

面板数据中方差的共同变点估计

赵军辉, 董翠玲*

(新疆师范大学 数学科学学院, 新疆 乌鲁木齐 830017)

摘要: 文章对面板数据中方差的共同变点提出了一个含有调节参数的CUSUM(Cumulative Sum)型估计量, 证明了变点估计量的相容性, 并结合二元分割法将其推广到多个方差共同变点的情形。蒙特卡洛模拟发现, 调节参数($\gamma \neq 0$)下CUSUM型估计量的精确度要高于无调节参数($\gamma = 0$)下CUSUM型估计量的精确度。应用外汇汇率进行实证分析, 结果也表明调节参数($\gamma \neq 0$)下CUSUM型估计方法是有效的。

关键词: 面板数据; 方差变点; CUSUM; 调节参数; 二元分割法

中图分类号: O212.1 **文献标识码:** A **文章编号:** 1008-9659(2024)01-0022-11

面板数据结合了时间序列与截面数据的特点, 是二维数据, 它扩大了样本的信息, 降低了变量之间的多重共线性, 提高了参数估计的准确性, 目前广泛应用于经济学、金融学、生命科学、医学、气象学等领域^[1-2]。面板数据的变点分析问题始于 Joseph 等人提出的随机变点模型^[3-4]。Bai 使用最小二乘法和拟极大似然方法估计了面板数据中均值与方差共同变点, 并得到了变点估计量的极限分布^[5]。自 Page 首次提出累积和(Cumulative Sum, CUSUM)方法对变点进行连续性检验后^[6], CUSUM 方法被许多统计学家改进并应用于变点的检测与估计。Horváth 等人在 Bai 的模型基础上, 关于面板数据中均值是否存在共同变点提出了一个基于平方累积和(Squared CUSUM)的检验统计量, 并在原假设 H_0 (即没有变点) 下得到了检验统计量的渐进分布^[7]。Li 等人和 Shi 分别使用 CUSUM 方法^[8]和似然方法^[9]对面板数据中方差是否存在共同变点进行了检验。徐小平等人使用拟极大似然方法和 CUSUM 方法对面板数据中方差的共同变点进行了估计, 并结合二元分割法将其推广到多变点情形^[10]。

这些关于面板数据变点分析的研究, 当观测时长 T 较长, 变点位置不在序列的端点附近时, 即 $k_0 \in (\rho T, (1 - \rho)T)$, (其中 $0 < \rho < \frac{1}{2}$), 估计都很有效, 但对于观测时长 T 较短 (即 $T < 30$), 或变点出现在序列端点附近时, 估计的精度大幅降低。Horváth 等人使用 CUSUM 方法对长相依序列中的变点进行估计时, 在数值模拟过程中发现调节参数对变点估计的精确度有显著影响^[11]。Chen 等人通过调节参数对面板数据中均值的共同变点提出了一个改进的 CUSUM 型估计量, 数值模拟给出了不同调节参数下变点估计的精确度^[12]。谭常春等人研究了 CUSUM 型统计量中调节参数对单变量序列中变点估计效果的影响^[13]。文章通过调节参数对面板数据中方差的共同变点提出了一个改进的 CUSUM 型估计量, 研究调节参数 $\gamma \in (0, 1)$ 对方差共同变点估计精确度的影响。蒙特卡洛模拟表明通过调节参数不仅使得变点位置在序列中间时得到很好的估计效果, 而且使得变点位置在序列端点附近时, 估计的精确度有了大幅度提升, 并结合二元分割法将其推广到多个方差共同变点的情形。最后, 应用 2018 年 1 月—2022 年 12 月外汇汇率进行实证分析, 结果表明调节

[收稿日期] 2023-07-21

[修回日期] 2023-09-08

[基金项目] 新疆维吾尔自治区自然科学基金项目(2023D01A37; 2022D01A219); 新疆师范大学重点实验室项目(XJNUSYS082018A01)。

[作者简介] 赵军辉(1995-), 男, 河南周口人, 硕士研究生, 主要从事面板数据变点的统计推断方面研究。

* [通讯作者] 董翠玲(1978-), 女, 新疆乌鲁木齐人, 副教授, 主要从事面板数据变点的统计推断、复杂数据分析方面研究。

参数($\gamma \neq 0$)下CUSUM型估计方法是有效的。

1 模型与主要结果

考虑面板数据中方差的共同变点模型

$$Y_{it} = \begin{cases} \mu_i + \sigma_{i1}\eta_{it}, & t = 1, 2, \dots, k_0 \\ \mu_i + \sigma_{i2}\eta_{it}, & t = k_0 + 1, \dots, T \end{cases} \quad i = 1, 2, \dots, N \quad (1)$$

其中, k_0 未知, Y_{it} ($i = 1, 2, \dots, N; t = 1, 2, \dots, T$) 是面板数据中第 i 个截面个体在 t 时刻的观测值, μ_i 是第 i 个截面个体的均值, η_{it} 是第 i 个截面个体在 t 时刻的误差项。在这个模型中, 若 $\sigma_{i1} \neq \sigma_{i2}$, 则未知时刻 k_0 ($1 \leq k_0 < T$) 称为面板数据中方差的共同变点, 即这 N 个截面个体方差共同变点。当 $k_0 = T$ 时, 表明面板数据中不存在方差共同变点。令

$$V_{NT}(k) = \left(\frac{k(T-k)}{T^2} \right)^{1-\gamma} \sum_{i=1}^N \left(\frac{1}{k} \sum_{t=1}^k (Y_{it} - \bar{Y}_i)^2 - \frac{1}{T-k} \sum_{t=k+1}^T (Y_{it} - \bar{Y}_i)^2 \right) \quad (2)$$

其中, γ 为调节参数, $\gamma \in (0, 1)$, 调节参数可以保证方差共同变点 k_0 在靠近序列端点时估计的有效性, $\gamma = 0$ 表示无调节参数。 $\bar{Y}_i = \frac{1}{T} \sum_{t=1}^T Y_{it}$ 表示第 i 个截面个体在 T 个不同时刻得到观测值的样本均值。记 $\tilde{Y}_{it} = Y_{it} - \bar{Y}_i$, 则 \tilde{Y}_{it} 为面板数据中心化后的结果, 从而

$$V_{NT}(k) = \left(\frac{k(T-k)}{T^2} \right)^{1-\gamma} \sum_{i=1}^N \left(\frac{1}{k} \sum_{t=1}^k \tilde{Y}_{it}^2 - \frac{1}{T-k} \sum_{t=k+1}^T \tilde{Y}_{it}^2 \right) \quad (3)$$

为了估计方差的共同变点, 需要下列假设条件:

假设 1: $E(\eta_{it}) = 0, \text{Var}(\eta_{it}) = 1$, 其中 $\eta_{it} = \sum_{j=0}^{\infty} b_{ij}\varepsilon_{i,t-j}, \varepsilon_{it} \sim i.i.d(0, 1), \sum_{j=1}^{\infty} j|b_{ij}| < \infty, \text{Var}(\eta_{it}) = \sum_{j=0}^{\infty} b_{ij}^2 = 1$;

假设 2: 存在正数 $M > 0$, 使得 $E(\varepsilon_{it}^4) \leq M$;

假设 3: 存在 $\tau_i \in (0, 1), \varepsilon > 0$, 使得 $k_i = [T\tau_i]$, 并且 $\tau_{i+1} - \tau_i > \varepsilon, i = 0, 1, 2, \dots, m$, 其中 $[\cdot]$ 为取整函数;

假设 4: $\lim_{N \rightarrow \infty} \frac{N}{\left| \sum_{i=1}^N \delta_i \right|^2} = 0$, 其中 $\delta_i = (\sigma_{ij} - \sigma_{i,j-1})^2 > 0$, 表示方差跳跃度的平方, 即方差的变化强度;

假设 5: 对任意的 $1 \leq k \leq s \leq T$, 都有 $\text{Var} \sum_{t=k}^s \tilde{Y}_{it}^2 \leq C(s-k+1)^h$, 其中 $h \in (0, 2)$, 注意在整篇文章中, 正数 C 可能会不同, 而且 C 与 N 和 T 是相互独立的;

假设 6: 存在 $\alpha \in (0, 1)$, 使得 $\lim_{N, T \rightarrow \infty} \frac{N}{T^\alpha} = 0$.

注: 假设 1~假设 3 为 Bai 的研究中关于面板数据的假设, 其中假设 1 能保证误差项 η_{it} 满足平稳性, 假设 2 要求误差项 η_{it} 的四阶矩有限, 假设 3 确保了模型的每两个变点之间有足够多的样本, 这是大数定律和中心极限定理成立的基本条件, 通常 ε 取 0.05, 0.01 等较小的数。假设 4 类似于 Bai 的假设 2, 这个假设既保证方差变化强度 δ_i 的非负性, 又合理描述截面个数 N 与方差变化强度 δ_i 之间的关系^[5]。假设 5 满足常见的平稳序列或者非平稳序列, 更多的案例参看文献[14]。假设 6 表示当 $N, T \rightarrow \infty$ 时, T^α 趋于无穷大的速度快于 N 。

设模型(1)存在一个方差的共同变点 $\tau_0 = \frac{k_0}{T}$, 则变点 τ_0 的CUSUM型估计量为

$$\hat{\tau} = \frac{\hat{k}}{T} = \frac{1}{T} \arg \max_{1 \leq k < T} |V_{NT}(k)| \quad (4)$$

为了得到变点估计量 $\hat{\tau}$ 的相合性, 需要以下引理。它是文献[14]的定理 3.1。

引理 1^[14] 设 Y_1, \dots, Y_n 是任意二阶矩有限的随机变量序列, C_1, C_2, \dots, C_n 为任意的非负常数, 则

$$\varepsilon^2 P \left\{ \max_k C_k \left| \sum_{j=1}^k Y_j \right| > \varepsilon \right\} \leq \sum_{k=1}^{n-1} |C_{k+1}^2 - C_k^2| \sum_{i,j=1}^k E(Y_i Y_j) + 2 \sum_{k=1}^{n-1} C_{k+1}^2 (E Y_{k+1}^2)^{\frac{1}{2}} \left(\sum_{i,j=1}^k E(Y_i Y_j) \right)^{\frac{1}{2}} + \sum_{k=0}^{n-1} C_{k+1}^2 E Y_{k+1}^2$$

定理 1 设面板数据模型(1)中存在一个方差共同变点 $\tau_0 = \frac{k_0}{T}$, 若假设 1~假设 5 都成立, 且 $\sum_{i=1}^N \delta_i \neq 0$, 则对于任意 $\varepsilon > 0$, 由式(3)定义的面板数据中方差的共同变点估计量 τ_0 满足

$$P(|\hat{\tau} - \tau_0| > \varepsilon) \leq \frac{C}{\varepsilon^2} \cdot \frac{N^3}{\left| \sum_{i=1}^N \delta_i \right|^2} \cdot T^{\frac{h}{2} - 2}$$

推论 1 设面板数据模型(1)中存在一个方差共同变点 $\tau_0 = \frac{k_0}{T}$, 若假设 1~假设 6 都成立, 且 $\sum_{i=1}^N \delta_i \neq 0$, 则对于任意 $\varepsilon > 0$, 由式(3)定义的面板数据中方差的共同变点估计量 τ_0 满足

$$\lim_{N, T \rightarrow \infty} P(|\hat{\tau} - \tau_0| > \varepsilon) = 0$$

2 结果的证明

定理 1 的证明 当面板数据模型(1)中存在方差共同变点 k_0 时, 可知

$$\tilde{Y}_{it} = Y_{it} - \bar{Y}_i = \frac{TY_{it}}{T} - \frac{(Y_{i1} + Y_{i2} + \cdots + Y_{ik_0}) + (Y_{ik_0+1} + \cdots + Y_{iT})}{T} = \begin{cases} \frac{T-k_0}{T}((\sigma_{i1} - \sigma_{i2})\eta_{it}), & t = 1, 2, \dots, k_0 \\ \frac{k_0}{T}((\sigma_{i2} - \sigma_{i1})\eta_{it}), & t = k_0 + 1, \dots, T \end{cases}$$

由假设 1, $E(\eta_{it}) = 0$, $Var(\eta_{it}) = 1$, 可得 $E(\eta_{it}^2) = 1$.

