中国科学: 数学　2025 年　第 55 卷　第 ? 期: 1 ～ 30
SCIENTIA SINICA Mathematica
《中国科学》杂志社
SCIENCE CHINA PRESS
CrossMark
click for updates
论 文

可加单指标协整时间序列模型的稳健 M 估计

董朝华[1], 周晨[1], 涂云东[2*]

1. 中南财经政法大学统计与数学学院, 数量经济研究中心, 武汉 430073;
2. 北京大学光华管理学院, 北京 100871
E-mail: dchaohua@zuel.edu.cn, zhouchen@stu.zuel.edu.cn, yundong.tu@gsm.pku.edu.cn

收稿日期: 2024-06-02; 接受日期: 2025-06-23; 网络出版日期: 2025-08-21; * 通信作者
国家自然科学基金 (批准号: 72473156, 72073143, 72073002, 72425009 和 72495123) 和中南财经政法大学中央高校基本科研业务费 (批准号: 2722022EG001) 资助项目

摘要　本文研究当误差的损失函数是一类非光滑函数时, 协整时间序列的可加单指标参数模型的稳健 M 估计. 这类非光滑函数可能具有纽结点, 包括但不局限于最小绝对偏差、分位数和 Huber 损失函数. 由于时间序列的非平稳性, 现有方法无法推导这类极值估计量的渐近理论. 损失函数的非光滑性是研究这类极值估计量渐近性质技术难题的根源, 而广义函数族提供了一种用正则光滑函数序列逼近非光滑函数的简洁方法, 从而克服了这一难题. 借助这种正则逼近, 本文得到非光滑损失函数对应的 M 估计量的渐近分布. 本文通过模拟数据和一个股票收益预测回归的实证分析来评估所提出的估计方法和理论的有限样本性能.

关键词　可加单指标模型　广义函数　非平稳时间序列　二次逼近　正则函数序列　稳健 M 估计

MSC (2020) 主题分类　62E20, 62F35, 62J02

1　引言

构建稳健且适用于估计和预测的模型是经济和金融中许多实证研究的重要组成部分. 一方面, 真实世界的数据经常表现出非线性和非平稳性等特征; 另一方面, 近年来随着经济、金融、气候、医学、社会学等学科数据的易获得性不断增加, 人们对这些学科领域的理论研究和实证数据分析的兴趣也越来越大. 因此, 模型构建的一个重要考量是如何将这些来源广泛的多种数据特征纳入模型, 尤其是计量经济学的一些内在问题, 例如, 许多原始数据看起来高度依赖, 甚至不平稳. 简单地使用原始数据的变换 (如差分) 可能会完全忽略原始数据的一些关键特征, 如趋势行为等, 这些特征也是气候学、能源学、流行病学、卫生和宏观经济学等时间序列数据建模中的关键兴趣点. 本文的研究涉及两个层面的稳健性. 首先, 提出一类模型以适用于多种类型的数据, 如具有确定性趋势的平稳时间序列数据、高度依赖和非平稳的时间序列数据. 其次, 构建的估计方法对具有异常值的数据具有稳健性, 且相关的计算算法对实证分析和从业者都是便于操作和易于实现的.

英文引用格式: Dong C H, Zhou C, Tu Y D. Robust M estimation for additive single-index cointegrating time series models (in Chinese). Sci Sin Math, 2025, 55: 1–30, doi: 10.1360/SSM-2024-0187

　www.scichina.com　mathcn.scichina.com

本文所提出的一类可加单指标参数模型形式如下:

$$y_t=\sum_{j=1}^{p_1}\gamma_{1j}^0\,g_{1j}(x_t^\top\theta_{1j}^0)+\sum_{j=1}^{p_2}\gamma_{2j}^0\,g_{2j}(z_t^\top\theta_{2j}^0)+e_t,\quad t=1,\dots,n, \tag{1.1}$$

其中对于 $j=1,\dots,p_i$ 和 $i=1,2$, $\gamma_{ij}^0\in\mathbb{R}$ 和 $\theta_{ij}^0\in\mathbb{R}^{d_i}$ 都是未知参数, $g_{ij}(\cdot)$ 是已知函数, p_1 和 p_2 是已知正整数, x_t 是 d_1 维的非平稳时间序列向量, z_t 是 d_2 维的趋势平稳时间序列向量, e_t 是误差项. 为了方便, 记 $\gamma_0=(\gamma_{11}^0,\dots,\gamma_{1p_1}^0,\gamma_{21}^0,\dots,\gamma_{2p_2}^0)^\top$, $\theta_0=(\theta_{11}^{0\top},\dots,\theta_{1p_1}^{0\top},\theta_{21}^{0\top},\dots,\theta_{2p_2}^{0\top})^\top$. 回归变量 x_t 是一阶单整的, 它和 z_t 分别满足

$$x_t=x_{t-1}+w_t\quad 和\quad z_t=h(\tau_t,v_t), \tag{1.2}$$

其中向量 w_t 是一个线性过程 (确切定义见下节), $\tau_t=t/n$, $h(\cdot)$ 是一个由时间趋势 τ_t 和平稳变量 v_t 驱动的已知向量函数. 众所周知, 一阶单整过程 x_t 可能描述 "大" 趋势行为, 而 z_t 是 (τ_t,v_t) 的函数, 可以描述 "小" 波动. 因为我们仅要求函数 $h(\cdot)$ 具有连续性, z_t 的形式包含了很多可能性. 例如, (1) $z_t=h(\tau_t)+v_t$, 一个与时间趋势向量叠加的平稳向量; (2) $z_t=\sigma(\tau_t)^\top v_t$, 具有变系数的 v_t 的一个组合; (3) $z_t=\sigma_1(\tau_t)+\sigma_2(\tau_t)^\top v_t$. 以上 $\sigma(\cdot)$, $\sigma_1(\cdot)$, $\sigma_2(\cdot)$ 和 v_t 的维数均与 z_t 相匹配. 这些 z_t 在均值和波动率分量中分别包含确定性和随机趋势行为.

与现有文献不同, 模型 (1.1) 不仅考虑了 $x_t^\top\theta_{1j}^0$ 驱动的 "大" 随机趋势, 而且考虑了以确定性时间趋势 τ_t 为主要代表的 "小" 缓慢变化 $z_t^\top\theta_{2j}^0$. 另外, 由于多个单指标结构的使用, 模型 (1.1) 可以处理较大维数的数据, 这在某种程度上迎合了大数据时代的需求. 注意 γ_{ij}^0 和 $g_{ij}(\cdot)$ 的乘积可能会引起识别问题. 例如, 如果 $g_{11}(u)=u$, 则 $\gamma_{11}^0g_{11}(x_t^\top\theta_{11}^0)=\gamma_{11}^0x_t^\top\theta_{11}^0$. 为了便于识别, 本文假设对于所有的 $1\leqslant j\leqslant p_i$ 和 $i=1,2$, 有 $\|\theta_{ij}^0\|=1$, 并且 θ_{ij}^0 的第一个元素是正的, 虽然这在 $g_{ij}(\cdot)$ 的非线性情形下是不必要的. 模型 (1.1) 属于非线性参数模型, 是模型构建里的一大类问题; 除了自身的价值以外, 此类模型的研究还为模型的识别性检验奠定了基础 (参见文献 [12, 15]). 因此, 我们认为在研究模型 (1.1) 的各种半参数形式的估计和推断问题之前, 研究参数模型 (1.1) 的估计是有必要的.

现有文献对含有非平稳时间序列的非线性参数模型的研究比较少见. Park 和 Phillips [35] 作了一个开创性工作, 他们研究的回归模型为 $y_t=f(x_t,\theta_0)+u_t$, 其中 f 已知, 单位根过程 x_t 是标量. Chang 等 [9] 以及 Chang 和 Park [8] 分别研究了单整过程和趋势项以及单指标结构的参数模型. Li 等 [31] 和 Bravo 等 [4] 分别研究了 Markov 链的非线性模型的一般估计和稳健估计. 相反, 此类非参数模型方面的研究比较多. 例如, Xiao [49] 研究了变系数半参数模型, 回归变量是单位根过程但其系数是未知函数; Dong 等 [14] 研究了单位根过程的半参数单指标模型和部分线性单指标模型; Dong 等 [15] 研究了一个非参数模型的检验问题, 其中的协变量有平稳和非平稳两种情形; Wang 等 [46] 考虑了非平稳模型的识别性检验问题; Lin 等 [32] 研究了双变换协整模型的估计问题; Dong 和 Linton [16] 考虑了非参数可加非平稳时间序列模型, Dong 等 [17] 研究了非参数非平稳时间序列模型. 文献 [16, 17] 的回归变量均包含时间趋势、平稳变量和单位根过程, 但它们都是标量. 据我们所知, 像模型 (1.1) 这样包含如此复杂数据的参数模型在文献中首次被提出.

本文采用稳健的 M 估计 (为简便起见, 以下简称稳健估计) 方法估计所有未知参数. 关于稳健估计已有大量的文献. 从文献 [23] 关于 Huber 损失的开创性工作开始, 稳健估计方法的研究逐渐扩展到一般的损失函数 (参见文献 [2, 4, 18, 50]). 为使用稳健估计, 先定义下列目标函数:

$$L_n(\theta,\gamma)=\sum_{t=1}^n\rho\bigg(y_t-\sum_{j=1}^{p_1}\gamma_{1j}g_{1j}(x_t^\top\theta_{1j})-\sum_{j=1}^{p_2}\gamma_{2j}g_{2j}(z_t^\top\theta_{2j})\bigg), \tag{1.3}$$

其中 $\rho(\cdot)$ 是某个损失函数, θ 和 γ 代表某个紧集 Θ 中的任意向量, 而真实向量 (θ_0,γ_0) 为该紧集的内点. 然后定义未知参数 (θ_0,γ_0) 的估计量 $(\widehat{\theta},\widehat{\gamma})$ 为

$$(\widehat{\theta},\widehat{\gamma})=\mathop{\arg\min}_{(\theta,\gamma)\in\Theta} L_n(\theta,\gamma). \tag{1.4}$$

应当指出, 在稳健估计文献中最重要的损失函数是, (1) 最小绝对偏差 (least absolute deviation, LAD), $\rho(u)=|u|$; (2) 分位数损失函数 [27], $\rho_\tau(u)=u(\tau-I(u<0))$, $\tau\in(0,1)$; (3) Huber 损失 [23],

$$\rho_c(u)=\begin{cases}\dfrac{1}{2}u^2, & |u|\leqslant c,\\ c|u|-\dfrac{1}{2}c^2, & |u|>c,\end{cases}$$

其中 $c>0$ 为预置常数. 前两个损失函数是不光滑的, 而第三个损失函数在 $\pm c$ 处没有二阶导数. 这些损失函数在数轴两端均为线性函数 (而不像最小二乘法使用的二次函数), 因此相应的估计量受异常值影响较小.

然而, 上述估计量渐近理论的建立非常具有挑战性. 我们知道, 极值估计量的渐近理论首先在可识别性条件下建立它的相合性; 一般而言这是最困难的一步, 因为往往极值估计量没有显性解, 需要建立目标函数的一致大数定律 (uniform law of large numbers). 然后, 在目标函数的一阶偏导数和二阶偏导数的极限建立的基础上, 推导极值估计量的中心极限定理 (参见文献 [44, 45]). 可见, 即使对平稳时间序列数据和常见的非光滑损失下参数模型的稳健估计有一些结果, 但是在时间序列非平稳的情形下, 现有方法不再适用, 新的渐近分析亟待研究.

为了解决这一理论难题, 本文采用广义函数方法来处理非光滑损失函数 $\rho(\cdot)$, 并通过二次逼近目标函数来建立 M 估计的渐近性质. 众所周知, 广义函数的任意阶导数都存在, 并且广义函数是无限次可微正则函数序列的极限 (参见文献 [21, 24]). 特别地, 在广义函数意义下, LAD 和分位数损失的二阶导函数是 Dirac 函数 $\delta(\cdot)$. 然而, 广义函数不能视为普通函数, 因为它们在某些点上可能是没有定义的, 但是在积分里它们的运算有意义. 这是调和分析里常用的技巧 (参见文献 [42]). 例如, 假设 $f_e(x)$ 是随机变量 e 的密度函数, 在 $x=0$ 处连续, 有

$$\mathbb{E}[\delta(e)]=\int\delta(x)f_e(x)dx=f_e(0).$$

因此, 在 M 估计中, 我们只在期望运算下使用广义函数; 如果必须考虑损失函数导数, 则使用附录 A 中定义的正则序列逼近广义函数. 对于不可微凸损失函数 $\rho(\cdot)$, 其次梯度 $\psi(\cdot)$ 在有效域上始终存在 (参见文献 [41]); 如果它在某点可微, 则该点的次梯度与一阶导数相同. 附录 A 中定义了一个逼近不可微 $\rho(\cdot)$ 的正则函数序列 $\{\rho_m(\cdot), m=1,2,\ldots\}$, 其中 $\rho_m(\cdot)$ 是无限可微的, 在广义函数的意义下, 对于任意的非负整数 k, 当 $m\to\infty$ 时, $\rho_m^{(k)}(u)\to\rho^{(k)}(u)$. 更重要的是, 本文建立了 $\rho_m(u)-\rho(u)$, $\rho_m'(e_t)-\psi(e_t)$ 和 $\mathbb{E}[\rho_m''(e_t)-\rho''(e_t)]$ 的收敛速度, 其中 $\rho''(\cdot)$ 为广义函数, 使得 $\mathbb{E}[\rho''(e_t)]$ 存在. 据我们所知, 这样的收敛速度在文献里是首次出现, 虽然 Phillips [36, 37] 在研究线性模型的稳健估计量的渐近性质时使用了广义函数方法, 但是该方法还不够成熟, 缺乏一些基础性结果的理论推导. 相比之下, 在这些收敛速度的基础上, 本文采用参数的二次型 Q_n 逼近公式 (1.3) 所定义的目标函数 L_n, 而这些二次型的最小值有显性表达式, 并且 L_n 与 Q_n 的最小值之间的差异可以忽略不计, 因此 L_n 的极值估计量与 Q_n 的最小值有相同的极限. 这个方法比现有文献中的方法更简单, 更有普适性 (参见文献 [2, 4, 18, 20, 26, 36, 37, 39, 44, 50]).

本文余下内容的结构如下. 第 2 节给出可加单指标模型的假设条件. 第 3 节在 $g_{1j}(\cdot)$ 分别为 H- 正则和 I- 正则两种情形下建立模型 (1.1) 中所有参数的估计量的相应渐近分布. 第 4 节是有关 Monte Carlo 模拟的结果, 以检验本文的估计量在有限样本情形下的表现. 第 5 节讨论一个关于股票收益的实证研究. 第 6 节给出结论. 重要的引理在附录 A 和 B 里给出, 主要结果的证明在附录 C 中给出, 其余的技术细节在补充材料的附录 D 和 E 中.

本文使用以下符号. $\|\cdot\|$ 是向量的 Euclid 范数或矩阵的 F 范数; $\equiv$ 意味着被定义为相等; $\int$ 表示在 $\mathbb{R}$ 上的积分, 对于多元函数 $f(u)$, $\int f(u)du$ 理解为多元积分; C 和 C_1 等表示绝对常数, 在每次出现时可能不同; 只要不会引起混淆, $\dot{f}$ 和 f' 都表示一阶导数, $\ddot{f}$ 和 f'' 都表示二阶导数; 上标 $\top$ 表示矩阵的转置; PR^2 是伪拟合度.

2 模型的假设

对于可加单指标参数模型 (1.1), 本文使用广义函数方法来建立所提出的稳健估计量的渐近极限. 为此, 给出以下假设.

假设 2.1 (i) 令 $\{\eta_j, j\in\mathbb{Z}\}$ 为 d_1 维独立同分布 (independent and identically distributed, i.i.d.) 连续随机向量序列, 满足 $\mathbb{E}\eta_0=0$, $\mathbb{E}\eta_0\eta_0^\top=I_{d_1}$ 和 $\mathbb{E}\|\eta_0\|^4<\infty$. 设 $\varphi(u)$ 为 η_0 的特征函数, 满足 $\int\|u\|\,|\varphi(u)|du<\infty$.

(ii) 定义线性过程 $w_t=\sum_{j=0}^{\infty}A_j\eta_{t-j}$, 其中 $\{A_j\}$ 是方阵序列, $A_0=I_{d_1}$, $\sum_{j=0}^{\infty}j\|A_j\|<\infty$. 此外, $A\equiv\sum_{j=0}^{\infty}A_j$ 是满秩的.

(iii) 对于 $t\geqslant 1$, 令 $x_t=x_{t-1}+w_t$, $x_0=O_P(1)$. 此外, 对于每个给定的 $1\leqslant j\leqslant p_1$, $x_t^\top\theta_{1j}^0$ 仍为单位根过程.

单位根过程 x_t 的这种结构在文献中广为使用, 可参见文献 [14,16,33,35]. 根据平稳时间序列 Wold 分解定理 (参见文献 [5, 第 187 页]), w_t 具有相当普遍的形式, 它涵盖了许多特殊情形. 如果 $A_j=0$, $j\geqslant 1$, 则 $w_t=\eta_t$, 此时 x_t 成为一个随机游走过程; 如果 $A_j=0$, $j\geqslant p$, 则 w_t 是 p 阶移动平均过程; 一般情形下, w_t 具有序列相关性. 计算可得 $d_t^2\equiv\mathbb{E}(x_tx_t^\top)=AA^\top t(1+o(1))$. 根据如上文献, 对于 $r\in[0,1]$, 当 $n\to\infty$ 时, $n^{-1/2}x_{[nr]}\to_D B(r)$, 这里 $B(r)$ 是 $[0,1]$ 上的 Brown 运动, 协方差为 $AA^\top$.

