相依样本下半参数变系数期望分位数风险度量模型统计推断

王子剑¹, 周勇^{2*}, 曾凡平³

1. 上海财经大学统计与管理学院, 上海 200433;

2. 统计与数据科学前沿理论及应用教育部重点实验室, 华东师范大学经管学部统计交叉科学研究院, 上海 200062;

3. 广西财经学院信息与统计学院, 南宁 530003

E-mail: wangzijian722@163.com, yzhou@amss.ac.cn, fpzeng@gxu.edu.cn

收稿日期: 2019-03-15; 接受日期: 2019-11-04; 网络出版日期: 2021-04-07; * 通信作者

国家自然科学基金 (批准号: 71331006, 71931004 和 91546202) 资助项目

摘要 本文在 α -混合序列假设下, 基于半参数变系数模型研究条件期望分位数风险价值 (expectile-based value at risk, EVaR) 的风险度量. 此模型不仅考虑了风险因素的影响, 还可以动态描述风险影响及交互效应. 同时, EVaR 比经典的风险在险价值 (quantile-based value at risk, QVaR) 具有更直观、更易于计算的良好性质, 而且对于资产分布的尾部损失更加敏感, 在度量极端风险情形下, 相对于 QVaR 更为有效和方便. 本文采用三阶段估计的方法, 分别对变系数部分和常系数部分的参数进行估计, 并且给出 3 个阶段中每个估计的相合性和渐近正态性. 为了节省计算时间, 提高计算效率, 本文采用一步估计的算法, 减少迭代所需的时间. 由于时间序列样本是非独立样本, 建立这些统计量的大样本性质时带来了更大的困难. 有别于独立同分布的观察数据, 本文利用大小块分割方法发展 α -混合序列的极限理论, 获得了基于金融时间序列数据建立的模型参数和非参数估计的统计渐近性质. 在数值模拟中, 本文给出 3 个模型假设下变系数曲线估计和常系数估计的结果, 无论是估计的精确度还是估计的稳健性, 模拟结果都表明本文所提出的估计方法有优良的性质. 实例则展示了本文所提出模型在上证指数的实际应用.

关键词 α -混合 半参数 变系数 期望分位数 在险价值 三阶段估计

MSC (2020) 主题分类 62G05, 91G70

1 引言

风险度量建模一直是统计学和金融计量学中重要的研究议题和前沿问题. 获得诺贝尔奖的美国经济学家 Markowitz^[1] 在投资组合理论中, 首次提出了收益与风险的度量理论, 即期望代表收益与方差代表风险的度量方法, 从而开创了风险的量化时代. 随后以 Black 和 Scholes^[2] 的资产定价理论为标

英文引用格式: Wang Z J, Zhou Y, Zeng F P. Semiparametric varying-coefficient expectile model for estimating value at risk on dependent samples (in Chinese). Sci Sin Math, 2021, 51: 1377–1406, doi: 10.1360/N012019-00079

志, 风险度量推动了金融计量学与风险管理的发展. 此外发展出更多风险管理的方法, 包括著名的一般均衡理论^[3,4] 和保费均衡原则^[5] 等.

根据风险度量的侧重点不同, 可以将常用的风险度量分为两大类: 基于矩的风险度量和基于分位点的风险度量. 有关矩的风险度量主要有方差 (variance)、半方差 (semi-variance)、绝对离差 (absolute deviation)、半绝对离差 (semi-absolute deviation)、低位部分矩 (lower partial moments, LPM) 和 G-均值等. 有关矩的风险度量, 主要采用概率统计中表示离散程度的矩 (moment) 来度量金融资产或投资组合的风险, 波动性越强, 风险越大.

另一类则是以风险价值 (value at risk, VaR) 和期望损益 (expected shortfall, ES) 为代表的分位数 (quantile) 风险度量. 风险价值和期望损益是资产组合常用的两个风险度量工具, 其中 VaR 是被如巴塞尔委员会等众多国外的金融机构所认可和提倡的风险度量方法. VaR 定义为在一定置信水平下, 在持有期内资产的最大可能损失, 本质上是资产损益分布的下端分位数. 但基于分位数的 VaR 无法有效度量出极端损失的情形, 例如, Danielsson 和 De Vries^[6] 得到结论: 无论取多高的置信度, 仍会出现难以预料的损失可能, 即 VaR 度量对损失分布尾部不敏感. Yamai 和 Yoshida^[7] 也指出, 若组合收益是非正态分布的, 则更高置信水平的 VaR 值也无法解决其非次可加性问题, 这是因为 VaR 仅依赖于尾部概率大小而不依赖于实际数值. 虽然 VaR 综合考虑了风险的所有可能来源, 易于在不同层面的参与者间进行交流, 但基于分位数的 VaR 度量没有一致性 (即次可加性), 并且对于尾部损失的测量不充分, 导致其对极端的尾部损失不敏感. 可以给出例子说明, 对两个不同分布的资产收益, 基于分位数的 VaR 可能计算出相同的数值, 即对损益分布尾部不敏感. 对具有灾难性的极端损失值的资产, 经常无法通过 VaR 衡量出它的风险大小, 而这样的灾难性损失, 恰恰是风险管理者所关心的.

为了克服 VaR 度量不具有一致性和尾部风险不敏感性的困难, 一种替代的风险度量是由 Acerbi 等^[8] 提出的 ES 度量, 即大于 VaR 的资产损益的条件均值. Chen^[9] 对 ES 进行了研究, 用核估计的方法给出了它的非参数估计. Cai 和 Wang^[10] 用加权双核局部线性估计方法, 得到了条件 VaR 和 ES 的估计. Frey 和 McNeil^[11] 利用 Bernoulli 混合模型描述损益的尾部特征估计 VaR 和 ES. Inui 等^[12] 采用外推法对 ES 进行了估计. Harmantzis 等^[13] 利用广义 Pareto 分布和稳态分布估计 ES. 在实证应用中也有一些研究, 例如, Pajor 和 Osiewalski^[14] 研究了一个大型投资组合的 Bayes VaR 和 ES, 比较了单变量与多变量间实证结果的差异. Chen 等^[15] 用 Bayes 半参数模型预测了金融市场的 ES, 得到了相比其他方法更好的效果. 但是 ES 的不足之处是, 不能进行有效的回测, 即没有可引出性^[16]; 而且 ES 依赖于 VaR 的计算数值, 还需要估计条件期望, 在计算上较复杂, 同时在风险管理中不易于计算风险因素的贡献度.

为了弥补 VaR 和 ES 的不足, Kuan 等^[17] 提出了基于期望分位数 (expectile) 的风险度量, 即期望分位数风险价值 (EVaR). EVaR 是金融中的一种重要的风险度量, 其意义同目前流行的基于分位数的 QVaR 相似. EVaR 通过最小化非对称加权均方误差来计算, 由于二次损失函数使 EVaR 对极端值很敏感, 因此能更好地衡量出极端损失, 从而有效地测量风险. 期望分位数最早由 Aigner 等^[18] 提出, 并由 Newey 和 Powell^[19] 进一步发展. 他们指出了期望分位数相对于 QVaR 的 3 个优势: 二次损失函数可微带来的计算优势; 在估计协方差矩阵时不需要对条件密度函数进行估计, 因此不会引入额外偏差; 对于尾部极端值敏感, 可以度量极端损失. 此外期望分位数同时具有一致性和可引出性, 故 EVaR 成为一个比 QVaR 更加有效且实用的风险度量. 本文研究基于条件 EVaR 的半参数变系数模型, 扩展了 Xie 等^[20] 的变系数期望分位数模型. 即在我们的研究模型中除了变系数部分外, 还考虑了参数部分, 使模型设定更加广泛.

在金融风险度量中, 人们更多关注的是资产收益率和时间的滞后性, 通常观察到的数据并不是独

立样本, 而是具有相依性的金融时间序列数据. 在相依性的研究中比较流行的假设是混合相依. 混合相依是指变量之间的相依关系是随着时间或空间的距离衰退的, 即当随机变量的指标的差趋于无穷时, 随机变量是渐近独立的. 假设 $\{X_n, n \geq 1\}$ 是同分布且定义在同一概率空间 $(\Omega, \mathcal{F}, \mathcal{P})$ 上的随机变量序列, 设 $\mathcal{A}$ 和 $\mathcal{B}$ 是 $\mathcal{F}$ 的两个子 σ -代数, 设

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |P(AB) - P(A)P(B)|,$$

令 $\mathcal{F}_a^b = \sigma(X_i, a \leq i \leq b, i \in \mathbb{Z})$. 如果

$$\alpha(n) = \sup_{k \in \mathbb{N}} \alpha(\mathcal{F}_{-\infty}^k, \mathcal{F}_{k+n}^{\infty}) \rightarrow 0, \quad n \rightarrow \infty,$$

则称序列 $\{X_n, n \geq 1\}$ 是 α -混合或强混合的序列^[21]. Cai 等^[22] 认为具有 α -混合性质的序列能很好地刻画相依关系下的金融时间序列数据. 因此, α -混合的数据假设是合理且重要的. 本文在 α -混合相依假设下展开对 EVaR 风险度量的研究.

另一方面, 半参数变系数模型拟合相依时间序列数据, 能更好地确定风险的影响因素, 对风险做出更合理的度量, 从而指导管理者的投资决策等经济行为. Zhou 和 Liang^[23] 认为半参数变系数时间序列模型使得拟合的数据类型更加广泛, 更加符合实际. 响应变量与解释变量间的关系不仅有非参数的部分, 即风险因素有动态的影响, 而且还有参数假设的部分, 即长期的固定影响. 而且对于金融时间序列, 非参数部分假设为变系数也是很自然的. 基于半参数变系数时间序列模型, 本文给出风险度量 EVaR 的估计, 从而得到更加精确与稳健的风险度量.

目前, 期望分位数模型的相关文献, 通常关注的是经济计量或估计问题, 而很少涉及风险度量方法. Yao 和 Tong^[24] 在 1996 年研究了非参数的期望分位数回归模型. Kuan 等^[17] 在 2009 年基于 Newey 和 Powell 的研究, 对条件自回归期望分位数 (conditional autoregressive expectile, CARE) 模型建模, 并做了分析与讨论. De Rossi 和 Harvey^[25] 用样条插值的非参数估计方法对分位数回归和期望分位数回归进行了对比分析. 实证中, Wang 和 Zhou^[26] 利用极值理论基于广义 Pareto 分布的 POT (peaks over threshold) 模型对上证综指日收益率的尾部数据直接建模, 进而计算出 VaR 和 ES. Su 和 Zhou^[27] 将条件自回归期望分位数模型应用于基金业绩的评价. Hamidi 等^[28] 将动态自回归期望分位数模型应用于时间不变的投资组合保护策略. 此外, Liu 和 Zhou^[29] 把加权复合的期望分位数自回归模型应用于股票指数分析. 最近, Su 等^[30] 综述了风险度量技术的发展历史, 介绍了最新的风险度量理论与方法.

基于风险度量 EVaR, Xie 等^[20] 建立了变系数期望分位数回归模型来估计 VaR, 并且找到了有效的迭代算法进行求解. Xie 等^[31] 扩展了 CARE 模型, 基于自回归条件异方差期望分位数 (autoregressive conditional heteroskedasticity-expectile) 方法对 VaR 和 ES 的尾部风险进行了测量. Gerlach 等^[32] 则拓展为 Bayes 半参数已实现的 CARE 模型, 对尾部风险进行了预测. 但这些文献并没有同时考虑参数和非参数的模型, 因而无法体现动态或风险因素的交互影响.

在独立同分布的样本下, 对半参数模型的研究已有一些相关的工作. Ahmad 等^[33] 给出了半参数部分线性变系数模型的有效估计, 证明了其相合性、渐近正态性和估计的半参有效性. Zhou 和 Liang^[23] 对有易错协变量的半参数变系数部分线性模型做了统计推断, 提出了基于截面最小二乘的假设检验方法. Sun 等^[34] 以及 Rodriguez-Poo 和 Soberon^[35] 对有固定效应的面板数据建立了半参数变系数模型, 并给出了相应的估计方法. Kai 等^[36] 对半参数变系数部分线性模型给出了新的估计和高维变量选择的方法. Liu 和 Zhou^[37] 研究了半参数变系数复合期望分位数回归模型, 仅考虑了在独立同分布

的数据下, 复合期望分位数估计以及模型的应用. Zhang 等^[38] 给出了半参数变系数测量误差分位数回归模型的估计方法. Truquet^[39] 对含时间变系数的线性模型进行了有效的半参数统计推断. Spiegel 等^[40] 针对半参数期望分位数模型用样条方法进行了模型选择. Cai 等^[41] 研究了与本文相似的部分变系数期望分位数模型, 他们要求序列是 β -混合严平稳的, 而本文 α -混合序列的假设条件更弱、更一般, 在金融时间序列数据下本文的假设与估计方法更为合理.

本文针对满足 α -混合条件的金融时间序列, 考虑尾部更敏感的 EVaR 作为资产的一种风险度量, 并且将 Xie 等^[20] 的变系数期望分位数模型拓展到半参数变系数期望分位数模型, 对变系数参数与常系数参数分别进行半参有效估计. 本文给出风险度量 EVaR 的估计, 进而计算 VaR 和 ES, 得到更为全面的风险度量, 并且把 Liu 和 Zhou^[37] 独立同分布的样本假设与 Cai 等^[41] 的 β -混合相依假设推广到满足 α -混合条件的时间序列数据, 利用大小块分割方法来发展 α -混合序列的极限理论, 获得了基于金融时间序列数据建立的模型参数和非参数估计的统计渐近性质, 证明了估计的相合性和渐近正态性.

本文余下内容安排如下. 第 2 节在 α -混合相依假设下, 给出半参数变系数期望分位数模型的统计推断方法. 在估计非参数的变系数时, 采用一步估计的方法, 在不过多损失估计精度的情形下, 减少循环的次数, 从而提高估计的效率. 而对变系数部分与常系数部分的估计, 则采用三阶段估计方法, 通过剖分 (profile) 的思想, 分别固定常系数和变系数部分进行参数估计提高估计精确程度. 然后, 在 α -混合的观测序列下, 本文给出统计量的相合性、渐近正态性及方差估计, 同时给出计算 VaR 和 ES 的公式. 第 3 节是数值模拟, 构建了 3 个具体的半参数变系数期望分位数时间序列模型, 并对模型中的变系数部分与常系数部分参数进行估计, 从而检验估计的效果. 第 4 节通过实际数据进行实证分析, 给出合理的分析结果. 最后一节为结论. 关于定理证明需要的引理均放在附录 A 中.

2 研究理论与方法

2.1 时间序列下半参数变系数期望分位数模型

设 $(Y_t, \mathbf{X}_t, \mathbf{W}_t, U_t)$ ($t = 1, 2, \dots, T$) 是一列严平稳的随机过程, 并且每个都与 $(Y, \mathbf{X}, \mathbf{W}, U)$ 同分布. 令 Y 为资产收益, $\mathbf{X} = (X_1, X_2, \dots, X_{p_1})^T$ 和 $\mathbf{W} = (W_1, W_2, \dots, W_{p_2})^T$ 是风险因素, 可能包含滞后的回报或者其他经济金融因素 (经济指数等). 协变量 U 则是单变量影响因素. 假设 $(Y_t, \mathbf{X}_t, \mathbf{W}_t, U_t)$ 满足 α -混合条件, 并且 $E(Y^2) < \infty$.

