

基于面板数据方差变点的股票收盘价分析

赵军辉, 董翠玲*

新疆师范大学数学科学学院, 新疆 乌鲁木齐

收稿日期: 2023年9月29日; 录用日期: 2023年10月30日; 发布日期: 2023年11月6日

摘要

文章应用含有调节参数的CUSUM (Cumulative Sum)型估计量对面板数据中方差的共同变点进行估计, 并对在美国上市且目前仍在市的教育机构的股票收盘价进行实证分析, 分析结果表明了基于含有调节参数的CUSUM型估计量寻找面板数据中方差共同变点的方法是有效的。在实证分析的基础上, 把时间序列中基于数据驱动的调节参数的选取方法推广到面板数据中, 使得面板数据中CUSUM型方差共同变点估计更可靠。

关键词

面板数据, 方差共同变点, CUSUM, 调节参数, 股票收盘价

Analysis of Stock Closing Price Based on Variance Change Point of Panel Data

Junhui Zhao, Cuiling Dong*

School of Mathematical Sciences, Xinjiang Normal University, Urumqi Xinjiang

Received: Sep. 29th, 2023; accepted: Oct. 30th, 2023; published: Nov. 6th, 2023

Abstract

In this paper, the CUSUM (Cumulative Sum)-type estimator with turning parameter is used to estimate the common change point of the variance of panel data, and the empirical analysis is carried out on the closing price of the stock of the educational institutions listed in the United States which still in the market. The analysis results show that the method of estimating the common change point of the variance based on the CUSUM-type estimator with turning parameter is effective. Based on the empirical analysis, the selection method of turning parameters based on da-

*通讯作者。

ta-driven in time series is extended to panel data, which makes the estimation of CUSUM-type variance common change point in panel data more reliable.

Keywords

Panel Data, Variance Common Change Point, CUSUM, Turning Parameters, Closing Price of Stock

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



1. 引言

面板数据的变点分析问题始于Joseph和Wolfson (1992, 1993) [1] [2]提出的随机变点模型。经过长期的研究发展, 广泛应用于经济学、金融学、生命科学、医学、气象学等领域[3] [4] [5] [6]。股票即属于经济学又属于金融学, 它的波动通常伴随着重大政策事件的发生, 国内外已有许多学者对股票进行了研究。刘建民和刘星(2006) [7]通过建立线性模型, 研究了我国深圳证券交易所的交易数据, 分析了公司总市值、账面市值比、债务权益比和销售价格比对股票横截面平均收益率的影响, 结果表明只有上市公司总市值与股票平均收益率具有显著的负相关性。Rapach和Wohar (2006) [8]针对美国市场季度的股票收益率, 使用SupF统计量等方法研究预测收益率模型的结构稳定性, 并且估计出不同政治制度影响下的结构变点。Tetlock (2007) [9]研究了与股票价格有关的新闻内容对投资者情绪的影响, 结果表明新闻事件背后所指向的政治、经济、社会等各个方面的变动都会影响投资者的情绪, 从而导致股票价格发生变动。黄蓉(2012) [10]通过建立自回归模型, 研究货币供应量对股票价格指数的影响, 结果表明股价涨幅的变化主要取决于人们对未来股市的预期。胡俊迎和谭常春等(2017) [11]采用面板数据中均值共同变点的检测和估计方法, 研究了股市收益率的均值共同变点, 并分析了变点产生的原因。张悟移和李杰(2019) [12]通过建立VAR模型、格兰杰因果检验和脉冲响应函数, 研究百度关注度指数与股票价格之间的关系, 结果表明投资者的关注行为与股票收盘价之间存在一定的影响。本文通过含有调节参数的CUSUM型估计量估计面板数据中方差共同变点的方法, 对新东方、好未来、51talk、博实乐、达内科技、ATA公司这六支在美国上市且目前仍在市的教育机构股票收盘价的方差共同变点进行估计, 分析影响方差共同变点的原因, 并把时间序列中基于数据驱动的调节参数的选取方法[13]推广到面板数据中, 使得面板数据中CUSUM型方差共同变点估计更可靠。

2. 面板数据中方差共同变点估计的介绍

面板数据 Y_{it} ($i=1,2,\dots,N;t=1,2,\dots,T$) 表示第 i 个截面个体在 t 时刻的观测值, 是一个 $N \times T$ 矩阵, 其表示形式如(1)所示:

$$\begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1T} \\ Y_{21} & Y_{22} & \cdots & Y_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{N1} & Y_{N2} & \cdots & Y_{NT} \end{pmatrix} \quad (1)$$