假设 2.2 (i) $h(r,v)$ 是关于 $r\in[0,1]$ 连续的 d_2 维向量函数.

(ii) 假设 $v_t=q(\eta_t,\ldots,\eta_{t-d_0+1};\tilde{v}_t)$, 其中 d_0 为固定的正整数, $\{\tilde{v}_t, 1\leqslant t\leqslant n\}$ 是严平稳且 α- 混合的时间序列向量, 与假设 2.1 中的 $\{\eta_j\}$ 独立, 并且向量函数 $q(\ldots)$ 是可测的, 使得

$$\sup_{r\in[0,1],j=1,\ldots,p_2}\mathbb{E}\|\dot{g}_{2j}^2(\theta_{2j}^{0\top}h(r,v_1))\,h(r,v_1)h(r,v_1)^\top\|^4<\infty.$$

另外, 假设 $\{\tilde{v}_t, 1\leqslant t\leqslant n\}$ 的 α- 混合系数 $\alpha(k)$ 满足 $\sum_{k=1}^{\infty}\alpha^{1/2}(k)<\infty$.

总体来说, $\{v_t\}$ 是一个严格平稳 α- 混合时间序列, 避免了经典的独立同分布假设, 在文献里被广泛采用 (参见文献 [16,17]). 由于 v_t 和 x_t 均含有 d_0 个 η, 所以 v_t 与 x_t 相关, 所以本假设应用范围很广. 假设 2.2(i) 对 $h(\cdot,v)$ 施加连续性是为了确保对于每个 v, $h(r,v)$ 在 $r\in[0,1]$ 中有界. 根据 Davydov 不等式, 假设 2.2(ii) 中 4 阶矩的存在和混合系数的条件保证了 $\dot{g}_{2j}^2(\theta_{2j}^{0\top}z_t)z_tz_t^\top$ 的样本平均依概率收敛. 本文中, 损失函数 $\rho(\cdot)$ 满足如下假设.

假设 2.3 (i) 假设 $\rho(\cdot)$ 是非负且凸的, 并且在原点处有唯一的最小值. 此外, 它是实直线上局部可积函数, 在无穷处的发散速度不快于某个多项式.

(ii) 对于任意 $x, u \in \mathbb{R}$, $\rho(\cdot)$ 满足 Lipschitz 条件, $|\rho(x+u)-\rho(u)| \leqslant C|x|$, 其中 C 是绝对常数.

假设 2.3 允许 $\rho(\cdot)$ 是不光滑的. 假设 2.3(i) (局部可积性以及发散不快于多项式) 确保 $\rho(\cdot)$ 为广义函数, 因此下面提出的方法行得通 (参见补充材料中 S 空间的定义), 而文献中所有的损失函数均满足这里的条件; 在广义函数的意义下, $\rho(\cdot)$ 的任意阶导数都存在. 由凸性可知, $\rho(\cdot)$ 是一个连续函数 (参见文献 [41, 第 83 页, 推论 10.1.1]). 假设 2.3(ii) 的 Lipschitz 条件在接下来的分析中起着核心作用, 它意味着次梯度函数或者导数 (如果存在的话) $\psi(\cdot)$ 的有界性, 从而我们提议的估计量受异常值的影响不大, 估计方法是稳健的. 值得注意的是, 前文提及的几个最常见的损失函数均满足 Lipschitz 条件. 例如, 当 $\rho(u)=|u|$ 时, $C=1$; 当 $\rho_\tau(u)$ 为分位数损失函数, $\tau \in (0,1)$, $C=\max(\tau, 1-\tau)$; 当 $\rho_c(u)$ 是 Huber 损失时, $C=c$. 注意到 $\rho(\cdot)$ 的次梯度函数 $\psi(\cdot)$ 在某些点的值可能不是唯一的, 在允许 $\psi(\cdot)$ 取其中任何一个值时后续的分析均成立, 所以不再重复这一点.

假设 2.4　(i) 假设 $(\psi(e_t), \mathcal{F}_t)$ 是一个鞅差序列, 其中 $\mathcal{F}_t$ 是递增的 σ 域序列, 使得 x_t 和 v_t 对于 $\mathcal{F}_{t-1}$ 适应. 另外, 存在常数 $0<a_1, a_2<\infty$ 使得 $\mathbb{E}[\psi^2(e_t) \mid \mathcal{F}_{t-1}]=a_1$ 和 $\mathbb{E}[\rho''(e_t) \mid \mathcal{F}_{t-1}]=a_2$ 以概率 1 成立, 其中当在普通意义上 $\rho''(\cdot)$ 不存在时, $\rho''(\cdot)$ 理解为广义函数. 此外, $\max_{1 \leqslant t \leqslant n} \mathbb{E}[\psi^4(e_t) \mid \mathcal{F}_{t-1}] < \infty$ 对 $n \geqslant 1$ 一致成立.

(ii) 用 $f_{t,e}(u)$ 表示给定 $\mathcal{F}_{t-1}$ 时 e_t 的条件密度 $(t=1, \ldots, n)$. 假设对于所有 n, 有 $\max_{1 \leqslant t \leqslant n} f_{t,e}(u) \leqslant f(u)$, 其中 $f(u)$ 属于检验函数空间 S (见补充材料的附录 D).

(iii) 当 $n \to \infty$ 时, 对于 $r \in [0,1]$, 有

$$\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \psi(e_t), \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \eta_t\right) \Rightarrow (U_\rho(r), W(r)),$$

这是 d_1+1 维的 Brown 运动, 其中 $U_\rho(r)$ 和 $W(r)$ 分别具有方差 a_1 和协方差 I_{d_1}.

在假设 2.4(i) 中, $\psi(\cdot)$ 和 $\rho''(\cdot)$ 分别是 $\rho(\cdot)$ 的次梯度函数和广义二阶导数, 而 σ 域可取为但不限于 $\mathcal{F}_t=\sigma(e_1, \ldots, e_t; x_s, z_s, s \leqslant t+1)$; 特别地, x_t 和 v_t 对 $\mathcal{F}_{t-1}$ 的适应性弱于 e_t 与 (x_t, z_t) 的相互独立性. 因为如果它们相互独立, 则对于自然信息流 $\mathcal{F}_t=\sigma(e_1, \ldots, e_t)$, 假设 $(\psi(e_t), \mathcal{F}_t)$ 是鞅差过程足够; 此时, 将信息流扩充为 $\mathcal{F}_t=\sigma(e_1, \ldots, e_t; x_s, z_s, s \leqslant t+1)$, 鞅差等式 $\mathbb{E}[\psi(e_t) \mid \mathcal{F}_{t-1}]=0$ 仍然成立, 同时适应性自动成立. 另外, 给定 $\mathcal{F}_{t-1}$, 如果 e_t 的条件分布是对称的, 对于 LAD 和 Huber 损失函数, 鞅差等式容易满足, 因为函数 $\psi(\cdot)$ 是奇函数, 同时, 对于分位数损失这个条件等价于 $F_e(0 \mid \mathcal{F}_{t-1})=\tau$ (参见文献 [10, 假设 2]). 常用的方法通过调整截距来使这个条件成立 (参见文献 [22, 第 124 页] 和 [48, 第 251 页]).

在假设 2.4(ii) 中, $f_{t,e}(u)$ 为 e_t 的条件密度. 由广义函数导数的定义可知

$$\mathbb{E}[\rho''(e_t) \mid \mathcal{F}_{t-1}]=\int \rho(u) f_{t,e}''(u) du.$$

尽管 $\rho(\cdot)$ 在某些点上不可微, 但如果 $f_{t,e}(u) \in S$ (检验函数空间, 见补充材料), 则假设中的条件期望 $\mathbb{E}[\psi^2(e_t) \mid \mathcal{F}_{t-1}]$ 和 $\mathbb{E}[\rho''(e_t) \mid \mathcal{F}_{t-1}]$ 是有定义的. 可以看到, He 和 Shao 在文献 [22, 第 125 页, 条件 D.2] 中以相同的方式定义 $\mathbb{E}[\rho''(e_t)]$. 此外, $\mathbb{E}[\rho''(e_t) \mid \mathcal{F}_{t-1}]=a_2>0$ 是由于所考虑的损失函数是凸的. 特别地, 在变量外生情形下, 对于分位数损失和 LAD 损失, 有 $a_2=f(0)>0$; 而对于 Huber 损失, 有 $a_2=\mathbb{P}(|e_t| \leqslant c)>0$. 假设 2.4(ii) 属于技术要求 (参见引理 B.3), 当回归变量独立于误差项时, 这一条件自动成立. 此外, 在协整回归文献中, 通常会施加类似于假设 2.4(iii) 的条件 (参见文献 [35, 46, 48]). $U_\rho(r)$ 的下标 ρ 表示联合收敛所对应的损失函数, 就像文献 [37, 式 (3)] 给出了 LAD 损

失的导数的泛函不变性原理, 该结论对应于符号函数 $\mathrm{sgn}(\cdot)$. 在不致引起混淆情形下将省略下标, 即用 $U(r)$ 表示 $U_\rho(r)$. 最后, 假设 2.4(iii) 蕴涵, 当 $n\to\infty$ 时,

$$\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\psi(e_t), \frac{1}{\sqrt{n}}x_{[nr]}\right) \Rightarrow (U(r), B(r)), \tag{2.1}$$

其中 $B(r)$ 是 Brown 运动的向量, 协方差矩阵为 $AA^\top$.

由于涉及单位根过程 x_t, 所以需要对函数 (如 H- 正则和 I- 正则) 进行分类研究 (参见文献 [33,35]). 在包含可加 H- 正则函数时, 我们的模型相较于文献 [35] 要更加精细, 而文献 [35] 中 H- 正则函数的低阶项对估计量的渐近分布没有任何贡献. 所以, 本文不需要考虑参考文献中的 H- 正则函数, 只考虑下列情形即可.

定义 2.1 假设函数 $g(\cdot)$ 定义在 $\mathbb{R}$ 上, 如果 $g(\lambda x)=\nu(\lambda)g(x)$ 对于所有 $\lambda>0$ 和 $x\in\mathbb{R}$ 成立, 则称其为 H- 正则, 其齐次阶为 $\nu(\cdot)$; 如果 $\int|g(x)|dx<\infty$, 则称其为 I- 正则.

在定义 2.1 中, H- 正则函数类主要包括幂函数, 而 I- 正则函数类为所有在实直线上可积的函数, 如概率密度函数. 关于这些定义更详细的讨论可参见文献 [33, 35].

假设 2.5 在模型 (1.1) 中, 对于 $i=1,2$, 假设函数 $g_{ij}(\cdot)$ $(1\leqslant j\leqslant p_i)$ 是线性无关且二阶连续可微的. 进一步, 对 H- 正则和 I- 正则的函数分别假设:

(i) 所有的 $g_{1j}(u)$, $\dot g_{1j}(u)$ 和 $\ddot g_{1j}(u)$ 均为 H- 正则, 齐次阶分别为 $\nu_j(\cdot)$, $\dot\nu_j(\cdot)$ 和 $\ddot\nu_j(\cdot)$, 使得

$$\lim_{\lambda\to+\infty}\frac{\dot\nu_j(\lambda)}{\nu_j(\lambda)}=0 \quad 和 \quad \lim_{\lambda\to+\infty}\frac{\ddot\nu_j(\lambda)}{\dot\nu_j(\lambda)}=0$$

成立, $j=1,\dots,p_1$;

(ii) 假设 $g_{1j}(u)$, $g_{1j}^2(u)$, $\dot g_{1j}(u)$, $\dot g_{1j}^2(u)$, $u\dot g_{1j}(u)$ 和 $[u\dot g_{1j}(u)]^2$ 都是 I- 正则, $\theta_{1j}^0=\theta_1^0$ $(j=1,\dots,p_1)$.

假设 2.5 要求对于 $i=1,2$, 所有 $g_{ij}(\cdot)$ $(j=1,\dots,p_i)$ 是线性无关的, 这是出于识别性的考虑, 而二阶连续可微是渐近分析的需要; 对于 $g_{1j}(u)$ $(j=1,\dots,p_1)$, 本文进一步考虑了两种可能性, H- 正则和 I- 正则. 这是因为, 当 $\theta^\top x_t$ 是单位根过程时, $g(\theta^\top x_t)$ 的行为在很大程度上取决于 $g(\cdot)$ 是 H- 正则还是 I- 正则 (参见文献 [14]). 更重要的是, 在 I- 正则的情形下, 我们要求所有的指标向量是相同的 $\theta_{1j}^0=\theta_1^0$ $(j=1,\dots,p_1)$. 也就是说, 所有可积函数都具有共同的随机趋势 $x_t^\top\theta_1^0$, 否则会存在涉及多维时间序列的局部时过程, 这在概率论里尚未建立. 在这个设定下, 回归函数的第一部分可以改写为

$$g_1(x_t^\top\theta_1^0;\gamma_1^0)\equiv\sum_{j=1}^{p_1}\gamma_{1j}^0 g_{1j}(x_t^\top\theta_1^0), \quad \gamma_1^0=(\gamma_{11}^0,\dots,\gamma_{1p_1}^0),$$

这里, 在 $g_1(x^\top\theta_1^0;\gamma_1^0)$ 中除了 θ_1^0 和 γ_1^0 其余都是已知的.

3 渐近理论

注意到回归方程中与非平稳部分相关的参数估计量的渐近行为与 $g_{1j}(\cdot)$ $(j=1,\dots,p_1)$ 的类别密切相关, 本节分别根据 H- 正则函数和 I- 正则函数建立估计量的渐近理论.

3.1 *H*- 正则函数

首先考虑 H- 正则函数的情形, 即假设 2.5(i) 成立. 为了简洁, 记 $x_{nt} \equiv n^{-1/2}x_t$, 并且

$$
\begin{aligned}
Z_1(x_{nt}) &\equiv (-\gamma_{11}^0\dot g_{11}(x_{nt}^\top\theta_{11}^0)x_{nt}^\top, \ldots, -\gamma_{1p_1}^0\dot g_{1p_1}(x_{nt}^\top\theta_{1p_1}^0)x_{nt}^\top, g_{11}(x_{nt}^\top\theta_{11}^0), \ldots, g_{1p_1}(x_{nt}^\top\theta_{1p_1}^0))^\top,\\
Z_2(z_t) &\equiv (-\gamma_{21}^0\dot g_{21}(z_t^\top\theta_{21}^0)z_t^\top, \ldots, -\gamma_{2p_2}^0\dot g_{2p_2}(z_t^\top\theta_{2p_2}^0)z_t^\top, g_{21}(z_t^\top\theta_{21}^0), \ldots, g_{2p_2}(z_t^\top\theta_{2p_2}^0))^\top.
\end{aligned}
\tag{3.1}
$$

由此, 定义如下对称矩阵 B 和向量 B_1:

$$
\begin{aligned}
B &\equiv a_2\begin{pmatrix} \int_0^1 Z_1(B(r))Z_1(B(r))^\top dr & \int_0^1 Z_1(B(r))[\mathbb{E}Z_2(h(r,v_1))]^\top dr \\ \int_0^1 [\mathbb{E}Z_2(h(r,v_1))]Z_1(B(r))^\top dr & \int_0^1 \mathbb{E}[Z_2(h(r,v_1))Z_2(h(r,v_1))]^\top dr \end{pmatrix},\\
B_1 &\equiv \begin{pmatrix} \int_0^1 Z_1(B(r))dU(r) \\ N(0,a_1\Sigma) \end{pmatrix}, \quad \text{其中 } \Sigma = \int_0^1 \mathbb{E}[Z_2(h(r,v_1))Z_2(h(r,v_1))]^\top dr,
\end{aligned}
\tag{3.2}
$$

这里正常数 a_1 和 a_2 由假设 2.4 给出, 而随机过程 $B(r)$ 和 $U(r)$ 由 (2.1) 给出. 在附录的证明中将用一个二次型来逼近目标函数, 记号 $Z_1(\cdot)$ 和 $Z_2(\cdot)$ 是帮助表达这个二次型中的变量, 而矩阵 B 和向量 B_1 是这些变量的部分和的渐近极限 (见引理 B.2). 另外, 记

$$
\begin{aligned}
\widehat\Lambda_1 &= [\dot\nu_1(\sqrt{n})n(\widehat\theta_{11}-\theta_{11}^0)^\top, \ldots, \dot\nu_{p_1}(\sqrt{n})n(\widehat\theta_{1p_1}-\theta_{1p_1}^0)^\top,\\
&\qquad \nu_1(\sqrt{n})\sqrt{n}(\widehat\gamma_{11}-\gamma_{11}^0), \ldots, \nu_{p_1}(\sqrt{n})\sqrt{n}(\widehat\gamma_{1p_1}-\gamma_{1p_1}^0)]^\top,\\
\widehat\Lambda_2 &= [\sqrt{n}(\widehat\theta_{21}-\theta_{21}^0)^\top, \ldots, \sqrt{n}(\widehat\theta_{2p_2}-\theta_{2p_2}^0)^\top, \sqrt{n}(\widehat\gamma_{21}-\gamma_{21}^0), \ldots, \sqrt{n}(\widehat\gamma_{2p_2}-\gamma_{2p_2}^0)]^\top,
\end{aligned}
\tag{3.3}
$$

其中 $\widehat\theta_{1j}$, $\widehat\theta_{2j}$, $\widehat\gamma_{1j}$ 和 $\widehat\gamma_{2j}$ 是由 (1.4) 所定义的估计量, 记号 $\widehat\Lambda_1$ 和 $\widehat\Lambda_2$ 为标准化后的估计量. 可以看到, 所有这些记号的引入使得定理的陈述变得简洁.

