



复合分位数下门限自回归模型的变点估计

张立文^{1*}, 程东坡², 薛文骏³, 杨廷干⁴

1. 上海财经大学统计与管理学院, 上海 200433;

2. 上海大学经济学院, 上海 200444;

3. 上海大学悉尼工商学院, 上海 200444;

4. 上海立信会计金融学院统计与数学学院, 上海 201209

E-mail: zhang.liwen@shufe.edu.cn, calm23@shu.edu.cn, wen.junx@ymail.com, yangtg@sfu.edu.cn

收稿日期: 2019-06-27; 接受日期: 2020-03-12; 网络出版日期: 2020-10-21; * 通信作者

国家自然科学基金 (批准号: 11601313) 和全国统计科学研究 (批准号: 2017LY32) 资助项目

摘要 金融领域中的突发事件是变点问题的一种体现, 往往由于其随机性和发生前信息量不足等因素造成突发事件难以识别和预测. 金融市场常常表现出非线性和异质性等特征, 门限分位数自回归模型作为金融领域变点问题研究的重要模型, 逐渐在经济和统计学界获得更多的关注. 本文结合不同的分位数对门限分位数自回归模型中的变点估计问题提出两种新的估计方法: 门限自回归分位数复合估计和分位数平均估计. 在一些实际数据分析中, 研究发现在不同分位数水平下门限分位数自回归模型中的变点非常接近. 基于变点的这种共同性, 首先通过最小化不同分位数下的联合损失函数得到更有效的复合变点估计, 并进一步推导复合变点估计量的渐近性质, 以及基于似然比和自助法构建所提出估计量的置信区间. 其次, 通过对不同分位数下的变点估计量求平均提出另一种复合分位数估计方法, 即门限分位数平均估计, 并给出相应的大样本性质. 数值模拟研究发现, 相比传统的门限最小二乘估计和分位数估计, 所提出的两种方法在有限样本条件下更加有效. 最后, 分析 2005–2014 年上证 A 股指数展示所提出方法的实际应用表现.

关键词 有效 变点 门限分位数自回归 复合分位数 分位数平均估计

MSC (2020) 主题分类 62F10, 62J99

1 引言

变点问题一直以来是统计学的热门课题之一, 其不仅在早期的工业自动控制中有大量应用, 而且在经济、金融、医学和计算机等诸多领域有着广泛的应用. 近年来, 很多学者对其研究过, 可参见文献 [1–3]. 门限模型作为变点模型因协变量引起结构变化的一种特殊情况, 自文献 [4] 提出以来, 越来越引起经济和统计学家的兴趣和重视. 早期大量文献研究集中于对条件均值建模. 例如, 文献 [5] 提出门限自回归 (threshold autoregressive, TAR) 模型, 其经过一系列的完善, 已经在参数估计和假设检验等

英文引用格式: Zhang L W, Cheng D P, Xue W J, et al. Combined change point estimation in threshold quantile autoregressive models (in Chinese). *Sci Sin Math*, 2022, 52: 63–84, doi: 10.1360/N012019-00167

方面发展出一套成熟的建模方法. 该模型广泛应用于各个领域, 包括经济学^[6]、金融学^[7]、水文学^[8]、精算学^[9]及其他领域. 现有文献大多侧重于均值回归模型, 它只能揭示响应变量条件均值的变动规律, 无法反映系统的异质效应. 为了更加全面描述解释变量对响应变量在不同分位点的异质影响, 本文考虑门限分位数自回归中的变点问题.

文献 [10] 中分位数回归 (quantile regression, QR) 的提出, 为进一步揭示系统规律提供了有力工具. 相对于最小二乘回归, 分位数回归不仅可以更加全面地描述被解释变量条件分布的全貌, 而且不需要对误差项要求很强的假设. 此外, 分位数回归估计的结果对离群值表现更加稳健. 近年来, 许多学者对分位数回归模型中的变点检测和估计进行了研究^[11-13]. 但对门限分位数自回归模型中的变点问题研究不多. 例如, 文献 [14, 15] 对自激励门限分位数自回归模型提出估计参数的方法, 并给出对应的渐近性质. 文献 [16] 结合分位数自回归和自激励门限自回归模型的思想, 提出了门限分位数自回归 (threshold quantile autoregressive, TQAR) 模型, 推导了相关回归系数的渐近性质并提出了变点存在性的检验过程. 文献 [17] 也在文献 [16] 的基础上考虑了异方差下门限分位数自回归模型的参数估计及其渐近性质, 并基于 Bayes Markov 链 Monte Carlo 模拟方法去估计参数.

在金融领域中, 人们特别关心的突发事件如金融风暴就是变点的一种, 特别是在金融全球化、信息高速的时代, 金融风险防范对从国家到企业的每一个单位都显得极其重要. 这对变点的估计和推断的精度提出了更高的要求. 在一些实际数据分析中, 我们发现在不同分位数水平下模型具有共同的变点. 因此, 我们可以结合多个分位数水平的信息提高变点的估计精度. 现有文献复合不同分位数信息提高估计效率有两种可行方法, 第一种是由文献 [18] 首次提出的复合分位数回归 (composite quantile regression, CQR) 估计量, 其主要通过最小化不同分位数水平下的联合损失函数得到; 第二种是由文献 [10] 首次提出的分位数平均估计量 (quantile average estimator, QAE), 其主要通过对不同分位数水平下回归模型的系数求平均得到. 近几年复合分位数回归研究多集中于前者, 具体可参见文献 [18-21]. 文献 [22] 结合多个分位点水平上的信息, 研究了线性和非线性分位数回归模型中的 CQR 估计和 QAE, 并给出了相应的渐近性质. 为了提高分位数回归尾部参数的估计效率, 文献 [23] 在此基础上利用极值理论考虑了尾部分位数回归模型中的 CQR 估计和 QAE. 但据我们所知, 复合分位数回归和变点问题的研究文献非常有限. 例如, 文献 [24] 首次基于复合分位数的思想研究了回归函数是连续时折线分位数回归中的变点问题, 并通过模拟和实证得到所提出估计的效率. 文献 [25] 结合在不同分位数水平下的信息, 研究了分位数回归中存在同一门限的检验和估计问题. 然而这些文献均是在独立数据条件下进行变点研究, 本文关心的是回归函数带有跳跃的时间序列模型, 即门限分位数自回归模型中复合变点的估计.

本文对门限分位数自回归模型中的变点估计问题提出两种新的估计方法: 门限自回归分位数复合估计 (composite threshold quantile autoregression estimator, CTQAE) 和门限分位数平均估计 (threshold quantile average estimator, TQAE). 这两种方法在不同分位数水平下门限相同的假设下, 结合多分位数水平信息提高门限参数估计精度和效率. 在一些实际数据分析中, 我们发现在不同分位数水平下门限分位数自回归模型中的变点非常接近. 基于变点的这种共同性, 首先通过最小化不同分位数下的联合损失函数得到更有效的复合变点估计, 并进一步推导了复合变点估计量的渐近性质, 以及基于似然比和自助法构建了所提出估计的置信区间. 此外, 我们通过对不同分位数下的回归模型系数求平均提出另一种复合分位数估计方法, 即门限分位数平均估计 (TQAE), 并给出了相应的大样本性质. 数值模拟研究发现, 相比传统的门限自回归估计和门限分位数自回归估计, 所提出的 CTQAE 和 TQAE 在有限样本条件下具有更高的估计效率. 实证研究中, 我们将本文所提的方法应用于分析 2005 年 1 月 4 日到 2014 年 12 月 31 日的上证综合指数. 研究结果发现不同分位数区间上门限自回归模型中的变点

具有共同性, 提出的方法相较于传统方法有更高的估计精度.

本文余下的内容安排如下: 第 2 节给出门限自回归分位数复合估计的方法和相应变点估计量的渐近性质, 并基于似然比法和自助法分别构建复合变点估计量的置信区间. 第 3 节给出门限自回归分位数平均估计方法及相应估计量的大样本性质. 第 4 节通过数值模拟研究验证所提出方法的有限样本表现. 第 5 节将该方法用于上证 A 股指数验证所提出方法的实用价值. 第 6 节给出一个简单的总结. 所有的证明见附录.

2 复合分位数回归下的门限估计和推断

2.1 模型介绍

假定 y_t 是本文感兴趣的响应变量, $\mathbf{x}_t = (1, y_{t-1}, \dots, y_{t-p})^\top$ 为 $p+1$ 维解释变量, u_t 为一维门限变量. 对于任意给定的分位数水平 $\tau \in (0, 1)$, 定义已知 $\mathbf{x}$ 和 u 条件下 y 的 τ 条件分位数为

$$Q_y(\tau | \mathbf{x}, u) = F^{-1}(\tau | \mathbf{x}, u) = \inf\{t : F(t | \mathbf{x}, u) \geq \tau\},$$

其中 $F(\cdot | \mathbf{x}, u)$ 是给定 $\mathbf{x}$ 和 u 条件下 y 的条件分布函数. 对于任意给定分位数水平区间

$$\tau \in \Gamma = [\omega_1, \omega_2],$$

其中 $0 < \omega_1 < \omega_2 < 1$, 考虑下面的门限分位数自回归模型:

$$Q_{y_t}(\tau | \mathcal{F}_{t-1}) = \begin{cases} \beta_{10,\tau}^0 + \beta_{11,\tau}^0 y_{t-1} + \dots + \beta_{1p,\tau}^0 y_{t-p}, & u_t \leq u^0, \\ \beta_{20,\tau}^0 + \beta_{21,\tau}^0 y_{t-1} + \dots + \beta_{2p,\tau}^0 y_{t-p}, & u_t > u^0, \end{cases} \quad (2.1)$$

其中 $\mathcal{F}_t$ 是 $\{y_s, s \leq t-1\}$ 上产生的 σ 域, 本文中的门限变量 u_t 选择 y_t 的滞后变量 y_{t-d} , d 为延迟参数. 定义参数 $\boldsymbol{\eta}_{1,\tau}^0 = (\beta_{10,\tau}^0, \beta_{11,\tau}^0, \dots, \beta_{1p,\tau}^0)^\top$ 和 $\boldsymbol{\eta}_{2,\tau}^0 = (\beta_{20,\tau}^0, \beta_{21,\tau}^0, \dots, \beta_{2p,\tau}^0)^\top$ 分别是在 τ 分位数水平下门限变量 u_t 前后的回归系数向量, 参数 u^0 表示门限值. 此外, 定义参数 $\boldsymbol{\eta}_\tau^0 = (\boldsymbol{\eta}_{1,\tau}^{0\top}, \boldsymbol{\eta}_{2,\tau}^{0\top})^\top$, $\boldsymbol{\theta}^0 = (\boldsymbol{\eta}^{0\top}, u^0)^\top$, 其中上标 0 表示参数的真值. 此外, 可从模型 (2.1) 发现回归函数在不同分位数水平下协变量的效应是变化的, 但在所有考虑的分位数下具有相同的门限值.

令 $\tau_1 \leq \tau_2 \leq \dots \leq \tau_K$ 为感兴趣的分位数水平的集合. 再令参数 $\boldsymbol{\eta}_{1,\tau} = (\beta_{10,\tau}, \beta_{11,\tau}, \dots, \beta_{1p,\tau})^\top$, $\boldsymbol{\eta}_{2,\tau} = (\beta_{20,\tau}, \beta_{21,\tau}, \dots, \beta_{2p,\tau})^\top$, $\boldsymbol{\eta}_\tau = (\boldsymbol{\eta}_{1,\tau}^\top, \boldsymbol{\eta}_{2,\tau}^\top)^\top$, $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top, \dots, \boldsymbol{\eta}_K^\top)^\top$ 和 $\boldsymbol{\theta} = (\boldsymbol{\eta}^\top, u)^\top$. 为方便表述, 模型 (2.1) 可以被重写为

$$Q_{y_t}(\tau_k | \mathcal{F}_{t-1}) = \mathbf{x}_t^\top(u) \boldsymbol{\eta}_{\tau_k}, \quad (2.2)$$

其中 $k = 1, \dots, K$, $\mathbf{x}_t(u) = \mathbf{x}_t \otimes (I(u_t \leq u), I(u_t > u))^\top$, $\otimes$ 表示 Kronecker 积, $I(\cdot)$ 为示性函数.

