., G.AndrewKarolyi, FrancisA. Longstaff,etal.（1992）。“短期利率备择模型的实证比较”。载于：《金融学杂志》47.3，页1209–1227。 DOI: https: // doi.org/10.1111/j.1540-6261.1992.tb04011.x（引于第1、3页）。Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross（1985）。“利率期限结构理论”。载于：Econometrica 53.2，页385–407。 ISSN: 00129682, 14680262。 URL: http://www.jstor.org/ stable/1911242（访问日期：2024年10月12日）（引于第1、3页）。

[1]教授，印度统计学院[2]教授，纽约市立大学

Dehtiar, Olena, Yuliya Mishura, and Kostiantyn Ralchenko（2022）。“Cox–Ingersoll–Ross过程漂移参数的两种估计方法：连续观测”。载于：Communications in Statistics – Theory and Methods 51.19，页6818–6833。 DOI: 10.1080/03610926.2020.1866611（引于第4页）。Duffie, Darrell, Jun Pan, and Kenneth Singleton（2000）。“仿射跳扩散的变换分析与资产定价”。载于：Econometrica 68.6，页1343–1376。 DOI: 10.1111/1468-0262.00164（引于第2页）。Duffie, Darrell and Kenneth J. Singleton（1993）。“资产价格马尔可夫模型的模拟矩估计”。载于：Econometrica 61.4，页929–952（引于第1页）。Feng, Xiaoxia and Dejun Xie（2022）。“CIR模型的贝叶斯估计”。载于：Journal of Data Science 10.2，页271–280。 ISSN: 1680-743X。 DOI: 10.6339/JDS.2012.10(2).746（引于第2页）。Ghosh, Abhik, Ayanendranath Basu, and Leandro Pardo（2013）。“广义密度幂散度与稳健估计”。载于：Journal of Statistical Planning and Inference 143.7，页1100–1114（引于第3页）。Heath, David, Robert Jarrow, and Andrew Morton（1992）。“债券定价与利率期限结构：一种新方法”。载于：Econometrica 60.1，页77–105（引于第1页）。Jacod, Jean（2008）。“实现幂变差的渐近性质”。载于：Annals of Probability（引于第2页）。Karlin, Samuel and Howard Edward Taylor（1981）。“随机过程教程（二）”。载于： URL: https://api.semanticscholar.org/CorpusID:118412345（引于第3页）。Kessler, Mathieu（1997）。“根据离散观测估计遍历扩散”。载于：Scandinavian Journal of Statistics 24.2，页211–229（引于第2、4页）。Kloeden, Peter E. and Eckhard Platen（1992）。《随机微分方程的数值解》。第23卷。Stochastic Modelling and Applied Probability。柏林，海德堡：Springer（引于第3页）。Lee, Sangyeol and Junmo Song（2013）。“扩散过程的最小密度幂散度估计量”。载于：Annals of the Institute of Statistical Mathematics 65.2，页213–236。 DOI: 10.1007/s10463-012-0366-9（引于第3、5页）。Lee, Suzanne S. and Per A. Mykland（2008）。“金融市场跳跃：一种新的非参数检验与跳跃动力学”。载于：The Review of Financial Studies 21.6，页2535–2563。 DOI: 10.1093/rfs/hhm056（引于第2页）。Li, Zenghu and Chunhua Ma（2013）。“稳定Cox-Ingersoll-Ross模型中估计量的渐近性质”。载于：arXiv: 1301.3243 [math.PR]。 URL: https://arxiv.org/abs/1301.3243（引于第2页）。Merton, Robert C.（1976）。“标的股票收益不连续时的期权定价”。载于：Journal of Financial Economics 3.1–2，页125–144。 DOI: 10.1016/0304-405X(76)90022-2（引于第2页）。Mishura, Yuliya, Kostiantyn Ralchenko, and Olena Dehtiar（2022）。“基于连续观测的CKLS模型参数估计”。载于：Statistics & Probability Letters 184.C。 DOI: 10.1016/j.spl.2022.109391。URL: https://ideas.repec.org/a/eee/stapro/v184y2022ics0167715222000153.html（引于第4页）。Overbeck, Ludger（1998）。“连续分枝过程的估计”。载于：Scandinavian Journal of Statistics 25.1，页111–126。 ISSN: 03036898, 14679469。 URL: http://www.jstor.org/stable/4616488（访问日期：2024年10月11日）（引于第2页）。Uchida, Masayuki and Nakahiro Yoshida（2012）。“基于抽样数据的遍历扩散过程自适应估计”。载于：Stochastic Processes and their Applications 122，页2885–2924（引于第2页）。Yoshida, Nobuaki（1988）。“扩散过程的M估计量及相关随机场的渐近行为”。载于：Annals of the Institute of Statistical Mathematics 40.2，页271–296。 DOI: 10.1007 / BF00052393（引于第2页）。