定理 3.1 在假设 2.1–2.4 和 2.5(i) 下, 当 $n\to\infty$, 有 $\widehat\Lambda = O_P(1)$, 其中 $\widehat\Lambda = (\widehat\Lambda_1^\top, \widehat\Lambda_2^\top)^\top$, 同时,

$$
\widehat\Lambda \to_D B^{-1}B_1. \tag{3.4}
$$

定理 3.1 的证明在附录 C 中给出. 在证明中使用引理 A.1 中定义的 $\rho(\cdot)$ 的正则序列, 证明目标函数可以被一个二次函数来逼近, 以帮助建立这个复杂的极限定理. 所有估计量的相合性和收敛速度都在结论 $\widehat\Lambda = O_P(1)$ 中体现. 由表达式 (3.3) 和 (3.4) 可以看出, 所有 $\widehat\theta_{1j}$ 和 $\widehat\gamma_{1j}$ 的收敛速度都受到 $g_{1j}(\cdot)$ $(j=1,\ldots,p_1)$ 的 H- 正则性的影响. 准确地说, $\widehat\theta_{1j}$ 的收敛速度为 $[\dot\nu_j(\sqrt{n})n]^{-1}$, 其中 $\dot\nu_j(\cdot)$ 是 $\dot g_{1j}(\cdot)$ 的齐次阶. 例如, 如果 $g_{11}(u)=u$, 则 $\dot\nu_1(\lambda)=1$, 从而 $\widehat\theta_{11}$ 的收敛速度为 n^{-1}, 即所谓非平稳情形下线性参数模型的超速度 (super rate); 如果 $g_{12}(u)=u^2$, 则 $\dot\nu_1(\lambda)=\lambda$, 因此 $\widehat\theta_{12}$ 的收敛速度是 $n^{-3/2}$.

注意, 由于有识别条件 $\|\theta_{1j}^0\|=1$, 所以 γ_{1j} 是可以识别的. 此外, $\widehat\gamma_{1j}$ 的收敛速度为 $[\nu_j(\sqrt{n})\sqrt{n}]^{-1}$, 也受到了 $g_{1j}(\cdot)$ 的影响. 由于每个 $x_t^\top\theta_{1j}^0$ 仍然是单位根过程, 因此 $g_{1j}(x_t^\top\theta_{1j}^0)$ 具有渐近阶 $\nu_j(\sqrt{n})$, 这导致 $\widehat\gamma_{1j}$ 的速度比平稳情形快一个因子 $[\nu_j(\sqrt{n})]^{-1}$.

尽管我们的回归模型中包含了平稳和非平稳变量, 但是所有的估计量 $\widehat\theta_{2j}$ 和 $\widehat\gamma_{2j}$ $(j=1,\ldots,p_2)$ 都有通常的 $\sqrt{n}$ 收敛速度. 这是因为回归函数是可加的, 这一结构使得平稳部分和非平稳部分分离, 从而平稳部分的参数保留通常的收敛速度. 然而, 如下文所示, 两部分中所有估计量的极限都包含了回归中的所有成分. 该发现与 Dong 和 Linton [16] 的发现类似.

为了分别得到 $\widehat{\Lambda}_1$ 和 $\widehat{\Lambda}_2$ 的渐近性质, 对矩阵 $B=(B_{ij})_{2\times 2}$ 和向量 $B_1=(b_{11}^\top, b_{12}^\top)^\top$ 进行维数相匹配的划分,

$$
\begin{aligned}
&B_{11}=a_2\int_0^1 Z_1(B(r))Z_1(B(r))^\top dr, \quad B_{12}=a_2\int_0^1 Z_1(B(r))[\mathbb{E}Z_2(h(r,z_1))]^\top dr,\\
&B_{21}=B_{12}^\top, \quad B_{22}=a_2\Sigma,\\
&b_{11}=\int_0^1 Z_1(B(r))dU(r), \quad b_{12}=\sqrt{a_1}\,N(0,\Sigma).
\end{aligned}
$$

那么, (3.4) 意味着

$$
\begin{aligned}
&\widehat{\Lambda}_1 \to_D (B_{11}-B_{12}B_{22}^{-1}B_{21})^{-1}b_{11}-B_{11}^{-1}B_{12}(B_{22}-B_{21}B_{11}^{-1}B_{12})^{-1}b_{12},\\
&\widehat{\Lambda}_2 \to_D (B_{22}-B_{21}B_{11}^{-1}B_{12})^{-1}b_{12}-(B_{22}-B_{21}B_{11}^{-1}B_{12})^{-1}B_{21}B_{11}^{-1}b_{11}.
\end{aligned} \tag{3.5}
$$

可以看到, 非对角线块 B_{12} 和 B_{21} 在估计量的极限里体现了平稳变量和非平稳变量对估计量的相互影响. 另外, 考虑特殊情形 $p_1=p_2=1$, 此时,

$$
y_t=\gamma_1^0 g_1(x_t^\top\theta_1^0)+\gamma_2^0 g_2(z_t^\top\theta_2^0)+e_t, \quad t=1,\dots,n,
$$

其中 $g_1(\cdot)$ 是 H- 正则, 其齐次阶为 $\nu_1(\cdot)$; 令 $\mathbb{E}[Z_2(h(r,z_1))]=0$, 这意味着 $B_{12}=B_{21}^\top=0$. 则 (3.5) 蕴涵

$$
\begin{pmatrix} \dot{\nu}_1(\sqrt{n})n(\widehat{\theta}_1-\theta_1^0)\\ \nu_1(\sqrt{n})\sqrt{n}(\widehat{\gamma}_1-\gamma_1^0)\end{pmatrix} \to_D B_{11}^{-1}b_{11}, \quad \sqrt{n}\begin{pmatrix}\widehat{\theta}_2-\theta_2^0\\ \widehat{\gamma}_2-\gamma_2^0\end{pmatrix} \to_D B_{22}^{-1}b_{12}. \tag{3.6}
$$

在此特殊情形下, 非平稳变量和平稳变量对估计量的相互影响在极限里不存在. 值得注意的是, 损失函数对估计的影响主要通过常数 a_1 和 a_2 实现; 可以看到 $\widehat{\Lambda}_1$ 和 $\widehat{\Lambda}_2$ 的极限分别有因子 $1/a_2$ 和 $\sqrt{a_1}/a_2$.

注 3.1 为了与相关文献进行比较, 这里假设变量是外生的.

(i) 当 $\rho(u)=|u|$ 时, $\rho'(u)=-H(-u)+H(u)$, 其中 $H(\cdot)$ 是 Heaviside 函数, 并且 $\rho''(u)=2\delta(u)$. 从而 $a_1=\mathbb{E}[\rho'(e_t)]^2=1$, $a_2=\mathbb{E}[\rho''(e_t)]=2f_e(0)>0$, 其中 $f_e(\cdot)$ 是 e_t 的密度. 此外, 如果 $g_1(u)=u$, $\gamma_1^0=1$ 不需要估计, 并且 $g_2(v)\equiv 0$, 则模型简化为一个通常的线性协整模型. 本文的结果表明

$$
n(\widehat{\theta}_1-\theta_1^0) \to_D [2f_e(0)]^{-1}\left[\int_0^1 B(r)B(r)^\top dr\right]^{-1}\int_0^1 B(r)dU(r),
$$

这与 Phillips [37] 的结果相似; 如果 $g_2(v)=v$, $\gamma_2^0=1$ 不需要估计, 并且 $g_1(u)\equiv 0$, $h(r,v)\equiv v$, 则模型简化为平稳线性回归模型. 本文的结果为 $\sqrt{n}(\widehat{\theta}_2-\theta_2^0) \to_D [2f_e(0)]^{-1}[\mathbb{E}(v_1v_1^\top)]^{-1/2}N(0,I_{d_2})$, 与 Pollard [39] 的结果一致.

(ii) 当 $\rho(u)$ 是分位数损失时, 有 $\rho'(u)=\tau H(u)+(\tau-1)H(-u)$ 和 $\rho''(u)=\delta(u)$. 则 $a_1=\mathbb{E}[\rho'(e_t)]^2=\tau(1-\tau)$ 和 $a_2=\mathbb{E}[\rho''(e_t)]=f_e(0)>0$, 其中 $f_e(\cdot)$ 是 e_t 的密度. 进一步地, 如果 $g_1(u)=u$, $\gamma_1^0=1$ 不需要估计, 并且 $g_2(v)\equiv 0$, 则模型简化为一个线性协整模型. 本文的结果表明 $n(\widehat{\theta}_1-\theta_1^0) \to_D [f_e(0)]^{-1}[\int_0^1 B(r)B(r)^\top dr]^{-1}\int_0^1 B(r)dU(r)$, 与 Xiao [48] 的结果相同; 如果 $g_2(v)=v$, $\gamma_2^0=1$ 不需要估计, 并且 $g_1(u)\equiv 0$, $h(r,v)\equiv v$, 则模型简化为平稳线性回归模型. 本文的结果表明

$$
\sqrt{n}(\widehat{\theta}_2-\theta_2^0) \to_D \sqrt{\tau(1-\tau)}[f_e(0)]^{-1}[\mathbb{E}(v_1v_1^\top)]^{-1/2}N(0,I_{d_2}),
$$

这是对平稳序列的线性模型进行分位数回归的结果 (参见文献 [27]).

虽然 Σ 可以容易通过函数 h 和 v_t 的观测值一致地估计, 但我们需要为 a_1 和 a_2 构建一致的估计量, 以便使用上面建立的渐近分布进行统计推断. 为此, 定义

$$\widehat{a}_1 = \frac{1}{n}\sum_{t=1}^{n}[\psi(\widehat{e}_t)]^2 \quad 和 \quad \widehat{a}_2 = \frac{1}{n}\sum_{t=1}^{n}\rho''_m(\widehat{e}_t), \tag{3.7}$$

其中 $\widehat{e}_t = y_t - \sum_{j=1}^{p_1}\widehat{\gamma}_{1j}g_{1j}(\widehat{\theta}_{1j}^\top x_t) - \sum_{j=1}^{p_2}\widehat{\gamma}_{2j}g_{2j}(\widehat{\theta}_{2j}^\top z_t)$ $(t=1,\ldots,n)$. 我们有以下推论.

推论 3.1 在定理 3.1 的条件下, 当 $n\to\infty$ 时有 $\widehat{a}_1 \to_P a_1$; 此外, 假设 $\{\rho''_m(e_t), t=1,\ldots,n\}$ 是遍历的, 则当 $m,n\to\infty$ 时有 $\widehat{a}_2 \to_P a_2$.

在证明 $\widehat{a}_1 \to_P a_1$ 的过程中, 我们使用了 $\{\psi(e_t)\}$ 的遍历性, 这是由它的鞅差序列性质推导出来的; 同样地, 在 $\widehat{a}_2 \to_P a_2$ 的证明中, 也需要正则序列 $\{\rho''_m(e_t)\}$ 对任意正整数 m 的遍历性. 当然, 这两个过程的遍历性可以由 $\{e_t\}$ 的遍历性保证.

3.2 *I*- 正则函数

本小节考虑所有 $g_{1j}(\cdot)$ 都满足假设 2.5(ii), 即 I- 正则. 这一假设将极大地改变渐近理论. 在假设 2.5(ii) 中, $\theta_{1j}^0 = \theta_1^0$ $(j=1,\ldots,p_1)$, 因此所有 I- 正则函数都以相同的随机趋势 $x_t^\top\theta_1^0$ 作为它们的变量. 此外, 极限理论依赖于标量 Brown 运动 $W(r)$ 的局部时间 (local time) 过程, 其定义为

$$L_W(p,s) = \lim_{\epsilon\to 0}\frac{1}{2\epsilon}\int_0^p I(|W(r)-s|<\epsilon)dr, \tag{3.8}$$

它度量了 $W(r)$ 在 $(0,p)$ 期间逗留在 s 附近的时间. 在本文中, 标量 Brown 运动是 $B_1(r)\equiv(\theta_1^0)^\top B(r)$, 其中 $B(r)$ 由 (2.1) 给出, 使用 $L_1(p,s)$ 表示它的局部时间过程.

这类回归函数的极限理论的另一个特点是, 首先在新的坐标系中推导估计量的渐近性, 然后还原为原坐标系中的极限. 这与 Park 和 Phillips[34] 以及 Dong 等[14] 的思想相同. 为此, 使用单位向量 θ_1^0, 构造正交矩阵 $P=(\theta_1^0, P_2)_{d_1\times d_1}$ 来旋转所有相关的向量, 包括 θ_1^0, x_t 和 $B(r)$. 因此, $P^\top\theta_1^0 = (1,0,\ldots,0)^\top$; $P^\top x_t$ 的第一个元素 $x_{1,t}\equiv(\theta_1^0)^\top x_t$, 其余的子向量 $x_{2,t}\equiv P_2^\top x_t$; 根据连续映射定理, 经 $\sqrt{n}$ 归一化后, 它们联合收敛如下:

$$\frac{1}{\sqrt{n}}x_{1,[nr]} \to_D B_1(r)\equiv(\theta_1^0)^\top B(r), \quad \frac{1}{\sqrt{n}}x_{2,[nr]} \to_D B_2(r)\equiv P_2^\top B(r), \quad r\in[0,1]. \tag{3.9}$$

下面, 记 $D_n = \mathrm{diag}(\sqrt[4]{n}, \sqrt[4]{n}^3 I_{d_1-1})$, $p_n=\sqrt{n}$ 和 $q_n=\sqrt[4]{n}$, 并定义

$$\begin{aligned} &\alpha_1 = D_nP^\top(\theta_1-\theta_1^0), \quad \beta_{1j} = q_n(\gamma_{1j}-\gamma_{1j}^0), \quad j=1,\ldots,p_1,\\ &\alpha_{2j} = p_n(\theta_{2j}-\theta_{2j}^0), \quad \beta_{2j} = p_n(\gamma_{2j}-\gamma_{2j}^0), \quad j=1,\ldots,p_2. \end{aligned} \tag{3.10}$$

这里, θ_1, θ_{2j}, γ_{1j} 和 γ_{2j} 是参数空间中的一般参数, 而 α_1, β_{1j}, α_{2j} 和 β_{2j} 是它们被中心化、旋转且被拉伸 (centered and rescaling) 后的向量. 旋转的作用仅用于渐近分析, 而在 Monte Carlo 实验和实证研究中, 旋转不是必要的; 更多的细节可以在相关文献中找到, 如文献 [14, 34].

为了简洁, 定义参数向量 $\Lambda\equiv(\Lambda_1^\top,\Lambda_2^\top)^\top$ 和它的估计量 $\widehat{\Lambda}=(\widehat{\Lambda}_1^\top,\widehat{\Lambda}_2^\top)^\top$ 如下:

$$\begin{aligned} &\Lambda_1 \equiv (\alpha_1^\top, \beta_{11},\ldots,\beta_{1p_1})^\top, \quad \Lambda_2 \equiv (\alpha_{21}^\top,\ldots,\alpha_{2p_2}^\top,\beta_{21},\ldots,\beta_{2p_2})^\top,\\ &\widehat{\Lambda}_1 \equiv (\widehat{\alpha}_1^\top, \widehat{\beta}_{11},\ldots,\widehat{\beta}_{1p_1})^\top, \quad \widehat{\Lambda}_2 \equiv (\widehat{\alpha}_{21}^\top,\ldots,\widehat{\alpha}_{2p_2}^\top,\widehat{\beta}_{21},\ldots,\widehat{\beta}_{2p_2})^\top. \end{aligned}$$

这里 $\widehat{\Lambda}_1$ 和 $\widehat{\Lambda}_2$ 中的估计量是通过 (3.10) 中的关系用 $\widehat{\theta}_1, \widehat{\theta}_{2j}, \widehat{\gamma}_{1j}$ 和 $\widehat{\gamma}_{2j}$ 定义的, 如 $\widehat{\alpha}_1 = D_n P^\top(\widehat{\theta}_1 - \theta_1^0)$, 而 $\widehat{\theta}_1, \widehat{\theta}_{2j}, \widehat{\gamma}_{1j}$ 和 $\widehat{\gamma}_{2j}$ 是 (1.4) 定义的估计量, 只不过目标函数 $L_n(\theta, \gamma)$ 中的函数满足假设 2.5(ii).