2.2 复合分位数回归模型下门限的估计问题

先定义 $\boldsymbol{\theta} = (\boldsymbol{\eta}^\top, u)^\top$, 再令

$$Q_n(\boldsymbol{\eta}, u) = \sum_{k=1}^K \sum_{t=p+1}^n \rho_{\tau_k} \{y_t - \mathbf{x}_t^\top(u) \boldsymbol{\eta}_{\tau_k}\}, \quad (2.3)$$

其中 $\rho_\tau(u) = u\{\tau - I(u < 0)\}$ 代表文献 [10] 中的损失函数, 参数 θ 可以通过下列参数来估计:

$$\hat{\theta} = (\hat{\eta}^\top, \hat{u})^\top = \underset{\eta \in \mathcal{B}, u \in (M_1, M_2)}{\operatorname{argmin}} Q_n(\eta, u), \quad (2.4)$$

其中 $\mathcal{B}$ 是 $\mathbb{R}^{2(p+1)K}$ 中的紧集合, M_1 和 M_2 是两个常数.

本文的目的是更有效地估计复合不同分位数水平的信息. 不同于文献 [11] 提出的基于门限分位数自回归模型下单个分位数的变点估计, 我们提出的变点估计是通过结合不同分位数下的信息得到的, 许多学者对复合分位数在没有变点的模型下的估计研究过, 如文献 [18–21]. 很少有学者在传统的分位数回归模型假设下结合不同分位数水平对复合变点进行研究 [24, 25], 我们主要关心时间序列的数据, 即门限自回归模型中的复合变点的估计和推断问题.

类似于文献 [16, 25], 用两步截面方法解决 (2.4) 中的最小化问题. 首先, 定义给定 u 下参数 η 的截面估计为

$$\hat{\eta}(u) = \underset{\eta \in \mathcal{B}}{\operatorname{argmin}} Q_n(\eta, u).$$

其次, 门限参数 u 可以通过如下最小化损失的方式进行估计:

$$\hat{u}_{\text{CTQAE}} = \underset{u \in (M_1 + \epsilon, M_2 - \epsilon)}{\operatorname{argmin}} Q_n\{\hat{\eta}(u), u\},$$

其中 ϵ 是很小的正常数, 目的是为了防止搜索 u 时涉及门限的边界效应问题. 从而可以得到最终的参数估计量为 $\hat{\theta} = (\hat{\eta}(\hat{u}_{\text{CTQAE}})^\top, \hat{u}_{\text{CTQAE}})^\top$.

在实际数据分析中, 我们需要选择 $\mathbf{\Gamma}$ 和 τ_k ($k = 1, \dots, K$). 一般来说, 区间 $\mathbf{\Gamma}$ 会由我们感兴趣的问题得到. 本文需要通过检验来保证变点在 $\mathbf{\Gamma}$ 上存在常数门限效应. 这里可以参考文献 [24] 中提出的参数估计量得到. 只要 $\mathbf{\Gamma}$ 确定下来, 分位数水平 τ_k 则可以简单均匀取于 $\mathbf{\Gamma}$ 中. 然而, 参数维度取决于 K 的选择, K 越大, 参数维度越高, 问题越复杂. 从模拟研究中发现复合变点估计在 K 取 19 附近时趋于稳定, 关于 K 选取的敏感性分析详见第 4 节.

注 1 我们在前面考虑的是固定的 K 个分位数水平下复合变点估计, 但当考虑的分位数区间 $\mathbf{\Gamma}$ 是连续的时, 参数估计如下:

$$\hat{\theta} = \hat{\theta}(\Pi) = \underset{u \in (M_1, M_2)}{\operatorname{argmin}} Q_{n, \Pi}(\eta, u),$$

其中

$$Q_{n, \Pi}(\eta, u) = \int_{\mathbf{\Gamma}} \sum_{t=p+1}^n \rho_\tau\{y_t - \mathbf{x}_t^\top(u) \eta_\tau\} d\Pi(\tau),$$

且 Π 是定义在 $\mathbf{\Gamma}$ 上的概率密度函数. 类似于 (2.4), 可以通过两步截面法得到回归系数和门限参数的估计.

2.3 渐近性质

本小节给出门限估计量的渐近分布. 为方便清晰表述, 首先引入一些记号. 定义

$$\Omega_0(u_1, u_2) = E\{\mathbf{x}_t(u_1) \mathbf{x}_t^\top(u_2)\} \quad \text{和} \quad \Omega_1(u) = \sum_{k=1}^K E[f\{\mathbf{x}_t^\top(u) \eta_{\tau_k}^0\} \mathbf{x}_t(u) \mathbf{x}_t^\top(u)].$$

特别地, 当 $u_1 = u_2 = u$ 时, 有 $\Omega_0(u) = E\{\mathbf{x}_t(u)\mathbf{x}_t^\top(u)\}$. 再定义

$$\hat{\Omega}_0(u_1, u_2) = \frac{1}{n-p} \sum_{t=p+1}^n \mathbf{x}_t(u_1)\mathbf{x}_t^\top(u_2),$$

以及

$$\hat{\Omega}_1(u) = \frac{1}{n-p} \sum_{k=1}^K \sum_{t=p+1}^n \hat{f}\{\mathbf{x}_t^\top(u)\hat{\boldsymbol{\eta}}_{\tau_k}\} \mathbf{x}_t(u)\mathbf{x}_t^\top(u).$$

假定 $g(\cdot)$ 为门限变量 u_t 的连续密度函数. 为了得到所提出估计量的渐近性质, 本文需要给出以下假设条件.

(A1) $\{y_t, u_t\}$ 为严平稳、遍历、 ρ -混合过程, 并且 ρ -混合系数满足 $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$.

(A2) 记 $F_t(\cdot) = F_t(\cdot | \mathcal{F}_{t-1})$ 表示给定事件域 $\mathcal{F}_{t-1}$ 条件下 y_t 的条件分布函数, $f_t(\cdot)$ 为对应的条件密度函数, $0 < f_t(\cdot) < \infty$ 且在定义域内一致可积.

(A3) $E\|\mathbf{x}_t\|^{2+\epsilon} < \infty$, 其中 $\|\cdot\|$ 表示常用 Euclid 范式, ϵ 表示一个很小的正常数.

(A4) 对所有 $\tau \in \Gamma$ 和 $u \in [M_1, M_2]$, $\boldsymbol{\eta}(u) = \operatorname{argmin}_{\boldsymbol{\eta}} E[\rho_{\tau}\{y_t - \mathbf{x}_t^\top(u)\boldsymbol{\eta}\}]$ 存在且唯一.

(A5) 对任意 u , $\Omega_1(u)$ 是正定矩阵.

这里条件 (A1)–(A5) 为分位数回归模型理论及门限回归中的常见假设^[10, 16]. 条件 (A1) 保证了序列的平稳性. 条件 (A2) 和 (A3) 给出了门限分位数回归模型的基本假设. 条件 (A4) 保证了门限分位数自回归解的存在唯一性. 条件 (A5) 保证了估计量协方差的可逆性. 下面的定理给出了回归系数和门限估计量的渐近性质.

定理 1 假设条件 (A1)–(A5) 成立, 则有

$$\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}^0.$$

定理 1 表明提出方法的回归系数和门限的估计 $\hat{\boldsymbol{\theta}}$ 具有相合性. 为了建立门限估计量的渐近性质, 我们增加如下额外假设.

(A6) 令 $\delta_{\tau_k}^0 = \boldsymbol{\eta}_{1, \tau_k}^0 - \boldsymbol{\eta}_{2, \tau_k}^0 = \nu_{\tau_k} n^{-a}$ ($k = 1, \dots, K$) 表示分位数 τ_k 下的门限效应, 其中 $\nu_{\tau_k} \neq \mathbf{0}$, $a \in (0, \frac{1}{2})$.

假设条件 (A6) 是为了建立门限估计量的渐近分布, 其给出了回归系数的变化幅度. 当样本量趋于正无穷时, $\delta_{\tau_k}^0$ 趋于 0, 具体内容可参见文献 [25, 26]. 因此可以得到下面的理论.

定理 2 假设条件 (A1)–(A6) 成立, 则有

- (i) $n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}(u^0));$
- (ii) $n^{1-2a}(\hat{u}_{\text{CTQAE}} - u^0) \xrightarrow{d} \frac{\lambda}{\mu^2} \operatorname{argmax}_{-\infty < r < \infty} \{B(r) - \frac{1}{2}|r|\},$

其中 $\boldsymbol{\Sigma}(u^0) = \Omega_1^{-1}(u^0)\tilde{\Omega}_0(u^0)\Omega_1^{-1}(u^0)$, $\tilde{\Omega}_0(u^0) = \sum_{k=1}^K \sum_{k'=1}^K \{\min(\tau_k, \tau_{k'}) - \tau_k \tau_{k'}\} \Omega_0(u^0)$, “ $\xrightarrow{d}$ ” 表示依分布收敛,

$$\lambda = \sum_{k=1}^K \sum_{k'=1}^K \{\min(\tau_k, \tau_{k'}) - \tau_k \tau_{k'}\} \boldsymbol{\nu}_{\tau_k}^\top E(\mathbf{x}_t \mathbf{x}_t^\top | u_t = u^0) \boldsymbol{\nu}_{\tau_{k'}} g(u^0),$$

$$\mu = \sum_{k=1}^K \boldsymbol{\nu}_{\tau_k}^\top E[f\{\mathbf{x}_t^\top(u^0)\boldsymbol{\eta}_{\tau_k}^0\} \mathbf{x}_t \mathbf{x}_t^\top | u_t = u^0] \boldsymbol{\nu}_{\tau_k} g(u^0),$$

$B(r)$ 是由 $B(r) = B_1(-r)I(r \leq 0) + B_2(r)I(r > 0)$ 定义的双边 Brown 运动, $B_1(r)$ 和 $B_2(r)$ 是在 $[0, \infty)$ 上的两个独立的标准 Brown 运动.

定理 2 给出了复合变点估计的渐近分布. 由上述定理可以发现复合变点估计 $\hat{u}$ 的收敛速度受 a 控制, a 越大, 门限效应越弱, 收敛速度越快. 相比较于文献 [16] 研究的单个分位数回归模型下的变点估计性质, 本文提出的复合变点估计具有更有效的估计, 这与文献 [24] 提出的折线分位数回归模型中的复合变点性质类似.

推论 1 在注 1 的条件下, 即当考虑的分位数区间 Γ 是连续时, 回归系数和门限估计量有类似于定理 1 和 2 的结论. 特别地, 我们有类似定理 2(ii) 的门限估计量的渐近分布, 其中 λ 和 μ 的连续形式分别为

$$\lambda = \left\{ \iint_S \sqrt{(\min(\tau, \tau') - \tau\tau') \boldsymbol{\nu}_\tau^\top \mathbb{E}(\mathbf{x}_t \mathbf{x}_t^\top | u_t = u^0) \boldsymbol{\nu}_{\tau'} d\Pi(\tau) d\Pi(\tau')} \right\}^2 g(u^0)$$

和

$$\mu = \int_{\Gamma} \boldsymbol{\nu}_\tau^\top \mathbb{E}[f\{\mathbf{x}_t^\top(u^0) \boldsymbol{\eta}_\tau^0\} \mathbf{x}_t \mathbf{x}_t^\top | u_t = u^0] \boldsymbol{\nu}_\tau d\Pi(\tau) g(u^0),$$

其中 $S = \{(\tau, \tau') | (\tau, \tau') \in \Gamma \times \Gamma\}$.