考虑半参数变系数期望分位数时间序列模型

$$Y = \mathbf{X}^T \boldsymbol{\alpha}_\theta(U) + \mathbf{W}^T \boldsymbol{\gamma}_\theta + \varepsilon, \quad (2.1)$$

其中 $\boldsymbol{\alpha}_\theta(U) = (\alpha_{1,\theta}(U), \dots, \alpha_{p_1,\theta}(U))^T$ 是 U 的光滑变系数函数向量, $\boldsymbol{\gamma}_\theta = (\gamma_{1,\theta}, \dots, \gamma_{p_2,\theta})^T$ 是常系数向量. 简记 $\alpha_{j,\theta}(U)$ 为 $\alpha_j(U)$, $\gamma_{i,\theta}$ 为 γ_i . 类似于 Y 的 τ -分位数, Y 的总体的 θ -期望分位数 (expectile) $\nu_\theta(Y)$ 定义为使得 $E[Q_\theta(Y - \nu)]$ 最小的 ν 值, 其中

$$Q_\theta(z) = \begin{cases} (1-\theta)z^2, & z \leq 0, \\ \theta z^2, & z > 0 \end{cases} \quad (2.2)$$

是一个非对称平方误差损失函数; $\theta \in (0, 1)$ 是非对称参数, 用来控制非对称的程度. 公式中的平方误差损失函数, 使得期望分位数对分布的尾部敏感性要高于分位数. 显然, 当 $\theta = 0.5$ 时, Y 的 θ -期望分

位数就是其均值. $\nu_\theta(\varepsilon | \mathbf{X}, \mathbf{W}, U) = 0$ 表示给定 $\mathbf{X}$ 、 $\mathbf{W}$ 和 U 下, ε 的期望分位数是 0. 故上面的模型 (2.1) 可以记为

$$\nu_\theta(Y | \mathbf{X}, \mathbf{W}, U) = \mathbf{X}^T \boldsymbol{\alpha}_\theta(U) + \mathbf{W}^T \boldsymbol{\gamma}_\theta. \quad (2.3)$$

半参数变系数期望分位数时间序列模型的设定十分灵活. 首先响应变量与一组解释变量线性相关, 与解释变量的线性关系分为两个部分的和, 第 1 部分, 回归系数可随着协变量 U 变化, 其中 U 可以与解释变量相关, 也可以和它们独立; 第 2 部分, 解释变量有固定的常系数. 当变系数恒为零时, 本文的模型退化为线性模型; 当常系数 $\boldsymbol{\gamma}_\theta$ 为零向量时, 退化为变系数模型. 因此, 这个半参数变系数模型有着十分广泛的应用. 在风险度量中, 这也是十分自然与合理的假设, 例如, 资产损益可能与近几日的标的股票价格间有一个变系数的线性关系, 同时还可能与某些经济指数存在固定系数的线性关系. 此时半参数变系数模型将更加准确地描述资产价值或风险与解释变量间的关系. 如何有效地对此期望分位数模型进行统计推断, 是本文主要研究的问题.

模型 (2.1) 是一个非常宽泛的模型, 包含了许多常见的半参数模型. 在独立观察样本下, Fan 和 Huang^[42] 在 2005 年系统研究了半参数变系数部分线性模型的估计与推断, 这是半参数变系数时间序列模型的一个特例, 即当协变量不包含响应变量的滞后项时, 我们给出半参数变系数部分线性模型

$$Y = \boldsymbol{\alpha}^T(U) \mathbf{X} + \boldsymbol{\beta}^T \mathbf{Z} + \varepsilon, \quad (2.4)$$

其中 ε 与 $(U, \mathbf{X}, \mathbf{Z})$ 独立, 且 $E(\varepsilon) = 0$, $\text{Var}(\varepsilon) = \sigma^2$.

在时间序列数据下, 模型 (2.1) 也包含了许多参数和非参数模型. 当模型 (2.1) 中的 $\boldsymbol{\gamma}_\theta = 0$ 时, 即为 Xie 等^[20] 讨论的变系数模型. 如果 $\mathbf{W} = \mathbf{0}$, 且 $p_1 = 1$, $X_1 = 1$, 则模型 (2.1) 变为普通的非参数期望分位数回归模型, 此模型由 Yao 和 Tong^[24] 在 1996 年的研究中提出. 如果 $\mathbf{X} = \mathbf{0}$, 即没有变系数部分, 则模型 (2.1) 变为 2009 年 Kuan 等^[17] 提出的 CARE 模型

$$y_t = a_{0,\theta} + \gamma_{1,\theta} y_{t-1}^+ + \gamma_{1,\theta}^* y_{t-1}^- + \cdots + \gamma_{q,\theta} y_{t-q}^+ + \gamma_{q,\theta}^* y_{t-q}^- + \varepsilon_t. \quad (2.5)$$

在模型 (2.1) 中当 $\theta = 0.5$ 时, 模型即为均值回归. 因此, 该模型还包含函数型系数自回归 (functional autoregressive, FAR) 模型、门限自回归 (threshold autoregressive, TAR) 模型和指数型自回归 (exponential autoregressive, EXPAR) 模型等. Chen 和 Tsay^[43] 在 1993 年研究了 FAR 模型

$$Y_t = a_1(Y_{t-d}^*) Y_{t-1} + \cdots + a_p(Y_{t-d}^*) Y_{t-p} + \varepsilon_t, \quad (2.6)$$

其中 ε_t 是一列独立同分布的随机变量, 且与 Y_{t-i} ($i > 0$) 独立; $Y_{t-d}^* = (Y_{t-i_1}, \dots, Y_{t-i_k})^T$, $i_j > 0$, $j = 1, \dots, k$.

Stine 和 Tong^[44] 在 1992 年给出了 TAR 模型的理论性质及应用:

$$y_t = \phi_{(1)}^i Y_{t-1} + \cdots + \phi_p^{(i)} Y_{t-p} + \varepsilon_t^{(i)}, \quad \text{如果 } y_{t-d} \in \Omega_i, \quad i = 1, \dots, k, \quad (2.7)$$

其中 Ω_i 构成实数轴的一个划分.

Haggan 和 Ozaki^[45] 在 1981 年研究了 EXPAR 模型

$$y_t = \sum_{i=1}^p \{ \alpha_i + (\beta_i + \gamma_i y_{t-d}) \exp(-\theta_i y_{t-d}^2) \} y_{t-i} + \varepsilon_t, \quad (2.8)$$

其中 $\theta_i > 0$, $i = 1, \dots, p$. 此模型也包含在模型 (2.1) 中.

由此看出本文中的模型有着十分广泛的适用性, 涵盖了很多典型的模型. 下面小节介绍半参数变系数期望分位数时间序列模型的估计方法.

2.2 估计方法

对于模型 (2.1), 注意到估计 $\alpha_j(U)$ ($j = 1, \dots, p_1$) 和 γ_i ($i = 1, \dots, p_2$) 就是求解如下优化问题:

$$\arg \min_{\alpha_j, \theta(U), \gamma_i} E \left\{ Q_\theta \left(Y - \sum_{j=1}^{p_1} \alpha_j(U) X_j - \sum_{i=1}^{p_2} \gamma_i W_i \right) \middle| \mathbf{X} = x, \mathbf{W} = w, U = u \right\}, \quad (2.9)$$

其中损失函数 $Q_\theta(z)$ 是上一小节定义的非对称平方误差损失.

显然, 当 $\theta = 0.5$ 时, (2.9) 是普通最小二乘目标函数; 当 $\theta \neq 0.5$ 时, (2.9) 变成广义的加权最小二乘问题. 然而, 直接求解上述的优化问题是不可行的, 因为 $\alpha_j(U)$ ($j = 1, 2, \dots, p_1$) 是一个无穷维的参数. 为了提高精度, 本文采用 3 个阶段估计方法: 在第 1 阶段进行预估计, 同时估计局部线性平滑后的变系数和常系数部分; 随后第 2 阶段利用剖分思想, 将预估计的变系数部分代入模型, 对常系数进行估计; 在第 3 阶段, 代入第 2 阶段常系数估计值到原模型中, 重新估计变系数部分. 由于预估计是在局部化的情形下进行的估计, 常系数的估计并未用到全部样本的信息. 而通过剖分思想, 在第 2 阶段的常系数估计利用了全部样本的信息, 因此从理论上会估计得更加精确, 即偏差更小. 将其代入模型, 进一步能得到更精确的变系数估计. 本文证明在 3 个阶段下的估计都具有相应的相合性和渐近正态性, 这保证了模型估计的理论基础和算法的收敛速度. 另外, 由于所得到的估计方程并没有显式解, 因此本文利用迭代的算法进行求解. 以下给出详细的估计方法.

第 1 阶段 首先, 对变系数做局部线性平滑处理. 设

$$\mathbf{a} = (\alpha_1(\cdot), \dots, \alpha_{p_1}(\cdot))^T, \quad \mathbf{b} = (\alpha'_1(\cdot), \dots, \alpha'_{p_1}(\cdot))^T,$$

并且假设 $\alpha_j(U)$ ($j = 1, \dots, p_1$) 是二阶连续可导的. 将 $\alpha_j(u)$ 在 u_0 点一阶局部线性近似可得

$$\alpha_j(U) \approx a_j + b_j(U - u_0).$$

记 $\beta = (\mathbf{a}^T, \mathbf{b}^T)^T$, 对给定 θ 水平下, u_0 点处的局部线性回归估计 $\hat{\beta}_\theta(u_0) = (\hat{\mathbf{a}}_\theta^T(u_0), \hat{\mathbf{b}}_\theta^T(u_0))^T$ 和 $\hat{\gamma}_\theta(u_0)$ 即可由最小化下式得到:

$$\arg \min_{\beta, \gamma} \frac{1}{T} \sum_{t=1}^T Q_\theta(Y_t - \beta^T \mathbf{Z}_t - \gamma^T \mathbf{W}_t) K_h(U_t - u_0), \quad (2.10)$$

其中 $\mathbf{Z}_t = (\mathbf{X}_t^T, \mathbf{X}_t^T(U_t - u_0))^T$, $K_h(\cdot) = K(\cdot/h)$, $K(\cdot)$ 是核函数, h 是窗宽.

对 $\alpha_j(U)$ ($j = 1, \dots, p_1$), 分别在预设的网格点 u_k ($k = 1, \dots, n_{\text{grid}}$) 处一阶线性展开, 做逐点估计, 即得到变系数的估计曲线. 但是因为对常系数 γ_θ 的估计时只用到了 u_0 附近的样本点, 因此一般得不到很精确的估计, 收敛的速度也比较慢. 因此就有下一阶段针对常系数的估计.

第 2 阶段 将第 1 阶段得到的估计 $\hat{\alpha}_j(U)$ 代入到原模型中, 得到新的期望分位数参数模型, 再由非对称加权最小二乘解得参数 γ_θ 新的估计. 在这一步不做局部平滑的处理, 直接最小化

$$\hat{\gamma}_\theta^* = \arg \min_{\gamma} \frac{1}{T} \sum_{t=1}^T Q_\theta(\tilde{Y}_t - \gamma^T \mathbf{W}_t), \quad (2.11)$$

其中

$$\tilde{Y}_t = Y_t - \sum_{j=1}^{p_1} \hat{\alpha}_j(U) X_{jt}.$$

此时得到的常系数参数估计为 $\hat{\gamma}_\theta^*$. 这里用到了所有的样本, 得到的常系数估计相比于第 1 阶段的预估计 $\hat{\gamma}_\theta(u_0)$ 更加有效, 有更小的偏差, 并且有着更快的收敛速度. 再由剖分 (profile) 的思想, 固定常系数, 对变系数做如下第 3 阶段的估计.

第 3 阶段 类似地, 将第 2 阶段得到的常系数估计 $\hat{\gamma}_\theta^*$ 代入到原模型中, 得到新的期望分位数变系数模型, 即不含有常系数部分:

$$\nu_\theta(Y_{\text{new}} | \mathbf{X}, \mathbf{W}, U) = \mathbf{X}^T \boldsymbol{\alpha}_\theta(U), \quad (2.12)$$

其中

$$Y_{\text{new}} = Y - \mathbf{W}^T \hat{\gamma}_\theta^*.$$

同样由上文估计变系数的方法再次估计 $\alpha_j(U)$, 关于 β 最小化下式:

$$\arg \min_{\beta} \frac{1}{T} \sum_{t=1}^T Q_\theta(Y_t - \beta^T \mathbf{Z}_t - \hat{\gamma}_\theta^{*T} \mathbf{W}_t) K_h(U_t - u_0), \quad (2.13)$$

这里同样用迭代加权局部最小二乘法方法进行求解. 这样重新估计出的变系数 $\hat{\alpha}_j^{\text{new}}(U)$ 从精度上要优于第一次估计出来的 $\hat{\alpha}_j(U)$, 并且有着更小的方差, 即更加有效.

2.3 渐近性质

对于以上 3 个阶段得到的估计量, 本文在下面给出了每个估计量的相合性和渐近正态性.

定理 2.1 (第 1 阶段估计相合性) 假设 $\{(Y_t, \mathbf{X}_t, \mathbf{W}_t, U_t)\}$ 中所有序列都是严平稳过程, 且满足 α -混合条件. 对任意的 $\theta \in (0, 1)$, 假设 $\hat{\beta}_\theta$ 和 $\hat{\gamma}_\theta$ 是第 1 阶段的估计, 其中 $\hat{\beta}_\theta = (\hat{\mathbf{a}}_\theta^T, \hat{\mathbf{b}}_\theta^T)^T$, 且在满足假设 (a.1)–(a.6) 的所有假设条件下, 当 $T \rightarrow \infty$ 时, 有

$$\|H(\hat{\beta}_\theta - \beta_\theta)\| + \|\hat{\gamma}_\theta - \gamma_\theta\| \xrightarrow{P} 0, \quad (2.14)$$

其中 $\|\cdot\|$ 是 Euclid 距离, H 是 $(2p_1) \times (2p_1)$ 的对角阵, 其前 p_1 个对角元为 1, 后 p_1 个对角元为 h , $\xrightarrow{P}$ 表示依概率收敛.

推论 2.1 在定理 2.1 的条件下, 对任意 $u \in \mathcal{U}$, 其中 $\mathcal{U}$ 是 u 的支撑集, 依概率有 $\hat{\alpha}_\theta(u)$ 是相合的, 即

$$\hat{\alpha}_\theta(u) \xrightarrow{P} \alpha_\theta(u). \quad (2.15)$$

以上定理和推论表明在第 1 阶段估计中可以保证模型参数和非参数的估计均是相合估计.