为了表达这些估计量的渐近性质, 需要定义几个复杂的向量和矩阵. 首先, 令 $x_{nt} \equiv D_n^{-1} P^\top x_t$, $\dot{g}_1(u) \equiv \sum_{j=1}^{p_1} \gamma_{1j}^0 \dot{g}_{1j}(u)$ 和 $Z_t \equiv (Z_1(x_{nt})^\top, Z_2(z_t)^\top)^\top$, 其中

$$
\begin{aligned}
&Z_1(x_{nt}) \equiv (-\dot{g}_1(x_t^\top \theta_1^0) x_{nt}^\top, q_n^{-1} g_{11}(x_t^\top \theta_1^0), \ldots, q_n^{-1} g_{1p_1}(x_t^\top \theta_1^0))^\top, \\
&Z_2(z_t) \equiv p_n^{-1}(-\gamma_{21}^0 \dot{g}_{21}(z_t^\top \theta_{21}^0) z_t^\top, \ldots, -\gamma_{2p_2}^0 \dot{g}_{2p_2}(z_t^\top \theta_{2p_2}^0) z_t^\top, g_{21}(z_t^\top \theta_{21}^0), \ldots, g_{2p_2}(z_t^\top \theta_{2p_2}^0))^\top,
\end{aligned}
\tag{3.11}
$$

这里 p_n 和 q_n 同 (3.10).

其次, 为了展示 $\sum_{t=1}^n Z_t Z_t^\top$ 和 $\sum_{t=1}^n \psi(e_t) Z_t$ 的极限, 需要定义一个矩阵. 令 $\mathcal{R}$ 是一个 $(d_1 + p_1) + p_2(d_2 + 1)$ 阶的方阵, 表示 $\sum_{t=1}^n Z_t Z_t^\top$ 的极限; 与 $Z_t Z_t^\top$ 中的子块 (即 $Z_1 Z_1^\top$, $Z_1 Z_2^\top$, $Z_2 Z_1^\top$ 和 $Z_2 Z_2^\top$) 一致, 将其分划为 $\mathcal{R} = (\mathcal{R}_{ij})_{2\times 2}$. 因此, $\mathcal{R}_{11} = (R_{ij}^{11})$ 是一个 $d_1 + p_1$ 维的对称矩阵, 包含以下元素:

$$
\begin{aligned}
&R_{11}^{11} = L_1(1,0) \int [\dot{g}_1(u)]^2 u^2 du, \quad R_{1,2:d_1}^{11} = \int_0^1 B_2(r)^\top dL_1(r,0) \int [\dot{g}_1(u)]^2 u du, \\
&R_{1,d_1+1}^{11} = -L_1(1,0) \int u g_{11}(u) \dot{g}_1(u) du, \quad \ldots, \quad R_{1,d_1+p_1}^{11} = -L_1(1,0) \int u g_{1p_1}(u) \dot{g}_1(u) du, \\
&R_{2:d_1,2:d_1}^{11} = \int_0^1 B_2(r) B_2(r)^\top dL_1(r,0) \int [\dot{g}_1(u)]^2 du, \\
&R_{2:d_1,d_1+1}^{11} = -\int_0^1 B_2(r) dL_1(r,0) \int g_{11}(u) \dot{g}_1(u) du, \quad \ldots, \\
&R_{2:d_1,d_1+p_1}^{11} = -\int_0^1 B_2(r) dL_1(r,0) \int g_{1p_1}(u) \dot{g}_1(u) du, \\
&R_{d_1+1,d_1+1}^{11} = L_1(1,0) \int [g_{11}(u)]^2 du, \quad \ldots, \quad R_{d_1+1,d_1+p_1}^{11} = L_1(1,0) \int g_{11}(u) g_{1p_1}(u) du, \quad \ldots, \\
&R_{d_1+p_1,d_1+1}^{11} = L_1(1,0) \int g_{11}(u) g_{1p_1}(u) du, \quad \ldots, \quad R_{d_1+p_1,d_1+p_1}^{11} = L_1(1,0) \int [g_{1p_1}(u)]^2 du,
\end{aligned}
$$

其中 $R_{2:d_1,2:d_1}^{11}$ 表示 $R_{i,j}^{11}$ $(i,j = 2,\ldots,d_1)$ 的所有元素; 其他符号, 如 $R_{2:d_1,d_1+1}^{11}, \ldots, R_{2:d_1,d_1+p_1}^{11}$ 的定义类似. 同时, $\mathcal{R}_{22} = \Sigma$, 由定理 3.1 给出, 而 $\mathcal{R}_{12} = \mathcal{R}_{12}^\top = 0$ 是 $(d_1 + p_1) \times (d_2 p_2 + p_2)$ 维零矩阵. 因此, $\mathcal{R}$ 是一个对角块状矩阵. 进一步可以看出, 如果 $g_{11}(u), \ldots, g_{1p_1}(u)$ 取自一个正交序列使得 $\int g_{1j}(u) g_{1k}(u) du = 0$ $(j \neq k)$, 则即使在 $\mathcal{R}_{11}$ 块中也会有许多零元素.

定理 3.2 在假设 2.1–2.4 和 2.5(ii) 下, 当 $n \to \infty$ 时, 有 $\widehat{\Lambda} = O_P(1)$, 同时,

$$
\widehat{\Lambda} \to_D \frac{\sqrt{a_1}}{a_2} \mathcal{R}^{-1/2} N(0, I),
\tag{3.12}
$$

其中 I 是一个 $d_1 + p_1 + p_2(d_2 + 1)$ 维的单位矩阵, 常数 a_1 和 a_2 由假设 2.4 给出.

定理 3.2 的证明在附录 C 中给出. 类似于定理 3.1, 这里所有估计量的一致性和收敛速度由 $\widehat{\Lambda} = O_P(1)$ 给出. 由于 $\mathcal{R}$ 包含 Brown 运动的局部时过程, 是随机矩阵, 所以极限 (3.12) 在相关文献中被称为混合正态分布 (mixed normal distribution), 可参见文献 [35]. 再者, 由于 $\mathcal{R}$ 的对角块状结构, (3.12) 表明, 当 $n \to \infty$ 时, 有

$$
\widehat{\Lambda}_1 \to_D \frac{\sqrt{a_1}}{a_2} \mathcal{R}_{11}^{-1/2} N(0, I) \quad \text{和} \quad \widehat{\Lambda}_2 \to_D \frac{\sqrt{a_1}}{a_2} \mathcal{R}_{22}^{-1/2} N(0, I).
\tag{3.13}
$$

从 (3.10) 得出, 所有估计量 $\widehat{\theta}_{2j}$ 和 $\widehat{\gamma}_{2j}$ $(j=1,\ldots,p_2)$ 具有传统的 $\sqrt{n}$- 收敛速度, 然而受 x_t 的单位根性质和函数 g_{1j} 的 I- 正则性的影响, $\widehat{\theta}_1$ 和 $\widehat{\gamma}_{1j}$ $(j=1,\ldots,p_1)$ 的速度被改变, 类似于文献 [14] 中的结果. 总体来讲, 这个模型的特征与文献 [16] 中的一致, 因为它们都具有可加形式, 并且非平稳变量和平稳变量被分离在不同的被加函数中.

特别地, 每个 $\widehat{\gamma}_{1j}$ 都有一个缓慢的收敛速度 $n^{-1/4}$, 因为它们是 $g_{1j}(x_t^\top\theta_1^0)$ 的系数, 而 $g_{1j}(x_t^\top\theta_1^0)$ 在 t 变大时衰减为零. 这个速度与经典文献中的相同, 如文献 [33,35]. 然而, 由于 D_n 和 P 的介入, $\widehat{\theta}_1$ 的速度在定理 3.2 中是不清楚的. 为了找出它的准确的速度和极限分布, 用 r_{11} 表示 $\mathcal{R}_{11}^{-1}$ 左上角 $d_1\times d_1$ 维的子矩阵. 由定理 3.2, 可得

$$D_nP^\top(\widehat{\theta}_1-\theta_1^0)\to_D \frac{\sqrt{a_1}}{a_2}N(0,r_{11}). \tag{3.14}$$

下面的推论从上述 $\widehat{\theta}_1$ 的旋转极限还原了 $\widehat{\theta}_1$ 的渐近分布.

推论 3.2　在定理 3.2 的条件下, 当 $n\to\infty$ 时, 有

$$\sqrt[4]{n}(\widehat{\theta}_1-\theta_1^0)\to_D \frac{\sqrt{a_1}}{a_2}N(0,r_{11}^{11}\theta_1^0(\theta_1^0)^\top),$$

其中 r_{11}^{11} 是 r_{11} 左上角元素.

推论 3.2 表明估计量 $\widehat{\theta}_1$ 的收敛速度为 $n^{-1/4}$. 这与文献 [14] 的速度相同, 并且文献 [14] 从几何的角度详细解释了为什么是这样的速度; 与文献 [14] 的不同之处在于, 本文的损失函数产生了两个常数 a_1 和 a_2, 出现在极限里. 另外, 如果将 $\widehat{\theta}_1$ 归一化为 $\widehat{\theta}_{1,\text{unit}}=\widehat{\theta}_1/\|\widehat{\theta}_1\|$, 我们发现 $\widehat{\theta}_{1,\text{unit}}$ 以 $n^{-3/4}$ 的速度收敛到 θ_1^0. 由于与文献 [14] 完全相同, 因此本文省略此推导, 读者可参阅相关文献. 最后, 与推论 3.1 一样, 可以定义估计量 $\widehat{a}_1$ 和 $\widehat{a}_2$, 然后建立它们的一致性, 由于类似, 故被省略.

4　Monte Carlo 模拟实验

本节通过模拟实验来评估可加单指标协整模型中稳健 M 估计量的有限样本性能. 出于篇幅考虑, 只考虑两个符合本文模型假设的例子, 第一个例子考虑齐次正则函数, 第二个例子研究可积正则函数的情形.

假设时间序列 $\{x_t,z_t\}$ 生成方式如下. $x_t=\rho_1x_{t-1}+\sigma_1w_t$, $z_t=h(t/n)+v_t$, $h(\tau)=(\tau,\tau)^\top$, $v_t=\rho_2v_{t-1}+\sigma_2\epsilon_t$, x_t 和 v_t 均为自回归二维向量过程, 其中 $x_0=0$, $v_0=0$, $\rho_1=I_2$, $\rho_2=0.5\times I_2$, $\sigma_1=\mathrm{diag}\{0.2,0.5\}$, $\sigma_2=I_2$, 并且 $(w_t^\top,\epsilon_t^\top)$ 是具有零均值、单位协方差矩阵的独立正态分布随机向量.

例 4.1　首先考虑模型 $y_t=\gamma_1(x_t^\top\theta_1)+\gamma_2(x_t^\top\theta_2)^2+\gamma_3(z_t^\top\theta_3)+e_t$, 其中不失一般性, 选取模为 1、分量等长度的向量 $\theta_1^\top=(\theta_{11},\theta_{12})=(1,1)/\sqrt{2}$, $\theta_2^\top=(\theta_{21},\theta_{22})=(1,1)/\sqrt{2}$, $\theta_3^\top=(\theta_{31},\theta_{32})=(1,1)/\sqrt{2}$, $\gamma_1=\gamma_2=2$, $\gamma_3=1$. 回归误差项 $e_t=0.5\cdot u_t$, 其中 u_t 根据 4 个分布独立生成: (D1) 标准正态 $N(0,1)$; (D2) 混合正态 $0.9\cdot N(0,1)+0.1\cdot N(0,4)$; (D3) $t(2)$, 自由度为 2 的 t 分布; (D4) $t(1)$, 标准 Cauchy 分布. 众所周知, 混合正态分布通常作为带异常值情形下的扰动项设定, 而 t 分布则作为厚尾分布, 它们可以有效地检验估计方法对异常值和厚尾扰动项的稳健性. 对于样本容量 n 分别等于 50, 100 和 200, 我们分别重复 $K=5{,}000$ 次模拟.

例 4.2　考虑如下模型:

$$y_t=\gamma_1\phi(x_t^\top\theta_1)+\gamma_2(z_t^\top\theta_3)+e_t',$$

其中 $\phi(\cdot)$ 是标准正态密度函数, $\theta_1^\top=(\theta_{11},\theta_{12})=(1,1)/\sqrt{2}$, $\theta_3^\top=(\theta_{31},\theta_{32})=(1,1)/\sqrt{2}$, $\gamma_1=2$, $\gamma_2=1$, $e_t'=\sigma\cdot e_t$, 并且 e_t 的生成方法与例 4.1 中的相同. 通过 $\sigma=1,2$ 来探究不同噪声水平对估计量表现的影响. 限于篇幅, 本例仅考虑 n 分别等于 50 和 200.

为了反映不同损失函数对估计的影响, 上面两个例子中都采用 3 种损失函数构造目标函数: (L1) Huber 损失 $\rho_c(e)=\frac{1}{2}e^2\cdot 1\{|e|\leqslant c\}+(c|e|-\frac{1}{2}c^2)\cdot 1\{|e|>c\}$, 其中 $c=1.25$; (L2) 绝对误差损失 $\rho(e)=|e|$; (L3) 分位数损失 $\rho_\tau(e)=e\cdot(\tau-I\{e<0\})$, 其中 $\tau=0.3$. 对于分位数回归, 将回归残差归一化, 使其 τ- 分位数为 0. 第一种损失函数可以在带有异常值情形下估计对称分布的均值, 而后两种则可以帮助我们得到中位数和 30% 分位数的估计. 为了衡量指标向量的非线性最小二乘估计的估计精度, 我们计算了指标的每个元素的偏差、估计标准差和均方误差. 由于篇幅限制, 在分别针对例 4.1 和 4.2 的表 1 和 2 中只报告均方误差. 以 θ_{11} 为例, 其估计的均方误差计算公式为 $\sum_{k=1}^K(\hat\theta_{11,k}-\theta_{11})^2/K$, 其中 $\hat\theta_{11,k}$

表 1　例 4.1 中 M 估计量的均方误差 ($\times 10^2$)

		(D1) 标准正态			(D2) 混合正态			(D3) $t(2)$			(D4) 标准 Cauchy 分布		
		(L1)	(L2)	(L3)	(L1)	(L2)	(L3)	(L1)	(L2)	(L3)	(L1)	(L2)	(L3)
$n=50$	γ_1	10.33	12.08	13.23	12.23	14.61	14.86	19.07	18.45	22.59	38.99	32.25	53.94
	θ_{11}	1.340	1.303	1.110	1.478	1.420	1.275	2.549	1.876	1.769	4.422	2.893	3.042
	θ_{12}	0.865	0.966	0.934	0.959	1.043	1.037	1.454	1.309	1.426	2.395	1.911	2.400
	γ_2	4.209	5.194	5.990	5.012	6.250	6.926	8.615	9.06	11.38	18.21	16.81	22.42
	θ_{21}	0.152	0.188	0.189	0.210	0.240	0.243	0.327	0.323	0.385	0.684	0.554	0.887
	θ_{22}	0.139	0.175	0.178	0.173	0.208	0.218	0.279	0.278	0.344	0.541	0.463	0.708
	γ_3	0.512	0.744	0.804	0.608	0.826	0.925	1.014	1.139	1.515	1.821	1.787	2.944
	θ_{31}	0.276	0.389	0.422	0.317	0.438	0.476	0.543	0.609	0.844	0.965	0.947	1.580
	θ_{32}	0.276	0.385	0.424	0.318	0.442	0.479	0.538	0.605	0.818	0.951	0.920	1.585
$n=100$	γ_1	2.370	3.204	3.264	2.755	3.654	3.821	4.561	4.813	5.796	7.061	6.593	10.40
	θ_{11}	0.271	0.330	0.336	0.339	0.429	0.389	0.564	0.533	0.611	0.902	0.734	1.018
	θ_{12}	0.226	0.283	0.295	0.277	0.355	0.336	0.398	0.411	0.481	0.613	0.558	0.797
	γ_2	0.497	0.696	0.739	0.577	0.776	0.853	0.962	1.012	1.299	1.590	1.513	2.518
	θ_{21}	0.017	0.023	0.025	0.020	0.027	0.030	0.032	0.033	0.045	0.051	0.049	0.092
	θ_{22}	0.017	0.023	0.024	0.020	0.026	0.029	0.031	0.032	0.044	0.049	0.048	0.083
	γ_3	0.216	0.308	0.338	0.260	0.361	0.412	0.408	0.445	0.618	0.679	0.652	1.134
	θ_{31}	0.118	0.169	0.185	0.143	0.197	0.224	0.218	0.240	0.331	0.363	0.351	0.607
	θ_{32}	0.118	0.169	0.185	0.144	0.198	0.221	0.216	0.240	0.329	0.370	0.359	0.622
$n=200$	γ_1	0.589	0.825	0.873	0.648	0.856	0.956	1.135	1.211	1.472	1.712	1.524	2.679
	θ_{11}	0.060	0.083	0.090	0.068	0.092	0.099	0.119	0.124	0.155	0.193	0.158	0.285
	θ_{12}	0.057	0.078	0.084	0.065	0.087	0.093	0.109	0.114	0.138	0.158	0.135	0.239
	γ_2	0.062	0.089	0.095	0.068	0.094	0.102	0.111	0.119	0.159	0.180	0.168	0.310
	θ_{21}	0.002	0.003	0.003	0.002	0.003	0.003	0.004	0.004	0.005	0.006	0.005	0.010
	θ_{22}	0.002	0.003	0.003	0.002	0.003	0.003	0.004	0.004	0.005	0.006	0.005	0.010
	γ_3	0.103	0.149	0.161	0.118	0.165	0.184	0.182	0.200	0.271	0.286	0.274	0.485
	θ_{31}	0.053	0.078	0.087	0.063	0.088	0.094	0.096	0.106	0.148	0.150	0.137	0.260
	θ_{32}	0.053	0.079	0.088	0.063	0.088	0.095	0.095	0.105	0.148	0.150	0.137	0.259