2.4 复合分位数自回归模型下门限的推断问题

为了构造门限参数 u^0 的置信区间, 若采用定理 2 中的统计量来反演, 则有限样本表现会很差 [5, 26]. 本小节提出两种不同的方法来构造门限变点 u^0 的置信区间, 包括基于似然比检验型和自助法型, 其中自助法型方法计算量大.

2.4.1 似然比检验型

类似于文献 [6, 25], 我们基于似然比检验提出一个更加稳定的方法. 考虑下面的假设检验问题:

$$H_0 : u = u^0, \quad \text{vs.} \quad H_1 : u \neq u^0,$$

原假设表明在所有考虑的分位数区间中变点相同且都等于 u^0 . 基于原假设构造如下似然比 (likelihood ratio, LR) 检验统计量:

$$\text{LR}_n(u^0) = Q_n(\hat{\boldsymbol{\eta}}, u^0) - Q_n(\hat{\boldsymbol{\eta}}, \hat{u}).$$

由于似然比统计量是由原假设下损失函数和备择假设的损失函数相减所得, 其中 $Q_n(\hat{\boldsymbol{\eta}}, u^0)$ 中的 $\boldsymbol{\eta}$ 是由 u 取真值 u^0 时估计而得, 即 $\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\eta}}(u^0)$; $Q_n(\hat{\boldsymbol{\eta}}, \hat{u})$ 是在备择假设下估计得到, 即假设 u 未知时利用截面方法估计得到 $\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\eta}}(\hat{u})$. 但为了表达方便, 此处我们在似然比统计量中用了相同的记号 $\hat{\boldsymbol{\eta}}$, 这也类似于文献 [25, 第 2.3 小节] 给出的记号. 若 $\text{LR}_n(u^0)$ 取很大的值, 我们有理由拒绝原假设, 由此可以得到如下理论.

定理 3 假设条件 (A1)–(A6) 成立, 在原假设下有

$$\text{LR}_n(u^0) \xrightarrow{d} \frac{\lambda}{2\mu} \xi,$$

其中

$$\xi = \sup_{r \in (-\infty, +\infty)} \{2B(r) - |r|\}.$$

注意到 ξ 的累积分布函数为 $P(\xi \leq x) = (1 - e^{-x/2})^2$ (参见文献 [26]). 此外, 实证中参数 $\lambda/(2\mu)$ 需要通过样本进行估计, 即

$$\frac{\hat{\lambda}}{2\hat{\mu}} = \frac{\sum_{k=1}^K \sum_{k'=1}^K \{\min(\tau_k, \tau_{k'}) - \tau_k \tau_{k'}\} \hat{\delta}_{\tau_k}^\top \hat{\mathbf{N}}_{\tau_k} \hat{\delta}_{\tau_{k'}}}{2 \sum_{k=1}^K \hat{\delta}_{\tau_k}^\top \hat{\mathbf{D}}_{\tau_k} \hat{\delta}_{\tau_k}},$$

其中 $\hat{\delta}_{\tau_k} = \hat{\eta}_{1,\tau_k} - \hat{\eta}_{2,\tau_k}$, $\hat{\mathbf{N}}_{\tau_k} = \hat{\mathbf{E}}(\mathbf{x}_t \mathbf{x}_t^\top | u_t = u_{\tau_k}^0)$ 可以基于可观测的 $\mathbf{x}_t \mathbf{x}_t^\top$ 和 u_t 使用窗宽 h_1 和核函数 $K(\cdot)$ 的局部线性方法进行估计, 这里 $u_{\tau_k}^0$ 是分位数水平 τ_k 下的真实门限位置. 同样, 可以使用局部线性方法估计 $\hat{\mathbf{D}}_{\tau_k}$, 即 $\hat{\mathbf{D}}_{\tau_k} = \hat{\mathbf{E}}[\hat{f}_{h_2}\{\mathbf{x}_t^\top(u_{\tau_k}^0)\hat{\eta}_{\tau_k}\}\mathbf{x}_t \mathbf{x}_t^\top | u_t = u_{\tau_k}^0]$, h_2 是核密度估计中的窗宽参数 [25]. 另外, 核函数 $K(\cdot)$ 可以选择为标准正态分布函数 $\Phi(\cdot)$. 窗宽选择可以依照文献 [27] 中的拇指原则, 即选取 $h = 1.06n^{-1/5}s_y$, 其中 s_y 是 $\{y_i\}$ 的标准差.

此外, 为了处理实际数据方便, 可以定义正则化统计量 $\text{NLR}_n(u^0) = 2\hat{\mu}\text{LR}_n(u^0)/\hat{\lambda}$. 由于我们有 $\hat{\lambda}/\hat{\mu} \xrightarrow{P} \lambda/\mu$, 其中 “ $\xrightarrow{P}$ ” 表示依概率收敛, 则其渐近分布为

$$\text{NLR}_n(u^0) \xrightarrow{d} \xi,$$

其中 ξ 在不同显著性水平下的临界值在很多研究中已给出, 如文献 [6, 26]. 同样, 类似于推论 1, 可以得到当分位数区间 $\mathbf{I}$ 是连续时更一般的似然比统计量及其渐近分布, 为了节约空间具体内容省略.

推论 2 定理 3 给出检验统计量在原假设下的渐近分布. 类似于文献 [12, 推论 2.1], 考虑如下局部备择假设, 即

$$H_{1n}: Q_{y_t}(\tau_k | \mathcal{F}_{t-1}) = \mathbf{x}_t^\top(u^0)\boldsymbol{\eta}_{1,\tau_k}^0 I(u_t \leq u^0) + n^{-1/2}a_n \mathbf{x}_t^\top(u^0)\boldsymbol{\eta}_{2,\tau_k}^0 I(u_t > u^0),$$

其中 $a_n \rightarrow \infty$. 对所有的 $t \geq 0$, 有

$$\lim_{n \rightarrow \infty} P(\|\text{NLR}_n(u_0)\| \geq t) = 1.$$

2.4.2 自助法型

考虑到门限估计量 u 的极限过程是非 Gauss 过程且不能简化为均值和方差核的形式, 我们可以采用自助法方法构建更精准的参数的置信区间, 其构建的区间收敛于真实区间的速度在某些条件下会比通常的正态近似要快很多. 对于极限分布难以获得的情况, 自助法依然可以利用样本分位数进行非参数的统计推断.

考虑文献 [28] 中的自助法. 首先从观测 $\{y_t, \mathbf{x}_t, u_t\}_{t=1}^T$ 中抽取新的样本, 并且基于计算参数 $\{\hat{\eta}_b^*, b = 1, \dots, B\}$ 和 $\{\hat{u}_b^*, b = 1, \dots, B\}$ 估计, 其中 B 为自助法重复的次数. 参数置信区间通过自助法样本分位数进行构造.

2.5 模型定阶

在实际数据中, 我们需要确定门限分位数自回归模型中复合变点估计涉及的门限变量 u_t (即 y_{t-d})、延迟参数 d 和滞后阶数 p . 文献 [29] 中的 3.4 部分 (model identification) 基于 GAIC (generalized Akaike information criterion) 方法给出双门限自回归模型的门限变量确定方法. 由于复合分位数回归是加权复合分位数回归 [30] 的一个特例, 因此, 类似于文献 [29], 我们基于 GAIC 方法给出门限变量 u_t 、延迟参数 d 和滞后阶数 p 的确定准则. 为简单起见, 假定 $d \leq d_{\max} = p$.

为了选择参数 d 和 p , 我们提出如下两步迭代算法. 第一步, 对每个 $d \in \{1, 2, \dots, d_{\max}\}$, 通过最小化 CTQAE 的目标函数 (参见 (2.4)), 最终选择目标函数最小值时对应的 d . 第二步, 通过 GAIC 准则选择滞后阶数 p , 这里 GAIC 定义为

$$\text{GAIC}(p) = (n - p)K \log(\hat{\sigma}^2) + 4(p + 1),$$

其中 $\hat{\sigma}^2 = Q_n(\boldsymbol{\eta}, u) / [(n - p)K]$, 且 GAIC 中的第一部分反映模型的拟合程度, 第二部分反映模型的复杂程度.

3 门限分位数自回归平均估计

第 2 节中考虑了一种复合分位数门限自回归模型的变点估计和推断问题, 其主要是通过最小化复合分位数自回归模型中的损失函数得到. 本节提出另一种复合分位数思想, 通过求单个门限分位数自回归系数下的平均估计得到, 这最早由文献 [10] 提出. 有些学者研究过分位数平均估计的问题. 文献 [22] 考虑了金融数据尖峰厚尾的分布特征, 并结合多分位点信息, 在分位数回归框架下提出了回归系数平均估计并给出所提出估计量的渐近性质. 文献 [23] 将分位数平均估计思想推广用于研究金融领域的极端分位数情况. 据我们所知, 分位数平均估计的思想在门限分位数变点估计中很少被讨论. 本文将首次将分位数平均思想应用到门限变点估计中去.

类似于文献 [10], 我们结合不同分位数 τ 下的门限估计提出门限分位数平均估计量. 令 $\hat{u}_\tau$ 和 u_τ^0 为单个分位数 τ 水平下的变点估计量和真实变点, $\hat{u}_{\text{TQAE}}$ 为分位数变点平均估计量. 于是对模型 (2.1) 提出如下门限分位数自回归平均估计量:

$$\hat{u}_{\text{TQAE}} = \frac{1}{K} \sum_{k=1}^K \hat{u}_{\tau_k}. \quad (3.1)$$

为给出变点的 TQAE 的渐近性质, 我们提出下面的假设条件.

(A7) 当 $n \rightarrow +\infty$ 时,

$$\sum_{t=p+1}^n \int_0^{-\Delta \mathbf{x}_t^\top(\omega_{2\tau}) \delta_\tau^0} [F\{\mathbf{x}_t^\top(u_\tau^0) \boldsymbol{\eta}_\tau^0 - s \Delta \mathbf{x}_t^\top(\omega_{2\tau}) \delta_\tau^0\} - F\{\mathbf{x}_t^\top(u_\tau^0) \boldsymbol{\eta}_\tau^0\}] ds = \frac{1}{2} \mu_\tau |\omega_{2\tau}| + o_p(1)$$

在 $\tau \in \Gamma \subset (0, 1)$ 上一致成立, 其中

$$\begin{aligned} \Delta \mathbf{x}_t(\omega_{2\tau}) &= \mathbf{x}_t \{I(u_t \leq u_\tau^0 + n^{-1+2a} \omega_{2\tau}) - I(u_t \leq u_\tau^0)\}, \\ \mu_\tau &= \nu_\tau^\top \mathbb{E}[f\{\mathbf{x}_t^\top(u_\tau^0) \boldsymbol{\eta}_\tau^0\} \mathbf{x}_t \mathbf{x}_t^\top \mid u_t = u_\tau^0] \nu_\tau g(u_\tau^0), \\ \delta_\tau^0 &= \boldsymbol{\eta}_{1,\tau}^0 - \boldsymbol{\eta}_{2,\tau}^0, \end{aligned}$$

$\omega_{2\tau}$ 为某个常数.