定理 2.2 (第 2 阶段估计相合性) 假设定理 2.1 中的条件满足, 对任意的 $\theta \in (0, 1)$, 假设 $\hat{\gamma}_\theta^*$ 是第 2 阶段常系数的估计. 当 $T \rightarrow \infty$ 时, 有

$$\hat{\gamma}_\theta^* \xrightarrow{P} \gamma_\theta. \quad (2.16)$$

定理 2.3 (第 3 阶段估计相合性) 假设定理 2.1 中的条件全部都满足, 对任意的 $\theta \in (0, 1)$, 假设 $\hat{\beta}_\theta^{\text{new}}$ 是第 3 阶段变系数的估计, 进一步得到 $\hat{\alpha}_\theta^{\text{new}}(u)$. 在满足一定的正则条件下, 当 $T \rightarrow \infty$ 时, 有

$$\hat{\alpha}_\theta^{\text{new}}(u) \xrightarrow{P} \alpha_\theta(u). \quad (2.17)$$

下面分别给 3 个阶段估计的渐近正态性. 从下面的 3 个定理可以看出, 由于定理 2.4 使用非参数方法, 在包括参数部分估计时, 我们仅获得非参数收敛速度 $\sqrt{Th}$, 而定理 2.5 表明经过剖分后参数的估计速度可以提高到 $\sqrt{T}$.

定理 2.4 (第 1 阶段估计渐近正态性) 假设定理 2.1 中所有的条件满足, 设 $\mu_l = \int u^l K(u) du$ 和 $\kappa_l = \int u^l K^2(u) du$ ($l = 0, 1, \dots$), $f_U(\cdot)$ 是 U 的边际密度, 并且假设 $F_{\mathbf{X}, \mathbf{W}, U}(\cdot)$ 是给定 $(\mathbf{X}, \mathbf{W}, U)$ 下 Y 的条件分布函数, 则当 $T \rightarrow \infty$ 时, 有

$$\sqrt{Th} \left[\begin{pmatrix} \hat{\mathbf{a}}_\theta(u_0) - \mathbf{a}_\theta(u_0) \\ h(\hat{\mathbf{b}}_\theta(u_0) - \mathbf{b}_\theta(u_0)) \\ \hat{\gamma}_\theta(u_0) - \gamma_\theta(u_0) \end{pmatrix} - \frac{h^2}{2} \mathbf{D}(\theta, u_0)^{-1} f_U(u_0) \begin{pmatrix} \mu_2 \mathbf{B}_1 \\ \mu_3 \mathbf{B}_1 \\ \mu_2 \mathbf{B}_2 \end{pmatrix} + o(h^2) \right] \\ \xrightarrow{\mathcal{L}} N(0, \mathbf{D}(\theta, u_0)^{-1} \boldsymbol{\Sigma}(\theta, u_0) \mathbf{D}(\theta, u_0)^{-1}), \quad (2.18)$$

其中

$$\mathbf{B}_1 = E[\mathbf{X} \mathbf{X}^T \boldsymbol{\alpha}''(u_0) (2(1-\theta)I(Y - \eta(U, \mathbf{X}, \mathbf{W}) \leq 0) + 2\theta I(Y - \eta(U, \mathbf{X}, \mathbf{W}) > 0))], \\ \mathbf{B}_2 = E[\mathbf{W} \mathbf{X}^T \boldsymbol{\alpha}''(u_0) (2(1-\theta)I(Y - \eta(U, \mathbf{X}, \mathbf{W}) \leq 0) + 2\theta I(Y - \eta(U, \mathbf{X}, \mathbf{W}) > 0))], \\ \mathbf{D}(\theta, u_0) = f_U(u_0) \begin{pmatrix} \mu_0 & \mu_1 & \mu_0 \\ \mu_1 & \mu_2 & \mu_1 \\ \mu_0 & \mu_1 & \mu_0 \end{pmatrix} * \boldsymbol{\Gamma}(\theta, u_0),$$

这里

$$\boldsymbol{\Gamma}(\theta, u_0) = 2E(\{\theta[1 - F_{\mathbf{X}, \mathbf{W}, U}(\eta(u, \mathbf{X}, \mathbf{W}))] \\ + (1-\theta)F_{\mathbf{X}, \mathbf{W}, U}(\eta(u, \mathbf{X}, \mathbf{W}))\}(\mathbf{X}^T, \mathbf{X}^T, \mathbf{W}^T)^T(\mathbf{X}^T, \mathbf{X}^T, \mathbf{W}^T) \mid U = u_0), \\ \boldsymbol{\Sigma}(\theta, u_0) = f_U(u_0) \begin{pmatrix} \kappa_0 & \kappa_1 & \kappa_0 \\ \kappa_1 & \kappa_2 & \kappa_1 \\ \kappa_0 & \kappa_1 & \kappa_0 \end{pmatrix} \otimes \mathbf{A}(\theta, u_0)$$

并且

$$\mathbf{A}(\theta, u_0) = E[Q_\theta'^2(\eta(u_0, \mathbf{X}, \mathbf{W}))(\mathbf{X}^T, \mathbf{X}^T, \mathbf{W}^T)^T(\mathbf{X}^T, \mathbf{X}^T, \mathbf{W}^T) \mid U = u_0],$$

其中

$$\eta(U, \mathbf{X}, \mathbf{W}) = \sum_{j=1}^{p_1} \alpha_j(U) X_j + \sum_{i=1}^{p_2} \gamma_i W_i,$$

* 表示 Khatri-Rao 乘积, $\otimes$ 表示 Kronecker 乘积.

定理 2.5 (第 2 阶段估计渐近正态性) 假设定理 2.1 中所有的条件都满足, 假设 $F_{\mathbf{W}}(\cdot)$ 是给定 $\mathbf{W}$ 下的 Y 的条件分布函数, 则当 $T \rightarrow \infty$ 时, 有

$$\sqrt{T}[\hat{\gamma}_\theta^* - \gamma_\theta] \xrightarrow{\mathcal{L}} N(0, \boldsymbol{\Gamma}^*(\theta)^{-1} \mathbf{A}^*(\theta) \boldsymbol{\Gamma}^*(\theta)^{-1}), \quad (2.19)$$

其中

$$\begin{aligned} \mathbf{A}^*(\theta) &= \mathbf{E}(Q_\theta'^2(\eta^*(\mathbf{W}))\mathbf{W}\mathbf{W}^T), \\ \mathbf{\Gamma}^*(\theta) &= 2\mathbf{E}(\{\theta[1 - F_{\mathbf{W}}(\eta^*(\mathbf{W}))] + (1 - \theta)F_{\mathbf{W}}(\eta^*(\mathbf{W}))\}\mathbf{W}\mathbf{W}^T), \end{aligned}$$

这里 $\eta^*(\mathbf{W}) = \sum_{i=1}^{p_2} \gamma_i W_i$.

定理 2.6 (第 3 阶段估计渐近正态性) 假设定理 2.1 中所有的条件都满足, 并且假设 $F_{\mathbf{X}, U}(\cdot)$ 是给定 $(\mathbf{X}, U)$ 下 Y 的条件分布函数, 则当 $T \rightarrow \infty$ 时, 有

$$\begin{aligned} \sqrt{Th} \left[\mathbf{H}(\hat{\beta}_\theta^{\text{new}}(u_0) - \beta_\theta(u_0)) - \frac{h^2}{2(\mu_2 - \mu_1^2)} \begin{pmatrix} (\mu_2^2 - \mu_1\mu_3)\alpha''_{u_0} \\ (\mu_3 - \mu_1\mu_2)\alpha''_{u_0} \end{pmatrix} + o(h^2) \right] \\ \xrightarrow{\mathcal{L}} N(0, \mathbf{D}^{\text{new}}(\theta, u_0)^{-1} \mathbf{\Sigma}^{\text{new}}(\theta, u_0) \mathbf{D}^{\text{new}}(\theta, u_0)^{-1}), \end{aligned} \quad (2.20)$$

其中

$$\begin{aligned} \mathbf{D}^{\text{new}}(\theta, u_0) &= f_U(u_0) \begin{pmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \otimes \mathbf{\Gamma}^{\text{new}}(\theta, u_0), \\ \mathbf{\Sigma}^{\text{new}}(\theta, u_0) &= f_U(u_0) \begin{pmatrix} \kappa_0 & \kappa_1 \\ \kappa_1 & \kappa_2 \end{pmatrix} \otimes \mathbf{A}^{\text{new}}(\theta, u_0), \\ \mathbf{\Gamma}^{\text{new}}(\theta, u_0) &= 2\mathbf{E}(\{\theta[1 - F_{\mathbf{X}, U}(\eta^{\text{new}}(u, \mathbf{X}))] \\ &\quad + (1 - \theta)F_{\mathbf{X}, U}(\eta^{\text{new}}(u, \mathbf{X}))\}\mathbf{X}\mathbf{X}^T \mid U = u_0), \\ \mathbf{A}^{\text{new}}(\theta, u_0) &= \mathbf{E}(Q_\theta'^2(\eta^{\text{new}}(u, \mathbf{X}))\mathbf{X}\mathbf{X}^T \mid U = u_0), \end{aligned}$$

这里 $\eta^{\text{new}}(u, \mathbf{X}) = \sum_{j=1}^{p_1} \alpha_j(u) X_j$.

推论 2.2 在定理 2.5 的条件下, 有

$$\sqrt{Th} \left[\hat{\alpha}_\theta^{\text{new}}(u_0) - \alpha_\theta(u_0) - \frac{(\mu_2^2 - \mu_1\mu_3)h^2}{2(\mu_2 - \mu_1^2)} \alpha''_{u_0} + o(h^2) \right] \xrightarrow{\mathcal{L}} N(0, \mathbf{\Sigma}), \quad (2.21)$$

其中

$$\mathbf{\Sigma} = \frac{c_0^2 \kappa_0 + 2c_0 c_1 \kappa_1 + c_1^2 \kappa_2}{f_U(u_0)} \mathbf{\Gamma}^{\text{new}}(\theta)^{-1} \mathbf{A}^{\text{new}}(\theta) \mathbf{\Gamma}^{\text{new}}(\theta)^{-1},$$

这里 $c_0 = \mu_2/(\mu_2 - \mu_1^2)$, $c_1 = -\mu_1/(\mu_2 - \mu_1^2)$.

2.4 方差估计

考虑 $(\hat{\beta}_\theta(u_0), \hat{\gamma}_\theta(u_0))$ 的协方差矩阵的估计, 可以先估计 $\mathbf{D}(\theta, u_0)$ 和 $\mathbf{\Sigma}(\theta, u_0)$. 由样本类似可以得到

$$\hat{\mathbf{D}}(\theta, u_0) = \frac{1}{Th} \sum_{t=1}^T \mathbf{Z}_t^{*\otimes 2} Q_\theta''(Y_t - \mathbf{Z}_t^T \hat{\beta}_\theta(u_0) - \mathbf{W}_t^T \hat{\gamma}_\theta(u_0)) K_t, \quad (2.22)$$

并且有

$$\hat{\Sigma}(\theta, u_0) = \frac{1}{Th} \sum_{t=1}^T \mathbf{Z}_t^{*\otimes 2} Q_\theta'^2(Y_t - \mathbf{Z}_t^T \hat{\beta}_\theta(u_0) - \mathbf{W}_t^T \hat{\gamma}_\theta(u_0)) K_t^2, \quad (2.23)$$

其中 $\mathbf{Z}_t^{*\otimes 2} = \mathbf{Z}_t^* \mathbf{Z}_t^{*\top}$, $Q_\theta'(\cdot)$ 和 $Q_\theta''(\cdot)$ 分别为 $Q_\theta(\cdot)$ 关于 θ 的一阶和二阶导数. 由引理 A.1 和 A.5 可知, $\hat{D}(\theta, u_0)$ 和 $\hat{\Sigma}(\theta, u_0)$ 都是其真实值的相合估计. 因此, $(\hat{\beta}_\theta(u_0), \hat{\gamma}_\theta(u_0))$ 协方差矩阵的相合估计可由下式得到:

$$\hat{D}(\theta, u_0)^{-1} \hat{\Sigma}(\theta, u_0) \hat{D}(\theta, u_0)^{-1}. \quad (2.24)$$

此相合估计避免了用嵌入 (plug-in) 方法对密度函数 $f_U(\cdot)$ 和分布函数 $F_{\mathbf{X}, \mathbf{W}, U}(\cdot)$ 等的估计. 为协方差矩阵的估计提供了一种简便的方法. 同时对于感兴趣的变系数参数 $\alpha(u_0)$ 的渐近协方差矩阵也可由 (2.24) 中左上 $p_1 \times p_1$ 的子矩阵得到.

类似地, 可以得到 $\mathbf{\Gamma}^*(\theta)$ 和 $\mathbf{A}^*(\theta)$ 以及 $\mathbf{D}^{\text{new}}(\theta, u_0)$ 和 $\Sigma^{\text{new}}(\theta, u_0)$ 的相合估计. 最终得到相应参数的方差估计.

对比 Liu 和 Zhou^[37] 在独立同分布数据假设下的估计结果, 本文基于时间序列数据所得到的估计在偏差项和方差项都更为复杂, 这是因为响应变量与解释变量间存在相依关系.

2.5 算法与一步估计

关于 β 和 γ 最小化 (2.10) 可以得到如下的估计方程:

$$\frac{1}{T} \sum_{t=1}^T L_\theta(Y_t - \beta^T \mathbf{Z}_t - \gamma^T \mathbf{W}_t) \mathbf{Z}_t^* K_h(U_t - u_0) = 0, \quad (2.25)$$

其中 $L_\theta(z) = 2z|\theta - I(z \leq 0)|$ 是损失函数的导数, $\mathbf{Z}_t^* = (\mathbf{X}_t^T, \mathbf{X}_t^T(U_t - u_0)/h, \mathbf{W}_t^T)$. 为了求解 (2.25) 的估计方程, 本文用迭代加权局部最小二乘 (iterative weighted local least squares, IWLLS) 和一步估计 (one-step estimation) 的方法.

而关于 γ 最小化 (2.11) 则得到的是如下估计方程:

$$\frac{1}{T} \sum_{t=1}^T L_\theta(\tilde{Y}_t - \gamma^T \mathbf{W}_t) \mathbf{W}_t = 0. \quad (2.26)$$

同样可以采用迭代加权最小二乘 (iterative weighted least squares, IWLS) 求解估计方程 (2.26).

类似地求解方程 (2.13), 可得

$$\frac{1}{T} \sum_{t=1}^T L_\theta(Y_t - \beta^T \mathbf{Z}_t - \hat{\gamma}^T \mathbf{W}_t) \mathbf{Z}_t^* K_h(U_t - u_0) = 0. \quad (2.27)$$

下面介绍求解估计方程使用的算法.