表 2　例 4.2 中 M 估计量的均方误差 ($\times 10$)

			(D1) 标准正态			(D2) 混合正态			(D3) $t(2)$			(D4) 标准 Cauchy 分布		
			(L1)	(L2)	(L3)	(L1)	(L2)	(L3)	(L1)	(L2)	(L3)	(L1)	(L2)	(L3)
$\sigma=1$	$n=50$	γ_1	0.658	0.928	0.670	0.780	1.054	0.734	1.150	1.270	0.967	1.734	1.612	1.843
		θ_{11}	0.123	0.136	0.121	0.145	0.158	0.137	0.213	0.185	0.184	0.311	0.232	0.247
		θ_{12}	0.124	0.135	0.124	0.149	0.159	0.135	0.222	0.189	0.186	0.327	0.230	0.248
		γ_2	0.017	0.025	0.024	0.019	0.026	0.026	0.030	0.032	0.038	0.050	0.044	0.073
		θ_{21}	0.011	0.017	0.016	0.014	0.018	0.018	0.020	0.020	0.025	0.032	0.028	0.047
		θ_{22}	0.011	0.017	0.016	0.014	0.018	0.018	0.020	0.021	0.025	0.032	0.028	0.047
	$n=200$	γ_1	0.237	0.364	0.297	0.290	0.385	0.333	0.455	0.482	0.495	0.708	0.603	0.830
		θ_{11}	0.018	0.026	0.026	0.023	0.030	0.030	0.032	0.034	0.039	0.048	0.041	0.064
		θ_{12}	0.018	0.026	0.025	0.023	0.029	0.030	0.032	0.034	0.040	0.050	0.042	0.064
		γ_2	0.004	0.005	0.005	0.004	0.005	0.005	0.006	0.007	0.008	0.009	0.008	0.013
		θ_{21}	0.002	0.003	0.004	0.003	0.004	0.004	0.004	0.004	0.006	0.006	0.005	0.010
		θ_{22}	0.002	0.003	0.004	0.003	0.004	0.004	0.004	0.005	0.006	0.006	0.005	0.010
$\sigma=2$	$n=50$	γ_1	2.329	3.288	2.361	2.694	3.580	2.680	3.993	4.317	3.362	6.277	5.500	5.290
		θ_{11}	0.367	0.341	0.276	0.407	0.357	0.292	0.518	0.416	0.346	0.558	0.437	0.390
		θ_{12}	0.384	0.340	0.265	0.435	0.364	0.293	0.537	0.408	0.337	0.592	0.425	0.373
		γ_2	0.064	0.093	0.089	0.077	0.105	0.108	0.114	0.125	0.145	0.184	0.167	0.247
		θ_{21}	0.045	0.064	0.063	0.052	0.070	0.067	0.072	0.077	0.100	0.128	0.112	0.172
		θ_{22}	0.045	0.064	0.062	0.052	0.070	0.069	0.072	0.078	0.101	0.127	0.112	0.169
	$n=200$	γ_1	0.974	1.404	1.099	1.158	1.566	1.286	1.631	1.770	1.814	2.644	2.285	2.939
		θ_{11}	0.075	0.094	0.088	0.086	0.102	0.096	0.120	0.116	0.129	0.169	0.134	0.173
		θ_{12}	0.078	0.100	0.089	0.089	0.100	0.097	0.120	0.120	0.131	0.171	0.138	0.176
		γ_2	0.015	0.021	0.020	0.017	0.024	0.020	0.024	0.026	0.029	0.037	0.033	0.048
		θ_{21}	0.010	0.015	0.015	0.011	0.015	0.017	0.017	0.019	0.024	0.026	0.023	0.039
		θ_{22}	0.010	0.014	0.015	0.011	0.015	0.017	0.017	0.019	0.025	0.026	0.024	0.039

表示在第 k ($k=1,\dots,K$) 次模拟中 θ_{11} 的估计值. 每个表包含 3 种损失函数 (L1)–(L3) 和 4 种类型的误差分布 (D1)–(D4) 下的估计结果.

表 1 的主要发现总结如下. 首先, 均方误差随着样本量 n 的增加而减小, 这表明稳健 M 估计量的一致性. 其次, 与平稳变量相关的估计量的标准 $\sqrt{n}$ 速度相比, 非平稳变量相关的指标参数估计在所有情形下都收敛得非常快, 与我们的理论结果吻合, 因为理论上, γ_1, θ_{11} 和 θ_{12} 的估计量的收敛速度为 n^{-2}, 而 γ_2, θ_{21} 和 θ_{22} 的估计量的收敛速度为 n^{-3}. 最后, 所考虑的 3 个估计在涉及不同的误差分布和异常值的情形下都非常稳健. 结果表明, 除了 Cauchy 误差 ($t(1)$) 外, Huber 估计量是最有效的. 当误差服从 Cauchy 分布时, 最小绝对误差估计成为最有效的估计, 这与理论一致.

对于协整函数是可积的情形, 从表 2 所展示的结果中有一些新的发现, 总结如下. 首先, γ_1 的估计量是一致的, 但收敛速度相对较慢. 这主要是由于协整函数的可积性所致. 其次, 与 γ_1 相比, θ_{11} 和 θ_{12} 的归一化估计量收敛速度相对较快. 这一发现印证了推论 3.2 的注解和 Dong 等 [14] 的发现. 此外, 估计量的精度随着噪声强度的增加而下降, 与理论结果一致.

5 收益可预测性的实证分析

为了展示本文所提出模型的实际应用价值, 本文研究了它在股票收益预测中的适用性. 在文献中使用线性预测均值回归很常见, 这导致了实证研究结果中关于股票收益是否可预测存在相当大的分歧 (如文献 [6,47]). 最近, Kasparis 等[25] 使用非参数均值回归, Lee[29] 以及 Fan 和 Lee[19] 考虑分位数线性回归, Tu 等[43] 采用非参数分位数框架, 重新研究这一重要问题. 本文通过所提出的具有平稳和非平稳预测变量的多指标模型来说明收益可预测性中的非线性特征.

所使用的收益预测数据集来自 Welch 和 Goyal[47], 这组数据集被广泛应用于预测文献中, 如 Ren 等[40] 的平衡预测均值回归模型, Lee[29] 以及 Fan 和 Lee[19] 的线性分位数预测回归, Koo 等[28] 的协整变量线性预测, Lee 等[30] 的最小绝对收缩与选择算子 (least absolute shrinkage and selection operator, LASSO) 预测回归, 以及 Tu 等[43] 的非参数分位数预测回归等. 数据是从 1927 年 1 月至 2005 年 12 月的月度数据. 因变量超额股票收益被定义为标准普尔 500 指数收益率 (包括股息) 与一个月国库券利率之间的差额. 根据上述文献, 非平稳预测变量包括股息价格 (dividend-price, dp)、股息派息率 (dividend-payout ratio, de)、长期收益 (long term yield, lty)、账面市值比 (book to market ratios, bm)、国库券利率 (T-bill rate, tbl)、默认收益率价差 (default yield spread, dfy)、净股本扩张 (net equity expansion, ntis)、收益价格 (earnings-price, ep) 和期限利差 (term spread, tms), 而平稳变量包括默认回报价差 (default return spread, dfr)、长期回报率 (long term rate of return, ltr)、股票方差 (stock variance, svar) 和通货膨胀 (inflation, inf). 这些变量的一阶序列自相关性估计、平稳性检验结果及其时间序列图参见文献 [30], 而每个序列的详细说明和数据构建参见文献 [47].

为了衡量回归预测模型的效果, 我们使用样本外伪拟合度 (pseudo-R^2, 简记为 PR^2), 其定义为

$$\mathrm{PR}^2 = 1 - \frac{\sum_{t=T_1+1}^{T_2} \rho(\widehat{e}_t)}{\sum_{t=T_1+1}^{T_2} \rho(\overline{e}_t)},$$

其中 $\widehat{e}_t$ 和 $\overline{e}_t$ 分别是一个给定模型和只有一个常数预测变量的基准模型的情形下, 在时间 t 的样本外预测误差, 且预测样本从 T_1+1 到 T_2. 对于 $\rho(\cdot)$, 采用 4 种损失函数: (SE) 平方误差损失 $\rho(e)=e^2$; (AE) 绝对误差损失 $\rho(e)=|e|$; (HL) Huber 损失 $\rho_c(e)=\frac{1}{2}e^2\cdot 1\{|e|\leqslant c\}+(c|e|-\frac{1}{2}c^2)\cdot 1\{|e|>c\}$, 其中 $c=1.25$; (QL) 分位数损失 $\rho_\tau(e)=e\cdot(\tau-I(e<0))$. 正的 PR^2 表示给定模型的性能优于基准模型, 该值越大, 给定模型的性能越优.

由于篇幅限制, 本文与 Kasparis 等[25]、Lee[29] 和 Tu 等[43] 一样考虑两个子样本: (i) 从 1927 年 1 月到 2005 年 12 月的时间段, 其中上述研究曾发现了显著的可预测性; (ii) 从 1952 年 1 月到 2005 年 12 月的平静期, 其中显著和不显著的可预测性都有发现. 我们采用滚动窗口的样本外收益预测, 而滚动窗口大小 (RWS) 分别为 120 (10 年)、240 (20 年) 和 360 (30 年). 例如, 当 RWS 等于 120 时, 第一个子样本的预测从 1927 年 1 月开始到 2005 年 12 月结束.

与 Tu 等[43] 相比, 本文提出的模型允许使用多元非平稳预测变量以及那些平稳预测变量一起来捕获收益可预测性中的非线性特征. 可以看到, 非平稳和平稳预测变量的组合有许多选择, 我们仅取平稳预测变量 $z=\{\mathrm{inf},\mathrm{svar}\}$, 而非平稳预测变量或者为情形 A ($x=\{\mathrm{bm},\mathrm{lty}\}$), 或者为情形 B ($x=\{\mathrm{dp},\mathrm{tms}\}$). 对于非平稳指标, 考虑一个线性变换 $g_1(u)=\gamma u$ 或者一个可积函数变换

$$g_1(u)=\ell(u)=\gamma_1\cdot\exp(-u^2)+\gamma_2\cdot u\cdot\exp(-u^2).$$

类似地, 对于平稳指标, 使用 $g_2(v)=\delta v$ 或者

$$g_2(v)=\ell(v)=\delta_1\cdot\exp(-v^2)+\delta_2\cdot v\cdot\exp(-v^2).$$

非线性变换的选择是基于筛分文献中的归一化的 Hermite 多项式, 它们在相关函数空间中构成一组基. 因此, 总共有 4 种可能的组合. 不难得出, 穷尽所有预测变量的组合和非线性链接函数的类别进行分析虽然有意义, 但也具有很大的挑战性. 限于篇幅, 我们将其他的组合尝试留作未来研究.

表 3 报告了针对上述两组预测变量和链接函数类别, 以及损失函数为 (SE), (AE) 和 (HL) 的情形下, 超额收益预测的样本外 PR^2. 对于每个损失函数和每个窗口期, 产生最大样本外 PR^2 的链接函数的组合以粗体显示. 有几个有趣的发现. 首先, 与只有一个截距项的基准模型相比, 本文的模型显示了显著的收益可预测性. 对于情形 A ($x=\{\mathrm{bm},\mathrm{lty}\}$), 在第一个窗口期 (1927 年 1 月至 2005 年 12 月, 窗口期 I), 所有 3 个损失函数的线性预测回归优于非线性的类别. 在第二个窗口期间 (1952 年 1 月至 2005 年 12 月, 窗口期 II), 与 Campbell 和 Yogo [7] 发现的一样, 虽然可预测性有所下降, 但我们还是发现了非线性可预测性. 对于情形 B ($x=\{\mathrm{dp},\mathrm{tms}\}$), 我们发现线性预测模型对两个样本时期都没有可预测性. 对于 (SE), (AE) 和 (HL) 损失, 窗口期 I 的样本外 PR^2 分别达到 10.8%, 4.35% 和 11.2%. 在平静期, 非线性可预测性也被认为有所下降, 但当两个指标都以非线性方式进入模型时, 非线性可预测性仍然显著, 样本外 PR^2 高达 10.2%. 这一发现与 Kasparis 等 [25] 发现的只有一个非平稳预测变量的弱非线性可预测性形成鲜明对比. 综上所述, 本文所引入的具有平稳和非平稳指标的非线性预测回归显示出显著的收益可预测性.

由于近期学者对分位数收益预测的兴趣日益浓厚 (参见文献 [19, 29, 43]), 接下来研究所提议的模型在预测收益分位数方面的效果. 为此, 取 $\tau=0.05,0.10,0.20,\dots,0.90,0.95$. 注意到表 3 中发现的非平稳指标的非线性效应, 对于前面提到的两个子样本, 只在表 4 中报告 $x=\{\mathrm{dp},\mathrm{tms}\}$ 和 $z=\{\mathrm{inf},\mathrm{svar}\}$ 在上述指定的所有分位数水平上的结果. 平稳分量和非平稳分量的变换类别同上. 从表 4 中可以看到,

表 3　超额收益预测样本外 PR^2

			窗口期 I						窗口期 II					
			$g_2(v)=v$			$\ell(v)$			$g_2(v)=v$			$\ell(v)$		
	$g_1(u)$	RWS	SE	AE	HL	SE	AE	HL	SE	AE	HL	SE	AE	HL
情形 A	u	120	**3.06**	**3.19**	**2.58**	0.80	0.36	−3.62	−2.74	0.14	−2.13	−2.48	0.33	−9.12
		240	**6.58**	**3.26**	**7.63**	3.48	1.02	2.35	1.62	**0.63**	**3.94**	**3.07**	−0.72	1.83
		360	**6.57**	**0.76**	**7.84**	3.21	0.48	5.08	1.77	−3.31	3.43	3.30	1.18	**5.58**
	$\ell(u)$	120	0.69	0.14	1.09	0.53	0.35	1.56	−1.67	0.08	−1.60	**−0.46**	**0.37**	**−0.31**
		240	2.34	1.13	1.91	0.47	0.28	0.44	1.99	−0.24	1.54	0.07	0.12	0.01
		360	5.52	0.64	3.43	0.74	0.36	0.59	**4.75**	**1.35**	4.04	0.51	0.23	0.73
情形 B	u	120	−4.73	−3.60	−5.35	4.53	1.05	5.23	−3.06	−2.63	−3.35	2.28	1.54	**4.39**
		240	−7.58	−4.51	−7.74	9.67	2.26	**11.2**	−5.80	−3.51	−5.79	6.70	1.68	**8.82**
		360	−6.65	−3.78	−7.16	10.7	**4.35**	11.1	−7.48	−3.79	−7.42	8.91	**2.99**	8.95
	$\ell(u)$	120	2.15	**2.39**	4.30	**5.53**	1.35	**5.37**	−0.41	**2.58**	3.04	**2.63**	0.94	3.85
		240	8.75	**3.40**	10.3	**10.0**	2.56	10.9	**7.56**	**1.99**	8.59	6.12	0.69	6.97
		360	8.48	3.53	10.5	**10.8**	3.91	**11.2**	6.18	1.58	8.79	**10.2**	1.97	**9.10**

情形 A: $x=\{\mathrm{bm},\mathrm{lty}\}$ 和 $z=\{\mathrm{inf},\mathrm{svar}\}$. 情形 B: $x=\{\mathrm{dp},\mathrm{tms}\}$ 和 $z=\{\mathrm{inf},\mathrm{svar}\}$.