定理 4 假设条件 (A1)–(A5) 和 (A7) 成立, 则有

$$n^{1-2a}(\hat{u}_{\text{TQAE}} - u^0) \xrightarrow{d} \zeta \arg\max_{-\infty < r < \infty} \left\{ B(r) - \frac{1}{2}|r| \right\},$$

其中

$$\zeta = \sum_{k=1}^K \left(\frac{\lambda_{\tau_k}}{\mu_{\tau_k}^2} \right), \quad \lambda_{\tau_k} = (\tau_k - \tau_k^2) \nu_{\tau_k}^\top \mathbb{E}(\mathbf{x}_t \mathbf{x}_t^\top \mid u_t = u_{\tau_k}^0) \nu_{\tau_k} g(u_{\tau_k}^0).$$

定理 4 表明 $\hat{u}_{\text{TQAE}}$ 的极限过程是非 Gauss 的. 第 4 和 5 节中的模拟数据和实证结果可以发现, 相比较于单个分位数下的变点估计, 门限分位数自回归平均估计更加准确.

4 模拟研究

本节在不同误差形式下通过 Monte Carlo 模拟研究验证所提出模型的有限样本表现. 考虑两类不同的数据产生过程: 条件同方差和条件异方差类型. 模型 1-3 的数据来自如下条件同方差数据产生过程:

$$y_t = \begin{cases} 2 + 0.3y_{t-1} + \phi(y_{t-1})\varepsilon_t, & y_{t-1} \leq 0, \\ -2 + 0.7y_{t-1} - \phi(y_{t-1})\varepsilon_t, & y_{t-1} > 0, \end{cases}$$

其中门限变量为 y_{t-1} ; 门限值为 0; ε_t 为随机扰动部分, 影响 y_t 与总体趋势之间的偏离. 模型 1-3 中 ε_t 分别服从正态分布、标准逻辑分布和对数正态分布, 即 $\varepsilon_t \sim N(0, 1)$, $\varepsilon_t \sim L(0, 1)$, $\varepsilon_t \sim LN(0, 1)$. 令 $\phi(y_{t-1}) = 1$ 表示条件同方差模型.

对于模型 4 和 5, 考虑如下条件异方差数据产生过程:

$$y_t = \begin{cases} 0.5y_{t-1} + \{1 - \phi(y_{t-1})\}\varepsilon_t, & y_{t-1} \leq 0, \\ 0.5y_{t-1} + \{1 + \phi(y_{t-1})\}\varepsilon_t, & y_{t-1} > 0, \end{cases}$$

在模型中 $\phi(y_{t-1}) = cy_{t-1}$ 对应于条件异方差情形. 模型 4 和 5 中 c 分别取值为 1.65 和 2.65, ε_t 分别服从分布 $\varepsilon_t \sim N(0, 1)$ 和 $\varepsilon_t \sim U(-1, 1)$.

在模拟研究中, 我们对比以下 4 种估计门限的方法: 文献 [26] 提出的门限最小二乘估计量 (TAR)、文献 [16] 提出的 $\tau = 0.5$ 下门限分位数自回归估计量 (TQAR)、门限自回归分位数复合估计量 (CTQAE) 和门限分位数平均估计量 (TQAE). 4 种估计方法对变点及回归系数估计的对比结果具体见表 1 和 2. 对 CTQAE 和 TQAE, 设定 $K = 19$ 并且等间隔划分分位数水平 $\tau_k = k/(K + 1)$ ($k = 1, \dots, 19$), 这主要是由后面表 3 和 4 中对 K 的敏感度分析而得的. 为了避免格子点在边界上搜索变点, 对所有的方法, 我们用 u 的上、下 10% 分位数作为搜索门限值的上、下界. 样本量 n 设置为 200 和 500, 在所有的模拟过程中, 模拟重复次数设定为 500 次.

表 1 总结了 4 种变点估计的平均偏差和均方误差. 我们发现所有估计都有一个可以忽略的偏差. 对条件同方差模型 1-3, 与 TAR 和 TQAR 相比, CTQAE 和 TQAE 具有更高的效率. 对具有条件异方差正态分布的模型 4 和条件异方差均匀分布模型 5, 与 TAR 门限估计相比, CTQAE 和 TQAE 更有效率. 同时, 复合估计在误差是条件异方差正态分布时 (模型 4) 体现地更加明显. 模型 4 和 5 由于存在较小的门限效应, 导致变点估计存在较大的波动, 最终引起估计偏差和均方误差偏大. 对所有的不同模型, CTQAE 都比 TAR 和 TQAR 更有效. 总体来说, 相比较于 TAR 和 TQAR, 提出的 CTQAE 和 TQAE 方法更加有效.

进一步, 考虑 4 种变点估计方法对回归系数估计的影响. 表 2 报告了回归系数估计的均方误差, 其中 TQAR、CTQAE 和 TQAE 估计结果均为 0.5 分位数处对应的系数估计结果. 参数 β_{11} 和 β_{21} 的真值为 0.5. 我们可以发现所有回归系数估计的均方误差都较小. 对于条件同方差正态模型 1, TAR 估计效果更有效率. 但对于具有条件同方差逻辑分布和对数正态分布的模型 2 和 3, 这 4 种方法表现相当, TQAR、CTQAE 和 TQAE 稍优于 TAR. 但对条件异方差模型 4 和 5, 相比 TAR 和 TQAR, 复合

表 1 不同方法下变点估计的偏差 (Bias) 和均方误差 (MSE), 所有的值都乘以因子 100

模型		TAR		TQAR		CTQAE		TQAE	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$n = 200$	1: Normal-HO	2.76	1.89	0.62	0.89	0.45	0.46	0.91	0.31
	2: Logit-HO	8.61	11.70	5.23	6.52	4.69	5.81	7.28	3.61
	3: LNormal-HO	7.04	3.68	8.14	3.50	6.86	2.99	9.82	2.69
	4: Normal-HE	28.26	135.58	29.45	119.49	16.46	51.77	21.49	54.09
	5: Uniform-HE	27.08	271.32	18.74	231.98	16.42	57.88	20.97	60.21
$n = 500$	1: Normal-HO	-0.14	0.09	-0.44	0.13	-0.45	0.07	-0.30	0.04
	2: Logit-HO	1.60	1.39	0.79	1.63	0.32	0.36	3.10	0.78
	3: LNormal-HO	5.57	1.99	6.78	2.15	5.48	1.78	6.51	1.05
	4: Normal-HE	2.14	98.78	22.77	73.56	4.80	28.59	4.22	28.69
	5: Uniform-HE	6.58	100.21	3.79	88.23	3.67	29.69	1.18	31.06

Normal-HO: 模型 1 具有正态的条件同方差; Logit-HO: 模型 2 具有标准逻辑分布的条件同方差; LNormal-HO: 模型 3 具有对数正态的条件同方差; Normal-HE: 模型 4 具有正态条件异方差; Uniform-HE: 模型 5 具有均匀分布条件异方差.

表 2 不同方法下回归系数估计的均方误差 (MSE), 所有的值都乘以因子 100

模型		TAR		TQAR		CTQAE		TQAE	
		β_{11}	β_{21}	β_{11}	β_{21}	β_{11}	β_{21}	β_{11}	β_{21}
$n = 200$	1: Normal-HO	5.68	1.32	2.67	1.65	2.45	1.68	2.38	1.62
	2: Logit-HO	6.80	1.96	5.34	2.02	5.35	2.10	5.77	2.12
	3: LNormal-HO	2.04	1.56	0.58	0.55	0.57	0.54	0.58	0.52
	4: Normal-HE	65.84	63.41	51.59	52.50	48.93	45.83	51.57	52.51
	5: Uniform-HE	48.77	51.80	51.29	53.80	50.96	53.27	51.04	53.25
$n = 500$	1: Normal-HO	0.89	0.51	1.00	0.71	0.88	0.71	0.86	0.70
	2: Logit-HO	1.99	0.63	1.49	0.82	1.30	0.76	1.86	0.76
	3: LNormal-HO	0.74	0.42	0.19	0.16	0.18	0.16	0.19	0.16
	4: Normal-HE	50.31	55.00	51.59	52.47	47.03	46.36	40.38	44.90
	5: Uniform-HE	40.96	48.51	40.61	46.35	50.02	39.11	40.57	46.33

CTQAE 和 TQAE 估计效率有很大程度的提高. 总体看, 对所有的不同的模型, CTQAE 和 TQAE 下的回归系数估计比 TQAR 的估计更加有效, 尤其是当误差项满足条件异方差条件时, 这就更加确认了结合不同分位数水平信息的变点估计比单个分位数下的估计更有效率.

我们提出的变点估计方法需要选择合适的 K , 为了研究其对估计 K 的敏感性, 我们在表 3 和 4 中分别给出 CTQAE 和 TQAE 在不同分位数下的均方误差, 分位数个数 K 设定为 1、3、6、9、19 和 29, 其中 $K = 1$ 对应于 $\tau = 0.5$, 等价于表 1 中 TQAR. 由表 3 和 4 可知, 对于所有考虑的模型, CTQAE 和 TQAE 的精度起初随着 K 增加而提高, 但当 K 取 19 附近时估计的精度趋向稳定. 为了在计算时间和估计精度上寻求平衡, 在进行数值模拟时将取 $K = 19$.

为了评估第 2.4 小节所提出的门限推断方法的有限样本表现, 表 5 给出不同方法下门限估计名义水平为 95% 下的覆盖率和平均区间长度, 其中 TQAR 方法下展示的是 $\tau = 0.5$ 下的模拟结果. 我们

表 3 不同 K 取值下回归系数 CTQAE 的均方误差 (MSE), 所有的值都乘以因子 100

模型		$K = 1$	$K = 3$	$K = 6$	$K = 9$	$K = 19$	$K = 29$
$n = 200$	1: Normal-HO	0.89	0.38	0.48	0.46	0.46	0.46
	2: Logit-HO	6.52	6.84	5.45	5.81	5.71	5.77
	3: LNormal-HO	3.50	3.02	2.94	2.99	3.05	3.02
	4: Normal-HE	116.95	52.65	52.85	51.59	51.77	51.90
	5: Uniform-HE	231.98	57.05	54.88	57.62	57.88	57.95
$n = 500$	1: Normal-HO	0.13	0.09	0.07	0.07	0.08	0.08
	2: Logit-HO	1.63	0.25	0.32	0.36	0.45	0.47
	3: LNormal-HO	2.15	1.80	1.85	1.78	1.78	1.80
	4: Normal-HE	72.88	28.66	28.83	28.82	28.59	28.66
	5: Uniform-HE	88.23	28.87	29.11	29.06	29.69	29.83

表 4 不同 K 取值下回归系数 TQAE 的均方误差 (MSE), 所有的值都乘以因子 100

模型		$K = 1$	$K = 3$	$K = 6$	$K = 9$	$K = 19$	$K = 29$
$n = 200$	1: Normal-HO	0.89	0.64	0.29	0.31	0.26	0.23
	2: Logit-HO	6.52	7.15	4.28	3.61	3.63	3.48
	3: LNormal-HO	3.50	4.90	2.95	2.70	2.26	2.51
	4: Normal-HE	116.95	61.95	55.08	54.32	54.09	54.20
	5: Uniform-HE	231.98	68.32	68.32	60.41	60.21	51.04
$n = 500$	1: Normal-HO	0.13	0.16	0.06	0.04	0.03	0.03
	2: Logit-HO	1.63	2.90	1.16	0.78	0.62	0.59
	3: LNormal-HO	2.15	2.96	1.25	1.05	0.81	0.77
	4: Normal-HE	72.88	42.29	30.67	27.96	28.69	28.38
	5: Uniform-HE	88.24	47.87	32.90	31.60	31.06	30.77

表 5 不同方法下门限估计名义水平为 95% 下的覆盖率和平均区间长度

模型		TAR		TQAR		CTQAE		TQAE	
		CP	Len	CP	Len	CP	Len	CP	Len
$n = 200$	1: Normal-HO	0.942	0.289	0.945	0.259	0.930	0.217	0.960	0.220
	2: Logit-HO	0.942	0.667	0.985	0.458	0.940	0.285	0.958	0.348
	3: LNormal-HO	0.954	0.442	0.960	0.550	0.880	0.317	0.930	0.348
	4: Normal-HE	0.150	420.701	1.000	381.594	0.840	1.600	0.815	1.800
	5: Uniform-HE	0.360	318.130	1.000	282.556	0.865	1.630	0.865	1.887
$n = 500$	1: Normal-HO	0.945	0.184	0.982	0.112	0.955	0.107	0.975	0.112
	2: Logit-HO	0.954	0.318	0.976	0.172	0.975	0.169	0.950	0.168
	3: LNormal-HO	0.982	0.306	0.942	0.195	0.960	0.196	0.960	0.180
	4: Normal-HE	0.200	192.545	1.000	157.361	0.950	1.215	0.940	1.070
	5: Uniform-HE	0.505	164.417	1.000	144.373	0.965	1.320	0.945	1.069

CP (coverage probability): 覆盖率, Len (Length): 平均区间长度.