2.5.1 迭代加权局部最小二乘 (IWLLS)

记估计方程 (2.25) 为

$$\frac{1}{T} \sum_{t=1}^T w(e_t(\theta); \theta) K_t \mathbf{Z}_t^* e_t(\theta) = 0, \quad (2.28)$$

其中 $K_t = K_h(U_t - u_0)$, $w(e_t(\theta); \theta) = |\theta - I_{\{e_t(\theta) \leq 0\}}|$, 并且 $e_t(\theta) = Y_t - \beta^T \mathbf{Z}_t - \gamma^T \mathbf{W}_t$. 因此, 迭代加权局部最小二乘估计即为

$$(\beta_\theta(u_0), \gamma_\theta(u_0))^T = \left[\sum_{t=1}^T w(e_t(\theta); \theta) K_t(\mathbf{Z}_t^T, \mathbf{W}_t^T)^T (\mathbf{Z}_t^T, \mathbf{W}_t^T) \right]^{-1} \left[\sum_{t=1}^T w(e_t(\theta); \theta) K_t(\mathbf{Z}_t^T, \mathbf{W}_t^T)^T Y_t \right], \quad (2.29)$$

其中

$$e_t(\theta) = Y_t - \beta_\theta^T(u_0) \mathbf{Z}_t - \gamma_\theta^T(u_0) \mathbf{W}_t.$$

可以看出 (2.29) 并不是一个显式解, 但它可以通过迭代的方式进行求解. 记 $(\hat{\beta}_\theta^j(u_0), \hat{\gamma}_\theta^j(u_0))$ 是第 j 次迭代的迭代加权局部最小二乘估计, 则第 $j+1$ 次的估计为

$$(\hat{\beta}_\theta^{j+1}(u_0), \hat{\gamma}_\theta^{j+1}(u_0))^T = \left[\sum_{t=1}^T w(\hat{e}_t(\theta); \theta) K_t(\mathbf{Z}_t^T, \mathbf{W}_t^T)^T (\mathbf{Z}_t^T, \mathbf{W}_t^T) \right]^{-1} \times \left[\sum_{t=1}^T w(\hat{e}_t(\theta); \theta) K_t(\mathbf{Z}_t^T, \mathbf{W}_t^T)^T Y_t \right], \quad (2.30)$$

其中

$$\hat{e}_t(\theta) = Y_t - \mathbf{Z}_t^T \hat{\beta}_\theta^j(u_0) - \mathbf{W}_t^T \hat{\gamma}_\theta^j(u_0).$$

这里假定矩阵的逆是存在的. 随着迭代次数的增加, 最终收敛得到迭代加权局部最小二乘估计. 初值的选取则选用 $\theta = 0.5$ 时的最小二乘估计, 因为这个估计很容易计算.

2.5.2 一步估计算法 (one-step estimation algorithm)

由于对变系数的估计是逐点估计, 即使在很小的区间都可能去估计大量的网格点, 并且每次估计都用到了迭代的方法, 因此计算的负担很大. 为了提高估计的效率、减轻计算的负担, 本文所用的方法是估计变系数模型中常用的一种算法 (参见文献 [46, 47]). 假设 $(\tilde{\beta}_0(u_0), \tilde{\gamma}_0(u_0))$ 是在 u_0 处估计的初值, 则一步估计为

$$(\hat{\beta}_{\text{OS}}(u_0), \hat{\gamma}_{\text{OS}}(u_0))^T = \left[\sum_{t=1}^T w(\tilde{e}_t(\theta); \theta) K_t(\mathbf{Z}_t^T, \mathbf{W}_t^T)^T (\mathbf{Z}_t^T, \mathbf{W}_t^T) \right]^{-1} \times \left[\sum_{t=1}^T w(\tilde{e}_t(\theta); \theta) K_t(\mathbf{Z}_t^T, \mathbf{W}_t^T)^T Y_t \right], \quad (2.31)$$

其中

$$\tilde{e}_t(\theta) = Y_t - \mathbf{Z}_t^T \tilde{\beta}_0(u_0) - \mathbf{W}_t^T \tilde{\gamma}_0(u_0).$$

根据 Fan 和 Chen^[48] 的研究, 不难证明, 在一定的条件下, 一步估计与全迭代估计有着同样的渐近性质. 可以看出, 为了提高估计效率, 一步估计只进行了一次迭代, 因此初值的选取对估计的准确度十分重要. 要估计变系数 $\alpha(u)$ 在整个支撑上的曲线, 本文先取少量平均分布在支撑上的网格点作为节点, 在节点处估计时用全迭代方法求出迭代加权局部最小二乘估计. 而其他的网格点处则用一步估计提高计算效率, 且所选初值为距离其最近的网格点的估计值. 从节点处开始分别向两边的网格点逐点估计, 最终估计出所有网格点的估计, 即得到变系数的估计曲线. 例如, 若 $n_{\text{grid}} = 200$ 时, 可以取

u_{20} 、 u_{60} 、 u_{100} 、 u_{140} 和 u_{180} 为节点. 在这 5 个点用全迭代方法求出变系数的估计部分. 例如, $\hat{\beta}(u_{60})$ 是全迭代得到的估计, 则 $\hat{\beta}(u_{59})$ 和 $\hat{\beta}(u_{61})$ 就是一步估计值, 且选 $\hat{\beta}(u_{60})$ 作为迭代的初值. 同理, $\hat{\beta}(u_{58})$ 和 $\hat{\beta}(u_{62})$ 则是分别选 $\hat{\beta}(u_{59})$ 和 $\hat{\beta}(u_{61})$ 为初值进行一步迭代得到的一步估计.

2.5.3 迭代加权最小二乘 (IWLS)

记估计方程 (2.26) 为

$$\frac{1}{T} \sum_{t=1}^T w(e_t^*(\theta); \theta) \mathbf{W}_t e_t^*(\theta) = 0, \quad (2.32)$$

其中 $w(e_t^*(\theta); \theta) = |\theta - I_{\{e_t^*(\theta) \leq 0\}}|$, 并且 $e_t^*(\theta) = \tilde{Y}_t - \gamma_\theta^T \mathbf{W}_t$. 这里常系数的迭代加权最小二乘 (IWLS) 估计是

$$\gamma_\theta = \left[\sum_{t=1}^T w(e_t^*(\theta); \theta) \mathbf{W}_t \mathbf{W}_t^T \right]^{-1} \left[\sum_{t=1}^T w(e_t^*(\theta); \theta) \mathbf{W}_t \tilde{Y}_t \right]. \quad (2.33)$$

同样用第 2.5.1 小节中的迭代方法求解, 最终得到估计 $\hat{\gamma}_\theta^*$. 在用上面两种算法估计变系数参数时, 每个网格点都对应得到了一个 $\hat{\gamma}_\theta(u_i)$ ($i = 1, \dots, n_{\text{grid}}$), 因此就将所有网格点的估计均值 $\hat{\gamma}_0$ 作为这里迭代加权最小二乘 (IWLS) 估计的初值.

2.6 窗宽的选择

变系数模型的回归有许多窗宽选择的方法, 这里采用 Cai 等^[22] 的研究中所采用的窗宽选择的方法, 这是一种简单且快速的窗宽选择方法, 可以看作是对于平稳时间序列数据的一种修正的多重交叉检验准则. 令 m 和 H 是两个给定的正整数, 且 $T > mH$. 首先用 H 个子序列, 每个序列的长度分别为 $T - mk$ ($k = 1, \dots, H$), 进行变系数部分的估计. 然后计算下一个长度为 m 的时间序列的一步预测误差. 更准确的是要选一个 h 使得下式最小:

$$\text{AMS}(h) = \sum_{k=1}^H \text{AMS}_k(h), \quad (2.34)$$

$$\text{AMS}_k(h) = \frac{1}{m} \sum_{t=km+1}^{T-km+m} Q_\theta \left\{ Y_t - \sum_{j=1}^p \hat{\alpha}_{j,k}(U_t) X_{t,j} - \hat{\gamma}_k^T \mathbf{W}_t \right\}, \quad (2.35)$$

其中 $\hat{\alpha}_{j,k}(\cdot)$ 和 $\hat{\gamma}_k$ 由子样本 $\{(Y_t, \mathbf{X}_t, \mathbf{W}_t, U_t), 1 \leq t \leq T - km\}$ 估计, 且可取最优窗宽, 即 $h \propto n^{-1/5}$. 实际应用中, 只要 mH 足够大, 使得预测误差稳定, m 和 H 的选择就不会对窗宽的选择造成很大的影响. 在随后的模拟实验中, 本文也将应用这种窗宽选择的方法.

2.7 VaR 与 ES 的计算

首先考虑 EVaR 与 QVaR 的关系. 给定分布和 θ 水平下, 存在一个 τ 使得 θ 水平下的 EVaR 与 τ 水平下的 QVaR 相等. 由这个特点, 可以通过 EVaR 计算 QVaR. 具体而言, 假设 $F(y)$ 是资产收益的分布函数, 对于任意的 $\tau \in (0, 1)$, 若 $\theta(\tau)$ 是使得 $\nu_{\theta(\tau)}(Y) = q_\tau(Y)$ 的期望分位数水平, 则在 1996 年, Yao 和 Tong^[24] 证明了 $\theta(\tau)$ 与 $q_\tau(Y)$ 之间的关系为

$$\theta(\tau) = \frac{\tau q_\tau(Y) - \int_{-\infty}^{q_\tau(Y)} y dF(y)}{E(Y) - 2 \int_{-\infty}^{q_\tau(Y)} y dF(y) - (1 - 2\tau) q_\tau(Y)}. \quad (2.36)$$

在 (2.36) 中, 期望分位数水平与 Y 的密度和 Y 的值都有关, 因此对于两个不同收益分布的资产很可能有相同的 QVaR, 但它们的 EVaR 会不一样, 因为期望分位数不仅依赖于 Y 的分布 $F(y)$, 还依赖于 Y 的极值. Kuan 等^[17] 研究了两个不同资产的分布, 其中一个有更高的尾部风险. 在正常的分位数水平下, 它们是有相同的 QVaR; 但是在正常的期望分位数水平下, 它们的 EVaR 却有很大的不同. 这说明, 对于它们尾部事件的风险暴露是不一样的. 因此, 相比于 QVaR 作为资产的风险度量, EVaR 对尾部概率更加敏感, 更能揭示出资产收益分布中存在极端损失的风险大小, 是更为有效的一种风险度量. 对大多数分布而言, 当 $\tau < 0.5$ 时, $\theta < \tau$, 并且对于尾部越厚的分布, 对应的 θ 越大.

这里考虑 ES 的计算. ES 相比 VaR 克服了尾部不敏感的不足之处, 成为了一个广泛应用的下端风险度量, 而通过期望分位数同样可以计算 ES. Taylor^[49] 证明了在一阶条件下, 有如下等式:

$$E(Y | Y < \nu_\theta(Y)) = \left(1 + \frac{\theta}{(1-2\theta)F(\nu_\theta(Y))}\right)\nu_\theta(Y) - \frac{\theta}{(1-2\theta)F(\nu_\theta(Y))}E(Y). \quad (2.37)$$

所以再由期望分位数与分位数的关系, 有 $F(\nu_\theta(Y)) = \tau$. 因此可得

$$ES(\tau) = \left(1 + \frac{\theta}{(1-2\theta)\tau}\right)\nu_\theta(Y) - \frac{\theta}{(1-2\theta)\tau}E(Y). \quad (2.38)$$

这个公式说明 ES 与 τ -分位数和相应的 θ -期望分位数有关. 而 (2.38) 计算的是分布下端尾部的 ES, 对于上端尾部的 ES 的计算, 只需要把 τ 和 θ 分别替换成 $(1-\tau)$ 和 $(1-\theta)$ 即可. 经验结果表明, 用期望分位数计算的 ES 与用其他方法计算出的 ES 非常接近. 此外, 条件的 $ES(\tau | \cdot)$ 也可以由 (2.37) 计算出来, 只需要将 $\nu_\theta(Y)$ 替换为 $\nu_\theta(Y | \cdot)$ 即可.

3 数值模拟

本节通过几个模型进行数值模拟实验, 展示半参数变系数期望分位数模型估计的有效性. 一方面是验证本文中的估计方法在有限样本下的估计准确性和估计效率; 另一方面是检验在不同的模型假设下, 参数估计有什么不同, 准确性是否有差别, 即在不同模型间估计的适用性. 同时还会对第 1 和 3 阶段的变系数估计作一个比较, 来检验第 3 阶段剖分 (profile) 后的估计是否在精确度上有提高.

3.1 评价指标

本文用均方误差的平方根 (root average squared error, RASE) 作为变系数估计好坏的判断指标, 其表达式如下:

$$RASE_{\alpha,j} = \left[\frac{1}{n_{\text{grid}}} \sum_{k=1}^{n_{\text{grid}}} (\hat{\alpha}_j(u_k) - \alpha_j(u_k))^2 \right]^{\frac{1}{2}}, \quad j = 0, 1, \dots, p_1, \quad (3.1)$$

其中 u_k ($k = 1, \dots, n_{\text{grid}}$) 是所有逐点估计的网格点. 同时给出每种情形下 RASE 的 500 次重复模拟估计的标准差.

另外, 本文给出 500 次重复模拟常系数参数的估计均值和标准差, 由结果可以清楚地知道估计与真实值之间非常接近.

3.2 模拟估计流程

本文考虑设定样本量大小 T 分别为 400 和 800, 并且采用最常用的 Epanechnikov 核函数 $K(u) = \frac{3}{4}(1-u^2)I(|u| \leq 1)$ 进行局部线性平滑. 而窗宽的选择方法则在第 2.6 小节中提出. 下面给出应用本文中的算法与一步估计的方法对此处数值模拟中的模型进行估计的流程:

步骤 1 首先设 $n_{\text{grid}} = 100$, 并且将 U 的支撑平均分成 100 个网格点 $\{u_1, \dots, u_{100}\}$.

步骤 2 先取 $\theta = 0.5$, 用最小二乘估计 $(\tilde{\beta}_0(u_0), \tilde{\gamma}_0(u_0))$ 作为迭代加权局部最小二乘的初始值, 在 u_0 分别等于 u_{10} 、 u_{30} 、 u_{50} 、 u_{70} 和 u_{90} 这 5 个点处用全迭代的方法求解估计方程 (2.25) 来逐点估计变系数部分. 求解出这 5 个点的估计分别为 $(\hat{\beta}(u_i), \hat{\gamma}(u_i))$, $i = 10, 30, 50, 70, 90$.

步骤 3 分别以步骤 2 中这 5 个估计的点为起始中心, 向两边进行一步估计, 即 u_{K-1} 和 u_{K+1} , $K = 10, 30, 50, 70, 90$ 点处用 $(\hat{\beta}(u_K), \hat{\gamma}(u_K))$ 作为迭代的初值进行一步估计. 而后在 u_{K-2} 和 u_{K+2} 点则用刚才 u_{K-1} 和 u_{K+1} 点处估计的值作为迭代的初始值进行估计, 直到 u_{K-9} 和 u_{K+10} 点, 即所有 100 个点均得到了估计值.

步骤 4 逐点估计后连接各点的估计值, 则得到变系数的估计曲线 $\hat{\alpha}_j(U)$, $j = 1, \dots, p_1$. 然后将其代入原模型中计算 Y_t^* 得到新的模型. 用最小二乘估计 $\tilde{\gamma}_0$ 作为初始值, 求解估计方程 (2.26), 得到迭代加权最小二乘估计 $\hat{\gamma}$.

步骤 5 再将所得到的估计 $\hat{\gamma}$ 代入原模型中, 计算 Y_t^{new} 得到新的模型, 用类似步骤 2 和 3 的方法估计变系数部分.

3.3 实验模型

下面本文拓展了文献 [20] 中数值模拟的 3 个模型来作为本文半参数变系数期望分位数模型的模拟案例.