表 4　分位数超额收益预测的样本外 $\mathbf{PR^2}$

	$g_2(v)$	$g_1(u)$	RWS	$\tau=0.05$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
窗口期 I	v	u	120	−1.87	0.32	−1.09	−0.99	−2.41	−3.60	−4.08	−4.51	−6.05	8.86	−7.96
			240	**3.89**	1.33	0.11	−1.08	−2.66	−4.51	−4.69	−6.77	−9.21	−11.2	−15.8
			360	0.30	3.05	0.93	−1.00	−1.87	−3.82	−4.55	−7.14	−8.95	−13.7	−18.3
		$\ell(u)$	120	**2.83**	−2.64	1.31	1.97	0.89	**2.39**	0.96	0.61	4.74	0.87	6.05
			240	1.95	2.93	4.06	4.39	3.69	**3.40**	1.67	3.65	5.51	**6.71**	−1.14
			360	4.40	4.40	**5.08**	6.10	3.15	3.53	3.85	2.71	**7.89**	**6.50**	**5.82**
	$\ell(v)$	u	120	1.20	2.17	1.57	1.64	**1.45**	1.05	**2.28**	**2.61**	4.08	**2.39**	4.33
			240	3.19	4.02	4.06	5.08	**4.14**	2.26	**2.01**	2.13	6.36	7.16	**2.99**
			360	**4.74**	4.94	4.75	**6.40**	**4.67**	**4.35**	**4.81**	4.33	7.36	5.64	4.49
		$\ell(u)$	120	1.65	**2.22**	**2.06**	**2.52**	1.42	1.35	1.46	2.16	**5.45**	2.16	**5.17**
			240	3.40	**4.20**	**4.28**	**5.51**	3.06	2.56	1.68	**4.80**	**7.29**	6.33	0.45
			360	4.48	**5.14**	4.91	5.58	3.96	3.73	3.48	**4.98**	7.29	5.50	3.61
窗口期 II	v	u	120	−1.69	−0.11	−0.80	−0.91	−1.05	−2.62	−2.51	−3.16	−4.26	−5.52	−6.29
			240	**4.59**	1.45	0.64	−0.93	−1.49	−3.50	−3.46	−5.18	−7.27	−10.7	−12.5
			360	4.06	3.20	1.25	−1.47	−2.97	−3.80	−4.54	−4.92	−7.09	−13.5	−18.1
		$\ell(u)$	120	**2.01**	−3.89	2.09	2.51	0.72	**2.58**	0.33	−1.51	2.99	3.88	**7.13**
			240	2.91	2.91	3.47	4.23	2.52	**1.99**	−1.00	0.36	0.68	**4.00**	3.00
			360	7.65	7.37	**7.18**	7.12	3.31	1.58	1.71	−1.17	**−0.70**	**−4.28**	**−4.34**
	$\ell(v)$	u	120	0.69	1.80	1.58	2.56	**1.70**	1.54	**2.13**	**0.46**	2.10	4.88	4.19
			240	3.43	4.59	3.69	5.31	**3.11**	1.68	**0.83**	−1.94	1.97	3.25	**4.90**
			360	9.30	7.43	6.69	**8.13**	**4.48**	**2.99**	**2.29**	**0.46**	−1.14	−4.84	−4.73
		$\ell(u)$	120	1.61	**2.01**	**2.43**	**2.90**	1.00	0.94	1.90	−0.66	**3.69**	**5.02**	6.10
			240	4.28	**5.35**	**4.14**	**5.82**	1.87	0.69	−0.21	**0.93**	**2.92**	3.07	2.60
			360	**9.64**	**8.53**	7.15	7.18	3.36	1.87	0.79	−0.33	−2.26	−5.65	−8.88

线性分位数可预测性似乎只存在于较低的分位数水平, 如 $\tau=0.05, 0.10, 0.20$; 而非线性分位数可预测性在考虑的所有分位数水平上普遍存在. 即使在较低的分位数下, 除了极少数情形外, 所提出的非线性指标模型也优于线性模型. 此外, 随着分位数水平向两端移动, 非线性可预测性似乎变得更强, 特别是对第一个窗口期来说.

6　结论

本文研究了一类具有多种回归变量的可加单指标参数模型, 它们包括单位根过程、时间趋势和 α-混合严格平稳过程. 为了构造目标函数, 我们提议了一类损失函数, 允许它们是非光滑的 (如 LAD、分位数损失和 Huber 损失), 使得估计量具有稳健性, 但同时为模型的估计带来挑战; 采用广义函数正则序列对此类损失函数的逼近, 建立了一个关于未知参数的二次型对目标函数的逼近, 方便了可加单指标协整时间序列模型的稳健 M 估计极限理论的推导. 我们发现, 单位根过程的回归函数形式, 即 H-正则和 I- 正则, 对于估计量的收敛速度和极限分布有很大影响, 与相关文献一致. 通过 Monte Carlo

仿真实验验证了所提出的估计量在有限样本情形下的良好性能; 同时, 我们还对一个股票收益进行了实证分析, 从而证明了本文所提出的模型和估计方法的重要性和实用性.

此类参数模型的估计和极限理论的建立为非参数模型的估计与检验奠定了基础; 下一步我们将对相应的非参数非平稳模型的稳健 M 估计进行研究, 以填补相关文献的空白.

致谢 非常感谢评审的建设性评论和修改意见. 感谢在北京大学举行的第二十一次中国统计科学讨论会和 2022 年厦门大学现代统计学讨论会的与会者提出的宝贵意见和建议. 同时, 感谢高集体、彭彬、Mervyn Silvapulle 和许若凡提出的建设性意见和建议. 感谢北京大学统计科学中心和数理经济与数理金融教育部重点实验室的支持.

补充材料

本文的补充材料见 mathcn.scichina.com. 补充材料为作者提供的原始数据, 作者对其学术质量和内容负责.

参考文献

1 Babu G J. Strong representations for LAD estimators in linear models. Probab Theory Related Fields, 1989, 83: 547–558
2 Bai Z D, Rao C R, Wu Y, et al. M-estimation of multivariate linear regression parameters under a convex discrepancy function. Statist Sinica, 1992, 2: 237–254
3 Bickel P J. Edgeworth expansions in nonparametric statistics. Ann Statist, 1974, 2: 1–20
4 Bravo F, Li D G, Tjøstheim D. Robust nonlinear regression estimation in null recurrent time series. J Econometrics, 2021, 224: 416–438
5 Brockwell P, Davis R. Time Series: Theory and Method. New York: Springer-Verlag, 1991
6 Campbell J Y, Thompson S B. Predicting excess stock returns out of sample: Can anything beat the historical average? Rev Financ Stud, 2008, 21: 1509–1531
7 Campbell J Y, Yogo M. Efficient tests of stock return predictability. J Financ Econ, 2006, 81: 27–60
8 Chang Y, Park J Y. Index models with integrated time series. J Econometrics, 2003, 114: 73–106
9 Chang Y, Park J Y, Phillips P C B. Nonlinear econometric models with cointegrated and deterministically trending regressors. Econom J, 2001, 4: 1–36
10 Chen X R, Li D G, Li Q, et al. Nonparametric estimation of conditional quantile functions in the presence of irrelevant covariates. J Econometrics, 2019, 212: 433–450
11 Davis R A, Knight K, Liu J. M-estimation for autoregressions with infinite variance. Stochastic Process Appl, 1992, 40: 145–180
12 Dong C H, Gao J T. Specification testing driven by orthogonal series for nonlinear cointegration with endogeneity. Econom Theory, 2018, 34: 754–789
13 Dong C H, Gao J T, Peng B, et al. Smoothing the nonsmoothness. arXiv:2309.16348, 2023
14 Dong C H, Gao J T, Tjøstheim D, et al. Estimation for single-index and partially linear single-index integrated models. Ann Statist, 2016, 44: 425–453
15 Dong C H, Gao J, Tjøstheim D, et al. Specification testing for nonlinear multivariate cointegrating regressions. J Econometrics, 2017, 200: 104–117
16 Dong C H, Linton O. Additive nonparametric models with time variable and both stationary and nonstationary regressors. J Econometrics, 2018, 207: 212–236
17 Dong C H, Linton O, Peng B. A weighted sieve estimator for nonparametric time series models with nonstationary variables. J Econometrics, 2021, 222: 909–932
18 Fan J Q, Hu T C, Truong Y K. Robust nonparametric function estimation. Scand J Stat, 1994, 21: 433–446
19 Fan R, Lee J H. Predictive quantile regressions under persistence and conditional heteroskedasticity. J Econometrics, 2019, 213: 261–280
20 Gao J T, Li D G, Lin Z Y. Robust estimation in parametric time series models under long- and short-range-dependent structures. Aust N Z J Stat, 2009, 51: 161–181
21 Gel'fand I M, Shilov G E. Generalized Functions. New York: Academic Press, 1964
22 He X M, Shao Q M. On parameters of increasing dimensions. J Multivariate Anal, 2000, 73: 120–135
23 Huber P J. Robust estimation of a location parameter. Ann of Math Stud, 1964, 35: 73–101
24 Kanwal R P. Generalized Functions: Theory and Technique. New York: Academic Press, 1983
25 Kasparis I, Andreou E, Phillips P C B. Nonparametric predictive regression. J Econometrics, 2015, 185: 468–494

26 Knight K. Limit theory for autoregressive-parameter estimates in an infinite-variance random walk. Canad J Statist, 1989, 17: 261–278
27 Koenker R, Bassett G. Regression quantiles. Econometrica, 1978, 46: 33–50
28 Koo B, Anderson H M, Seo M H, et al. High-dimensional predictive regression in the presence of cointegration. J Econometrics, 2020, 219: 456–477
29 Lee J H. Predictive quantile regression with persistent covariates: IVX-QR approach. J Econometrics, 2016, 192: 105–118
30 Lee J H, Shi Z, Gao Z. On LASSO for predictive regression. J Econometrics, 2022, 229: 322–349
31 Li D G, Tjøstheim D, Gao J T. Estimation in nonlinear regression with Harris recurrent Markov chains. Ann Statist, 2016, 44: 1957–1987
32 Lin Y Q, Tu Y D, Yao Q W. Estimation for double-nonlinear cointegration. J Econometrics, 2020, 216: 175–191
33 Park J Y, Phillips P C B. Asymptotics for nonlinear transformations of integrated time series. Econom Theory, 1999, 15: 269–298
34 Park J Y, Phillips P C B. Nonstationary binary choice. Econometrica, 2000, 68: 1249–1280
35 Park J Y, Phillips P C B. Nonlinear regressions with integrated time series. Econometrica, 2001, 69: 117–161
36 Phillips P C B. A shortcut to LAD estimator asymptotics. Econom Theory, 1991, 7: 450–463
37 Phillips P C B. Robust nonstationary regression. Econom Theory, 1995, 11: 912–951
38 Phillips P C B, Solo V. Asymptotics for linear processes. Ann Statist, 1992, 20: 971-1001
39 Pollard D. Asymptotics for least absolute deviation regression estimators. Econom Theory, 1991, 7: 186–199
40 Ren Y, Tu Y D, Yi Y P. Balanced predictive regressions. J Empir Finance, 2019, 54: 118–142
41 Rockafellar R T. Convex Analysis. Princeton: Princeton Univ Press, 1970
42 Stein E M, Shakarchi R. Fourier Analysis: An Introduction. Princeton: Princeton Univ Press, 2003
43 Tu Y D, Liang H Y, Wang Q. Nonparametric inference for quantile cointegrations with stationary covariates. J Econometrics, 2022, 230: 453–482
44 van der Vaart A W. Weak Convergence and Empirical Processes with Applications to Statistics. New York: Springer, 1996
45 van der Vaart A W. Asymptotic Statistics. Cambridge: Cambridge Univ Press, 1998
46 Wang Q Y, Wu D S, Zhu K. Model checks for nonlinear cointegrating regression. J Econometrics, 2018, 207: 261–284
47 Welch I, Goyal A. A comprehensive look at the empirical performance of equity premium prediction. Rev Financ Stud, 2008, 21: 455–508
48 Xiao Z. Quantile cointegrating regression. J Econometrics, 2009, 150: 248–260
49 Xiao Z. Functional-coefficient cointegration models. J Econometrics, 2009, 152: 81–92
50 Zhou W X, Bose K, Fan J Q, et al. A new perspective on robust M-estimation: Finite sample theory and applications to dependence-adjusted multiple testing. Ann Statist, 2018, 46: 1904–1931

附录

本附录包括 3 部分. 附录 A 介绍广义函数正则序列的收敛性, 这为稳健估计奠定了理论基础; 附录 B 给出一些有用的引理, 出于篇幅考虑, 它们的证明都放在补充材料中. 所有主要结果的证明都包含在附录 C 中.

附录 A　广义函数正则序列

本节主要介绍广义函数正则序列的收敛性, 所涉及的基本概念可以在文献 [13] 及其参考文献中找到, 其证明见补充材料.

引理 A.1　令 $\rho(u)$ 满足假设 2.3. 定义 $\rho(u)$ 的正则序列,

$$\rho_m(u) = \int \rho(x)\phi_m(x-u)dx, \quad \text{其中 } \phi_m(x) = \sqrt{m/\pi}\exp(-mx^2), \quad m = 1, 2, \dots \tag{A.1}$$

(1) $\rho_m(u)$ $(m = 1, 2, \dots)$ 是凸函数, 并且任意阶可微, 特别地,

$$\rho_m'(u) = -\int \rho(x)\phi_m'(x-u)dx, \quad \rho_m''(u) = \int \rho(x)\phi_m''(x-u)dx. \tag{A.2}$$

另外, 当 $m \to \infty$ 时, $\rho_m'(x) \to \psi(x)$ 除了一个零测度集都成立, 并且 $\rho_m''(x) \to \rho''(x)$ 在广义函数意义下成立.

(2) $\sup_{u\in\mathbb{R}} |\rho_m(u) - \rho(u)| \leqslant Cm^{-1/2}$, 其中 C 为绝对常数. 因此, 对于任意随机变量 e, $\rho_m(e) - \rho(e) = O(m^{-1/2})$ 以概率 1 成立.

(3) 记 $f(u)$ 为随机变量 e 的密度函数. 假设 $f(u)$ 可微, $\int |f'(u)|du < \infty$, 且当 $|u| \to \infty$ 时, $\rho(u)f(u) \to 0$, 则 $\rho_m'(e) - \psi(e) = O_P(m^{-1/2})$.

(4) 假设 $\mathbb{E}[\rho''(e)]$ 存在. 另外假设 $\int |f''(u)|du < \infty$, 且当 $|u| \to \infty$ 时, $\rho(u)f(u) \to 0$, $\psi(u)f'(u) \to 0$. 则 $\mathbb{E}[\rho_m''(e) - \rho''(e)] = O(m^{-1/2})$.

(5) 对于所有给定的 ϵ, 有 $\mathbb{E}[\rho_m''(e+\epsilon) - \rho_m''(e)] = O(m^{-1/2} + \epsilon)$.

注 A.1 每个 $\phi_m(x)$ 是正态分布 $N(0, (2m)^{-1})$ 的密度, 它们形成一个 δ- 收敛序列, 即当 $m \to \infty$ 时, 在广义函数意义下, $\phi_m(x) \to \delta(x)$ (参见文献 [24, 42]). 另外, 由于 $\rho(u)$ 的凸性以及在广义函数意义下 $\rho_m''(u) \to \rho''(u)$, 所以对于较大的 m, 有 $\mathbb{E}[\rho_m''(e)] > 0$.

注 A.2 引理 A.1 的重要性在于收敛速度: $\sup_{u\in\mathbb{R}} |\rho_m(u) - \rho(u)| \leqslant Cm^{-1/2}$. 这个速度为下面的分析提供了坚实的理论基础. 从证明中注意到假设 2.3 中的 Lipschitz 条件可以放松为

$$|\rho(x) - \rho(y)| \leqslant C|x-y|^{\alpha}, \quad \alpha \in (0,1].$$

在这种情形下除了 m 的选择发生变化, 所有的结果都保持不变.

附录 B 引理

在假设 2.1 下, 不失一般性, 令 $x_0 = 0$ 以概率 1 成立. 类似于文献 [15, (A.5) 和 (A.6)], 有

$$x_t = \sum_{i=1}^{t} w_i = \sum_{j=-\infty}^{t} B_{t,j}\eta_j, \quad \text{其中 } B_{t,j} = \sum_{i=\max(1,j)}^{t} A_{i-j}, \tag{B.1}$$

并且对于 $t > s > 0$, 有

$$x_t = \sum_{i=1}^{t} w_i = \sum_{i=s+1}^{t} w_i + x_s = x_{ts} + x_{ts}^*, \tag{B.2}$$

其中

$$x_{ts} = \sum_{j=s+1}^{t} B_{t,j}\eta_j, \quad x_{ts}^* = x_s + \bar{x}_{ts}, \quad \bar{x}_{ts} = \sum_{j=-\infty}^{s}\left(\sum_{i=s+1}^{t} A_{i-j}\right)\eta_j.$$

因此, x_{ts} 独立于 x_{ts}^*. 记 $d_{ts}^2 := \mathbb{E}(x_{ts}x_{ts}^{\top})$, 并且由 B-N 分解 [38] 可知, 当 $t-s \to \infty$ 时, $d_{ts}^2 \sim t-s$. 表达式 (B.1) 和 (B.2) 以及下面的引理, 将用于渐近分析.

引理 B.1 在假设 2.1 成立的前提下, 设 $x_{j,t} = \theta_{1j}^{0\top} x_t$ 是单位根过程, 对于 $t > s$, 记 $x_{j,ts} = \theta_{1j}^{0\top} x_{ts}$ $(j = 1, \ldots, p_1)$, 其中 x_{ts} 由 (B.2) 给出; 令 $d_{j,t}^2 = \mathbb{E}(x_{j,t}^2)$ 和 $d_{j,ts}^2 = \mathbb{E}(x_{j,ts}^2)$.

(1) 对于充分大的 t, $d_{j,t}^{-1}x_{j,t}$ 的密度 $f_{jt}(u)$ 对 $u \in \mathbb{R}$ 和 t 一致有界. 同时, $f_{jt}(u)$ 的导数也是一致有界的. 因此, 一致 Lipschitz 条件

$$\sup_{u\in\mathbb{R}} |f_{jt}(u+\Delta u) - f_{jt}(u)| \leqslant C\,|\Delta u| \tag{B.3}$$

对于某个 $C > 0$ 和任意 $\Delta u \in \mathbb{R}$ 成立.

(2) 对于充分大的 $t-s$, $d_{j,ts}^{-1}x_{j,ts}$ 的密度 $f_{j,ts}(u)$ 对所有的 $u \in \mathbb{R}$ 和 (t,s) 一致有界. 此外, $f_{j,ts}(u)$ 具有有界导数, 并且也满足类似 (B.3) 的 Lipshitz 条件.