发现对条件同方差模型 1-3, 与 TAR 和 TQAR 相比, CTQAE 和 TQAE 下的覆盖率在不同的样本量下更接近于名义水平 95%, 且具有更短的置信区间长度. 而由于在条件异方差模型 4 和 5 中 TAR 和 TQAR 方法下不存在门限, 这导致模型在这两种方法下估计完全失效. 与此同时, 我们可以发现本文提出的 CTQAE 和 TQAE 仍然有效, 尤其在样本量 $n = 500$ 时, 覆盖率接近于名义水平 95%, 这表明了我们提出的方法在门限推断下的有效性.

从表 1-5 可以得出, 总体上 CTQAE 和 TQAE 具有相似的有限样本表现. 如从表 1, 我们发现条件同方差模型 1-3 中, 有限样本下 CTQAE 具有更小的 Bias, TQAE 具有更小的 MSE (mean squared error), 但模型 2 在 $n = 500$ 时 CTQAE 具有更小的 MSE. 在条件异方差模型 4 和 5 中, 样本量 $n = 200$ 时, 可以看出 CTQAE 具有更小的 Bias 和 MSE, 而当样本量增大至 $n = 500$ 时, CTQAE 和 TQAE 的 MSE 逐渐接近. 同时我们发现, 随着样本量的增大, 两种方法的 Bias 和 MSE 的差距逐渐变小, 尤其当 $n \geq 1,000$ 时 (为节约空间没有附上模拟结果). 此外, 从表 2 和 5 中可以得到两种方法下系数的均方误差和门限变点估计的覆盖率和平均长度. 所提出的 CTQAE 和 TQAE 方法随着样本量的增大, 估计和推断的差距逐渐变小. 由此可看出, 我们提出的两种估计方法的效率在不同的情形下是可比的.

5 实证分析

股市收益率的自相关性是金融市场研究的热点问题, 其对于股票价格预测和股市风险度量具有极其重要的意义. 传统金融观点认为股票收益率之间不存在相关性, 故而金融市场不可预测. 文献 [31] 提出的有效市场假说被诸多学者批评. 因为这些学者通过实证研究发现了收益率的自相关性和可预测性. 这可以使投资者系统获得超额收益. 随着统计学的发展, 很多研究方法被引入到金融系统中, 以期更好地逼近和揭示真实金融系统的运行规律, 更深入描述股市收益率自相关性特征. 由于金融市场的复杂性, 尤其是非线性和异质性特征, 在揭示金融系统运行规律时, 应用最广泛的是门限模型. 文献 [5] 在均值回归框架下提出了 TAR 模型, 但该模型无法反应系统的异质性. 为了更加全面描述解释变量在不同分位点处的异质影响, 文献 [16] 结合分位数自回归和自激励门限自回归模型^[14], 提出了门限分位数自回归模型, 推导了相关回归系数的渐近性质并提出了变点存在性检验的程序. 此后有较多学者研究 TQAR 模型并将其用于金融研究, 提高了对金融数据的挖掘深度. 例如, 文献 [32] 通过上证 A 股指数、不同市场特点以及投资者情绪探究收益率自相关性和可预测性; 文献 [33] 考虑两种类型的非线性 (变点和分位数), 应用 TQAR 模型解释 Tobin 税对汇率波动的作用; 文献 [34] 提出一种基于资产定价模型的门限分位数自回归模型, 该模型在广义自回归条件异方差的假设下, 检验超额股票收益与“异常交易量”之间的关系. 中国股市是世界上最大的新兴市场, 本节使用所提出的 CTQAE 和 TQAE 模型重新研究上证 A 股指数的自相关性.

图 1 给出上证综合指数日收益序列图. 由图 1 可以看出, 收益率序列具有非对称的异质性特征, 并且在一些时期波动较大, 如 2008-2009 年金融危机时期, 股市出现明显的大起大落.

5.1 数据选取与数据描述

本文选取 2005 年 1 月 4 日至 2014 年 12 月 31 日上证综合指数收盘价 2,427 个数据来研究中国股市收益率的自相关性. 本文样本选取的逻辑在于: 第一, 上证综合指数覆盖面广、市场影响力大, 能够充分刻画中国股票市场的动态变化; 第二, 时间跨度较长, 由于 2015 年是中国股市极为特殊

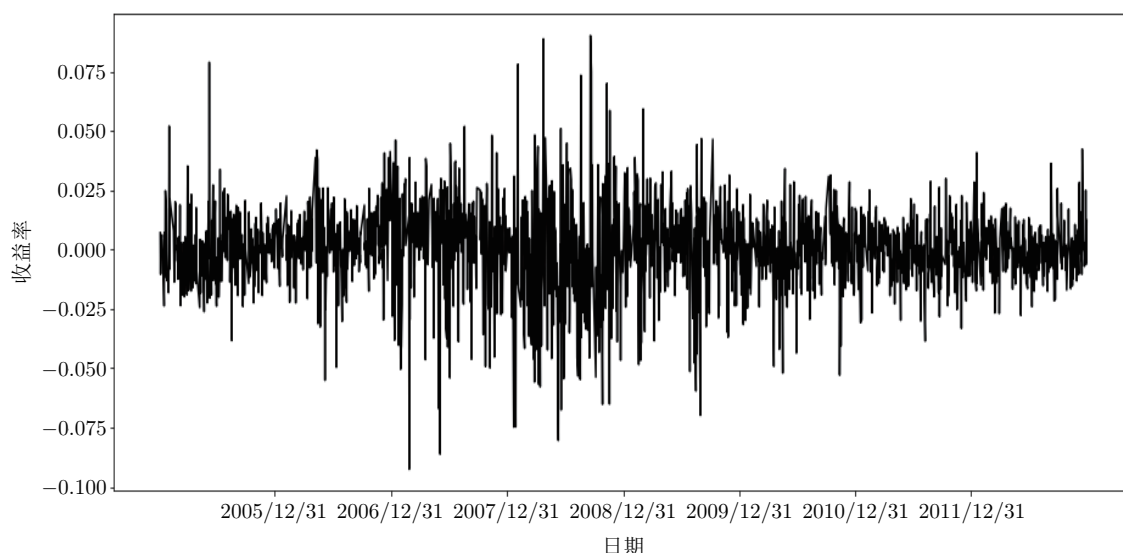


图 1 上证综合指数 2005–2014 年收益率时间序列图

的一年, 国家调控金融市场, 为了减少该外生因素对股市的影响, 从而选取 2015 年前的 10 年数据; 第三, 相比月度和年度数据, 日度数据更能揭示中国股市波动的细节. 定义收益率为对数收益率, 即 $y_t = \ln(P_t/P_{t-1})$, 其中 P_t 为日收盘价. 数据来源于 Wind 数据库.

表 6 给出上证综合指数日收益率的描述统计. 通过表 6 相关统计量可以看出中国股市收益率序列不服从正态分布, 且具有平稳、非对称及厚尾分布特征. 故而有理由相信文中所提出的模型可以很好地拟合该数据.

5.2 模型的估计与检验

首先需要确定模型中的门限变量 y_{t-d} 、延迟参数 d 和滞后参数 p . 利用第 2.5 小节提出的方法, 我们可以得到 $d = 1$ 且 $p = 3$ 时 GAIC 统计量达到最小值. 因此, 考虑下面的模型:

$$y_t = \begin{cases} \beta_{10} + \beta_{11}y_{t-1} + \beta_{12}y_{t-2} + \beta_{13}y_{t-3}, & y_{t-1} \leq u, \\ \beta_{20} + \beta_{21}y_{t-1} + \beta_{22}y_{t-2} + \beta_{23}y_{t-3}, & y_{t-1} > u, \end{cases} \quad (5.1)$$

其中 $t = 1, \dots, T$, y_t 表示在 t 时刻的收益率, y_{t-2} 和 y_{t-3} 表示控制变量. 由于两个门限值的分位数回归不显著, 所以仅考虑一个变点的门限分位数自回归模型. 在利用复合分位数门限自回归模型分析实际数据时, 我们将遵循以下次序. 首先, 检测感兴趣的单个分位点上是否存在门限效应; 其次, 在单个分位数变点存在的条件下, 检验变点在不同的分位数下是否相同; 最后, 结合多分位数信息, 利用提出的 CTQAE 和 TQAE 方法进行估计.

首先, 根据文献 [16] 构建的 Ave-W 和 K-S AveW 统计量分别检验在单个分位点和整个分位区间

表 6 上证综合指数日收益率的描述统计

均值	中位数	标准差	偏度系数	J-B 检验	峰度	ADF	PP
0.000	0.001	0.017	-0.328	3.631	1380.000***	-11.571***	-2492.800***

*** 表示检验结果在 1% 水平下显著, ADF: Augmented Dickey-Fuller, PP: Phillips-Perron.

上变点的存在性. 考虑在 9 个单独分位点 ($\tau = 0.1, 0.2, \dots, 0.9$)、0.1–0.9 整个分位数区间上是否存在门限效应. 变点搜索上下界为上证综合指数收益率第 10% 和 90% 分位数. 表 7 报告了 TQAR 模型参数估计结果及门限效应检验统计量的 p 值. p 值结果表明, 上证综合指数收益率在单分位点都存在明显的门限效应. 利用 K-S AveW 在 0.1–0.9 整个分位数区间上检验的 p 值为 0, 表明上证综合指数收益率在整个分位数区间上同样存在明显的门限效应. 我们可以发现, 变点估计值在区间 $[0.1, 0.3]$ 、 $[0.4, 0.6]$ 和 $[0.7, 0.9]$ 上非常接近. 其次, 我们利用文献 [25] 中对不同分位数下是否存在常数门限效应提出基于似然比检验, 可以得到 3 个分位区间, 即 Γ 分别为 $[0.1, 0.3]$ 、 $[0.4, 0.6]$ 和 $[0.7, 0.9]$ 时都存在常数门限效应. 这表明在统计意义上, 在区间 Γ 分别为 $[0.1, 0.3]$ 、 $[0.4, 0.6]$ 和 $[0.7, 0.9]$ 时存在共同的变点. 最后, 在此条件下考虑变点的共同性, 利用复合分位数回归方法提高变点的估计精度.