模型一: 半参数变系数 EXPAR 模型

考虑如下模型:

$$Y_t = \alpha_1(Y_{t-1})Y_{t-1} + \alpha_2(Y_{t-1})Y_{t-2} + \gamma_1 W_{1t} + \gamma_2 W_{2t} + 2\alpha_1(Y_{t-1})Y_{t-1}I(Y_{t-1} > 0)\epsilon_t,$$

其中 $\alpha_1(u) = 0.138 + (0.316 + 0.982u)e^{-3.89u^2}$, $\alpha_2(u) = -0.437 - (0.659 + 0.1260u)e^{-3.89u^2}$; $\gamma_1 = 1.0$, $\gamma_2 = 0.5$; W_{1t} 和 W_{2t} 都与 Y_t 独立, 这里 W_{1t} 是 $[0, 1]$ 区间内的均匀分布, W_{2t} 是 $[-2, 0]$ 区间内的均匀分布; ϵ_t 是独立同分布的正态分布 $N(0, 0.2^2)$. 考虑 $\theta = 0.25, 0.50, 0.75$, 则相应 ϵ_t 的 θ 水平下的期望分位数 $\nu_\theta(\epsilon)$ 分别是 -0.09 、 0 和 0.09 . 期望分位数回归模型可以记作

$$\nu_\theta(Y_t | Y_{t-1}, Y_{t-2}, W_{1t}, W_{2t}) = (1 + 2\nu_\theta(\epsilon)I(Y_{t-1} > 0))\alpha_1(Y_{t-1})Y_{t-1} + \alpha_2(Y_{t-1})Y_{t-2} + \gamma_1 W_{1t} + \gamma_2 W_{2t}.$$

最优窗宽由第 2.6 小节的方法可以得出, $\theta = 0.50$ 时, 当 $T = 400$ 时, $h_{\text{opt}} = 0.1365$; 当 $T = 800$ 时, $h_{\text{opt}} = 0.1507$.

模型二: 半参数变系数 TAR 模型

考虑如下模型:

$$Y_t = \alpha_1(Y_{t-2})Y_{t-1} + \alpha_2(Y_{t-2})Y_{t-2} + \gamma_1 W_{1t} + \gamma_2 W_{2t} + \epsilon_t,$$

其中 $\alpha_1(u) = 0.4I(u \leq 1) - 0.8I(u > 1)$, $\alpha_2(u) = -0.6I(u \leq 1) + 0.2I(u > 1)$; $\gamma_1 = 1.0$, $\gamma_2 = 0.5$; W_{1t} 和 W_{2t} 都与 Y_t 独立, 这里 W_{1t} 是 $[0, 2]$ 区间内的均匀分布, W_{2t} 是 $[-4, 0]$ 区间内的均匀分布; ϵ_t 是独立

同分布的正态分布 $N(0, 0.2^2)$. 同样考虑 $\theta = 0.25, 0.50, 0.75$, 此时相应的 ϵ_t 分布下 θ 水平的期望分位数 $\nu_\theta(\epsilon)$ 分别是 0.44、0 和 0.44. 模型二的期望分位数回归模型则记作

$$\nu_\theta(Y_t | Y_{t-1}, Y_{t-2}, W_{1t}, W_{2t}) = \alpha_1(Y_{t-2})Y_{t-1} + \alpha_2(Y_{t-2})Y_{t-2} + \gamma_1 W_{1t} + \gamma_2 W_{2t} + \nu_\theta(\epsilon).$$

最优窗宽的选择与模型一相同.

模型三: 半参数变系数外生变量模型

考虑如下模型:

$$Y_t = \alpha_1(U_t)Y_{t-1} + \alpha_2(U_t)Y_{t-2} + \gamma_1 W_{1t} + \gamma_2 W_{2t} + \epsilon_t,$$

其中 $\alpha_1(u) = \sin(\sqrt{2}\pi u)$, $\alpha_2(u) = \cos(\sqrt{2}\pi u)$; $\gamma_1 = 1.0$, $\gamma_2 = 0.5$; W_{1t} 和 W_{2t} 都与 Y_t 独立, 这里 W_{1t} 和 W_{2t} 相互独立, 且都是服从 $N(0, 4)$ 的正态分布; U_t 是 $[0, 3]$ 区间内的均匀分布独立产生的; ϵ_t 是以概率 $1 - P$ 独立产生于正态分布 $N(0, 1/\sqrt{1-P})$ 中, 或者以概率 P 产生于正态分布 $N(c, 1/\sqrt{P})$. 在这里取 $c = -5$ 和 $P = 0.05$, 这是为了观察比较在遇到灾难性的损失时 QVaR 与 EVaR 的表现有什么不同. 模型三的期望分位数回归模型记作

$$\nu_\theta(Y_t | Y_{t-1}, Y_{t-2}, W_{1t}, W_{2t}, U_t) = \alpha_1(U_t)Y_{t-1} + \alpha_2(U_t)Y_{t-2} + \gamma_1 W_{1t} + \gamma_2 W_{2t} + \nu_\theta(\epsilon_t).$$

最优窗宽的选择和之前模型相同. 下面给出数值模拟实验的结果分析.

3.4 模拟结果分析

图 1 展示了模型一中变系数 $\alpha_1(u)$ 和 $\alpha_2(u)$ 的估计结果, 给出了当 $\theta = 0.25$ 、 $\theta = 0.50$ 和 $\theta = 0.75$ 时的估计曲线. 如图 1(a), 我们看到, 当 $\theta = 0.25$ 时估计曲线略低于真实曲线; 当 $\theta = 0.75$ 时, 估计曲线略高于真实曲线; 而当进行均值回归, 即 $\theta = 0.50$ 时, 估计曲线基本与真实曲线重合. 这说明在四分之三期望分位点处估计会略微偏高, 相对的四分之一期望分位点则估计不足, 这也能从估计方法中很直接地预料到, 对于高于回归曲线的数值给了更大的权重, 就很可能产生这样如图 1(a) 的结果. 图 1(b) 的估计曲线则十分稳定, 几乎与真实曲线重合, 而边缘部分的估计会产生偏差波动, 这是因为两侧的数据样本量明显较小, 在作局部平滑估计时会使得估计值出现较大的偏差. 但是本文所采用的估计方法在绝大部分变系数函数曲线上的估计都是较为精确和有效的.

图 2 展示了模型二中变系数 $\alpha_1(u)$ 和 $\alpha_2(u)$ 的估计结果, 给出了当 $\theta = 0.25$ 、 $\theta = 0.50$ 和 $\theta = 0.75$ 时的估计曲线. 图 2(a) 的变系数估计曲线在 $\theta = 0.25$ 、 $\theta = 0.50$ 和 $\theta = 0.75$ 处都有着较好的估计效果. 而图 2(b) 与模型一中的 $\alpha_1(u)$ 类似, 对于不同的 θ 值, 有着相应过高或者过低的估计曲线.

图 3 展示了模型三中变系数 $\alpha_1(u)$ 和 $\alpha_2(u)$ 的估计结果, 给出了当 $\theta = 0.25$ 、 $\theta = 0.50$ 和 $\theta = 0.75$ 的估计曲线. 可以看到估计拟合出的变系数函数效果很好, 几乎与真实的变系数函数曲线重合. 由于所估计的 $\alpha_1(u)$ 和 $\alpha_2(u)$ 分别为 $\sin$ 函数和 $\cos$ 函数, 均为光滑函数, 本身就具有可导等良好的性质, 因此这里的估计偏差较小.

图 4 给出了模型一中变系数部分估计的 RASE 箱线图. 这里所展示的均是在 $T = 800$ 时的结果, 取 θ 分别等于 0.25、0.50 和 0.75. 可以看出, 估计的 RASE 都很小, 表明逐点估计变系数所用的方法是十分有效的. 对变系数 $\alpha_1(u)$ 和 $\alpha_2(u)$ 分别在第 1 和 3 阶段的估计的 RASE 作了对比, 可以看出, 第 3 阶段重新进行的变系数部分的估计效果都要优于第 1 阶段的估计, 估计值的误差和方差都减小了. 这也说明了代回常系数部分的估计进行第 3 阶段重新估计变系数的必要性, 这样不仅提高了模型拟合的精确程度还增加了估计的稳定性.

图 1 (网络版彩图) 模型一中变系数 $\alpha_1(u)$ (a) 和 $\alpha_2(u)$ (b) 分别在 $\theta = 0.25$ (方点)、 $\theta = 0.50$ (圆点)、 $\theta = 0.75$ (三角形点) 的估计曲线图, 其中黑色实线为真实曲线

图 2 (网络版彩图) 模型二中变系数 $\alpha_1(u)$ (a) 和 $\alpha_2(u)$ (b) 分别在 $\theta = 0.25$ (方点)、 $\theta = 0.50$ (圆点)、 $\theta = 0.75$ (三角形点) 的估计曲线图, 其中黑色实线为真实曲线

图 3 (网络版彩图) 模型三中变系数 $\alpha_1(u)$ (a) 和 $\alpha_2(u)$ (b) 分别在 $\theta = 0.25$ (方点)、 $\theta = 0.50$ (圆点)、 $\theta = 0.75$ (三角形点) 的估计曲线图, 其中黑色实线为真实曲线

图 5 和 6 给出了模型二和三的 RASE 比较的箱线图, 都有着类似的结论. 尤其是图 4(a) 和 5(b), 估计效果在第 3 阶段估计上有了明显的提高.

随后, 表 1 给出了模型一 500 次 Monte Carlo 模拟的详细数据结果, 展示了变系数参数估计 $\hat{\alpha}_i$ 和 $\hat{\alpha}_i^{\text{new}}$ ($i = 1, 2$) 的 RASE 在 θ 分别等于 0.25、0.50 和 0.75 水平下的均值和标准差. 首先可以看出 RASE 的均值都较小, 说明变系数的估计偏差较小, 逐点估计出的曲线十分接近真实曲线; 标准差也处在一个很小的水平下, 说明本文的估计是较为稳健的, 偏差的浮动较小. 然后, 对比 $\hat{\alpha}_i$ 和 $\hat{\alpha}_i^{\text{new}}$ 可以发

图 4 (网络版彩图) 模型一中 $\alpha_1(u)$ (a) 和 $\alpha_2(u)$ (b) 估计的 RASE 分别在 $\theta = 0.25$ (1,2)、 $\theta = 0.50$ (3,4)、 $\theta = 0.75$ (5,6) 处的箱线图, 其中 1、3 和 5 号为第 1 阶段估计的 RASE, 2、4 和 6 号为第 3 阶段估计的 RASE

图 5 模型二中 $\alpha_1(u)$ (a) 和 $\alpha_2(u)$ (b) 估计的 RASE 分别在 $\theta = 0.25$ (1,2)、 $\theta = 0.50$ (3,4)、 $\theta = 0.75$ (5,6) 处的箱线图, 其中 1、3 和 5 号为第 1 阶段估计的 RASE, 2、4 和 6 号为第 3 阶段估计的 RASE

图 6 模型三中 $\alpha_1(u)$ (a) 和 $\alpha_2(u)$ (b) 估计的 RASE 分别在 $\theta = 0.25$ (1,2)、 $\theta = 0.50$ (3,4)、 $\theta = 0.75$ (5,6) 处的箱线图, 其中 1、3 和 5 号为第 1 阶段估计的 RASE, 2、4 和 6 号为第 3 阶段估计的 RASE

现, $\hat{\alpha}_i^{\text{new}}$ 的 RASE 均值都要小于 $\hat{\alpha}_i$ 的 RASE 均值, 同时, 基本都有更小的标准差. 这说明第 3 阶段进行的估计在精确程度和稳健性上都要优于第 1 阶段对变系数部分的估计, 提高了估计的效果.

表 2 和 3 给出了模型二和三的 500 次 Monte Carlo 模拟的详细数据结果. 模型二的变系数为阶梯

表 1 模型一的变系数参数估计的 RASE 结果

估计参数	θ	RASE 均值	标准差
$\hat{\alpha}_1$	0.25	0.0526	0.0132
	0.50	0.0336	0.0078
	0.75	0.0519	0.0177
$\hat{\alpha}_1^{\text{new}}$	0.25	0.0432	0.0232
	0.50	0.0215	0.0116
	0.75	0.0413	0.0132
$\hat{\alpha}_2$	0.25	0.0295	0.0216
	0.50	0.0305	0.0241
	0.75	0.0327	0.0241
$\hat{\alpha}_2^{\text{new}}$	0.25	0.0266	0.0140
	0.50	0.0271	0.0240
	0.75	0.0293	0.0213

表 2 模型二的变系数参数估计的 RASE 结果

估计参数	θ	RASE 均值	标准差
$\hat{\alpha}_1$	0.25	0.1592	0.0097
	0.50	0.1579	0.0100
	0.75	0.1603	0.0106
$\hat{\alpha}_1^{\text{new}}$	0.25	0.1583	0.0094
	0.50	0.1572	0.0097
	0.75	0.1593	0.0104
$\hat{\alpha}_2$	0.25	0.1669	0.0418
	0.50	0.1291	0.0283
	0.75	0.1739	0.0428
$\hat{\alpha}_2^{\text{new}}$	0.25	0.1327	0.0364
	0.50	0.0968	0.0177
	0.75	0.1404	0.0381

表 3 模型三的变系数参数估计的 RASE 结果

估计参数	θ	RASE 均值	标准差
$\hat{\alpha}_1$	0.25	0.0329	0.0061
	0.50	0.0303	0.0041
	0.75	0.0293	0.0035
$\hat{\alpha}_1^{\text{new}}$	0.25	0.0329	0.0060
	0.50	0.0299	0.0040
	0.75	0.0289	0.0033
$\hat{\alpha}_2$	0.25	0.0333	0.0059
	0.50	0.0305	0.0041
	0.75	0.0295	0.0033

(续表)

估计参数	θ	RASE 均值	标准差
$\hat{\alpha}_2^{\text{new}}$	0.25	0.0330	0.0059
	0.50	0.0302	0.0040
	0.75	0.0292	0.0033

表 4 模型一的常系数参数估计结果

估计参数	θ	均值	标准差	最小值	最大值
$\hat{\gamma}_1$	0.25	0.9903	0.0157	0.7707	1.0404
	0.50	0.9966	0.0111	0.9341	1.0639
	0.75	1.0018	0.0119	0.9690	1.1120
$\hat{\gamma}_2$	0.25	0.5017	0.0060	0.4705	0.5356
	0.50	0.4993	0.0061	0.4391	0.5455
	0.75	0.4969	0.0055	0.4586	0.5214

表 5 模型二的常系数参数估计结果

估计参数	θ	均值	标准差	最小值	最大值
$\hat{\gamma}_1$	0.25	0.9939	0.0317	0.9014	1.0856
	0.50	0.9998	0.0303	0.9001	1.0923
	0.75	1.0068	0.0325	0.9065	1.1100
$\hat{\gamma}_2$	0.25	0.5031	0.0160	0.4601	0.5601
	0.50	0.4997	0.0147	0.4410	0.5394
	0.75	0.4973	0.0164	0.4527	0.5430

表 6 模型三的常系数参数估计结果

估计参数	θ	均值	标准差	最小值	最大值
$\hat{\gamma}_1$	0.25	0.9990	0.0246	0.9263	1.0877
	0.50	1.0020	0.0179	0.9541	1.0573
	0.75	1.0012	0.0160	0.9514	1.0555
$\hat{\gamma}_2$	0.25	0.5001	0.0241	0.4140	0.6143
	0.50	0.5006	0.0184	0.4407	0.5710
	0.75	0.4999	0.0154	0.4349	0.5446

函数, 估计结果偏差稍大, 但其偏差还是在接受的范围内. 而其他的结果与模型一的结果显示的基本相似, 都有着较小的 RASE 均值和标准差, 并且第 3 阶段的估计优于第 1 阶段的估计效果.