再由假设 4 可得

$$E(\tilde{Y}_{it})^2 = E(Y_{it} - \bar{Y}_i)^2 = \begin{cases} \left(\frac{T-k_0}{T}\right)^2 \delta_i, & t = 1, 2, \dots, k_0 \\ \left(\frac{k_0}{T}\right)^2 \delta_i, & t = k_0 + 1, \dots, T \end{cases}$$

于是

$$\begin{aligned} E(V_{NT}(k)) &= \begin{cases} \left(\frac{k(T-k)}{T^2}\right)^{1-\gamma} \sum_{i=1}^N E \left\{ \frac{1}{k} \sum_{t=1}^k \tilde{Y}_{it}^2 - \frac{1}{T-k} \left(\sum_{t=k_0+1}^{k_0} \tilde{Y}_{it}^2 + \sum_{t=k_0+1}^T \tilde{Y}_{it}^2 \right) \right\}, & k \leq k_0 \\ \left(\frac{k(T-k)}{T^2}\right)^{1-\gamma} \sum_{i=1}^N E \left\{ \frac{1}{k} \left(\sum_{t=1}^{k_0} \tilde{Y}_{it}^2 + \sum_{t=k_0+1}^k \tilde{Y}_{it}^2 \right) - \frac{1}{T-k} \sum_{t=k_0+1}^T \tilde{Y}_{it}^2 \right\}, & k > k_0 \end{cases} \\ &= \begin{cases} \left(\frac{k(T-k)}{T^2}\right)^{1-\gamma} \sum_{i=1}^N \delta_i \left\{ \left(1 - \frac{k_0}{T}\right)^2 - \left(\frac{k_0-k}{T-k} \left(1 - \frac{k_0}{T}\right)^2 + \frac{T-k_0}{T-k} \left(\frac{k_0}{T}\right)^2 \right) \right\}, & k \leq k_0 \\ \left(\frac{k(T-k)}{T^2}\right)^{1-\gamma} \sum_{i=1}^N \delta_i \left\{ \frac{k_0}{k} \left(\frac{T-k_0}{T}\right)^2 + \frac{k-k_0}{k} \left(\frac{k_0}{T}\right)^2 - \left(\frac{k_0}{T}\right)^2 \right\}, & k > k_0 \end{cases} \\ &= \begin{cases} \tau^{1-\gamma} (1-\tau)^{-\gamma} (1-\tau_0) (1-2\tau_0) \sum_{i=1}^N \delta_i, & k \leq k_0 \\ \tau^{-\gamma} (1-\tau)^{1-\gamma} \tau_0 (1-2\tau_0) \sum_{i=1}^N \delta_i, & k > k_0 \end{cases} \end{aligned}$$

注意到当 $k = k_0$ 时

$$E(V_{NT}(k_0)) = \tau_0^{1-\gamma} (1-\tau_0)^{1-\gamma} (1-2\tau_0) \sum_{i=1}^N \delta_i$$

所以

$$|E(V_{NT}(k_0))| - |E(V_{NT}(k))| = \begin{cases} |(1-2\tau_0)|(1-\tau_0)^{1-\gamma} \left| \sum_{i=1}^N \delta_i \right| (\tau_0^{1-\gamma} - \tau^{1-\gamma}(1-\tau)^{-\gamma}(1-\tau_0)^\gamma), & k \leq k_0 \\ |(1-2\tau_0)|\tau_0^{1-\gamma} \left| \sum_{i=1}^N \delta_i \right| ((1-\tau_0)^{1-\gamma} - \tau^{-\gamma}(1-\tau)^{1-\gamma}\tau_0^\gamma), & k > k_0 \end{cases}$$

$$\geq \begin{cases} |(1-2\tau_0)|(1-\tau_0)^{1-\gamma} \left| \sum_{i=1}^N \delta_i \right| (\tau_0^{1-\gamma} - \tau^{1-\gamma}), & k \leq k_0 \\ |(1-2\tau_0)|\tau_0^{1-\gamma} \left| \sum_{i=1}^N \delta_i \right| ((1-\tau_0)^{1-\gamma} - (1-\tau)^{1-\gamma}), & k > k_0 \end{cases}$$

若令 $f_1(\tau) = \tau^{1-\gamma}$, ($0 \leq \gamma < 1$), $f_2(\tau) = (1-\tau)^{1-\gamma}$, ($0 \leq \gamma < 1$), 由拉格朗日中值定理知

$$|E(V_{NT}(k_0))| - |E(V_{NT}(k))| \geq \begin{cases} |(1-2\tau_0)|(1-\tau_0)^{1-\gamma}(1-\gamma)\tau_0^{-\gamma}(\tau_0 - \tau) \left| \sum_{i=1}^N \delta_i \right|, & k \leq k_0 \\ |(1-2\tau_0)|\tau_0^{1-\gamma}(1-\gamma)(1-\tau_0)^{-\gamma}(\tau_0 - \tau) \left| \sum_{i=1}^N \delta_i \right|, & k > k_0 \end{cases}$$

记

$$\bar{\tau} = |(1-2\tau_0)|(1-\gamma)\tau_0^{-\gamma}(1-\tau_0)^{-\gamma} \min\{\tau_0, 1-\tau_0\}$$

则

$$|E(V_{NT}(k_0))| - |E(V_{NT}(k))| \geq \bar{\tau} |\tau - \tau_0| \left| \sum_{i=1}^N \delta_i \right|$$

由三角不等式

$$|E(V_{NT}(k_0))| - |E(V_{NT}(k))| \leq 2 \max_{1 \leq k < T} |V_{NT}(k) - E(V_{NT}(k))| + |V_{NT}(k_0)| - |V_{NT}(k)|$$

在上式中用 \hat{k} 代替 k , 由式(3)知 $|V_{NT}(k_0)| \leq |V_{NT}(\hat{k})|$, 从而

$$|E(V_{NT}(k_0))| - |E(V_{NT}(\hat{k}))| \leq 2 \max_{1 \leq k < T} |V_{NT}(k) - E(V_{NT}(k))|$$

即

$$\begin{aligned} \bar{\tau} |\hat{\tau} - \tau_0| \left| \sum_{i=1}^N \delta_i \right| &\leq 2 \max_{1 \leq k < T} |V_{NT}(k) - E(V_{NT}(k))| \\ &= 2 \max_{1 \leq k < T} \left| \sum_{i=1}^N \left\{ \left(\frac{k(T-k)}{T^2} \right)^{1-\gamma} \left(\frac{1}{k} \sum_{t=1}^k (\tilde{Y}_{it}^2 - E\tilde{Y}_{it}^2) - \frac{1}{T-k} \sum_{t=k+1}^T (\tilde{Y}_{it}^2 - E\tilde{Y}_{it}^2) \right) \right\} \right| \\ &\leq 2 \sum_{i=1}^N \max_{1 \leq k < T} k^{-1} \left| \sum_{t=1}^k (\tilde{Y}_{it}^2 - E\tilde{Y}_{it}^2) \right| + 2 \sum_{i=1}^N \max_{1 \leq k < T} (T-k)^{-1} \left| \sum_{t=k+1}^T (\tilde{Y}_{it}^2 - E\tilde{Y}_{it}^2) \right| \end{aligned}$$