A 附录

定理 *1* 的证明。 对 (6) 在 $(t_{i-1}, t_i]$ 上积分，得到

$$\Delta_i^n X = \int_{t_{i-1}}^{t_i} \kappa(\theta - X_s)\, ds + \int_{t_{i-1}}^{t_i} \sigma X_s^\gamma\, dW_s + \Delta_i^n J.$$

一阶 Itˆo–Taylor 展开给出

$$\Delta_i^n X = \kappa(\theta - X_{t_{i-1}})\Delta_n + \sigma X_{t_{i-1}}^\gamma \Delta_i^n W + \Delta_i^n J + O_p(\Delta_n).$$

减去 估计漂移分量并利用参数的一致性，得到 s

$$\Delta_i^n X - \hat{\kappa}_n(\hat{\theta}_n - X_{t_{i-1}})\Delta_n = \sigma X_{t_{i-1}}^\gamma \Delta_i^n W + \Delta_i^n J + o_p(\sqrt{\Delta_n}).$$

由于

$$\hat{\sigma}_n X_{t_{i-1}}^\gamma \sqrt{\Delta_n} = \sigma X_{t_{i-1}}^\gamma \sqrt{\Delta_n}(1 + o_p(1)),$$

两边相除可得（9），由此即可得出所述结论。 □

定理*2*的证明。令

$$M_n := \max_{i \in \mathcal{C}_n} |Z_{i,n}|.$$

由参数估计量的一致性,

$$Z_{i,n} = \frac{\Delta_i^n X - b(X_{t_{i-1}})\Delta_n}{\sigma(X_{t_{i-1}})\sqrt{\Delta_n}} + o_p(1) = \xi_i + o_p(1),$$

对 $i \in \mathcal{C}_n$ 一致成立，其中 $\{\xi_i\}$ 是相互独立的 $N(0,\ 1)$ 变量。由于 $|\mathcal{C}_n|/n \xrightarrow{P} 1$，因此

$$M_n = \max_{1 \le i \le n} |\xi_i| + o_p(1).$$

因此，对任意 $u > 0$,

$$\Pr(M_n \le u) = [2\Phi(u) - 1]^n + o(1).$$

定义标准化序列

$$a_n = \sqrt{2\log n} - \frac{\log\log n + \log \pi}{2\sqrt{2\log n}}, \qquad b_n = \frac{1}{\sqrt{2\log n}},$$

并令 $u_n(x) = a_n + b_n x$。利用高斯尾部展开

$$1 - \Phi(u) = \frac{\phi(u)}{u}(1 + o(1)), \qquad u \to \infty,$$

我们得到

$$n\,[1 - \Phi(u_n(x))] \to e^{-x}.$$

因此,

$$\Pr\left(\frac{M_n - a_n}{b_n} \le x\right) \to \exp(-e^{-x}),$$

这证明了其收敛至标准耿贝尔定律。

最后，由于 $(M_n - a_n)/b_n = O_p(1)$ 且 $a_n = \sqrt{2\log n} + o(1)$,

$$M_n = \sqrt{2\log n} + O_p(1).$$

□

定理3的证明。令

$$\mathcal{C}_n = \{i : \Delta_i^n J = 0\}, \qquad \mathcal{J}_n = \{i : \Delta_i^n J \neq 0\},$$

分别表示连续增量集和跳跃增量集，并定义阈值

$$\xi_n = a_n + b_n G^{-1}(1-q),$$

其中 a_n， b_n 由定理 2 给出，且 G^{-1} 为 Gumbel 分位数函数。

我们首先控制误检。根据定理2,

$$\frac{\max_{i\in\mathcal{C}_n} |Z_{i,n}| - a_n}{b_n} \Rightarrow \Lambda,$$

其中 Λ 服从标准 Gumbel 分布。因此,

$$\Pr\Big(\max_{i\in\mathcal{C}_n} |Z_{i,n}| > \xi_n\Big) = \Pr\big(\Lambda > G^{-1}(1-q)\big) + o(1) = q + o(1).$$

由于 $q > 0$ 可以任意小，误检渐近消失。

接下来，考虑备择假设下的检测。对于任意固定的 $i \in \mathcal{J}_n$，定理 1 给出

$$|Z_{i,n}| \xrightarrow{P} \infty.$$

由于 $\xi_n = O(\sqrt{\log n})$，可得

$$\Pr(|Z_{i,n}| \leq \xi_n) \to 0,$$

因此，每个跳跃被检测到的概率趋于 1。

最后，定义估计跳跃集

$$\widehat{\mathcal{J}}_n = \{i : |Z_{i,n}| > \xi_n\}.$$

由前述界可知

$$\Pr\Big(\widehat{\mathcal{J}}_n = \mathcal{J}_n\Big) \to 1,$$

这就确立了分类一致性。 □