引理 B.2 令 $\psi(u)$ 是 $\rho(u)$ 的次梯度函数. 在假设 2.1–2.4 和 2.5(i) 下, 当 $n\to\infty$ 时,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi(e_t)\begin{pmatrix}Z_1(x_{nt})\\ Z_2(z_t)\end{pmatrix}\to_D B_1\equiv\begin{pmatrix}\int_0^1 Z_1(B(r))dU(r)\\ N(0,a_1\Sigma)\end{pmatrix}, \tag{B.4}$$

其中 $x_{nt}\equiv n^{-1/2}x_t$, $Z_1(x_{nt})$ 和 $Z_2(z_t)$ 由 (3.1) 定义, $\Sigma=\int_0^1[\mathbb{E}Z_2(h(r,v_1))\,Z_2(h(r,v_1))^\top]dr$, 而 $a_1>0$ 由假设 2.4 给出.

此外, 当 $n\to\infty$ 时,

$$\frac{1}{n}\sum_{t=1}^{n}Z_2(z_t)Z_2(z_t)^\top\to_P\Sigma, \tag{B.5}$$

并且

$$\frac{1}{n}\sum_{t=1}^{n}\begin{pmatrix}Z_1(x_{nt})Z_1(x_{nt})^\top\\ Z_2(z_t)Z_1(x_{nt})^\top\end{pmatrix}\to_D\begin{pmatrix}\int_0^1 Z_1(B(r))Z_1(B(r))^\top dr\\ \int_0^1[\mathbb{E}Z_2(h(r,v_1))]Z_1(B(r))^\top dr\end{pmatrix}. \tag{B.6}$$

由此可得, 当 $n\to\infty$ 时,

$$\frac{1}{n}\sum_{t=1}^{n}Z_tZ_t^\top\to_D\frac{1}{a_2}B, \tag{B.7}$$

并且 (B.4) 和 (B.7) 是联合收敛的, 其中 $Z_t=(Z_1(x_{nt})^\top,Z_2(z_t)^\top)^\top$, 矩阵 B 由 (3.2) 定义.

引理 B.3 令 Z_t 如引理 B.2 所定义. 在假设 2.1–2.4 和 2.5(i) 下, 当 $n\to\infty$ 时,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}[\rho_m'(e_t)-\psi(e_t)]Z_t=O_P(m^{-1/2}n^{1/2}), \tag{B.8}$$

$$\frac{1}{n}\sum_{t=1}^{n}[\rho_m''(e_t)-\mathbb{E}\rho_m''(e_t)]Z_tZ_t^\top=O_P(n^{-1/2}), \tag{B.9}$$

其中 $m=[n^{1+\varepsilon}]$, $\varepsilon>0$.

引理 B.4 记 $Z_t=(Z_1(x_{nt})^\top,Z_2(z_t)^\top)^\top$, 其中 Z_1 和 Z_2 由 (3.11) 定义. 在假设 2.1–2.4 和 2.5(ii) 下, 当 $n\to\infty$ 时,

$$\sum_{t=1}^{n}[\rho_m'(e_t)-\psi(e_t)]Z_1(x_{nt})=O_P(n^{1/4}m^{-1/2}), \tag{B.10}$$

$$\sum_{t=1}^{n}[\rho_m'(e_t)-\psi(e_t)]Z_2(z_t)=O_P(n^{1/2}m^{-1/2}), \tag{B.11}$$

并且

$$\sum_{t=1}^{n}[\rho_m''(e_t)-\mathbb{E}\rho_m''(e_t)]Z_1(x_{nt})Z_1(x_{nt})^\top=O_P(n^{-1/4}), \tag{B.12}$$

$$\sum_{t=1}^{n}[\rho_m''(e_t)-\mathbb{E}\rho_m''(e_t)]Z_2(z_t)Z_2(z_t)^\top=O_P(n^{-1/2}), \tag{B.13}$$

$$\sum_{t=1}^{n}[\rho_m''(e_t)-\mathbb{E}\rho_m''(e_t)]Z_1(x_{nt})Z_2(z_t)^\top=O_P(n^{-1/2}), \tag{B.14}$$

其中 $m=[n^{1+\varepsilon}]$, $\varepsilon>0$.

引理 B.5 对于引理 B.4 中定义的 Z_t, 在假设 2.1–2.4 和 2.5(ii) 下, 当 $n\to\infty$ 时, 有联合收敛

$$\mathcal{R}_n \equiv \sum_{t=1}^n Z_t Z_t^\top \to_D \mathcal{R}, \tag{B.15}$$

$$\sum_{t=1}^n \psi(e_t) Z_t \to_D \mathcal{R}^{1/2} N(0, a_1 I), \tag{B.16}$$

其中 a_1 由假设 2.4 给出, 并且单位矩阵的维数是 $d_1+p_1+d_2p_2+p_2$. 因此, 当 $n\to\infty$ 时,

$$\mathcal{R}_n^{-1/2} \sum_{t=1}^n \psi(e_t) Z_t \to_D N(0, a_1 I). \tag{B.17}$$

附录 C 主要结果的证明

定理 3.1 的证明 假定 λ_{ij} 和 q_{ij} $(i=1,2,\ j=1,\dots,p_i)$ 依赖于 n, 且当 $n\to\infty$ 时, $\lambda_{ij}, q_{ij}\to\infty$, 它们的具体关系随后给出; 令 $\beta_{1j}=q_{1j}(\gamma_{1j}-\gamma_{1j}^0)$, $\beta_{2j}=q_{2j}(\gamma_{2j}-\gamma_{2j}^0)$, $\alpha_{1j}=\lambda_{1j}(\theta_{1j}-\theta_{1j}^0)$, $\alpha_{2j}=\lambda_{2j}(\theta_{2j}-\theta_{2j}^0)$. 为了方便, 将目标函数减去一个常数, 然后重新参数化:

$$\begin{aligned}
L_n(\theta,\gamma) &= \sum_{t=1}^n \left[\rho\left(y_t - \sum_{j=1}^{p_1}\gamma_{1j}g_{1j}(x_t^\top\theta_{1j}) - \sum_{j=1}^{p_2}\gamma_{2j}g_{2j}(z_t^\top\theta_{2j})\right) - \rho(e_t)\right] \\
&= \sum_{t=1}^n \left[\rho\left(e_t + \sum_{j=1}^{p_1}[\gamma_{1j}^0 g_{1j}(x_t^\top\theta_{1j}^0) - \gamma_{1j}g_{1j}(x_t^\top\theta_{1j})]\right.\right. \\
&\quad \left.\left. + \sum_{j=1}^{p_2}[\gamma_{2j}^0 g_{2j}(z_t^\top\theta_{2j}^0) - \gamma_{2j}g_{2j}(z_t^\top\theta_{2j})]\right) - \rho(e_t)\right] \\
&= \sum_{t=1}^n \left[\rho\left(e_t + \sum_{j=1}^{p_1}[\gamma_{1j}^0 g_{1j}(x_t^\top\theta_{1j}^0) - \gamma_{1j}g_{1j}(x_t^\top\theta_{1j}^0 + x_t^\top\lambda_{1j}^{-1}\alpha_{1j})]\right.\right. \\
&\quad \left.\left. + \sum_{j=1}^{p_2}[\gamma_{2j}^0 g_{2j}(z_t^\top\theta_{2j}^0) - \gamma_{2j}g_{2j}(z_t^\top\theta_{2j}^0 + z_t^\top\lambda_{2j}^{-1}\alpha_{2j})]\right) - \rho(e_t)\right] \\
&= \sum_{t=1}^n \left[\rho\left(e_t + \sum_{j=1}^{p_1}[\gamma_{1j}^0(g_{1j}(x_t^\top\theta_{1j}^0) - g_{1j}(x_t^\top\theta_{1j}^0 + x_t^\top\lambda_{1j}^{-1}\alpha_{1j}))\right.\right. \\
&\quad + (\gamma_{1j}^0-\gamma_{1j})g_{1j}(x_t^\top\theta_{1j}^0 + x_t^\top\lambda_{1j}^{-1}\alpha_{1j})] + \sum_{j=1}^{p_2}[\gamma_{2j}^0(g_{2j}(z_t^\top\theta_{2j}^0) - g_{2j}(z_t^\top\theta_{2j}^0 + z_t^\top\lambda_{2j}^{-1}\alpha_{2j})) \\
&\quad \left.\left. + (\gamma_{2j}^0-\gamma_{2j})g_{2j}(z_t^\top\theta_{2j}^0 + z_t^\top\lambda_{2j}^{-1}\alpha_{2j})]\right) - \rho(e_t)\right] \\
&= \sum_{t=1}^n \left[\rho\left(e_t + \sum_{j=1}^{p_1}[\gamma_{1j}^0(g_{1j}(x_t^\top\theta_{1j}^0) - g_{1j}(x_t^\top\theta_{1j}^0 + x_t^\top\lambda_{1j}^{-1}\alpha_{1j})) + q_{1j}^{-1}\beta_{1j}g_{1j}(x_t^\top\theta_{1j}^0 + x_t^\top\lambda_{1j}^{-1}\alpha_{1j})]\right.\right. \\
&\quad \left.\left. + \sum_{j=1}^{p_2}[\gamma_{2j}^0(g_{2j}(z_t^\top\theta_{2j}^0) - g_{2j}(z_t^\top\theta_{2j}^0 + z_t^\top\lambda_{2j}^{-1}\alpha_{2j})) + q_{2j}^{-1}\beta_{2j}g_{2j}(z_t^\top\theta_{2j}^0 + z_t^\top\lambda_{2j}^{-1}\alpha_{2j})]\right) - \rho(e_t)\right] \\
&\equiv \tilde{L}_n(\alpha,\beta).
\end{aligned}$$

这种重新参数化的方法在文献中经常被采用 (参见文献 [1,3,11,37]). 然后考虑 $\tilde{L}_n(\alpha,\beta)$ 对 (α,β) 的

优化. 显然, 如果 $(\widehat{\alpha},\widehat{\beta})$ 使得 $\tilde{L}_n(\alpha,\beta)$ 达到最小, 则 $(\widehat{\theta},\widehat{\gamma})$ 使得 $L_n(\theta,\gamma)$ 达到最小, 其中

$$\widehat{\beta}_{1j}=q_{1j}(\widehat{\gamma}_{1j}-\gamma_{1j}^0),\quad \widehat{\beta}_{2j}=q_{2j}(\widehat{\gamma}_{2j}-\gamma_{2j}^0),\quad \widehat{\alpha}_{1j}=\lambda_{1j}(\widehat{\theta}_{1j}-\theta_{1j}^0),\quad \widehat{\alpha}_{2j}=\lambda_{2j}(\widehat{\theta}_{2j}-\theta_{2j}^0).$$

为方便起见, 记

$$\Delta_{1t}(\alpha_{1j},\beta_{1j})=\gamma_{1j}^0(g_{1j}(x_t^\top\theta_{1j}^0)-g_{1j}(x_t^\top\theta_{1j}^0+x_t^\top\lambda_{1j}^{-1}\alpha_{1j}))+q_{1j}^{-1}\beta_{1j}g_{1j}(x_t^\top\theta_{1j}^0+x_t^\top\lambda_{1j}^{-1}\alpha_{1j})$$

和

$$\Delta_{2t}(\alpha_{2j},\beta_{2j})=\gamma_{2j}^0(g_{2j}(z_t^\top\theta_{2j}^0)-g_{2j}(z_t^\top\theta_{2j}^0+z_t^\top\lambda_{2j}^{-1}\alpha_{2j}))+q_{2j}^{-1}\beta_{2j}g_{2j}(z_t^\top\theta_{2j}^0+z_t^\top\lambda_{2j}^{-1}\alpha_{2j}).$$

因此, $\tilde{L}_n(\alpha,\beta)$ 被简记为

$$\tilde{L}_n(\alpha,\beta)=\sum_{t=1}^n\bigg[\rho\bigg(e_t+\sum_{j=1}^{p_1}\Delta_{1t}(\alpha_{1j},\beta_{1j})+\sum_{j=1}^{p_2}\Delta_{2t}(\alpha_{2j},\beta_{2j})\bigg)-\rho(e_t)\bigg].\tag{C.1}$$

另外, 使用 Taylor 展开公式, 可得

$$\begin{aligned}\Delta_{1t}(\alpha_{1j},\beta_{1j})&=\gamma_{1j}^0(g_{1j}(x_t^\top\theta_{1j}^0)-g_{1j}(x_t^\top\theta_{1j}^0+x_t^\top\lambda_{1j}^{-1}\alpha_{1j}))+q_{1j}^{-1}\beta_{1j}g_{1j}(x_t^\top\theta_{1j}^0)\\&\quad+q_{1j}^{-1}\beta_{1j}[g_{1j}(x_t^\top\theta_{1j}^0+x_t^\top\lambda_{1j}^{-1}\alpha_{1j})-g_{1j}(x_t^\top\theta_{1j}^0)]\\&=-\gamma_{1j}^0\bigg[\dot{g}_{1j}(x_t^\top\theta_{1j}^0)x_t^\top\lambda_{1j}^{-1}\alpha_{1j}+\frac{1}{2}\ddot{g}_{1j}(x_t^\top\theta_{1j}^{0*})(x_t^\top\lambda_{1j}^{-1}\alpha_{1j})^2\bigg]\\&\quad+q_{1j}^{-1}\beta_{1j}g_{1j}(x_t^\top\theta_{1j}^0)\\&\quad+q_{1j}^{-1}\beta_{1j}\bigg[\dot{g}_{1j}(x_t^\top\theta_{1j}^0)x_t^\top\lambda_{1j}^{-1}\alpha_{1j}+\frac{1}{2}\ddot{g}_{1j}(x_t^\top\theta_{1j}^{0*})(x_t^\top\lambda_{1j}^{-1}\alpha_{1j})^2\bigg],\end{aligned}$$

其中 $x_t^\top\theta_{1j}^{0*}$ 位于连接 $x_t^\top\theta_{1j}^0$ 和 $x_t^\top\theta_{1j}^0+x_t^\top\lambda_{1j}^{-1}\alpha_{1j}$ 的线段上.

假定假设 2.5(i) 成立. 令 $\lambda_{1j}=\dot{\nu}_j(\sqrt{n})n$, $q_{1j}=\nu_j(\sqrt{n})\sqrt{n}$ $(j=1,\ldots,p_1)$. 注意 $x_{nt}=x_t/\sqrt{n}$. 在 $\Delta_{1t}(\alpha_{1j},\beta_{1j})$ 中观察到

$$\begin{aligned}\ddot{g}_{1j}(x_t^\top\theta_{1j}^{0*})(x_t^\top\lambda_{1j}^{-1}\alpha_{1j})^2&=\ddot{g}_{1j}(x_{nt}^\top\theta_{1j}^{0*})(x_{nt}^\top\alpha_{1j})^2\frac{\ddot{\nu}_j(\sqrt{n})}{\dot{\nu}_j(\sqrt{n})^2n}=o_P\bigg(\frac{1}{\dot{\nu}_j(\sqrt{n})n}\bigg),\\q_{1j}^{-1}\dot{g}_{1j}(x_t^\top\theta_{1j}^0)x_t^\top\lambda_{1j}^{-1}\alpha_{1j}&=\frac{1}{\nu_j(\sqrt{n})n}\dot{g}_{1j}(x_{nt}^\top\theta_{1j}^0)x_{nt}^\top\alpha_{1j}=O_P\bigg(\frac{1}{\nu_j(\sqrt{n})n}\bigg),\\q_{1j}^{-1}\ddot{g}_{1j}(x_t^\top\theta_{1j}^{0*})(x_t^\top\lambda_{1j}^{-1}\alpha_{1j})^2&=\ddot{g}_{1j}(x_{nt}^\top\theta_{1j}^{0*})(x_{nt}^\top\alpha_{1j})^2\frac{\ddot{\nu}_j(\sqrt{n})}{\nu_j(\sqrt{n})\dot{\nu}_j(\sqrt{n})^2n\sqrt{n}}\\&=o_P\bigg(\frac{1}{\nu_j(\sqrt{n})\dot{\nu}_j(\sqrt{n})n\sqrt{n}}\bigg),\end{aligned}$$

由此得出结论

$$\Delta_{1t}(\alpha_{1j},\beta_{1j})=\frac{1}{\sqrt{n}}[-\gamma_{1j}^0\dot{g}_{1j}(x_{nt}^\top\theta_{1j}^0)x_{nt}^\top\alpha_{1j}+\beta_{1j}g_{1j}(x_{nt}^\top\theta_{1j}^0)](1+o_P(1)).\tag{C.2}$$

类似地,

$$\Delta_{2t}(\alpha_{2j},\beta_{2j})=\gamma_{2j}^0(g_{2j}(z_t^\top\theta_{2j}^0)-g_{2j}(z_t^\top\theta_{2j}^0+z_t^\top\lambda_{2j}^{-1}\alpha_{2j}))+q_{2j}^{-1}\beta_{2j}g_{2j}(z_t^\top\theta_{2j}^0)$$

$$
\begin{aligned}
&+ q_{2j}^{-1}\beta_{2j}[g_{2j}(z_t^\top\theta_{2j}^0 + z_t^\top\lambda_{2j}^{-1}\alpha_{2j}) - g_{2j}(z_t^\top\theta_{2j}^0)]\\
={}& -\gamma_{2j}^0\left[\dot g_{2j}(z_t^\top\theta_{2j}^0)z_t^\top\lambda_{2j}^{-1}\alpha_{2j} + \frac{1}{2}\ddot g_{2j}(z_t^\top\theta_{2j}^{0*})(z_t^\top\lambda_{2j}^{-1}\alpha_{2j})^2\right]\\
&+ q_{2j}^{-1}\beta_{2j}g_{2j}(z_t^\top\theta_{2j}^0)\\
&+ q_{2j}^{-1}\beta_{2j}\left[\dot g_{2j}(z_t^\top\theta_{2j}^0)z_t^\top\lambda_{2j}^{-1}\alpha_{2j} + \frac{1}{2}\ddot g_{2j}(z_t^\top\theta_{2j}^{0*})(z_t^\top\lambda_{2j}^{-1}\alpha_{2j})^2\right],
\end{aligned}
$$

其中 $z_t^\top\theta_{2j}^{0*}$ 位于连接 $z_t^\top\theta_{2j}^0$ 和 $z_t^\top\theta_{2j}^0 + z_t^\top\lambda_{2j}^{-1}\alpha_{2j}$ 的线段上.