于是

$$\begin{aligned} P\left(\bar{\tau} |\hat{\tau} - \tau_0| \left| \sum_{i=1}^N \delta_i \right| > \varepsilon\right) &\leq \sum_{i=1}^N P\left(2 \max_{1 \leq k < T} k^{-1} \left| \sum_{t=1}^k (\tilde{Y}_{it}^2 - E\tilde{Y}_{it}^2) \right| > \frac{\varepsilon}{2N}\right) + \sum_{i=1}^N P\left(2 \max_{1 \leq k < T} (T-k)^{-1} \left| \sum_{t=k+1}^T (\tilde{Y}_{it}^2 - E\tilde{Y}_{it}^2) \right| > \frac{\varepsilon}{2N}\right) \\ &= \sum_{i=1}^N P\left(\max_{1 \leq k < T} k^{-1} \left| \sum_{t=1}^k (\tilde{Y}_{it}^2 - E\tilde{Y}_{it}^2) \right| > \frac{\varepsilon}{4N}\right) + \sum_{i=1}^N P\left(\max_{1 \leq k < T} (T-k)^{-1} \left| \sum_{t=k+1}^T (\tilde{Y}_{it}^2 - E\tilde{Y}_{it}^2) \right| > \frac{\varepsilon}{4N}\right) \end{aligned} \quad (5)$$

对于式(5)中的第一项, 使用引理1, 令 $C_k = k^{-1}$, $Z_{it} = \tilde{Y}_{it}^2 - E\tilde{Y}_{it}^2$, 对于 $i = 1, 2, \dots, N$, 有

$$\begin{aligned} \left(\frac{\varepsilon}{4N}\right)^2 \cdot P\left(\max_{1 \leq k < T} k^{-1} \left| \sum_{t=1}^k (\tilde{Y}_{it}^2 - E\tilde{Y}_{it}^2) \right| > \frac{\varepsilon}{4N}\right) &= \left(\frac{\varepsilon}{4N}\right)^2 \cdot P\left(\max_{1 \leq k < T} k^{-1} \left| \sum_{t=1}^k Z_{it} \right| > \frac{\varepsilon}{4N}\right) \\ &\leq \sum_{k=1}^{T-1} |(k+1)^{-2} - k^{-2}| \sum_{m,l=1}^k E(Z_{im}Z_{il}) + 2 \sum_{k=1}^{T-1} \frac{(E(Z_{i,k+1}^2))^{1/2}}{(k+1)^2} \left(\sum_{m,l=1}^k E(Z_{im}Z_{il})\right)^{1/2} + \sum_{k=0}^{T-1} \frac{E(Z_{i,k+1}^2)}{(k+1)^2} \end{aligned}$$

$$= \sum_{k=1}^{T-1} |(k+1)^{-2} - k^{-2}| \sum_{m,l=1}^k \text{Cov}(\tilde{Y}_{im}^2, \tilde{Y}_{il}^2) + \sum_{k=1}^{T-1} \frac{(\text{Var}(\tilde{Y}_{i,k+1}^2))^{\frac{1}{2}}}{(k+1)^2} \left(\sum_{m,l=1}^k \text{Cov}(\tilde{Y}_{im}^2, \tilde{Y}_{il}^2) \right)^{\frac{1}{2}} + \sum_{k=0}^{T-1} \frac{\text{Var}(\tilde{Y}_{i,k+1}^2)}{(k+1)^2}$$

由假设5,对 $i = 1, 2, \dots, N$, 都有

$$\max_{1 \leq k < T} \text{Var} \tilde{Y}_{ik}^2 \leq C \quad (6)$$

$$\sum_{m,l=1}^k \text{Cov}(\tilde{Y}_{im}^2, \tilde{Y}_{il}^2) = \text{Var} \left(\sum_{m=1}^k \tilde{Y}_{im}^2 \right) \leq Ck^h \quad (7)$$

$$\sum_{m,l=1}^k \text{Cov}(\tilde{Y}_{i,T-m+1}^2, \tilde{Y}_{i,T-l+1}^2) = \text{Var} \left(\sum_{m=T-k+1}^T \tilde{Y}_{im}^2 \right) \leq Ck^h \quad (8)$$

这里 $h \in (0, 2)$, 令 $f(k) = k^{-2}$, 由拉格朗日中值定理知

$$k^{-2} - (k+1)^{-2} \leq \frac{2}{k^3} \quad (9)$$

综合式(6)~式(9), 可得

$$\left(\frac{\varepsilon}{4N} \right)^2 \cdot P \left(\max_{1 \leq k < T} k^{-1} \left| \sum_{i=1}^k (\tilde{Y}_{ii}^2 - E\tilde{Y}_{ii}^2) \right| > \frac{\varepsilon}{4N} \right) \leq \sum_{k=1}^{T-1} \frac{2}{k^3} Ck^h + 2 \sum_{k=1}^{T-1} (k+1)^{-2} Ck^{\frac{h}{2}} + \sum_{k=0}^{T-1} (k+1)^{-2} C \leq \sum_{k=1}^{T-1} Ck^{\frac{h}{2}-2}$$

即

$$P \left(\max_{1 \leq k < T} k^{-1} \left| \sum_{i=1}^k (\tilde{Y}_{ii}^2 - E\tilde{Y}_{ii}^2) \right| > \frac{\varepsilon}{4N} \right) \leq \frac{1}{\left(\frac{\varepsilon}{4N} \right)^2} \sum_{k=1}^{T-1} Ck^{\frac{h}{2}-2}$$

同理对于式(5)中的第二项, 使用引理1, 令 $C_k = (T-k)^{-1}$, $Z_{ii} = \tilde{Y}_{ii}^2 - E\tilde{Y}_{ii}^2$, 对于 $i = 1, 2, \dots, N$, 有

$$P \left(\max_{1 \leq k < T} (T-k)^{-1} \left| \sum_{i=k+1}^T (\tilde{Y}_{ii}^2 - E\tilde{Y}_{ii}^2) \right| > \frac{\varepsilon}{4N} \right) \leq \frac{1}{\left(\frac{\varepsilon}{4N} \right)^2} \sum_{k=1}^{T-1} Ck^{\frac{h}{2}-2}$$

因此

$$\begin{aligned} P \left(\bar{\tau} \left| \hat{\tau} - \tau_0 \right| \left| \sum_{i=1}^N \delta_i \right| > \varepsilon \right) &\leq P \left(2 \max_{1 \leq k < T} |V_{NT}(k) - E(V_{NT}(k))| > \varepsilon \right) \\ &\leq \sum_{i=1}^N \frac{1}{\left(\frac{\varepsilon}{4N} \right)^2} \sum_{k=1}^{T-1} Ck^{\frac{h}{2}-2} + \sum_{i=1}^N \frac{1}{\left(\frac{\varepsilon}{4N} \right)^2} \sum_{k=1}^{T-1} Ck^{\frac{h}{2}-2} \leq \frac{CN^3}{\varepsilon^2} \sum_{k=1}^{T-1} k^{\frac{h}{2}-2} \end{aligned}$$