上述选取的 $[0.1, 0.3]$ 、 $[0.4, 0.6]$ 和 $[0.7, 0.9]$ 不同分位区间代表股市不同的状态, 其分别对应低迷市场 (lower quantile, LQ)、温和市场 (middle quantile, MQ) 和繁荣市场 (higher quantile, HQ). 表 8 报告所提出的 TAR 模型和 TQAR 模型在 0.1、LAD (least absolute deviation) 和 0.9 分位数水平, 以及 CTQAE 和 TQAE 在 LQ、MQ 和 HQ 下的参数估计和区间估计. 考虑到样本较小, 我们通过自助法型方法, 给出回归系数和门限值的置信区间. 由表 8 可知, 考虑变点的共同性, 本文所提出方法与 TQAR 模型相比, 参数估计的置信区间更短, 说明 CTQAE 和 TQAE 具有更高的估计精度.

模型设定为两阶段, 其中阶段 1: $y_{t-1} \leq u$ 表示前期收益 (低收益) 小于门限值, 阶段 2: $y_{t-1} \geq u$ 表示前期收益 (高收益) 大于门限值. 表 8 报告了参数估计的结果. 表中仅报告了截距、一阶自回归系数和门限估计量. 通过表 8 中 CTQAE 模型估计的系数可知, 中国股市存在明显的非线性和异质性特征. 非线性特征具体表现在不同阶段收益率序列的自相关性差异较大. 对于阶段 1, 股票收益率存在自相关性并随分位点的增加, 收益率序列由正相关转为负相关. 阶段 2 所得结论与阶段 1 相反. 异质性特征具体表现在, 不同阶段不同分位点自相关性强度有所差异. 对于阶段 1, 自相关程度随着分位点增加而逐渐提高. 对阶段 2, 自相关性在 0.1 和 0.9 分位点处强于 0.5 分位点处. 可见在极端市场条件下, 收益率具有更强的可预测性. 正自相关性表明价格对信息反应不足; 相反, 则表示反应过度. 市场反应过度意味着投资者对信息过分关注, 产生过激反应. 反应不足意味着投资者对新信息持保守态度, 这种心理偏差导致收益沿着原来的方向运动, 称为“动量效应”. 反应过度表示对新信息所持态度太过分乐观, 容易产生非理性投资, 使得收益发生反转, 这种现象称为“反转效应”. 由表 8 中模型回归系数的估计结果可以看出, 门限值控制着“反转效应”与“动量效应”之间的转换. 在 2005–2014 年间, 中国股市两种效应具体表现在: 当市场处于低迷状态时, 投资者对新信息反应不足, 造成“动量效应”到“反转效应”的现象转变. 随着股市向繁荣过度, 投资者对市场普遍看好, 这种心理偏差造成“反转效应”到“惯性效应”的转变. 中国股市表现出复杂的自相关特征, 投资者可以结合市场环境和门限效

表 7 单个分位数和整个分位数区间下门限分位数自回归模型 (TQAR) 参数估计及显著性检验

τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\beta}_{10}$	-0.018	-0.010	-0.006	-0.003	0.000	0.001	0.006	0.008	0.016
$\hat{\beta}_{11}$	0.087	0.049	-0.083	-0.098	-0.094	-0.171	-0.144	-0.230	-0.244
$\hat{\beta}_{20}$	-0.005	-0.003	-0.006	-0.001	0.001	0.002	0.005	0.010	0.016
$\hat{\beta}_{21}$	-0.470	-0.215	-0.127	0.127	0.150	0.122	0.171	0.206	0.333
$\hat{u}$	0.011	0.011	0.010	-0.004	-0.003	-0.004	0.003	0.003	0.003
Ave-W 检验	0.025	0.005	0.005	0.010	0.005	0.005	0.000	0.000	0.000
K-S AveW 检验	0.000								

表 8 不同方法下的参数估计和区间估计

TAR	TQAR			CTQAE			TQAE		
	0.1	LAD	0.9	LQ	MQ	HQ	LQ	MQ	HQ
$\hat{\beta}_{10}$	-0.006	-0.018	0.016	-0.018	0.002	0.009	-0.011	-0.001	0.008
$\hat{\beta}_{11}$	-0.236	0.087	-0.243	0.090	-0.152	-0.231	-0.009	-0.119	-0.244
Boot	$[-0.491, -0.016]$	$[-0.401, 0.363]$	$[-0.678, -0.116]$	$[-0.191, 0.245]$	$[-0.267, -0.032]$	$[-0.409, -0.074]$	$[-0.229, 0.141]$	$[-0.277, -0.012]$	$[-0.409, -0.083]$
$\hat{\beta}_{20}$	0.000	-0.005	0.016	-0.005	0.002	0.009	-0.005	0.001	0.010
$\hat{\beta}_{21}$	0.061	-0.470	0.332	-0.461	0.138	0.204	-0.138	0.083	0.170
Boot	$[-0.097, 0.206]$	$[-0.646, 0.068]$	$[-0.104, 0.673]$	$[-0.608, 0.057]$	$[-0.013, 0.204]$	$[0.004, 0.334]$	$[-0.471, 0.157]$	$[-0.159, 0.188]$	$[-0.003, 0.382]$
$\hat{u}$	-0.018	0.011	0.003	0.011	-0.003	0.003	0.011	-0.004	0.003
Boot	$[-0.019, 0.011]$	$[-0.016, 0.016]$	$[-0.012, 0.011]$	$[-0.016, 0.014]$	$[-0.015, 0.010]$	$[-0.013, 0.010]$	$[-0.016, 0.014]$	$[-0.015, 0.012]$	$[-0.010, 0.010]$

应, 准确地把握市场自相关性特征、预测股市动向和理性投资.

6 结论

本文复合不同的分位数对门限分位数自回归模型中的变点估计问题提出两种新的估计方法: 门限自回归分位数复合估计和门限分位数平均估计, 并给出相应的渐近性质. 数值模拟和实证数据显示, 相比较于传统的 TAR 和 TQAR 方法, 本文所提出的 CTQAE 和 TQAE 具有更高的计算精度. 然而, 本文尚存在一些值得改进的地方. 例如, 可以研究门限分位数回归下面板分位数模型中的变点估计问题. 另外, 本文是基于传统的先检验变点存在性、再估计变点的位置的方式去识别变点, 需要搜索所有的变点可能存在的区间, 也可以尝试基于 LASSO (least absolute shrinkage and selection operator) 方法同时检验变点存在性以及估计变点的位置 (参见文献 [35]). 这超出了本文的研究范围, 我们将在未来开展相应研究.

致谢 作者非常感谢审稿人以及副主编提出的宝贵修改意见和建议, 这让文章的质量有了很大的提升.

参考文献

- 1 Csörgő M, Horváth L. Limit Theorems in Change-Point Analysis. New York: John Wiley & Sons, 1997
- 2 Bai J. Testing for parameter constancy in linear regressions: An empirical distribution function approach. *Econometrica*, 1996, 64: 597–622
- 3 Zou C, Yin G, Feng L, et al. Nonparametric maximum likelihood approach to multiple change-point problems. *Ann Statist*, 2014, 42: 970–1002
- 4 Tong H. On a threshold model. In: *Pattern Recognition and Signal Processing*. Amsterdam: Sijthoff & Noordhoff, 1978, 575–586
- 5 Tong H. Threshold Models in Non-linear Time Series Analysis. New York: Springer-Verlag, 1983
- 6 Hansen B E. Threshold autoregression in economics. *Stat Interface*, 2011, 4: 123–127
- 7 Tsay R S. Analysis of Financial Time Series. New York: John Wiley & Sons, 2015
- 8 Tong H, Thanoon B, Gudmundsson G. Threshold time series modeling of two icelandic riverflow systems. *J Am Water Resour Assoc*, 1985, 21: 651–662
- 9 Chan W S, Wong A C S, Tong H. Some nonlinear threshold autoregressive time series models for actuarial use. *N Am Actuar J*, 2004, 8: 37–61
- 10 Koenker R, Bassett Jr G. Regression quantiles. *Econometrica*, 1978, 46: 33–50
- 11 Qu Z. Testing for structural change in regression quantiles. *J Econometrics*, 2008, 146: 170–184
- 12 Zhang L, Wang H J, Zhu Z. Testing for change points due to a covariate threshold in quantile regression. *Statist Sinica*, 2014, 24: 1859–1877
- 13 Lee S, Liao Y, Seo M H, et al. Oracle estimation of a change point in high-dimensional quantile regression. *J Amer Statist Assoc*, 2018, 113: 1184–1194
- 14 Cai Y, Stander J. Quantile self-exciting threshold autoregressive time series models. *J Time Ser Anal*, 2008, 29: 186–202
- 15 Cai Y. Forecasting for quantile self-exciting threshold autoregressive time series models. *Biometrika*, 2010, 97: 199–208
- 16 Galvao Jr A F, Montes-Rojas G, Olmo J. Threshold quantile autoregressive models. *J Time Ser Anal*, 2011, 32: 253–267
- 17 Chen C W S, Gerlach R. Semi-parametric quantile estimation for double threshold autoregressive models with heteroskedasticity. *Comput Statist*, 2013, 28: 1103–1131
- 18 Zou H, Yuan M. Composite quantile regression and the oracle model selection theory. *Ann Statist*, 2008, 36: 1108–1126
- 19 Kai B, Li R, Zou H. Local composite quantile regression smoothing: An efficient and safe alternative to local polynomial regression. *J R Stat Soc Ser B Stat Methodol*, 2010, 72: 49–69
- 20 Lv S, He X, Wang J. Gradient-induced model-free variable selection with composite quantile regression. *Statist Sinica*, 2018, 28: 1521–1538
- 21 Xie Q, Sun Q, Liu J. Local weighted composite quantile estimation and smoothing parameter selection for nonparametric derivative function. *Econometric Rev*, 2020, 39: 215–233

- 22 Zhao Z, Xiao Z. Efficient regressions via optimally combining quantile information. *Econom Theory*, 2014, 30: 1272–1314
- 23 Wang K, Wang H J. Optimally combined estimation for tail quantile regression. *Statist Sinica*, 2016, 26: 295–311
- 24 Zhang L, Wang H J, Zhu Z. Composite change point estimation for bent line quantile regression. *Ann Inst Statist Math*, 2017, 69: 145–168
- 25 Su L, Xu P. Common threshold in quantile regressions with an application to pricing for reputation. *Econometric Rev*, 2019, 38: 417–450
- 26 Hansen B E. Sample splitting and threshold estimation. *Econometrica*, 2000, 68: 575–603
- 27 Silverman B W. *Density Estimation for Statistics and Data Analysis*. London: Routledge, 2018
- 28 Freedman D A. Bootstrapping regression models. *Ann Statist*, 1981, 9: 1218–1228
- 29 Jiang J, Jiang X, Song X. Weighted composite quantile regression estimation of DTARCH models. *Econom J*, 2014, 17: 1–23
- 30 Jiang X, Jiang J, Song X. Oracle model selection for nonlinear models based on weighted composite quantile regression. *Statist Sinica*, 2012, 4: 1479–1506
- 31 Malkiel B G, Fama E F. Efficient capital markets: A review of theory and empirical work. *J Finance*, 1970, 25: 383–417
- 32 Xue W J, Zhang L W. Stock return autocorrelations and predictability in the Chinese stock market—evidence from threshold quantile autoregressive models. *Econ Model*, 2017, 60: 391–401
- 33 Damette O, Park B J. Tobin tax and volatility: A threshold quantile autoregressive regression framework. *Rev Int Econ*, 2015, 23: 996–1022
- 34 Chen C W S, So M K P, Chiang T C. Evidence of stock returns and abnormal trading volume: A threshold quantile regression approach. *Jpn Econ Rev*, 2016, 67: 96–124
- 35 Li D, Qian J, Su L. Panel data models with interactive fixed effects and multiple structural breaks. *J Amer Statist Assoc*, 2016, 111: 1804–1819
- 36 Chernozhukov V, Hansen C. Instrumental quantile regression inference for structural and treatment effect models. *J Econometrics*, 2006, 132: 491–525

附录 A

首先引入一些记号. 令

$$Q_n(\boldsymbol{\eta}, u) = \frac{1}{n} \sum_{t=p+1}^n q_t(\boldsymbol{\eta}, u), \quad \bar{Q}_n(\boldsymbol{\eta}, u) = \frac{1}{n} \sum_{t=p+1}^n E\{q_t(\boldsymbol{\eta}, u)\},$$

其中

$$q_t(\boldsymbol{\eta}, u) = \sum_{k=1}^K [\rho_{\tau_k}\{y_t - \mathbf{x}_t^\top(u) \boldsymbol{\eta}_{\tau_k}\} - \rho_{\tau_k}\{y_t - \mathbf{x}_t^\top(u^0) \boldsymbol{\eta}_{\tau_k}^0\}].$$

另记

$$\Delta(\boldsymbol{\eta}, u) = \sum_{k=1}^K \{\mathbf{x}_t^\top(u) \boldsymbol{\eta}_{\tau_k} - \mathbf{x}_t^\top(u^0) \boldsymbol{\eta}_{\tau_k}^0\}.$$

推论 2 的证明类似于定理 3, 我们在文中省略.