表 4-6 分别给出了模型一、二和三常系数部分的估计数据结果. 在参数设置时, 将 γ_1 和 γ_2 分别设为 1.0 和 0.5. 从表中可以看出, 估计 $\hat{\gamma}_i$ ($i = 1, 2$) 均值与真实值十分接近, 并且都有着较小的标准差, 最小值与最大值也没有异常的偏差. 这说明所用的方法估计精度较高, 有着较好的估计效果.

4 股票市场周内效应的实证研究

4.1 样本选取及数据说明

随着国际贸易的全球化发展, 国内股票市场逐渐受到其他各国股票市场和国际经济形势的影响. 本文选取 2015 年 3 月 30 日至 2018 年 3 月 16 日共 726 个交易日的上证综指 (股票代码: 000001) 作为研究对象, 想要了解代表国内股市的上证综指与美国股市及国际经济之间的相关性. 具体地, 选用前一天的标准普尔 500 指数 (S&P 500) 作为代表美国股票市场的指标, 定其为影响变系数的外生变量, 同时做对数处理, 研究指数收益率之间的联系, 即设 p_t 和 p_t^* 为 t 时刻上证综指和标准普尔 500 指数, 则

$$Y_t = 100 \log \left(\frac{p_t}{p_{t-1}} \right), \quad U_t = 100 \log \left(\frac{p_t^*}{p_{t-1}^*} \right).$$

同时选取相同时间段内的无风险利率、人民币兑美元汇率和世界黄金现货市场价格收益率作为常系数部分的解释变量. 所选取的数据分别代表国内经济形势、与美国间的经济关系和国际经济的影响因素. 实际上, 模型建立了美国股市的回报值与国内市场风险的关系. 由此引入 5 个虚拟变量代表一周内的 5 个交易日, 建立如下半参数期望分位数模型:

$$\nu_\theta(Y_t | D1_t, D2_t, D3_t, D4_t, D5_t, U_{t-1}) = \alpha_{0,\theta}(U_{t-1}) + \sum_{j=1}^5 \alpha_{j,\theta}(U_{t-1}) D_j t + \sum_{j=1}^3 \gamma_{j,\theta} W_{j,t}, \quad (4.1)$$

其中 D_j ($j = 1, \dots, 5$) 表示周内星期日的虚拟变量. 用前 707 个交易日数据作为样本进行参数估计, 后 20 个交易日数据来做预测. 这里选取 Epanechnikov 核函数 $K(u) = \frac{3}{4}(1 - u^2)I(|u| \leq 1)$ 进行局部线性平滑.

4.2 结果分析

图 7 展示了变系数 $\alpha_{j,\theta}(U_{t-1})$ ($j = 1, \dots, 5$) 的估计曲线. 可以看出, 星期一的变系数估计曲线呈现对勾型, 星期二和星期四呈现“几”字型, 星期三和星期五呈现 M 型. 这与文献 [20] 中的模型估计结果较为相似. 例如, 在星期一, 对于前一个交易日的 S&P 500 有较高的涨幅是对上证综指有正向的影响效应; 而相反较大的跌幅则对上证综指有负向效应, 若前一个交易日 S&P 500 出现更大的跌幅, 则负向效应并不明显. 对于星期二和星期四, 变系数曲线整体呈现负向效应, 对于过高的涨跌幅意味着更大的负向效应. 而星期三和星期五的变系数曲线中间低两边高, 表示 S&P 500 涨跌幅较小时对上证综指为负向效应, 而有过高涨跌幅情形下会对上证综指有正向的影响效应. 图中分别展示了 3 条估计曲线, 分别是在 θ 等于 0.25、0.50 和 0.75 的期望分位数水平下的结果, 都呈现出相同的模式. 在 $\theta = 0.25$ 时估计偏低, 而在 $\theta = 0.75$ 时估计偏高.

表 7 给出了拟合的半参数变系数期望分位数模型中常系数部分的估计结果, 这里取 $\theta = 0.50$, 其中黄金收益率的系数估计为 $\hat{\gamma}_1 = -0.5264$, 美元汇率的系数估计为 $\hat{\gamma}_2 = 0.2297$, 无风险利率的系数估计为 $\hat{\gamma}_3 = 2.3646$. 黄金的收益率与美元汇率的估计系数分别在 0.01 和 0.10 水平下显著, 引入变量有着较好的解释效果. 而无风险利率的系数估计则不够显著.

表 8 给出了应用前面 707 个交易日数据作为样本拟合模型后得到的预测结果. 我们分别用 $\theta = 0.025$ 和 $\theta = 0.975$ 的估计值进行预测, 构建出一个基于期望分位数的 95% 预测区间. 除了 2 月 13 日和 2 月 22 日由于过年的节日效应而导致的预测失败, 其他的预测结果均表现较好, 真实值基本上都落入了预测区间内. 这表明本文构建的模型有一定的解释意义, 并且有着较为有效的预测效果.

图 7 (网络版彩图) 模型 (4.1) 中星期一到星期五的变系数在不同 θ 水平下参数估计的曲线图, 其中 $\theta = 0.25$ (长虚线), $\theta = 0.50$ (实线), $\theta = 0.75$ (短虚线)

表 7 模型 (4.1) 常系数估计结果

变量	估计值	标准误	t 值	$\Pr(> t)$
黄金收益率	-0.5264	0.1841	-2.86	0.0044
美元汇率	0.2297	0.1184	1.94	0.0527
无风险利率	2.3646	1.7397	1.36	0.1745

表 8 上证综指 95% 的预测区间结果

日期	真实值	预测区间	日期	真实值	预测区间
2018.2.12	0.7756	(-0.4780, 3.8714)	2018.3.5	0.0737	(-1.7329, 1.9529)
2018.2.13	0.9776	(-1.7254, -1.6943)	2018.3.6	1.0045	(-2.7858, 2.7846)
2018.2.14	0.4458	(-1.6917, 1.2499)	2018.3.7	-0.5464	(-1.5992, 0.9799)
2018.2.22	2.1693	(-7.4296, 0.5314)	2018.3.8	0.5116	(-1.8268, 1.7349)
2018.2.23	0.6261	(-1.9809, 1.2293)	2018.3.9	0.5705	(-2.0402, 1.3943)
2018.2.26	1.2329	(0.4856, 3.5999)	2018.3.12	0.5906	(1.4627, 3.2485)
2018.2.27	-1.1265	(-2.7827, 2.6508)	2018.3.13	-0.4948	(-1.3428, 1.5746)
2018.2.28	-0.9921	(-1.3153, -0.0383)	2018.3.14	-0.5697	(-1.6121, 2.0906)
2018.3.1	0.4402	(-0.1958, 1.5989)	2018.3.15	-0.0082	(-1.1800, 1.4613)
2018.3.2	-0.5873	(-0.3688, 1.5416)	2018.3.16	-0.6451	(-2.1534, 1.7196)

5 结论

本文研究了期望分位数的基本性质以及期望分位数风险价值 (EVaR), 阐述了它相比于 QVaR 的优点, 并且拓展了已有的变系数期望分位数模型. 本文引入常数项系数, 建立了适用性更加广泛的半参数变系数期望分位数时间序列模型. 该模型可以用于分析风险因素的来源、风险因素的动态和静态的交互效应. 本文对模型中的参数进行了三阶段估计, 利用非对称局部最小二乘和一步估计等方法, 证明了给出的变系数和常系数估计的相合性及渐近正态性. 三个具体模型的数值模拟展示了本文估计方法的有效性及精确性. 实际数据的分析结果也说明了本文提出的模型有较好的适用性. 这对金融风险管理的有着重要的经济意义与统计意义. 首先, 由于模型为变系数的估计, 估计的精确度和算法的运行效率之间的权衡问题会自然而然地产生. 若数据维度较高, 算法运行的时间会比较长, 一定时间内精确度会受到挑战和影响. 其次, 对于风险因素的动态效应在解释上较为复杂, 需要做更进一步细致的分析才能解释其意义. 此外, 模型估计要求数据满足一定的平稳性和其他一些假设, 在实际应用中较难验证. 因此, 本文的研究仍有可以改进之处, 一方面可以对模型做假设检验, 找到真正对关心的变量有影响的风险因素项, 做出更加完善和合理的解释; 另一方面, 可以放宽假设条件, 考虑更加平凡的条件假设, 如删失数据和异方差性等, 来进行更加完善的模型估计.

致谢 作者非常感谢审稿专家提出的宝贵意见和建议.

参考文献

1 Markowitz H M. Portfolio selection. J Finance, 1952, 7: 77-91
2 Black F, Scholes M. The pricing of options and corporate liabilities. J Political Economy, 1973, 81: 637-654
3 Sharpe W F. Capital asset prices: A theory of market equilibrium under conditions of risk. J Finance, 1964, 19: 425-442
4 Gerber H U, Shiu E S W. On the time value of ruin. N Am Actuar J, 1998, 2: 48-72
5 Bühlmann H. An economic premium principle. Astin Bull, 1980, 11: 52-60
6 Danielsson J, De Vries C G. Value-at-risk and extreme returns. Ann Économ Statist, 2000, 60: 239-270
7 Yamai Y, Yoshida T. Comparative analyses of expected shortfall and value-at-risk: Their estimation error, decomposition, and optimization. Monetary Econom Study, 2002, 20: 87-121
8 Acerbi C, Nordio C, Sirtori C. Expected shortfall as a tool for financial risk management. Quant Finance, 2001, 31: 379-388
9 Chen S X. Nonparametric estimation of expected shortfall. J Financ Econom, 2008, 6: 87-107

- 10 Cai Z, Wang X. Nonparametric estimation of conditional VaR and expected shortfall. *J Econometrics*, 2008, 147: 120–130
- 11 Frey R, McNeil A J. VaR and expected shortfall in portfolios of dependent credit risks: Conceptual and practical insights. *J Bank Finance*, 2002, 26: 1317–1334
- 12 Inui K, Kijima M, Kitano A. VaR is subject to a significant positive bias. *Statist Probab Lett*, 2005, 72: 299–311
- 13 Harmantzis F C, Miao L, Chien Y. Empirical study of value-at-risk and expected shortfall models with heavy tails. *J Risk Finance*, 2006, 7: 117–135
- 14 Pajor A, Osiewalski J. Bayesian value-at-risk and expected shortfall for a large portfolio (multi- and univariate approaches). *Acta Phys Polon A*, 2012, 121: B-101–B-109
- 15 Chen Q, Gerlach R, Lu Z. Bayesian value-at-risk and expected shortfall forecasting via the asymmetric Laplace distribution. *Comput Statist Data Anal*, 2012, 56: 3498–3516
- 16 Gneiting T. Making and evaluating point forecasts. *J Amer Statist Assoc*, 2011, 106: 746–762
- 17 Kuan C M, Yeh J H, Hsu Y C. Assessing value at risk with CARE, the conditional autoregressive expectile models. *J Econometrics*, 2009, 150: 261–270
- 18 Aigner D J, Amemiya T, Poirier D J. On the estimation of production frontiers: Maximum likelihood estimation of the parameters of a discontinuous density function. *Internat Econom Rev*, 1976, 17: 377–396
- 19 Newey W K, Powell J L. Asymmetric least squares estimation and testing. *Econometrica*, 1987, 55: 819–847
- 20 Xie S, Zhou Y, Wan A T K. A varying-coefficient expectile model for estimating value at risk. *J Bus Econom Statist*, 2014, 32: 576–592
- 21 Rosenblatt M. A central limit theorem and a strong mixing condition. *Proc Natl Acad Sci USA*, 1956, 42: 43–47
- 22 Cai Z, Fan J, Yao Q. Functional-coefficient regression models for nonlinear time series. *J Amer Statist Assoc*, 2000, 95: 941–956
- 23 Zhou Y, Liang H. Statistical inference for semiparametric varying-coefficient partially linear models with error-prone linear covariates. *Ann Statist*, 2009, 37: 427–458
- 24 Yao Q, Tong H. Asymmetric least squares regression estimation: A nonparametric approach. *J Nonparametr Stat*, 1996, 6: 273–292
- 25 De Rossi G, Harvey A. Quantiles, expectiles and splines. *J Econometrics*, 2009, 152: 179–185
- 26 Wang Y X, Zhou Y. The empirical study of China's stock market risk in extreme situations (in Chinese). *Chinese J Manag Sci*, 2012, 20: 20–27 [王艺馨, 周勇. 极端情形下对我国股市风险的实证研究. *中国管理科学*, 2012, 20: 20–27]
- 27 Su X, Zhou Y. New insight into application in mutual funds' performance evaluation on conditional auto regressive expectile models (in Chinese). *Chinese J Manag Sci*, 2013, 21: 22–29 [苏辛, 周勇. 条件自回归 expectile 模型及其在基金业绩评价中的应用. *中国管理科学*, 2013, 21: 22–29]
- 28 Hamidi B, Maillet B, Prigent J L. A dynamic autoregressive expectile for time-invariant portfolio protection strategies. *J Econom Dynam Control*, 2014, 46: 1–29
- 29 Liu X Q, Zhou Y. Weighted composite expectile regression estimate of autoregressive models with application (in Chinese). *Systems Eng Theory Practice*, 2016, 36: 1089–1098 [刘晓倩, 周勇. 自回归模型的加权复合 Expectile 回归估计及其应用. *系统工程理论与实践*, 2016, 36: 1089–1098]
- 30 Su X, Xie S Y, Zhou Y. Some statistical models and inferences in measurement of financial risk and their applications (in Chinese). *Oper Res Manage Sci*, 2018, 27: 185–199 [苏辛, 谢尚宇, 周勇. 金融风险度量的建模理论与方法的一些进展及其应用. *运筹与管理*, 2018, 27: 185–199]
- 31 Xie S Y, Yao H W, Zhou Y. VaR and ES measurements based on ARCH-Expectile model (in Chinese). *Chinese J Manag Sci*, 2014, 22: 1–9 [谢尚宇, 姚宏伟, 周勇. 基于 ARCH-Expectile 方法的 VaR 和 ES 尾部风险测量. *中国管理科学*, 2014, 22: 1–9]
- 32 Gerlach R, Walpole D, Wang C. Semi-parametric Bayesian tail risk forecasting incorporating realized measures of volatility. *Quant Finance*, 2017, 17: 199–215
- 33 Ahmad I, Leelahanon S, Li Q. Efficient estimation of a semiparametric partially linear varying coefficient model. *Ann Statist*, 2005, 33: 258–283
- 34 Sun Y, Carroll R, Li D. Semiparametric estimation of fixed-effects panel data varying coefficient models. In: *Nonparametric Econometric Methods. Advances in Econometrics*, vol. 25. Bingley: Emerald Group Publishing Limited, 2009, 101–129
- 35 Rodriguez-Poo J M, Soberon A. Direct semi-parametric estimation of fixed effects panel data varying coefficient models. *Econom J*, 2014, 17: 107–138
- 36 Kai B, Li R, Zou H. New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Ann Statist*, 2011, 39: 305–332
- 37 Liu X Q, Zhou Y. The semiparametric varying-coefficient composite expectile regression model in risk measurement