由于 $0 < h < 2$, $\frac{h}{2} - 2 \in (-2, -1)$, 所以

$$\sum_{k=1}^{T-1} k^{\frac{h}{2}-2} \leq C \cdot T^{\frac{h}{2}-2}$$

由于假设4表明 $\sum_{i=1}^N \delta_i \neq 0$ 以及 $\bar{\tau} \neq 0$, 所以

$$\left| \hat{\tau} - \tau_0 \right| \leq \frac{2 \max_{1 \leq k < T} |V_{NT}(k) - E(V_{NT}(k))|}{\bar{\tau} \left| \sum_{i=1}^N \delta_i \right|}$$

因此

$$P \left(\left| \hat{\tau} - \tau_0 \right| > \varepsilon \right) \leq P \left(\frac{2 \max_{1 \leq k < T} |V_{NT}(k) - E(V_{NT}(k))|}{\bar{\tau} \left| \sum_{i=1}^N \delta_i \right|} > \varepsilon \right) \leq \frac{CN^3}{\varepsilon^2 \left| \sum_{i=1}^N \delta_i \right|^2} \sum_{k=1}^{T-1} k^{\frac{h}{2}-2} \leq \frac{C}{\varepsilon^2} \cdot \frac{N^3}{\left| \sum_{i=1}^N \delta_i \right|^2} \cdot T^{\frac{h}{2}-2}$$

从而定理1得证。

推论 1 的证明 当 $h \in (0,2)$, $\frac{h}{2} - 2 \in (-2, -1)$ 时

$$T^{\frac{h}{2}-2} = \frac{1}{T^{2-\frac{h}{2}}} \triangleq \frac{1}{T^\theta}$$

令 $\theta = \theta_1 + \theta_2 \in (1,2)$, 这里 $\theta_i \in (0,1)$, $i = 1,2$. 由假设 6 可知

$$\frac{N^2}{T^{2-\frac{h}{2}}} = \frac{N^2}{T^\theta} = \frac{N}{T^{\theta_1}} \cdot \frac{N}{T^{\theta_2}} \rightarrow 0$$

结合假设 4

$$P(|\hat{\tau} - \tau_0| > \varepsilon) \leq P\left(\frac{2 \max_{1 \leq k < T} |V_{NT}(k) - E(V_{NT}(k))|}{\bar{\tau} \left| \sum_{i=1}^N \delta_i \right|} > \varepsilon\right) \leq \frac{C}{\varepsilon^2} \cdot \frac{N}{\left| \sum_{i=1}^N \delta_i \right|^2} \cdot \frac{N^2}{T^{2-\frac{h}{2}}} = 0$$

即 $\lim_{N,T \rightarrow \infty} P(|\hat{\tau} - \tau_0| > \varepsilon) = 0$, 推论 1 得证。

3 面板数据中方差多变点的估计步骤

若模型(1)中存在 m 个变点, 且变点个数 m 已知, 则模型(1)转化为

$$Y_{it} = \begin{cases} \mu_i + \sigma_{i1} \eta_{it}, & t = 1, 2, \dots, k_1 \\ \mu_i + \sigma_{i2} \eta_{it}, & t = k_1 + 1, \dots, k_2 \\ \vdots & \\ \mu_i + \sigma_{im} \eta_{it}, & t = k_{m-1} + 1, \dots, k_m \\ \mu_i + \sigma_{i,m+1} \eta_{it}, & t = k_m + 1, \dots, T \end{cases} \quad i = 1, 2, \dots, N \quad (10)$$

再结合二元分割法将上述方法推广到多变点的情形, 则模型(10)中的方差多变点估计具体步骤如下:

第一步: 利用式(4)估计出第一个变点 \hat{k}_1 ;

第二步: 在 \hat{k}_1 处将整个面板数据一分为二, 得到两个子样本, 第一部分为 $Y_{i1}, Y_{i2}, \dots, Y_{i\hat{k}_1}$, 第二部分为 $Y_{i,\hat{k}_1+1}, Y_{i,\hat{k}_1+2}, \dots, Y_{iT}$, $i = 1, 2, \dots, N$, 再利用式(4)分别估计这两个部分的变点 k'_2, k''_2 ;

第三步: 在前一部分面板数据中计算出 $V_{NT}(k'_2)$;

$$V_{NT}(k'_2) = \max_{1 < k < \hat{k}_1} \left| \left(\frac{k(T-k)}{T^2} \right)^{1-\gamma} \sum_{i=1}^N \left(\frac{1}{k} \sum_{t=1}^k \tilde{Y}_{it}^2 - \frac{1}{\hat{k}_1 - k} \sum_{t=k+1}^{\hat{k}_1} \tilde{Y}_{it}^2 \right) \right|$$

第四步: 在后一部分面板数据中计算出 $V_{NT}(k''_2)$;

$$V_{NT}(k''_2) = \max_{\hat{k}_1 < k < T} \left| \left(\frac{k(T-k)}{T^2} \right)^{1-\gamma} \sum_{i=1}^N \left(\frac{1}{k - \hat{k}_1} \sum_{t=\hat{k}_1+1}^k \tilde{Y}_{it}^2 - \frac{1}{T-k} \sum_{t=k+1}^T \tilde{Y}_{it}^2 \right) \right|$$

第五步: 比较 $V_{NT}(k'_2)$ 与 $V_{NT}(k''_2)$ 的大小, 若 $V_{NT}(k'_2) > V_{NT}(k''_2)$ 则 $\hat{k}_2 = k'_2$, 否则 $\hat{k}_2 = k''_2$;

第六步: 将 \hat{k}_1 与 \hat{k}_2 进行排序, 然后基于这两个变点将整个面板分成三个部分, 类似第二、三、四步估计出第三个变点 \hat{k}_3 , 重复使用上述方法, 直至估计出 m 个变点。

4 数值模拟

应用 MATLAB 软件, 通过蒙特卡洛模拟研究调节参数 γ 的取值对面板数据中方差的共同变点估计精确度的影响。对于模型(1), 简单起见, 只考虑一个变点的情形, 令 $u_i = 1$, $\eta_{it} \sim N(0,1)$, $\sigma_{i1} = 0.1$, $\sigma_{i2} = 0.2$, 这里方差跳跃度并不大。