考虑面板数据中方差的共同多变点模型:

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

其中: 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, 满足 $E(\eta_{it})=0$, $Var(\eta_{it})=1$, $\eta_{it} = \sum_{j=0}^{\infty} b_{ij}\varepsilon_{i,t-j}$, $\varepsilon_{it} \sim i.i.d(0,1)$, $\sum_{j=1}^{\infty} j|b_{ij}| < \infty$, $Var(\eta_{it}) = \sum_{j=0}^{\infty} b_{ij}^2 = 1$, 且存在正数 $M > 0$, 使得 $E(\varepsilon_{it}^4) \leq M$ 。若 $\sigma_{ij} \neq \sigma_{i,j+1}$, $j=1,2,\dots,m$, 则称未知时刻 k_1, k_2, \dots, k_m 为 m 个变点。

记 $\tau_1 = \frac{k_1}{T}$, $0 < \tau_1 < 1$, 则变点 τ_1 的CUSUM型估计量为

$$\hat{\tau} = \frac{\hat{k}}{T} = \frac{1}{T} \arg \max_{1 \leq k < T} |V_{NT}(k)| \quad (3)$$

式(3)中

$$V_{NT}(k) = \left(\frac{k(T-k)}{T^2} \right)^{1-\gamma} \sum_{i=1}^N \left(\frac{1}{k} \sum_{t=1}^k \tilde{Y}_{it}^2 - \frac{1}{T-k} \sum_{t=k+1}^T \tilde{Y}_{it}^2 \right) \quad (4)$$

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_1 在靠近序列端点时估计的有效性, $\gamma=0$ 表示无调节参数。 \tilde{Y}_{it} 为面板数据中心化后的结果, 即 $\tilde{Y}_{it} = Y_{it} - \bar{Y}_i$, $\bar{Y}_i = \frac{1}{T} \sum_{t=1}^T Y_{it}$ 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。

3. 面板数据中基于数据驱动的调节参数的选取方法

Step 1: 设 $Y_{i1}, Y_{i2}, \dots, Y_{i,k_1}, Y_{i,k_1+1}, \dots, Y_{iT}$ 是第 i 个截面个体带有变点 k_1 的样本观测值, 记 $A = \{\gamma_1, \gamma_2, \dots, \gamma_L\}$ 为调节参数 γ 的可能取值集, 给定 γ_l , 得到一个变点估计量, 记为 \hat{k}_{γ_l} , 有

$$\hat{k}_{\gamma_l} = \arg \max_{1 \leq k < T} |V_{NT}(k)|, 1 \leq l \leq L \quad (5)$$

Step 2: 计算 $\hat{k}_{\gamma_1}, \hat{k}_{\gamma_2}, \dots, \hat{k}_{\gamma_L}$ 的均值 \hat{k} :

$$\hat{k} = \frac{1}{L} \sum_{l=1}^L \hat{k}_{\gamma_l} \quad (6)$$

Step 3: 定义

$$\gamma_s = \arg \min_{\gamma_l} \left| \hat{k}_{\gamma_l} - \hat{k} \right| \quad (7)$$

即为面板数据中基于数据驱动方法选取的合适的调节参数。

定理1: 设 $Y_{i1}, Y_{i2}, \dots, Y_{i,k_1}, Y_{i,k_1+1}, \dots, Y_{iT}$ 是取自模型(2)的样本观测值, 其中 k_1 为变点, 记 $A = \{\gamma_1, \gamma_2, \dots, \gamma_L\}$ 为调节参数 γ 的可能取值集, 则至少存在一个 $\gamma^* \in A$, 使得

$$\left| \hat{k}_{\gamma_s} - k_1 \right| \leq \left| \hat{k}_{\gamma^*} - k_1 \right| \quad (8)$$

证明: 设变点 $k_1 \geq \hat{k}$, 则

$$\left| \hat{k}_{\gamma_s} - k_1 \right| = \left| \hat{k}_{\gamma_s} - \hat{k} + \hat{k} - k_1 \right| \leq \left| \hat{k}_{\gamma_s} - \hat{k} \right| + \left| \hat{k} - k_1 \right|$$