- and its application (in Chinese). Systems Eng Theory Practice, 2020, 40: 2176–2192 [刘晓倩, 周勇. 风险度量半参数变系数复合 expectile 回归模型及应用. 系统工程理论与实践, 2020, 40: 2176–2192]
- 38 Zhang J, Zhou Y, Cui X, et al. Semiparametric quantile estimation for varying coefficient partially linear measurement errors models. Braz J Probab Stat, 2018, 32: 616–656
- 39 Truquet L. Efficient semiparametric estimation in time-varying regression models. Statistics, 2018, 52: 590–618
- 40 Spiegel E, Sobotka F, Kneib T. Model selection in semiparametric expectile regression. Electron J Stat, 2017, 11: 3008–3038
- 41 Cai Z, Fang Y, Tian D. Assessing tail risk using expectile regressions with partially varying coefficients. J Manage Sci Eng, 2018, 3: 183–213
- 42 Fan J, Huang T. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli, 2005, 11: 1031–1057
- 43 Chen R, Tsay R S. Functional-coefficient autoregressive models. J Amer Statist Assoc, 1993, 88: 298–308
- 44 Stine R A, Tong H. Non-linear time series: A dynamical system approach. J Amer Statist Assoc, 1992, 87: 903–904
- 45 Haggan V, Ozaki T. Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. Biometrika, 1981, 68: 189–196
- 46 Cai Z, Fan J, Li R. Efficient estimation and inferences for varying-coefficient models. J Amer Statist Assoc, 2000, 95: 888–902
- 47 Cai J, Fan J, Zhou H, et al. Hazard models with varying coefficients for multivariate failure time data. Ann Statist, 2007, 35: 324–354
- 48 Fan J, Chen J. One-step local quasi-likelihood estimation. J R Stat Soc Ser B Stat Methodol, 1999, 61: 927–943
- 49 Taylor J W. Estimating value at risk and expected shortfall using expectiles. J Financ Econom, 2007, 6: 231–252
- 50 Hall P, Heyde C C. Martingale Limit Theory and Its Application. New York: Academic Press, 1980
- 51 Pollard D. Asymptotics for least absolute deviation regression estimators. Econom Theory, 1991, 7: 186–199

附录 A

首先给出一些技术假设, 以方便推导出本文的结果. 假设如下:

- (a.1) $\alpha_j(u)$ 在 $u \in \mathcal{U}$ 上是二阶连续可导的, $j = 1, \dots, p_1$;
- (a.2) 核函数 $K(\cdot)$ 在一个紧集的支撑上是有界非负对称函数;
- (a.3) $\{(Y_t, \mathbf{X}_t, \mathbf{W}_t, U_t)\}$ 中所有序列都满足 α -混合条件, 即对于 $\delta > 2$ 和 $c > 1 - 2/\delta$, 有

$$\sum_l l^c [\alpha(l)]^{1-2/\delta} \leq \infty,$$

其中 $\alpha(l) = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : -\infty < t < \infty, A \in \{x\}_{-\infty}^t, B \in \{x\}_{t+l}^{\infty}\}$, $\{x\}$ 为任意时间序列;

(a.4) 当 $T \rightarrow \infty$ 和 $h \rightarrow 0$ 时有 $Th \rightarrow \infty$; 另外, 存在一列正整数 d_T 使得 $d_T \rightarrow \infty$, $d_T = o((Th)^{\frac{1}{2}})$, 并且当 $T \rightarrow \infty$ 时, $(Th^{-1})^{\frac{1}{2}}\alpha(d_T) \rightarrow 0$;

(a.5) 令 $f_{\mathbf{X}, \mathbf{W}, U}(y)$ 是给定 $\mathbf{X} = x$, $\mathbf{W} = w$ 和 $U = u$ 下 Y 的条件密度函数, 并且假设 $f_{\mathbf{X}, \mathbf{W}, U}(y) \leq M$;

(a.6) 令 $f(\mu, \nu | x_0, x_l)$ 是给定 $(\mathbf{X}_0, \mathbf{X}_l)$ 下 (U_0, U_l) 的条件密度函数, 并且对所有的 $l \geq 1$ 都有 $f(\mu, \nu | x_0, x_l) \leq M < \infty$.

第 2 节有 $\boldsymbol{\beta} = (\mathbf{a}^T, \mathbf{b}^T)^T$, 其中 $\mathbf{a} = (\alpha_1(\cdot), \dots, \alpha_{p_1}(\cdot))^T$, $\mathbf{b} = (\alpha'_1(\cdot), \dots, \alpha'_{p_1}(\cdot))^T$. 现在设 α'_u 和 α''_u 分别为 α_u 的一阶导数和二阶导数. 又设

$$\begin{aligned} \eta(u, \mathbf{X}, \mathbf{W}) &= \sum_{j=1}^{p_1} \alpha_j(u) X_j + \sum_{i=1}^{p_2} \gamma_i W_i, \\ \eta^*(\mathbf{W}) &= \sum_{i=1}^{p_2} \gamma_i W_i, \quad \eta^{\text{new}}(u, \mathbf{X}) = \sum_{j=1}^{p_1} \alpha_j(u) X_j, \end{aligned}$$

$$\bar{\eta}(u_0, U_t, \mathbf{X}_t, \mathbf{W}_t) = \sum_{j=1}^{p_1} (\alpha_j(u_0) + \alpha'_j(u_0)(U_t - u_0))X_{tj} + \sum_{i=1}^{p_2} \gamma_i W_{ti}.$$

由 Taylor 展开有

$$\eta(u, \mathbf{X}_t, \mathbf{W}_t) = \bar{\eta}(u_0, U_t, \mathbf{X}_t, \mathbf{W}_t) + \frac{1}{2} \sum_{j=1}^{p_1} \alpha''_j(u_0) X_{tj} (U_t - u_0)^2 + o(h^2). \quad (\text{A.1})$$

为了方便处理, 本文将估计再参数化为

$$\begin{aligned} \hat{\xi} = & \sqrt{Th}(\hat{a}_1(u_0) - a_1(u_0), \dots, \hat{a}_{p_1}(u_0) - a_{p_1}(u_0), \\ & h(\hat{b}_1(u_0) - b_1(u_0)), \dots, h(\hat{b}_{p_1}(u_0) - b_{p_1}(u_0)), \\ & \hat{\gamma}_1(u_0) - \gamma_1(u_0), \dots, \hat{\gamma}_{p_2}(u_0) - \gamma_{p_2}(u_0))^T, \end{aligned} \quad (\text{A.2})$$

则有

$$\sum_{j=1}^{p_1} (\hat{a}_j + \hat{b}_j(U_t - u_0))X_{tj} + \sum_{i=1}^{p_2} \hat{\gamma}_i W_{ti} = \beta_\theta^T(u_0) \mathbf{Z}_t + \gamma_\theta^T(u_0) \mathbf{W}_t + \frac{1}{\sqrt{Th}} \hat{\xi}^T \mathbf{Z}_t^*, \quad (\text{A.3})$$

并且 $\hat{\xi}$ 最小化下面这个函数:

$$G_T(\xi) \equiv G_T(\xi; \theta, \mathbf{X}, \mathbf{W}, U, u_0) = \sum_{t=1}^T \left[Q_\theta \left(Y_t^* - \frac{1}{\sqrt{Th}} \xi^T \mathbf{Z}_t^* \right) - Q_\theta(Y_t^*) \right] K_t, \quad (\text{A.4})$$

其中 $Y_t^* = Y_t - \beta_\theta^T(u_0) \mathbf{Z}_t - \gamma_\theta^T(u_0) \mathbf{W}_t = Y_t - \bar{\eta}(u_0, U_t, \mathbf{X}_t, \mathbf{W}_t)$.

基于上述的假设条件, 有如下几个引理和定理:

引理 A.1 若假设 (a.1)–(a.6) 的条件都满足, 则当 $T \rightarrow \infty$ 时, 有下面的式子成立:

$$G_T(\xi) = \frac{1}{2} \xi^T \mathbf{D}(\theta, u_0) \xi - \frac{1}{\sqrt{Th}} S_T^T \xi + R_T(\xi), \quad (\text{A.5})$$

其中 $R_T(\xi) = o_p(1)$ 和

$$S_T = \sum_{t=1}^T Q'_\theta(Y_t^*) K_t \mathbf{Z}_t^*. \quad (\text{A.6})$$

证明 记

$$\begin{aligned} G_T(\xi) = & \mathbb{E}(G_T(\xi) \mid \mathbf{X}, \mathbf{W}, U) \\ & - \frac{1}{\sqrt{Th}} \left(\sum_{t=1}^T Q'_\theta(Y_t^*) \mathbf{Z}_t^* K_t - \mathbb{E}(Q'_\theta(Y_t^*) \mid \mathbf{X}_t, \mathbf{W}_t, U_t) \mathbf{Z}_t^* K_t \right)^T \xi + R_T^*(\xi). \end{aligned} \quad (\text{A.7})$$

现在只需证

$$\mathbb{E}(G_T(\xi) \mid \mathbf{X}, \mathbf{W}, U) = -\frac{1}{\sqrt{Th}} \mathbb{E}(Q'_\theta(Y_t^*) \mid \mathbf{X}_t, \mathbf{W}_t, U_t) \mathbf{Z}_t^{*T} K_t \xi + \frac{1}{2} \xi^T \mathbf{D}(\theta, u_0) \xi + o_p(1). \quad (\text{A.8})$$

首先设

$$\psi(a \mid \mathbf{x}, w, u) = \mathbb{E}[Q_\theta(Y - \eta(u, \mathbf{X}, \mathbf{W}) + a) \mid \mathbf{X} = \mathbf{x}, \mathbf{W} = \mathbf{w}, U = u].$$

由 (A.4), 可得

$$\begin{aligned} E(G_T(\xi) \mid \mathbf{X}, \mathbf{W}, U) &= \sum_{t=1}^T \left[\psi \left(\eta(u, \mathbf{X}_t, \mathbf{W}_t) - \bar{\eta}(u_0, U_t, \mathbf{X}_t, \mathbf{W}_t) - \frac{1}{\sqrt{Th}} \xi^T \mathbf{Z}_t^* \mid \mathbf{X}_t, \mathbf{W}_t, U_t \right) \right. \\ &\quad \left. - \psi(\eta(u, \mathbf{X}_t, \mathbf{W}_t) - \bar{\eta}(u_0, U_t, \mathbf{X}_t, \mathbf{W}_t) \mid \mathbf{X}_t, \mathbf{W}_t, U_t) \right] K_t. \end{aligned} \quad (\text{A.9})$$

由 Taylor 展开, 可得

$$\begin{aligned} E(G_T(\xi) \mid \mathbf{X}, \mathbf{W}, U) &= - \sum_{t=1}^T \psi'(\eta(u, \mathbf{X}_t, \mathbf{W}_t) - \bar{\eta}(u_0, U_t, \mathbf{X}_t, \mathbf{W}_t) \mid \mathbf{X}_t, \mathbf{W}_t, U_t) \frac{1}{\sqrt{Th}} \xi^T \mathbf{Z}_t^* K_t \\ &\quad + \frac{1}{2} \sum_{t=1}^T \psi''(\eta(u, \mathbf{X}_t, \mathbf{W}_t) - \bar{\eta}(u_0, U_t, \mathbf{X}_t, \mathbf{W}_t) \mid \mathbf{X}_t, \mathbf{W}_t, U_t) \left(\frac{1}{\sqrt{Th}} \xi^T \mathbf{Z}_t^* \right)^2 K_t \{1 + o_p(1)\} \\ &= - \frac{1}{\sqrt{Th}} \sum_{t=1}^T E(Q'_\theta(Y_t^*) \mid \mathbf{X}_t, \mathbf{W}_t, U_t) \mathbf{Z}_t^{*T} K_t \xi \\ &\quad + \frac{1}{2Th} \xi^T \left(\sum_{t=1}^T K_t \psi''(0 \mid \mathbf{X}_t, \mathbf{W}_t, U_t) \mathbf{Z}_t^* \mathbf{Z}_t^{*T} \right) \xi \{1 + o_p(1)\}. \end{aligned} \quad (\text{A.10})$$

由引理 A.3 可以证明 $R_T^*(\xi) = o_p(1)$, 又由 (A.7)、(A.10) 和引理 A.4 即可得到 (A.5) 成立, 其中 $R_T(\xi) = R_T^*(\xi) + o_p(1) = o_p(1)$. 证毕. $\square$

引理 A.2 对任意的 $x, y \in \mathbb{R}^{2p_1+p_2}$ 和 $\theta \in (0, 1)$, 有下面的不等式成立:

$$|Q_\theta(x+y) - Q_\theta(x) - Q'_\theta(x)y| \leq 4y^2 \quad (\text{A.11})$$

并且

$$|Q'_\theta(x+y) - Q'_\theta(x) - Q''_\theta(x)y| \leq 4|y|, \quad (\text{A.12})$$

其中 $Q'_\theta = \frac{dQ_\theta(x)}{dx}$, 对 $x \neq 0$ 有 $Q''_\theta = \frac{d^2Q_\theta(x)}{dx^2}$, 并且 $Q''_\theta(0) = 0$.

证明 参见文献 [24]. $\square$

引理 A.3 对 $\xi \in \mathbb{R}^{2p_1+p_2}$, 依概率有 $R_T^*(\xi) \xrightarrow{P} 0$.

证明 由 (A.7) 可得

$$R_T^*(\xi) = \sum_{t=1}^T [V_t - E(V_t \mid \mathbf{X}_t, \mathbf{W}_t, U_t)], \quad (\text{A.13})$$

其中

$$V_t = \left[Q_\theta \left(Y_t^* - \frac{1}{\sqrt{Th}} \xi^T \mathbf{Z}_t^* \right) - Q_\theta(Y_t^*) + Q'_\theta(Y_t^*) \frac{1}{\sqrt{Th}} \xi^T \mathbf{Z}_t^* \right] K_t. \quad (\text{A.14})$$

注意到 $ER_T^*(\xi) = 0$, 由引理 A.2 可得

$$EV_1^2 \leq 16E \left[\left(\frac{1}{\sqrt{Th}} \xi^T \mathbf{Z}_1^* \right)^4 K_1^2 \right] = O \left(\frac{1}{T^2 h} \right). \quad (\text{A.15})$$

因此有

$$ER_T^{*2} \leq TEV_1^2 + 2 \sum_{l=1}^{T-1} (T-l) \text{Cov}(V_1, V_{1+l}), \quad (\text{A.16})$$

其中 $TEV_1^2 = O(\frac{1}{Th})$.