定理 1 的证明 通过一些简单直接计算可以得到

$$E|q_t(\boldsymbol{\eta}, u)| \leq 2E|\Delta(\boldsymbol{\eta}, u)| \leq 2E\|\mathbf{x}_t\| \sum_{k=1}^K (\|\boldsymbol{\eta}_{1,\tau_k}\| + \|\boldsymbol{\eta}_{2,\tau_k}\|) < +\infty.$$

对任意固定 $\boldsymbol{\eta}$ 应用大样本定理有

$$Q_n(\boldsymbol{\eta}, u) = \bar{Q}_n(\boldsymbol{\eta}, u) + o_p(1).$$

为了建立参数估计量 $\hat{\boldsymbol{\theta}}$ 的强一致性, 我们需要证明 $\mathcal{F} = \{q_t(\boldsymbol{\eta}, u) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ 满足 Glivenko 定理. 这里记 $\varepsilon_{t\tau_k}^0 = y_t - \mathbf{x}_t^\top(u^0)\boldsymbol{\eta}_{\tau_k}^0$. 我们可以将 $q_t(\boldsymbol{\eta}, u)$ 改写为

$$\begin{aligned} l_t(\boldsymbol{\eta}, u) = & - \sum_{k=1}^K [\rho_{\tau_k}\{y_t - \mathbf{x}_t^\top(u)\boldsymbol{\eta}_{\tau_k}\} - \rho_{\tau_k}(\varepsilon_{t\tau_k}^0)]I(u_t \leq u \wedge u^0) \\ & - \sum_{k=1}^K [\rho_{\tau_k}\{y_t - \mathbf{x}_t^\top(u)\boldsymbol{\eta}_{\tau_k}\} - \rho_{\tau_k}(\varepsilon_{t\tau_k}^0)]I(u < u_t \leq u^0) \\ & - \sum_{k=1}^K [\rho_{\tau_k}\{y_t - \mathbf{x}_t^\top(u)\boldsymbol{\eta}_{\tau_k}\} - \rho_{\tau_k}(\varepsilon_{t\tau_k}^0)]I(u^0 < u_t \leq u) \\ & - \sum_{k=1}^K [\rho_{\tau_k}\{y_t - \mathbf{x}_t^\top(u)\boldsymbol{\eta}_{\tau_k}\} - \rho_{\tau_k}(\varepsilon_{t\tau_k}^0)]I(u_t > u \vee u^0). \end{aligned}$$

根据文献 [16] 的证明可得 $\mathcal{F} = \{q_t(\boldsymbol{\eta}, u) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ 满足 Glivenko 定理. 因此, 通过应用最大值理论和文献 [36, 引理 B1] 可以得到

$$\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0.$$

从而完成了定理 1 的证明. □

定理 2 的证明 类似于文献 [25, 定理 2.3] 可以得到

$$n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^0) = O_p(1)$$

和

$$n^{1-2a}(\hat{u}_{\text{CTQAE}} - u^0) = O_p(1),$$

则在参数的紧致集 $\boldsymbol{\Theta} = \{(\boldsymbol{\eta}, u) : n^{1/2}\|\boldsymbol{\eta} - \boldsymbol{\eta}^0\| \leq C, n^{1-2a}|u - u^0| \leq C\}$ 上, 其中 C 是一个很大且固定的正数, 我们可以得到

$$D_n(\boldsymbol{\eta}^0 + n^{-1/2}\boldsymbol{\omega}_1, u^0 + n^{-1+2a}\omega_2) = D_{1n}(\boldsymbol{\omega}_1) + \bar{D}_{2n}(\boldsymbol{\omega}_1, \omega_2), \quad (\text{A.1})$$

其中 $\boldsymbol{\omega}_1 = n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^0)$, $\omega_2 = n^{1-2a}(\hat{u}_{\text{CTQAE}} - u^0)$, $\bar{D}_{2n}(\boldsymbol{\omega}_1, \omega_2) = D_{2n}(\boldsymbol{\eta}^0 + n^{-1/2}\boldsymbol{\omega}_1, u^0 + n^{-1+2a}\omega_2)$. 类似于文献 [25], 对 $D_{1n}(\boldsymbol{\omega}_1)$ 有

$$\sup_{\|\boldsymbol{\omega}_1\| \leq C} \left| D_{1n}(\boldsymbol{\omega}_1) + n^{-1/2}\boldsymbol{\omega}_1^\top \sum_{k=1}^K \sum_{t=p+1}^n \mathbf{x}_t(u^0)\psi_{\tau_k}(\varepsilon_{t\tau_k}^0) - \frac{1}{2}\boldsymbol{\omega}_1^\top \boldsymbol{\Omega}_1(u^0)\boldsymbol{\omega}_1 \right| = o_p(1), \quad (\text{A.2})$$

其中 C 是任意一个正常数,

$$\boldsymbol{\Omega}_1(u^0) = \sum_{k=1}^K \mathbb{E}[f\{\mathbf{x}_t^\top(u^0)\boldsymbol{\eta}_{\tau_k}^0\}\mathbf{x}_t(u^0)\mathbf{x}_t^\top(u^0)].$$

定义

$$D_n(\boldsymbol{\eta}, u) = Q_n(\boldsymbol{\eta}, u) - Q_n(\boldsymbol{\eta}^0, u^0),$$

$$\begin{aligned} D_{1n}(\boldsymbol{\omega}_1) &= D_n(\boldsymbol{\eta}^0 + n^{-1/2}\boldsymbol{\omega}_1, u^0) - Q_n(\boldsymbol{\eta}^0, u^0), \\ D_{2n}(\boldsymbol{\eta}, u) &= Q_n(\boldsymbol{\eta}, u) - Q_n(\boldsymbol{\eta}, u^0). \end{aligned}$$

类似于文献 [25], 对 $\bar{D}_{2T}(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ 有

$$\sup_{\|\boldsymbol{\omega}_1\| \leq C} |\bar{D}_{2n}(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) - \bar{D}_{2n}(0, \boldsymbol{\omega}_2)| = o_p(1). \quad (\text{A.3})$$

结合 (A.1)–(A.3) 可得

$$D_n(\boldsymbol{\eta}_0 + n^{-1/2}\boldsymbol{\omega}_1, u_0 + n^{-1+2a}\boldsymbol{\omega}_2) = \bar{D}_{1n}(\boldsymbol{\omega}_1) + \bar{D}_{2n}(0, \boldsymbol{\omega}_2) + o_p(1)$$

在 $\|\boldsymbol{\omega}_1\| \leq C$ 和 $\|\boldsymbol{\omega}_2\| \leq C$ 时成立, 其中

$$\bar{D}_{1n}(\boldsymbol{\omega}_1) = -n^{-1/2}\boldsymbol{\omega}_1^\top \sum_{k=1}^K \sum_{t=p+1}^n \mathbf{x}_t(u^0) \psi_{\tau_k}(\varepsilon_{t\tau_k}^0) + \frac{1}{2}\boldsymbol{\omega}_1^\top \boldsymbol{\Omega}_1(u^0)\boldsymbol{\omega}_1.$$

下面分别证明 (i) 和 (ii).

(i) 由于 $\bar{D}_{1n}(\boldsymbol{\omega}_1)$ 在 $\boldsymbol{\omega}_1$ 上是凸的, 由假设 (A5) 有

$$\boldsymbol{\omega}_1 = \frac{1}{\sqrt{n}}\boldsymbol{\Omega}_1^{-1}(u^0) \sum_{k=1}^K \sum_{t=p+1}^n \mathbf{x}_t(u^0) \psi_{\tau_k}(\varepsilon_{t\tau_k}^0) + o_p(1) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}(u^0)), \quad (\text{A.4})$$

其中

$$\begin{aligned} \boldsymbol{\Sigma}(u^0) &= \text{Var}\left(\frac{1}{\sqrt{n}}\boldsymbol{\Omega}_1^{-1}(u^0) \sum_{k=1}^K \sum_{t=p+1}^n \mathbf{x}_t(u^0) \psi_{\tau_k}(\varepsilon_{t\tau_k}^0)\right) \\ &= \boldsymbol{\Omega}_1^{-1}(u^0) \text{Var}\left(\frac{1}{\sqrt{n}} \sum_{k=1}^K \sum_{t=p+1}^n \mathbf{x}_t(u^0) \psi_{\tau_k}(\varepsilon_{t\tau_k}^0)\right) \boldsymbol{\Omega}_1^{-1}(u^0) \\ &= \boldsymbol{\Omega}_1^{-1}(u^0) \tilde{\boldsymbol{\Omega}}_0(u^0) \boldsymbol{\Omega}_1^{-1}(u^0), \end{aligned}$$

其中

$$\begin{aligned} \tilde{\boldsymbol{\Omega}}_0(u^0) &= \sum_{k=1}^K \sum_{k'=1}^K \{\min(\tau_k, \tau_{k'}) - \tau_k \tau_{k'}\} \text{E}\{\mathbf{x}_t(u^0) \mathbf{x}_t^\top(u^0)\}, \\ \boldsymbol{\Omega}_1(u^0) &= \sum_{k=1}^K \text{E}[f\{\mathbf{x}_t^\top(u^0) \boldsymbol{\eta}_{\tau_k}^0\} \mathbf{x}_t(u^0) \mathbf{x}_t^\top(u^0)]. \end{aligned}$$

(ii) 通过一些简单直接的计算, 可得

$$\begin{aligned} \bar{D}_{2n}(0, \boldsymbol{\omega}_2) &= \sum_{k=1}^K \sum_{t=p+1}^n [\rho_{\tau_k}\{\varepsilon_{t\tau_k}^0 + \mathbf{x}_t^\top(u^0) \boldsymbol{\eta}_{\tau_k}^0 - \mathbf{x}_t^\top(u^0 + n^{-1+2a}\boldsymbol{\omega}_2) \boldsymbol{\eta}_{\tau_k}^0\} - \rho_{\tau_k}(\varepsilon_{t\tau_k}^0)] \\ &= \sum_{k=1}^K \sum_{t=p+1}^n [\rho_{\tau_k}\{\varepsilon_{t\tau_k}^0 + \Delta \mathbf{x}_t^\top(\boldsymbol{\omega}_2) \boldsymbol{\eta}_{\tau_k}^0\} - \rho_{\tau_k}(\varepsilon_{t\tau_k}^0)]. \end{aligned}$$

根据 Knight 等式 $\rho_\tau(u - v) - \rho_\tau(u) = -v\psi_\tau(u) + \int_0^v \{I(u \leq s) - I(u \leq 0)\}ds$ 有

$$\begin{aligned}\bar{D}_{2n}(0, \omega_2) &= \sum_{k=1}^K \sum_{t=p+1}^n \psi_{\tau_k}(\varepsilon_{t\tau_k}^0) \Delta \mathbf{x}_t^\top(\omega_2) \delta_{\tau_k}^0 \\ &\quad + \sum_{k=1}^K \sum_{t=p+1}^n \int_0^{-\Delta \mathbf{x}_t^\top(\omega_2) \delta_{\tau_k}^0} \Delta \mathbf{x}_t(\omega_2) [I(\varepsilon_{t\tau_k}^0 \leq s) - I(\varepsilon_{t\tau_k}^0 \leq 0)] ds \\ &=: \bar{D}_{2n,1}(\omega_2) + \bar{D}_{2n,2}(\omega_2),\end{aligned}\tag{A.5}$$

其中 $\delta_{\tau_k}^0$ 定义为门限效应, 见假设 (A6).