对任意的 $\gamma_l \in A, 1 \leq l \leq L$, 由式(7)中 γ_s 的定义知, $|\hat{k}_{\gamma_s} - \bar{k}| \leq |\hat{k}_{\gamma_l} - \bar{k}|, 1 \leq l \leq L$, 故有

$$|\hat{k}_{\gamma_s} - k_1| \leq |\hat{k}_{\gamma_l} - \bar{k}| + |\bar{k} - k_1|, 1 \leq l \leq L$$

又因为在 $A = \{\gamma_1, \gamma_2, \dots, \gamma_L\}$ 中至少存在一个 $\gamma^* \neq \gamma_s$, 使得 $\hat{k}_{\gamma^*} \leq \min(\hat{k}_{\gamma_s}, \bar{k})$, 所以

$$|\hat{k}_{\gamma_s} - k_1| \leq |\hat{k}_{\gamma^*} - \bar{k}| + |\bar{k} - k_1| \leq \bar{k} - \hat{k}_{\gamma^*} + k_1 - \bar{k} = |\hat{k}_{\gamma^*} - k_1| \quad (9)$$

对于 $k_1 < \bar{k}$ 的情形, 同理可证得式(9)。

4. 实证分析

下面将时间序列中基于数据驱动的调节参数的选取方法推广到面板数据中, 并通过实证分析验证面板数据中基于数据驱动选择调节参数 γ_s 的可靠性。在美上市的教育机构公司曾有很多, 但目前只有新东方、好未来、51talk、博实乐、达内科技、ATA公司这六家, 其他的教育机构公司已经退市或转市, 选取2017年6月~2023年8月这六支教育机构股票的月收盘价数据(数据来源于<https://cn.investing.com/currencies/>)作为研究对象, 共有6个不同的截面个体, 每个截面个体含有75个历史数据, 即 $N = 6, T = 75$ 。采用式(3)中的CUSUM型估计量估计方差共同变点的位置, 调节参数 γ 的取值集 $A = \{0.1, 0.25, 0.5, 0.75, 0.9\}$, 估计的第一个变点位置都是32, 对应的实际时间是2020年1月, 所得结果见表1所列。由于第一个变点位置都是32, 故使用面板数据中基于数据驱动选取调节参数的效果相同, 此时最优调节参数 γ_s 取0.1, 0.25, 0.5, 0.75, 0.9中任一个都是合适的。出现这一方差变点的主要原因是受疫情影响, 2020年1月27日, 教育部正式下发《教育部关于2020年春季学期延期开学的通知》, 根据相关要求, “停课不停学”成为各地教育行政部门和大中小学的主要防控措施, 在线教育成为了“刚需”, 一大波“免费直播课”迅速上线, 很多在线教育机构因此受益, 流量暴涨, 股票出现了很大的波动, 这也是造成新东方、好未来、51talk、博实乐、达内科技、ATA公司这六支教育机构股票的收盘价出现方差变点的主要原因。

Table 1. The change point position estimated by the CUSUM estimator under different adjustment parameters γ and the actual time corresponding to the first change point

表1. 不同的调节参数 γ 下CUSUM型估计量估计的变点位置及第一个变点对应的实际时间

变点对应的实际时间	调节参数 γ 的取值	变点位置 \hat{k}
2020年1月	0.1	32
2020年1月	0.25	32
2020年1月	0.5	32
2020年1月	0.75	32
2020年1月	0.9	32

估计第一个变点位置后, 结合二元分割法, 将2017年6月~2023年8月的月收盘价数据一分为二, 采用上述方法对2020年2月~2023年8月的收盘价数据进行变点估计, 调节参数 γ 的取值集 $A = \{0.1, 0.25, 0.5, 0.75, 0.9\}$, 估计的第二个变点位置分别为16, 16, 16, 16, 14, 所得结果见表2所列。此时变点位置不同, 应用面板数据中基于数据驱动的调节参数的选取方法中Step 2, 得到 $\bar{k} = 15.6$, 结合Step 3, 适用于这一过程的最优调节参数 γ_s 取0.1, 0.25, 0.5, 0.75中任一个都是合适的, 变点位置都是16, 对应的实际时间是2021年5月。出现这一方差变点的主要原因是由于在2021年5月召开的中央会议中, 强