首先, 研究观察时长较短时的情况, 取 $T = 10$, 变点位置在端点附近及中间位置时, 不同的截面个体数

量($N = 10, 20, 50, 80, 100, 150, 200, 250, 300$)情况下,调节参数 γ 的取值对面板数据中方差的共同变点估计精确度的影响, MATLAB模拟10000次。图1和图2分别展示了变点位置为 $k_0 = 2$ 和 $k_0 = 9$ 时,调节参数 γ 的取值对面板数据中方差的共同变点估计精确度的影响。 $\gamma = 0$ 表示无调节参数,这里“精确度”指的是数值模拟中变点估计量包含真实变点的频率,模拟结果表明调节参数($\gamma \neq 0$)下CUSUM型估计量的精确度高于无调节参数($\gamma = 0$)下CUSUM型估计量的精确度,并且随着截面个体数量 N 的增加,精确度会大幅度上升。图3展示了变点位置在中间时($k_0 = 5$),调节参数 γ 的取值对面板数据中方差的共同变点估计精确度的影响,模拟结果表明调节参数($\gamma \neq 0$)下CUSUM型估计量的精确度与无调节参数($\gamma = 0$)下CUSUM型估计量的精确度几乎相当,并且随着截面个体数量 N 的增加,精确度会大幅度上升,当 $N = 150$ 时,调节参数($\gamma \neq 0$)下CUSUM型估计量的精确度与无调节参数($\gamma = 0$)下CUSUM型估计量的精确度几乎都达到100%。

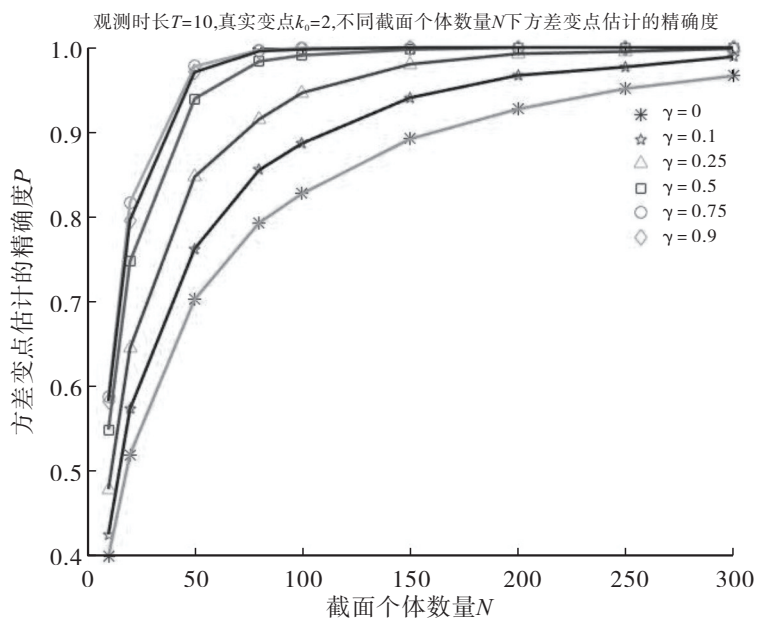


图1 不同的调节参数 γ 下CUSUM型估计量的精确度($T = 10$,左端变点 $k_0 = 2$)

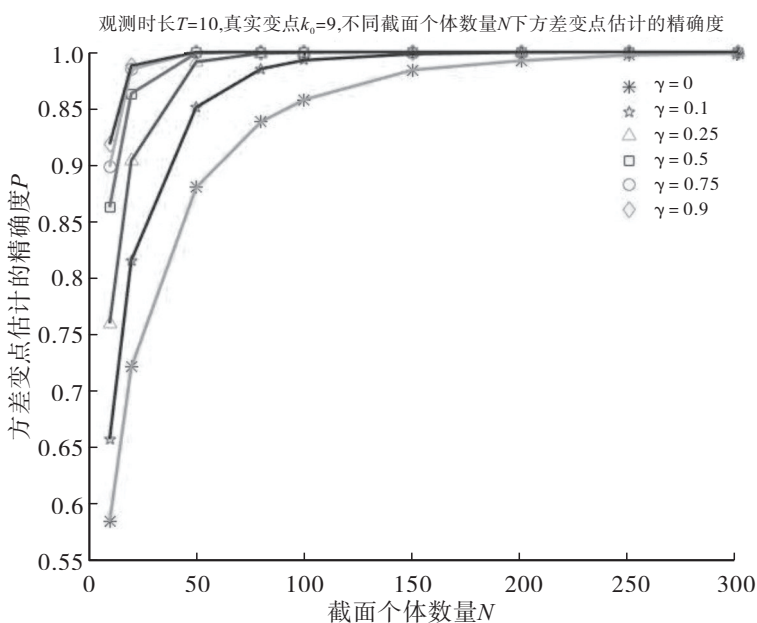
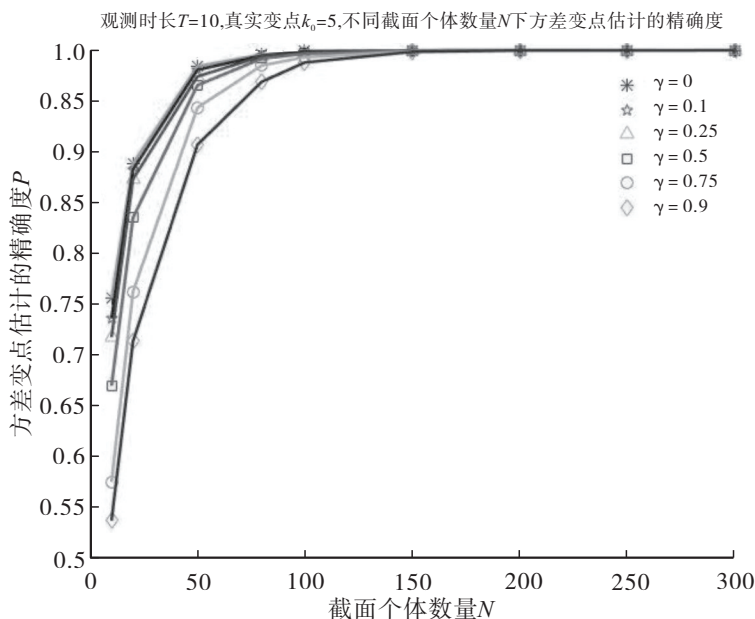
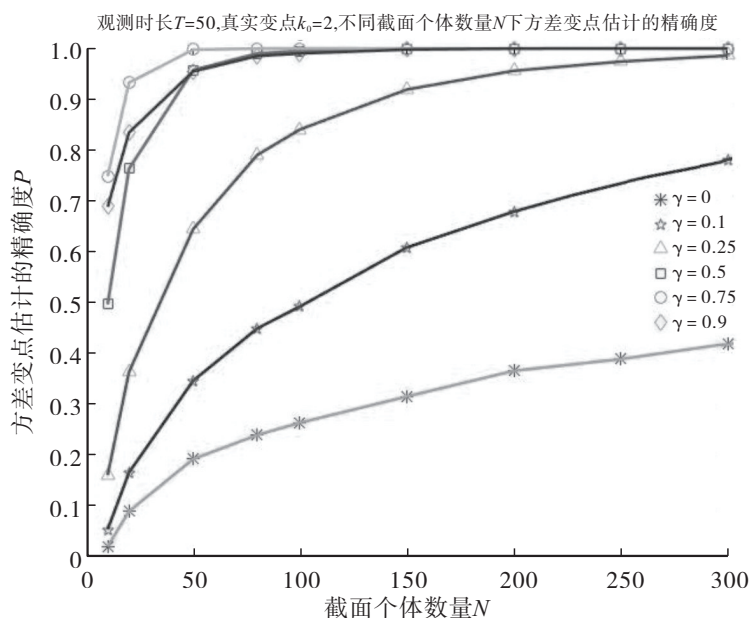