对 $\bar{D}_{2n,1}(\omega_2)$, 参照文献 [26, 引理 A.11], 有

$$\bar{D}_{2n,1}(\omega_2) \Rightarrow B(\omega_2),\tag{A.6}$$

其中 $B(\omega_2)$ 为 Brown 运动, 其方差为

$$\begin{aligned}\mathbb{E}[B(1)^2] &= \sum_{t=p+1}^n n^{-2a} \sum_{k=1}^K \sum_{k'=1}^K \{\min(\tau_k, \tau_{k'}) - \tau_k \tau_{k'}\} \nu_{\tau_k}^\top \mathbb{E}[\mathbf{x}_t \mathbf{x}_t^\top \{I(u_t \leq u^0 + n^{-1+2a}\omega_2) - I(u_t \leq u^0)\}] \nu_{\tau_{k'}} \\ &= \sum_{t=p+1}^n n^{-1} \sum_{k=1}^K \sum_{k'=1}^K \{\min(\tau_k, \tau_{k'}) - \tau_k \tau_{k'}\} \nu_{\tau_k}^\top g(u^0) \mathbb{E}(\mathbf{x}_t \mathbf{x}_t^\top | u_t = u^0) \nu_{\tau_{k'}} \\ &\rightarrow \sum_{k=1}^K \sum_{k'=1}^K \{\min(\tau_k, \tau_{k'}) - \tau_k \tau_{k'}\} \nu_{\tau_k}^\top g(u^0) \mathbb{E}(\mathbf{x}_t \mathbf{x}_t^\top | u_t = u^0) \nu_{\tau_{k'}},\end{aligned}$$

其中 $g(\cdot)$ 是门限变量 u_t 的概率密度函数.

对 $\bar{D}_{2n,2}(\omega_2)$, 参见文献 [26, 引理 A.10] 中证明过程及条件 (A6) 有

$$\begin{aligned}\bar{D}_{2n,2}(\omega_2) &= \sum_{k=1}^K \sum_{t=p+1}^n \int_0^{-\Delta \mathbf{x}_t^\top(\omega_2) \delta_{\tau_k}^0} [F\{\mathbf{x}_t^\top(u^0) \boldsymbol{\eta}_{\tau_k}^0 - s \Delta \mathbf{x}_t^\top(\omega_2) \delta_{\tau_k}^0\} - F\{\mathbf{x}_t^\top(u^0) \boldsymbol{\eta}_{\tau_k}^0\}] ds + o_p(1) \\ &= \frac{1}{2} \sum_{k=1}^K \sum_{t=p+1}^n f\{\mathbf{x}_t^\top(u^0) \boldsymbol{\eta}_{\tau_k}^0\} \delta_{\tau_k}^{0\top} \Delta \mathbf{x}_t(\omega_2) \Delta \mathbf{x}_t^\top(\omega_2) \delta_{\tau_k}^0 + o_p(1) \\ &= \frac{1}{2} \sum_{k=1}^K \sum_{t=p+1}^n f\{\mathbf{x}_t^\top(u^0) \boldsymbol{\eta}_{\tau_k}^0\} \delta_{\tau_k}^{0\top} \mathbf{x}_t \mathbf{x}_t^\top \delta_{\tau_k}^0 \{I(u_t \leq u^0 + n^{-1+2a}\omega_2) - I(u_t \leq u^0)\}^2 + o_p(1) \\ &= \frac{1}{2} \sum_{k=1}^K \sum_{t=p+1}^n n^{-1} f\{\mathbf{x}_t^\top(u^0) \boldsymbol{\eta}_{\tau_k}^0\} \nu_{\tau_k}^\top \mathbf{x}_t \mathbf{x}_t^\top \nu_{\tau_k} g(u^0) |\omega_2| + o_p(1) \\ &\rightarrow \frac{1}{2} \mu |\omega_2|,\end{aligned}\tag{A.7}$$

其中

$$\mu = \sum_{k=1}^K \nu_{\tau_k}^\top \mathbb{E}\{f(\mathbf{x}_t^\top(u^0) \boldsymbol{\eta}_{\tau_k}^0) \mathbf{x}_t \mathbf{x}_t^\top\} \nu_{\tau_k} g(u^0).$$

结合 (A.5)–(A.7), 有

$$\bar{D}_{2n}(0, \omega_2) \Rightarrow \frac{1}{2} \mu |\omega_2| - \sqrt{\lambda} B(\omega_2).$$

通过连续映射定理, 对变量 $\omega_2 = (\lambda/\mu^2)r$ 进行映射, 有

$$\begin{aligned}\omega_2 &\rightarrow \operatorname{argmax}_{-\infty < \omega_2 < \infty} \left\{ -\frac{1}{2}\mu|\omega_2| + \sqrt{\lambda}B(\omega_2) \right\} \\ &= \frac{\lambda}{\mu^2} \operatorname{argmax}_{-\infty < r < \infty} \left\{ -\frac{1}{2}\mu \left| \frac{\lambda}{\mu^2}r \right| + \sqrt{\lambda}B\left(\frac{\lambda}{\mu^2}r\right) \right\} \\ &= \frac{\lambda}{\mu^2} \operatorname{argmax}_{-\infty < r < \infty} \left\{ B(r) - \frac{1}{2}|r| \right\}.\end{aligned}\quad (\text{A.8})$$

至此完成了定理 2 的证明. $\square$

定理 3 的证明 由定义 $\omega_1 = n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^0)$ 和 $\omega_2 = n^{1-2a}(\hat{u}_{\text{CTQAE}} - u^0)$, 有

$$\text{LR}_n(u^0) = \bar{D}_{2n}(\omega_1, 0) - \bar{D}_{2n}(\omega_1, \omega_2). \quad (\text{A.9})$$

根据定理 2 和文献 [25], 通过映射变量 $R = (\lambda/\mu^2)r$ 有

$$\begin{aligned}\text{LR}_n(u^0) &= -\bar{D}_{2n}(0, \omega_2) + o_p(1) \\ &\rightarrow \sup_R \left\{ -\frac{1}{2}\mu|R| + \sqrt{\lambda}B(R) \right\} \\ &= \sup_r \left\{ -\frac{1}{2}\mu \left| \frac{\lambda}{\mu^2}r \right| + \sqrt{\lambda}B\left(\frac{\lambda}{\mu^2}r\right) \right\} \\ &= \frac{\lambda}{2\mu} \sup_r \{2B(r) - |r|\}.\end{aligned}$$

至此完成了定理 3 的证明. $\square$

定理 4 的证明 先令 $\omega_{1n\tau_k} = n^{1/2}(\hat{\boldsymbol{\eta}}_{\tau_k} - \boldsymbol{\eta}_{\tau_k}^0)$ 和 $\omega_{2n\tau_k} = n^{1-2a}(\hat{u}_{\tau_k} - u_{\tau_k}^0)$, 类似于定理 2 的证明中的 (A.7), 在固定分位数水平 τ_k 下有

$$\begin{aligned}\bar{D}_{2n\tau_k, 2}(\omega_{2\tau_k}) &= \sum_{t=p+1}^n \int_0^{-\Delta \mathbf{x}_t^\top(\omega_2)\boldsymbol{\delta}_{\tau_k}^0} [F\{\mathbf{x}_t^\top(u_{\tau_k}^0)\boldsymbol{\eta}_{\tau_k}^0 - s\Delta \mathbf{x}_t^\top(\omega_2)\boldsymbol{\delta}_{\tau_k}^0\} - F\{\mathbf{x}_t^\top(u_{\tau_k}^0)\boldsymbol{\eta}_{\tau_k}^0\}]ds + o_p(1) \\ &= \frac{1}{2} \sum_{t=p+1}^n n^{-1} f\{\mathbf{x}_t^\top(u_{\tau_k}^0)\boldsymbol{\eta}_{\tau_k}^0\} \nu_{\tau_k}^\top \mathbf{x}_t \mathbf{x}_t^\top \nu_{\tau_k} g(u_{\tau_k}^0) |\omega_{2\tau_k}| + o_p(1) \\ &\rightarrow \frac{1}{2} \mu_{\tau_k} |\omega_{2\tau_k}|,\end{aligned}$$

其中

$$\mu_{\tau_k} = \sum_{k=1}^K \nu_{\tau_k}^\top \mathbb{E}\{f(\mathbf{x}_t^\top(u_{\tau_k}^0)\boldsymbol{\eta}_{\tau_k}^0) \mathbf{x}_t \mathbf{x}_t^\top\} \nu_{\tau_k} g(u_{\tau_k}^0).$$

由此可得

$$\bar{D}_{2n\tau_k}(0, \omega_{2\tau_k}) \Rightarrow \frac{1}{2} \mu_{\tau_k} |\omega_{2\tau_k}| - \sqrt{\lambda_{\tau_k}} B(\omega_{2\tau_k}).$$

因此, 通过连续映射定理和变量替换及假设 (A7), 有

$$n^{1-2a}(\hat{u}_{\text{TQAE}} - u^0) \xrightarrow{d} \zeta \operatorname{argmax}_{-\infty < r < \infty} \left\{ B(r) - \frac{1}{2}|r| \right\}.$$

定理 4 证明完毕. $\square$

Combined change point estimation in threshold quantile autoregressive models

Liwen Zhang, Dongpo Cheng, Wenjun Xue & Tinggan Yang

Abstract The financial market is a typical complex system, and we can find its expressions in the complexity of nonlinearity and heterogeneity of the market. In order to mine for more meaningful information, the threshold quantile autoregressive model regarded as an effective method has become very popular in econometrics and statistics. We propose two new methods for estimating the change point, which are the composite quantile regression estimator and the quantile average estimator. By combining quantile regression over multiple quantiles, these proposed methods improve estimation accuracy and efficiency of change points under the assumption of constant thresholds. In some actual data analysis, we find that the change points in the threshold quantile autoregressive model are very close on different quantile levels. Due to the commonality of the change points, we first obtain a more efficient change point estimator by minimizing the combined quantile objective function across different quantiles. We further derive the asymptotic properties and build the confidence interval based on likelihood ratio and bootstrap methods of the composite change point estimator. Second, we propose an alternative estimator that is the quantile average estimator by averaging the quantile-specific slope estimators involved in the estimation procedure at different quantiles. In general, the results of numerical simulation show that the proposed methods have higher estimation efficiency compared with the traditional ordinary least squares estimator and the quantile threshold estimator under limited sample conditions. Finally, we analyze the empirical application performance of the proposed method with the A-share index of Shanghai Stock Exchange from 2005 to 2014.

Keywords efficiency, change points, threshold quantile autoregression, composite quantile, quantile average estimator

MSC(2020) 62F10, 62J99

doi: 10.1360/N012019-00167