调了中国义务教育阶段学生压力过大的问题, 并且突出强调了校外培训机构的问题, 因此国家要取缔某些不合规的校外培训机构, 为中国义务教育阶段的学生减负。新东方、好未来、51talk、博实乐、达内科技、ATA公司正是以线下培训、在线培训为主要业务, 因此股票受到了影响, 出现了很大的波动。

Table 2. The change point position estimated by the CUSUM estimator under different adjustment parameters γ and the actual time corresponding to the second change point

表2. 不同的调节参数 γ 下CUSUM型估计量估计的变点位置及第二个变点对应的实际时间

变点对应的实际时间	调节参数 γ 的取值	变点位置 \hat{k}
2021年5月	0.1	16
2021年5月	0.25	16
2021年5月	0.5	16
2021年5月	0.75	16
2021年3月	0.9	14

采用上述方法对2021年5月~2023年8月的月收盘价数据进行变点估计, 调节参数 γ 的取值集 $A = \{0.1, 0.25, 0.5, 0.75, 0.9\}$, 估计的第三个变点位置分别为3, 3, 3, 2, 2, 所得结果见表3所列。此时变点位置不同, 应用面板数据中基于数据驱动的调节参数的选取方法中Step 2, 得到 $\hat{k} = 2.6$, 结合Step 3, 适用于这一过程的最优调节参数 γ_s 取0.1, 0.25, 0.5中任一个都是合适的, 变点位置都是3, 对应的实际时间是2021年7月。而出现这一方差变点的主要原因是由于2021年7月《关于进一步减轻义务教育阶段学生作业负担和校外培训负担的意见》(简称“双减”意见)国家“双减”政策出台, 精准打击各类教培行业, 教培行业股票均出现大幅下跌, 新东方、好未来、51talk、博实乐、达内科技、ATA公司这六支教育机构股票的收盘价出现很大的波动, 这也是在美国上市的很多其它教育机构选择退市或转市的主要原因。

Table 3. The change point position estimated by the CUSUM estimator under different adjustment parameters γ and the actual time corresponding to the third change point

表3. 不同的调节参数 γ 下CUSUM型估计量估计的变点位置及第三个变点对应的实际时间

变点对应的实际时间	调节参数 γ 的取值	变点位置 \hat{k}
2021年7月	0.1	3
2021年7月	0.25	3
2021年7月	0.5	3
2021年6月	0.75	2
2021年6月	0.9	2

综合分析, 适用于以上过程的最优调节参数 γ_s 取0.1, 0.25, 0.5中任一个都是合适的, 并由此得到了六支教育股票收盘价的方差变点(2020年1月, 2021年5月, 2021年7月)如图1所示, 从图中可以看出基于调节参数的CUSUM型的变点估计量符合实际情况。

5. 结论

本文采用含有调节参数的 CUSUM 型估计量估计面板数据中方差共同变点的方法, 对新东方、好未来、51talk、博实乐、达内科技、ATA 公司这六家在美国上市且目前仍在市的教育机构股票的收盘价进