图2 不同的调节参数 γ 下CUSUM型估计量的精确度($T = 10$,右端变点 $k_0 = 9$)

图3 不同的调节参数 γ 下CUSUM型估计量的精确度($T = 10$,中间变点 $k_0 = 5$)

其次,研究观察时长较长时的情况,取 $T = 50$,变点位置在端点附近及中间位置时,不同的截面个体数量($N = 10, 20, 50, 80, 100, 150, 200, 250, 300$)情况下,调节参数 γ 的取值对面板数据中方差的共同变点估计精确度的影响, MATLAB模拟10000次。图4和图5分别展示了变点位置为 $k_0 = 2$ 和 $k_0 = T - 1$ 时,调节参数 γ 的取值对面板数据中方差的共同变点估计精确度的影响,模拟结果表明调节参数($\gamma \neq 0$)下CUSUM型估计量的精确度明显高于无调节参数($\gamma = 0$)下CUSUM型估计量的精确度。图6展示了变点位置在中间时($k_0 = T/2$),调节参数 γ 的取值对面板数据中方差的共同变点估计精确度的影响,模拟结果表明调节参数($\gamma \neq 0$)下CUSUM型估计量的精确度与无调节参数($\gamma = 0$)下CUSUM型估计量的精确度几乎相当,并且随着截面个体数量 N 的增加,精确度会大幅度上升,当 $N = 100$ 时,调节参数($\gamma \neq 0$)下CUSUM型估计量的精确度与无调节参数($\gamma = 0$)下CUSUM型估计量的精确度几乎都达到100%。

综合分析,无论变点位置在中间还是端点附近时,调节参数($\gamma \neq 0$)下CUSUM型估计量都会有非常好的表现,这与前面的理论结果也相吻合。

图4 不同的调节参数 γ 下CUSUM型估计量的精确度($T = 50$,左端变点 $k_0 = 2$)

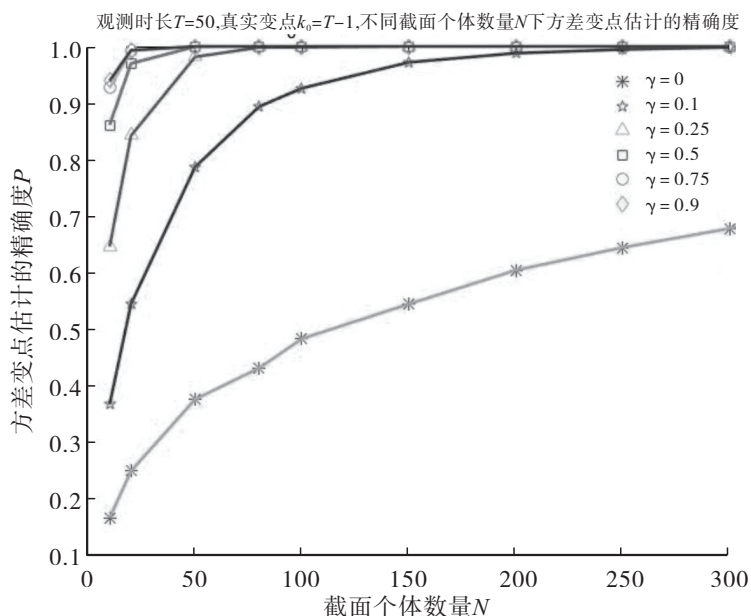


图5 不同的调节参数 γ 下CUSUM型估计量的精确度($T = 50$, 右端变点 $k_0 = T - 1$)

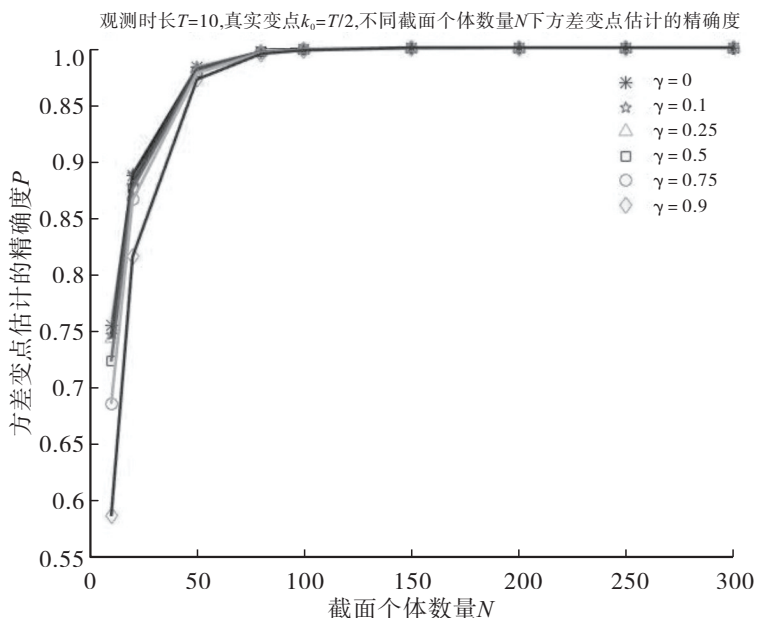


图6 不同的调节参数 γ 下CUSUM型估计量的精确度($T = 50$, 中间变点 $k_0 = T/2$)

5 实证分析

文章选取2018年1月—2022年12月10个国家的货币(澳大利亚元(AUD)、加拿大元(CAD)、瑞士法郎(CHF)、欧元(EUR)、英镑(GBP)、美元(USD)、新西兰元(NZD)、新加坡元(SGD)、巴西雷亚尔(BRL)、波兰兹罗提(PLN))兑换人民币(CNY)的外汇月度汇率数据(数据来源于<https://cn.investing.com/currencies/>),共有10个不同的截面个体,每个截面个体含有60个历史数据,即 $N = 10, T = 60$,首先对选取的数据进行去均值化处理,然后采用不同调节参数($\gamma = 0.1, 0.25, 0.5, 0.75, 0.9$)情况下CUSUM型估计量的方法估计变点,估计的变点位置都是46,对应的实际时间是2021年10月。造成这种现象的原因主要是2021年8月国际货币基金组织(IMF)批准了史上规模最大的一轮新增特别提款权(SDR)分配计划。结合二元分割法,该变点将2018年1月—2022年12月外汇汇率数据一分为二,采用上述方法分别对2018年1月—2021年10月和2021年11月