设 $d_T \rightarrow \infty$ 是一列正整数, 并且使得 $d_T h_T \rightarrow 0$. 定义

$$J_1 = \sum_{l=1}^{d_T-1} T |\text{Cov}(V_1, V_{1+l})|, \quad J_2 = \sum_{l=d_T}^{T-1} T |\text{Cov}(V_1, V_{1+l})|. \quad (\text{A.17})$$

由 Davydov 引理 [50], 可以得到如下的结论:

$$|\text{Cov}(V_1, V_{1+l})| \leq k[\alpha(l)]^{1-2/\delta} [E|V_1|^\delta]^{2/\delta}. \quad (\text{A.18})$$

由引理 A.2 和假设 (a.2) 可知, 存在常数 M , 使得

$$E|V_1|^\delta \leq 4^\delta E\left(\frac{1}{\sqrt{Th}} \xi^T \mathbf{Z}_1^*\right)^{2\delta} K_1^\delta \leq MT^{-\delta} h^{1-\delta}. \quad (\text{A.19})$$

而由假设 (a.3), 有

$$\begin{aligned} J_2 &\leq T k M^{2/\delta} T^{-2} h^{2/\delta-2} \sum_{l=d_T}^{\infty} [\alpha(l)]^{1-2/\delta} \\ &\leq k M^{2/\delta} T^{-1} h^{2/\delta-2} d_T^{-c} \sum_{l=d_T}^{\infty} l^c [\alpha(l)]^{1-2/\delta} \\ &= o\left(\frac{1}{Th}\right). \end{aligned} \quad (\text{A.20})$$

选取 d_T 使得对 $\delta > 2$, $d_T^c h^{1-2/\delta} = M_0$ 并且满足 $d_T h \rightarrow 0$, 则由 Cauchy-Schwarz 不等式, 有

$$|\text{Cov}(V_1, V_{1+l})| \leq MEV_1^2 = O\left(\frac{1}{T^2 h}\right). \quad (\text{A.21})$$

再由假设 (a.4), 可以得到

$$J_1 \leq T d_T O\left(\frac{1}{T^2 h}\right) = o\left(\frac{1}{Th^2}\right). \quad (\text{A.22})$$

因此, 由 Chebyshev 不等式, 有

$$P\{|R_T^*(\xi)| > \epsilon\} \leq \frac{E[R_T^{*2}(\xi)]}{\epsilon^2} = o(1). \quad (\text{A.23})$$

证毕. □

引理 A.4 若假设 (a.1)–(a.6) 的条件都满足, 则当 $T \rightarrow \infty$ 时, 有下面的式子成立:

$$\frac{1}{Th} \sum_{l=1}^T K_l \psi''(0 | \mathbf{X}_t, \mathbf{W}_t, U_t) \mathbf{Z}_t^* \mathbf{Z}_t^{*\top} \rightarrow \mathbf{D}(\theta, u_0). \quad (\text{A.24})$$

证明 由 ψ 的定义, 有

$$\psi''(0 | \mathbf{X}, \mathbf{W}, U) = 2\{\theta[1 - F_{\mathbf{X}, \mathbf{W}, U}(\eta(u, \mathbf{X}, \mathbf{W}))] + (1 - \theta)F_{\mathbf{X}, \mathbf{W}, U}(\eta(u, \mathbf{X}, \mathbf{W}))\}. \quad (\text{A.25})$$

由平稳序列的遍历性可以得到

$$\frac{1}{Th} \sum_{t=1}^T K_t \psi''(0 | \mathbf{X}_t, \mathbf{W}_t, U_t) \mathbf{Z}_t^* \mathbf{Z}_t^{*\text{T}} - \frac{1}{h} \mathbb{E}[K_t \psi''(0 | \mathbf{X}_t, \mathbf{W}_t, U_t) \mathbf{Z}_t^* \mathbf{Z}_t^{*\text{T}}] = o_p(1). \quad (\text{A.26})$$

再经简单计算可得

$$\begin{aligned} & \mathbb{E}[K_t \psi''(0 | \mathbf{X}_t, \mathbf{W}_t, U_t) \mathbf{Z}_t^* \mathbf{Z}_t^{*\text{T}}] \\ &= \mathbb{E} \left[K_t \psi''(0 | \mathbf{X}_t, \mathbf{W}_t, U_t) \begin{pmatrix} 1 & \frac{U_t - u_0}{h} & 1 \\ \frac{U_t - u_0}{h} & \frac{(U_t - u_0)^2}{h^2} & \frac{U_t - u_0}{h} \\ 1 & \frac{U_t - u_0}{h} & 1 \end{pmatrix} * (\mathbf{X}^{\text{T}}, \mathbf{X}^{\text{T}}, \mathbf{W}^{\text{T}})^{\text{T}} (\mathbf{X}^{\text{T}}, \mathbf{X}^{\text{T}}, \mathbf{W}^{\text{T}}) \right] \\ &= h f_U(u_0) \begin{pmatrix} \mu_0 & \mu_1 & \mu_0 \\ \mu_1 & \mu_2 & \mu_1 \\ \mu_0 & \mu_1 & \mu_0 \end{pmatrix} * \boldsymbol{\Gamma}(\theta, u_0) \{1 + o(1)\}. \end{aligned} \quad (\text{A.27})$$

因此有 (A.24) 成立. $\square$

引理 A.5 若假设 (a.1)–(a.6) 的条件都满足, 则当 $T \rightarrow \infty$ 时, 有

$$\frac{1}{\sqrt{Th}} \left[S_T - \frac{h^2}{2} f_U(u_0) \begin{pmatrix} \mu_2 \mathbf{B}_1 \\ \mu_3 \mathbf{B}_1 \\ \mu_2 \mathbf{B}_2 \end{pmatrix} + o(h^2) \right] \xrightarrow{\mathcal{L}} N(0, \boldsymbol{\Sigma}(\theta, u_0)), \quad (\text{A.28})$$

其中

$$\mathbf{B}_1 = \mathbb{E}[\mathbf{X} \mathbf{X}^{\text{T}} \boldsymbol{\alpha}''(u_0) (2(1 - \theta)I(Y - \eta(U, \mathbf{X}, \mathbf{W}) \leq 0) + 2\theta I(Y - \eta(U, \mathbf{X}, \mathbf{W}) > 0))], \quad (\text{A.29})$$

$$\mathbf{B}_2 = \mathbb{E}[\mathbf{W} \mathbf{X}^{\text{T}} \boldsymbol{\alpha}''(u_0) (2(1 - \theta)I(Y - \eta(U, \mathbf{X}, \mathbf{W}) \leq 0) + 2\theta I(Y - \eta(U, \mathbf{X}, \mathbf{W}) > 0))], \quad (\text{A.30})$$

$$\boldsymbol{\Sigma}(\theta, u_0) = f_U(u_0) \begin{pmatrix} \kappa_0 & \kappa_1 & \kappa_0 \\ \kappa_1 & \kappa_2 & \kappa_1 \\ \kappa_0 & \kappa_1 & \kappa_0 \end{pmatrix} * \mathbf{A}(\theta, u_0) \quad (\text{A.31})$$

并且

$$\mathbf{A}(\theta, u_0) = \mathbb{E}[Q_\theta^2(\eta(u_0, \mathbf{X}, \mathbf{W})) (\mathbf{X}^{\text{T}}, \mathbf{X}^{\text{T}}, \mathbf{W}^{\text{T}})^{\text{T}} (\mathbf{X}^{\text{T}}, \mathbf{X}^{\text{T}}, \mathbf{W}^{\text{T}}) | U = u_0]. \quad (\text{A.32})$$

证明 首先证明, 当 $T \rightarrow \infty$ 时, 有

$$\mathbb{E}[\mathbf{Z}_1^* Q'_\theta(Y_1^*) K_1] = \frac{h^3}{2} f_U(u_0) \begin{pmatrix} \mu_2 B_1 \\ \mu_3 B_1 \\ \mu_2 B_2 \end{pmatrix} \{1 + o(1)\}, \quad (\text{A.33})$$

$$E[\mathbf{Z}_1^* Q'_\theta(Y_1^*) K_1]^{\otimes 2} = hf_U(u_0) \begin{pmatrix} \kappa_0 & \kappa_1 & \kappa_0 \\ \kappa_1 & \kappa_2 & \kappa_1 \\ \kappa_0 & \kappa_1 & \kappa_0 \end{pmatrix} * \mathbf{A}(\theta, u_0) \{1 + o(1)\}. \quad (\text{A.34})$$

注意到 $E[\mathbf{Z}_1^* Q'_\theta(Y_1 - \eta(U_1, \mathbf{X}_1, \mathbf{W}_1)) K_1] = 0$ 和 (A.1), 则有

$$\begin{aligned} & E[\mathbf{Z}_1^* Q'_\theta(Y_1^*) K_1] \\ &= E[\mathbf{Z}_1^* Q'_\theta(Y_1 - \bar{\eta}(u_0, U_1, \mathbf{X}_1, \mathbf{W}_1)) K_1] \\ &= E\left[\mathbf{Z}_1^* Q''_\theta(Y_1 - \eta(U_1, \mathbf{X}_1, \mathbf{W}_1)) K_1 \mathbf{X}_1^T \left(\frac{1}{2} \alpha''(u_0)(U_1 - u_0)^2\right)\right] \{1 + o(1)\} \\ &= E\left[\frac{K_1}{2} \left(\mathbf{X}_1^T, \frac{U_1 - u_0}{h} \mathbf{X}_1^T, \mathbf{W}_1^T\right)^T \mathbf{X}_1^T \alpha''(u_0)(U_1 - u_0)^2\right. \\ &\quad \left. \times (2(1 - \theta)I(Y_1 - \eta(U_1, \mathbf{X}_1, \mathbf{W}_1) \leq 0) + 2\theta I(Y_1 - \eta(U_1, \mathbf{X}_1, \mathbf{W}_1) > 0))\right] \{1 + o(1)\} \\ &= \frac{h^3}{2} f_U(u_0) \begin{pmatrix} \mu_2 B_1 \\ \mu_3 B_1 \\ \mu_2 B_2 \end{pmatrix} \{1 + o(1)\}. \end{aligned} \quad (\text{A.35})$$

同样地, 有

$$\begin{aligned} & E[\mathbf{Z}_1^* Q'_\theta(Y_1^*) K_1]^{\otimes 2} \\ &= E\left[K_1^2 Q'^2_\theta(Y_1^*) \begin{pmatrix} \mathbf{X}_1 \mathbf{X}_1^T & \frac{U_1 - u_0}{h} \mathbf{X}_1 \mathbf{X}_1^T & \mathbf{X}_1 \mathbf{W}_1^T \\ \frac{U_1 - u_0}{h} \mathbf{X}_1 \mathbf{X}_1^T & \frac{(U_1 - u_0)^2}{h^2} \mathbf{X}_1 \mathbf{X}_1^T & \frac{U_1 - u_0}{h} \mathbf{X}_1 \mathbf{W}_1^T \\ \mathbf{W}_1 \mathbf{X}_1^T & \frac{U_1 - u_0}{h} \mathbf{W}_1 \mathbf{X}_1^T & \mathbf{W}_1 \mathbf{W}_1^T \end{pmatrix}\right] \\ &= hf_U(u_0) \begin{pmatrix} \kappa_0 & \kappa_1 & \kappa_0 \\ \kappa_1 & \kappa_2 & \kappa_1 \\ \kappa_0 & \kappa_1 & \kappa_0 \end{pmatrix} * \mathbf{A}(\theta, u_0) \{1 + o(1)\}. \end{aligned} \quad (\text{A.36})$$

而 S_T 的渐近正态性可以用 Cramer-Wold 方法进行证明, 即对任意单位向量 $\mathbf{e} \in \mathbb{R}^{2p_1 + p_2}$, 设

$$L_T = \frac{\mathbf{e}^T}{\sqrt{Th}} (S_T - ES_T), \quad (\text{A.37})$$

有如下的结果:

$$L_T \xrightarrow{\mathcal{L}} N(0, \mathbf{e}^T \boldsymbol{\Sigma}(\theta, u_0)). \quad (\text{A.38})$$

详细证明参见文献 [20]. 由 (A.38) 可知, (A.28) 成立. $\square$

定理 2.1 的证明 由凸引理^[51]可知, 凸函数 $G_T(\xi)$ 的极小值点 $\hat{\xi}$ 依概率收敛到凸函数

$$\frac{1}{2} \xi^T \mathbf{D}(\theta, u_0) \xi - \frac{1}{\sqrt{Th}} S_T^T \xi$$

的极小值点 $-\frac{\mathbf{D}(\theta, u_0)^{-1}S_T}{\sqrt{Th}}$. 由引理 A.1 可知, 在紧集 $\xi \in \mathcal{K}$ 上一致地有

$$\hat{\xi} = -\frac{\mathbf{D}(\theta, u_0)^{-1}S_T}{\sqrt{Th}} + o_p(1). \quad (\text{A.39})$$

因此, 在 $u_0 \in \mathcal{U}$ 上一致地有

$$\begin{pmatrix} \hat{\mathbf{a}}_\theta - \mathbf{a}_\theta \\ h(\hat{\mathbf{b}}_\theta - \mathbf{b}_\theta) \\ \hat{\gamma}_\theta - \gamma_\theta \end{pmatrix} = -\frac{\mathbf{D}(\theta, u_0)^{-1}S_T}{Th} + o_p(1). \quad (\text{A.40})$$

再由引理 A.5 即可得证定理 2.1. □

定理 2.4 的证明 由 (A.40)、引理 A.5 和 Slutsky 定理即可以证明

$$\begin{aligned} & \sqrt{Th} \left[\begin{pmatrix} \hat{\mathbf{a}}_\theta(u_0) - \mathbf{a}_\theta(u_0) \\ h(\hat{\mathbf{b}}_\theta(u_0) - \mathbf{b}_\theta(u_0)) \\ \hat{\gamma}_\theta(u_0) - \gamma_\theta(u_0) \end{pmatrix} - \frac{h^2}{2} \mathbf{D}(\theta, u_0)^{-1} f_U(u_0) \begin{pmatrix} \mu_2 \mathbf{B}_1 \\ \mu_3 \mathbf{B}_1 \\ \mu_2 \mathbf{B}_2 \end{pmatrix} + o(h^2) \right] \\ & \xrightarrow{\mathcal{L}} N(0, \mathbf{D}(\theta, u_0)^{-1} \boldsymbol{\Sigma}(\theta, u_0) \mathbf{D}(\theta, u_0)^{-1}). \end{aligned} \quad (\text{A.41})$$

证毕. □

因为定理 2.2、2.3、2.5 和 2.6 都退化为前人研究中的估计, 其相合性和渐近正态性的证明详见文献 [17, 20].

Semiparametric varying-coefficient expectile model for estimating value at risk on dependent samples

Zijian Wang, Yong Zhou & Fanping Zeng

Abstract We introduce a semiparametric varying-coefficient model to study the expectile-based value at risk (EVaR) based on the α -mixing assumption. The model considers the risk factors as well as the dynamic structure and the interaction of these factors. Meanwhile, EVaR not only has great advantages such as the subadditivity or easy to calculate, but also is very sensitive to the scale of losses compared with the conventional quantile-based value at risk (QVaR), which makes it as a more mature and efficient way of the risk measure in extreme risk situations. We develop a three-stage method to estimate the parameters of the varying-coefficient part and the constant-coefficient part. We also establish the challenging consistency and the asymptotic normality of all the resultant estimation in three stages under the time series samples. To save computing time, we propose to use a one-step procedure to compute the estimation. Financial time series samples are not independent data. It has more challenge to establish the large sample properties of the estimators. We introduce the dividing methods to prove the limit theory of the α -mixing series, and obtain the asymptotic properties of the parametric constant coefficients and the nonparametric varying-coefficients. In our simulation, three models' results show the accuracy and robustness of our estimation. The real financial data example performs as an application of our model at the end of our study.

Keywords α -mixing, semiparametric, varying-coefficient, expectile, value at risk, three-stage estimation

MSC(2020) 62G05, 91G70

doi: 10.1360/N012019-00079