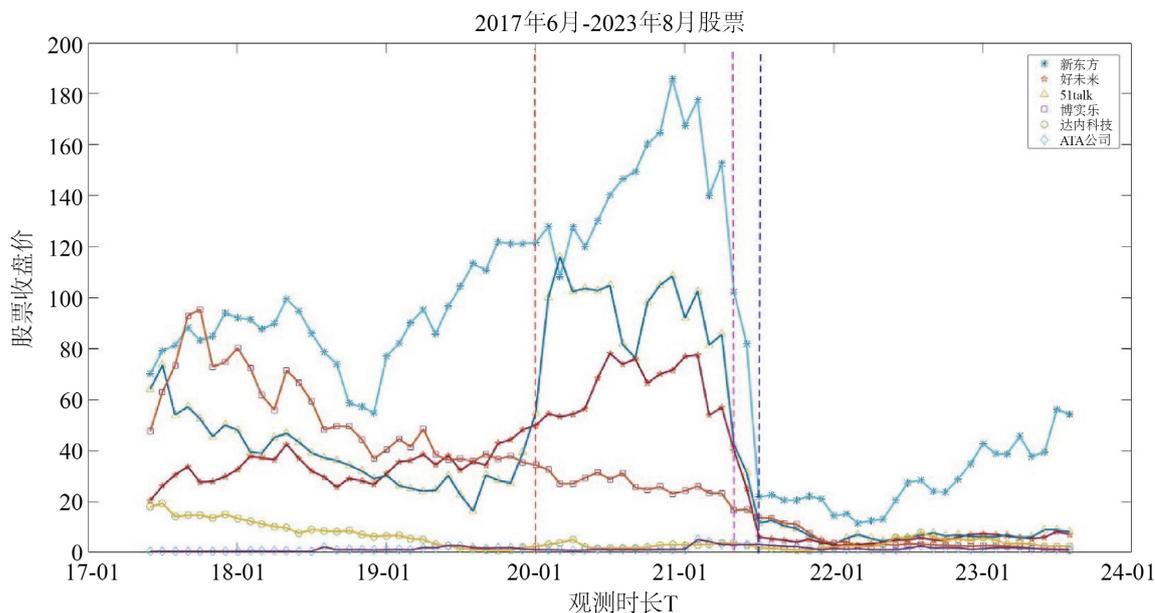


Figure 1. Monthly data on stock closing prices of six educational institutions
图1. 六支教育机构股票收盘价月数据

行实证分析, 估计方差共同变点的位置, 分析了变点产生的原因, 在实证分析的基础上, 把时间序列中基于数据驱动的调节参数的选取方法推广到面板数据中, 使得面板数据中 CUSUM 型方差共同变点估计更可靠, 丰富了面板数据中方差共同变点估计在实际中的应用。

基金项目

国家自然科学基金项目(11801488), 新疆师范大学重点实验室项目(XJNUSYS082018A01), 新疆维吾尔自治区自然科学基金项目(2023D01A37, 2022D01A219)。

参考文献

- [1] Joseph, L. and Wolfson, D.B. (1992) Estimation in Multi-Path Change-Point Problems. *Communications in Statistics—Theory and Methods*, **21**, 897-913. <https://doi.org/10.1080/03610929208830822>
- [2] Joseph, L. and Wolfson, D.B. (1993) Maximum Likelihood Estimation in the Multi-Path Change-Point Problem. *Annals of the Institute of Statistical Mathematics*, **45**, 511-530. <https://doi.org/10.1007/BF00773352>
- [3] Skates, S.J., Pauler, D.K. and Jacobs, I.J. (2001) Screening Based on the Risk of Cancer Calculation from Bayesian Hierarchical Change-point and Mixture Models of Longitudinal Markers. *Journal of the American Statistical Association*, **96**, 429-439. <https://doi.org/10.1198/016214501753168145>
- [4] Atak, A., Linton, O. and Xiao, Z. (2011) A Semiparametric Panel Model for Unbalanced Data with an Application to Climate Change in the United Kingdom. *Journal of Econometrics*, **164**, 92-115. <https://doi.org/10.1016/j.jeconom.2011.02.008>
- [5] Li, D., Qian, J. and Su, L. (2016) Panel Data Models with Interactive Fixed Effects and Multiple Structural Breaks. *Journal of the American Statistical Association*, **111**, 1804-1819. <https://doi.org/10.1080/01621459.2015.1119696>
- [6] 任燕燕, 王纬, 严晓东. 具有组群异方差结构的面板数据模型及其应用研究[J]. *统计研究*, 2021, 38(11): 141-149.
- [7] 刘建民, 刘星. 我国股市股票收益率的横截面数据分析[J]. *统计与决策*, 2006(2): 97-100.
- [8] Rapach, D.E. and Wohar, M.E. (2006) Structural Breaks and Predictive Regression Models of Aggregate U.S. Stock Returns. *Journal of Financial Econometrics*, **4**, 238-274. <https://doi.org/10.1093/jfinec/nbj008>
- [9] Tetlock, P.C. (2007) Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, **62**, 1139-1168. <https://doi.org/10.1111/j.1540-6261.2007.01232.x>

- [10] 黄蓉. 自回归模型货币供应量对股票价格指数影响分析[J]. 统计与决策, 2012(12): 150-152.
- [11] 胡俊迎, 谭常春, 张雪莲. 基于面板数据均值变点的股市收益率分析[J]. 合肥工业大学学报(自然科学版), 2017, 40(11): 1577-1580.
- [12] 张悟移, 李杰. 百度关注度指数与股票价格关系研究[J]. 西南师范大学学报(自然科学版), 2019, 44(2): 75-83.
- [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(7): 920-928.