—2022年12月数据进行变点估计。得到的变点位置为7(前一部分)和10(后一部分),对应的实际时间分别是2018年7月和2022年8月,前一部分变点出现主要与2018年6月美联储的加息政策以及央行的降准政策有关。后一部分变点出现主要与2022年8月美联储的加息政策以及国内经济复苏缓慢有关。

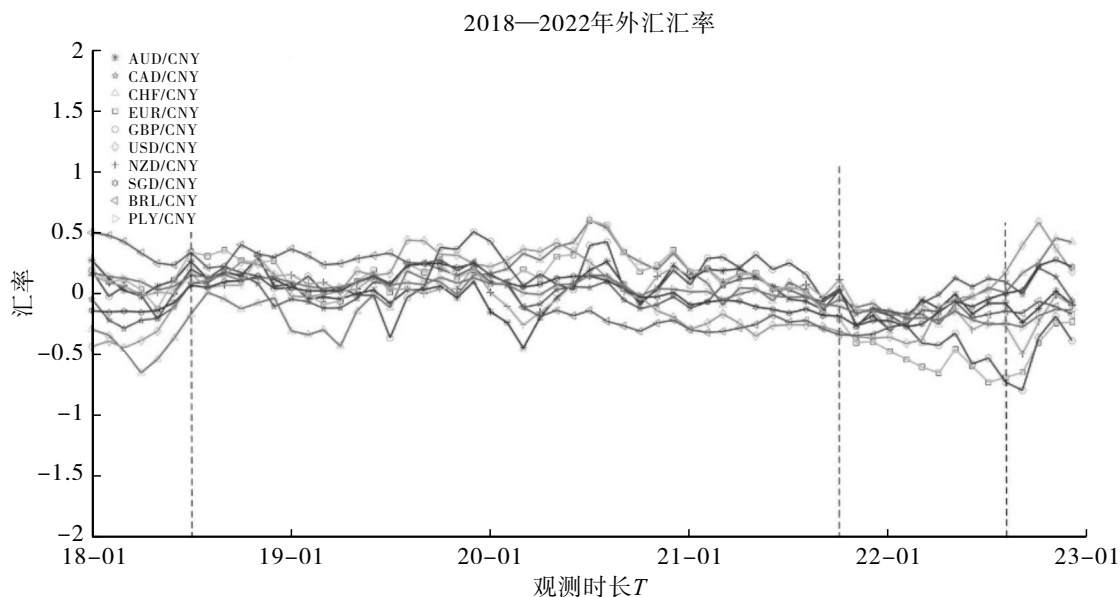


图7 10个国家的货币兑换人民币的外汇月度汇率数据图

6 结论

文章通过调节参数对面板数据中方差的共同变点提出了一个改进的CUSUM型估计量,研究调节参数对面板数据中方差的共同变点估计效果的影响,并结合二元分割法将其推广到多个方差共同变点的情形。模拟结果表明调节参数对面板数据中方差的共同变点估计有显著影响,同时发现此方法不仅适合样本小的面板数据,也适合样本大的面板数据,丰富了面板数据中方差的共同变点估计的研究方法。

参考文献:

- [1] 白仲林. 面板数据的计量经济分析[M]. 天津:南开大学出版社,2008.
- [2] HSIAO C. Analysis of Panel Data[M]. Beijing:Peking University Press,2003.
- [3] JOSEPH L, WOLFSON B D. Estimation in Multi-path Change-point Problems[J]. Communications in Statistics—Theory and Methods,1992,21(04):897-913.
- [4] JOSEPH L, WOLFSON B D. Maximum Likelihood Estimation in the Multi-path Change-point Problem[J]. Annals of the Institute of Statistical Mathematics,1993,45(03):511-530.
- [5] BAI J. Common Breaks in Means and Variances for Panel Data[J]. Journal of Econometrics,2010,157(01):78-92.
- [6] PAGE E S. Continuous Inspection Schemes[J]. Biometrika,1954,41(01-02):100-115.
- [7] HORVÁTH L, HUŠKOVÁ M. Change-point Detection in Panel Data[J]. Journal of Time Series Analysis,2012,33(04):631-648.
- [8] LI F, TIAN Z, XIAO Y, et al. Variance Change-point Detection in Panel Data Models[J]. Economics Letters,2015,126(02):140-143.
- [9] SHI Y. Testing Change in Volatility Using Panel Data[J]. Economics Letters,2015,134(09):107-110.
- [10] 徐小平,刘君,李拂晓. 面板数据中方差多变点的估计[J]. 统计与决策,2021,37(12):10-14.
- [11] HORVÁTH L, KOKOSZKA P. The Effect of Long-range Dependence on Change-point Estimators[J]. Journal of Statistical Planning and Inference,1997,64(01):57-81.
- [12] CHEN Z, HU Y. Cumulative Sum Estimator for Change-point in Panel Data[J]. Statistical Paers,2017,58(03):707-728.

- [13] 谭常春,江敏. CUSUM型统计量中调节参数对变点估计效果的影响分析[J]. 中国科学技术大学学报,2020,50(07): 920-928.
- [14] KOKOSZKA P, LEIPUS R. Change-point in the Mean of Dependent Observations [J]. Statistics and Probability Letters, 1998, 40(04): 385-393.
- [15] 张阳,曾福庚. 数据压缩与人工智能的数学研究[J]. 新疆师范大学学报(自然科学版),2022,41(01):1-7.

Common Change Point Estimation of Variance in Panel Data

ZHAO Jun-hui, DONG Cui-ling*

(School of Mathematical Sciences, Xinjiang Normal University, Urumqi, Xinjiang, 830017, China)

Abstract: This paper proposes a CUSUM (Cumulative Sum) estimator with a turning parameter for common variance change point in panel data, proves the consistency of the estimator, and extends it to the case of multiple common variance change points combining with binary segmentation. Monte Carlo simulation shows that the estimation accuracy of CUSUM estimator with turning parameter ($\gamma \neq 0$) is higher than that of CUSUM estimator without turning parameter ($\gamma = 0$). The application analysis of foreign exchange rate also shows that the CUSUM estimation method under turning parameter ($\gamma \neq 0$) is effective.

Keywords: Panel data; Variance change point; CUSUM; Turning parameters; Binary segmentation