令 $\lambda_{2j} = \sqrt{n}$ 和 $q_{2j} = \sqrt{n}$ $(j = 1, \dots, p_2)$. 于是在 $\Delta_{2t}(\alpha_{2j}, \beta_{2j})$ 中,

$$
\begin{aligned}
&\ddot g_{2j}(z_t^\top\theta_{2j}^{0*})(z_t^\top\lambda_{2j}^{-1}\alpha_{2j})^2 = \frac{1}{n}\ddot g_{2j}(z_t^\top\theta_{2j}^{0*})(z_t^\top\alpha_{2j})^2 = O_P(n^{-1}),\\
&q_{2j}^{-1}\beta_{2j}\dot g_{2j}(z_t^\top\theta_{2j}^0)z_t^\top\lambda_{2j}^{-1}\alpha_{2j} = O_P(n^{-1}),\\
&q_{2j}^{-1}\beta_{2j}\ddot g_{2j}(z_t^\top\theta_{2j}^{0*})(z_t^\top\lambda_{2j}^{-1}\alpha_{2j})^2 = O_P(n^{-3/2}),
\end{aligned}
$$

由此可得

$$
\Delta_{2t}(\alpha_{2j}, \beta_{2j}) = \frac{1}{\sqrt{n}}[-\gamma_{2j}^0\dot g_{2j}(z_t^\top\theta_{2j}^0)z_t^\top\alpha_{2j} + \beta_{2j}g_{2j}(z_t^\top\theta_{2j}^0)](1 + o_P(1)). \tag{C.3}
$$

因此, (C.2) 和 (C.3) 分别给出 $\Delta_{1t}(\alpha_{1j}, \beta_{1j})$ 和 $\Delta_{2t}(\alpha_{2j}, \beta_{2j})$ 的主项, 而它们都是 $O_P(n^{-1/2})$. 于是,

$$
\Delta_t(\alpha, \beta) \equiv \sum_{j=1}^{p_1}\Delta_{1t}(\alpha_{1j}, \beta_{1j}) + \sum_{j=1}^{p_2}\Delta_{2t}(\alpha_{2j}, \beta_{2j}), \tag{C.4}
$$

阶为 $O_P(n^{-1/2})$, 从而 Taylor 展开能够成立.

因为 $\rho(\cdot)$ 满足假设 2.3, 对于引理 A.1 中定义的正则序列 $\rho_m(u)$, 有

$$
\sup_{u\in\mathbb{R}}|\rho_m(u) - \rho(u)| \leqslant Cm^{-1/2}.
$$

因此, $\sup_{\alpha,\beta}|\tilde L_n(\alpha, \beta) - \tilde L_{mn}(\alpha, \beta)| = O(nm^{-1/2})$ 依概率 1 成立, 其中

$$
\tilde L_{mn}(\alpha, \beta) = \sum_{t=1}^{n}[\rho_m(e_t + \Delta_t(\alpha, \beta)) - \rho_m(e_t)]. \tag{C.5}
$$

注意, 如果令 m 充分大, 如 $m = n^{2+\varepsilon}$, $\varepsilon > 0$, 则 $\tilde L_n(\alpha, \beta)$ 与 $\tilde L_{mn}$ 的差异可以忽略不计. $\tilde L_{mn}$ 的好处是 $\rho_m(\cdot)$ 无限次可微. 进而考虑 $\rho_m(\cdot)$ 的 Taylor 展开, 可得

$$
\tilde L_{mn}(\alpha, \beta) = \sum_{t=1}^{n}\left\{\rho_m'(e_t)\Delta_t(\alpha, \beta) + \frac{1}{2}\rho_m''(e_t)[\Delta_t(\alpha, \beta)]^2 + O_P(n^{-3/2})\right\},
$$

因为 $\rho_m'''(e_t) = f_e'(0) + o_P(1)$ 对于 m 一致成立.

回顾正文里记号, $Z_t = (Z_1(x_{nt})^\top, Z_2(z_t)^\top)^\top$, $\Lambda = (\Lambda_1^\top, \Lambda_2^\top)^\top$, 其中

$$
\Lambda_1 = (\alpha_{11}, \dots, \alpha_{1p_1}, \beta_{11}, \dots, \beta_{1p_1})^\top, \quad \Lambda_2 = (\alpha_{21}, \dots, \alpha_{2p_2}, \beta_{21}, \dots, \beta_{2p_2})^\top,
$$

从而有

$$\Delta_t(\alpha,\beta)=\frac{1}{\sqrt{n}}[Z_1(x_{nt})^\top\Lambda_1+Z_2(z_t)^\top\Lambda_2]=\frac{1}{\sqrt{n}}Z_t^\top\Lambda.$$

于是, 忽略高阶无穷小项, 可得

$$\begin{aligned}
\tilde{L}_{mn}(\alpha,\beta)&=\sum_{t=1}^{n}\left\{\rho_m'(e_t)\Delta_t(\alpha,\beta)+\frac{1}{2}\rho_m''(e_t)[\Delta_t(\alpha,\beta)]^2\right\}\\
&=\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi(e_t)Z_t^\top\right)\Lambda+\frac{1}{2}\Lambda^\top\left(\frac{1}{n}\sum_{t=1}^{n}\rho_m''(e_t)Z_tZ_t^\top\right)\Lambda\\
&\quad+\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}[\rho_m'(e_t)-\psi(e_t)]Z_t^\top\right)\Lambda\\
&=\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi(e_t)Z_t^\top\right)\Lambda+\frac{1}{2}\mathbb{E}[\rho_m''(e_t)]\Lambda^\top\left(\frac{1}{n}\sum_{t=1}^{n}Z_tZ_t^\top\right)\Lambda\\
&\quad+\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}[\rho_m'(e_t)-\psi(e_t)]Z_t^\top\right)\Lambda\\
&\quad+\frac{1}{2}\Lambda^\top\left(\frac{1}{n}\sum_{t=1}^{n}[\rho_m''(e_t)-\mathbb{E}\rho_m''(e_t)]Z_tZ_t^\top\right)\Lambda\\
&=\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi(e_t)Z_t^\top\right)\Lambda+\frac{1}{2}a_2\Lambda^\top\left(\frac{1}{n}\sum_{t=1}^{n}Z_tZ_t^\top\right)\Lambda\\
&\quad+\frac{1}{2}\{\mathbb{E}[\rho_m''(e_t)]-a_2\}\Lambda^\top\left(\frac{1}{n}\sum_{t=1}^{n}Z_tZ_t^\top\right)\Lambda\\
&\quad+\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}[\rho_m'(e_t)-\psi(e_t)]Z_t^\top\right)\Lambda\\
&\quad+\frac{1}{2}\Lambda^\top\left(\frac{1}{n}\sum_{t=1}^{n}[\rho_m''(e_t)-\mathbb{E}\rho_m''(e_t)]Z_tZ_t^\top\right)\Lambda.
\end{aligned}$$

记

$$Q_n(\alpha,\beta)\equiv\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi(e_t)Z_t^\top\right)\Lambda+\frac{a_2}{2}\Lambda^\top\left(\frac{1}{n}\sum_{t=1}^{n}Z_tZ_t^\top\right)\Lambda,\tag{C.6}$$

然后由引理 A.1 和 B.3, 对于所有的 (α,β), 有 $\tilde{L}_{mn}(\alpha,\beta)-Q_n(\alpha,\beta)\to_P 0$; 对于任意的 $c>0$, 有

$$\begin{aligned}
&\sup_{\|\alpha\|^2+\|\beta\|^2\leqslant c}|\tilde{L}_{mn}(\alpha,\beta)-Q_n(\alpha,\beta)|\\
&\leqslant\frac{c}{2}|\mathbb{E}[\rho_m''(e_t)]-a_2|\left\|\frac{1}{n}\sum_{t=1}^{n}Z_tZ_t^\top\right\|+c\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}[\rho_m'(e_t)-\psi(e_t)]Z_t\right\|\\
&\quad+\frac{c}{2}\left\|\frac{1}{n}\sum_{t=1}^{n}[\rho_m''(e_t)-\mathbb{E}\rho_m''(e_t)]Z_tZ_t^\top\right\|\\
&=O_P(m^{-1/2}n^{1/2}+n^{-1/2}).
\end{aligned}$$

因为 $a_2>0$, 所以二次型 (C.6) 有唯一的最小值,

$$\widehat{\Lambda}_Q=a_2^{-1}\left(\frac{1}{n}\sum_{t=1}^{n}Z_tZ_t^{\top}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi(e_t)Z_t. \tag{C.7}$$

注意, 向量 $\widehat{\Lambda}$ 包含所有感兴趣的估计量, 即 $\widehat{\Lambda}=(\widehat{\Lambda}_1^{\top},\widehat{\Lambda}_2^{\top})^{\top}$, 其中 $\widehat{\Lambda}_1$ 和 $\widehat{\Lambda}_2$ 在 (3.3) 中定义.

类似于文献 [13], 我们可以证明 $\widehat{\Lambda}=\widehat{\Lambda}_Q+o_P(1)$. 因此, 利用引理 B.2 中的联合收敛性, 当 $n\to\infty$ 时, $\widehat{\Lambda}=O_P(1)$, 并且

$$\widehat{\Lambda}\to_D B^{-1}B_1,$$

其中矩阵 B 和向量 B_1 都在 (3.2) 中给出. □

推论 3.1 的证明 观察到

$$\begin{aligned}
\widehat{e}_t&=y_t-\sum_{j=1}^{p_1}\widehat{\gamma}_{1j}g_{1j}(\widehat{\theta}_{1j}^{\top}x_t)-\sum_{j=1}^{p_2}\widehat{\gamma}_{2j}g_{2j}(\widehat{\theta}_{2j}^{\top}z_t)\\
&=e_t+\sum_{j=1}^{p_1}[\gamma_{1j}^0g_{1j}(\theta_{1j}^{0\top}x_t)-\widehat{\gamma}_{1j}g_{1j}(\widehat{\theta}_{1j}^{\top}x_t)]+\sum_{j=1}^{p_1}[\gamma_{2j}^0g_{2j}(\theta_{2j}^{0\top}z_t)-\widehat{\gamma}_{2j}g_{2j}(\widehat{\theta}_{2j}^{\top}z_t)]\\
&=e_t+\sum_{j=1}^{p_1}\Delta_{1t}(\widehat{\alpha}_{1j},\widehat{\beta}_{1j})+\sum_{j=1}^{p_2}\Delta_{2t}(\widehat{\alpha}_{2j},\widehat{\beta}_{2j})\\
&=e_t+\Delta_t(\widehat{\alpha},\widehat{\beta}),
\end{aligned}$$

同时, 由定理 3.1 的证明中 $\widehat{\Lambda}$ 的收敛性可得, 对于 $\epsilon\in(0,1/2)$, $\Delta_t(\widehat{\alpha},\widehat{\beta})=n^{-1/2}Z_t^{\top}\widehat{\Lambda}=O_P(n^{-1/2+\epsilon})$ 对 t 一致成立. 注意

$$\begin{aligned}
\widehat{a}_1-a_1&=\frac{1}{n}\sum_{t=1}^{n}\{[\psi(\widehat{e}_t)]^2-\mathbb{E}[\psi(e_t)]^2\}\\
&=\frac{1}{n}\sum_{t=1}^{n}\{[\psi(\widehat{e}_t)]^2-[\psi(e_t)]^2\}+\frac{1}{n}\sum_{t=1}^{n}\{[\psi(e_t)]^2-\mathbb{E}[\psi(e_t)]^2\}\\
&=\frac{1}{n}\sum_{t=1}^{n}\{\psi(\widehat{e}_t)-\psi(e_t)\}\{\psi(\widehat{e}_t)+\psi(e_t)\}+\frac{1}{n}\sum_{t=1}^{n}\{[\psi(e_t)]^2-\mathbb{E}[\psi(e_t)]^2\}\\
&\equiv I_1+I_2.
\end{aligned}$$

由于 $\{\psi(e_t)\}$ 是鞅差序列, 因此它是遍历的, 进而 $\{\psi(e_t)^2\}$ 也是如此, 故第二项 $I_2=o_P(1)$. 在第一项中, 因为 $\rho(\cdot)$ 满足 Lipschitz 条件, 所以 $|\psi(u)|\leqslant C$ 一致成立, 且

$$\begin{aligned}
|I_1|&\leqslant\frac{c}{n}\sum_{t=1}^{n}|\psi(\widehat{e}_t)-\psi(e_t)|\\
&\leqslant\frac{c}{n}\sum_{t=1}^{n}|\psi(\widehat{e}_t)-\rho_m'(\widehat{e}_t)|+\frac{c}{n}\sum_{t=1}^{n}|\rho_m'(\widehat{e}_t)-\rho_m'(e_t)|+\frac{c}{n}\sum_{t=1}^{n}|\rho_m'(e_t)-\psi(e_t)|\\
&\equiv I_{11}+I_{12}+I_{13}.
\end{aligned}$$

根据引理 A.1, 可得 $\mathbb{E}|\rho_m'(e_t)-\psi(e_t)|=O(m^{-1/2})$, 所以 $I_{13}=o_P(1)$. 同样地, 根据引理 A.1(3), 可得 $\widehat{e}_t=e_t+\Delta_t(\widehat{\alpha},\widehat{\beta})$, 对 $\Delta_t(\widehat{\alpha},\widehat{\beta})$ 取条件期望, 有 $I_{11}=o_P(1)$. 为了证明 $I_{12}=o_P(1)$, 注意到

$\rho'_m(\widehat{e}_t)-\rho'_m(e_t)=\rho''_m(e_t^*)\Delta_t(\widehat{\alpha},\widehat{\beta})$, 其中 e_t^* 在连接 $\widehat{e}_t$ 和 e_t 的线段上. 对于充分大的 m, 由 $\rho(\cdot)$ 的凸性可知 $\rho''_m(\cdot)$ 是非负的. 再根据引理 A.1, 当 $m,n\to\infty$ 时, $\mathbb{E}\rho''_m(e_t^*)\to\mathbb{E}\rho''(e_t)$. 于是

$$I_{12}=O_P(n^{-1/2+\epsilon})\frac{c}{n}\sum_{t=1}^{n}\rho''_m(e_t^*)=O_P(n^{-1/2+\epsilon}),$$

从而得到 $I_{12}=o_P(1)$. 因此, $I_1=o_P(1)$, 所以有 $\widehat{a}_1\to_P a_1$.

为了证明第二个结论, 注意

$$\begin{aligned}\widehat{a}_2-a_2&=\frac{1}{n}\sum_{t=1}^{n}\{\rho''_m(\widehat{e}_t)-\mathbb{E}[\rho''(e_t)]\}\\&=\frac{1}{n}\sum_{t=1}^{n}[\rho''_m(\widehat{e}_t)-\rho''_m(e_t)]+\frac{1}{n}\sum_{t=1}^{n}\{\rho''_m(e_t)-\mathbb{E}[\rho''_m(e_t)]\}+\mathbb{E}\{\rho''_m(e_t)-\rho''(e_t)\}\\&\equiv I_3+I_4+O(m^{-1/2}).\end{aligned}$$

由 ρ''_m 的可微性和 $\widehat{e}_t-e_t=\Delta_t(\widehat{\alpha},\widehat{\beta})$ 有 $I_3=o_P(1)$, 由 $\{e_t\}$ 的遍历性知 $I_4=o_P(1)$. 因此 $\widehat{a}_2\to_P a_2$. $\square$

定理 3.2 的证明 定义 $D_n=\mathrm{diag}(\sqrt[4]{n},\sqrt[4]{n}^3I_{d_1-1})$, $p_n=\sqrt{n}$, $q_n=\sqrt[4]{n}$,

$$\begin{aligned}&\alpha_1=D_nP^\top(\theta_1-\theta_1^0),\quad \beta_{1j}=q_n(\gamma_{1j}-\gamma_{1j}^0),\quad j=1,\ldots,p_1,\\&\alpha_{2j}=p_n(\theta_{2j}-\theta_{2j}^0),\quad \beta_{2j}=p_n(\gamma_{2j}-\gamma_{2j}^0),\quad j=1,\ldots,p_2.\end{aligned}\tag{C.8}$$

于是, 有

$$\begin{aligned}&x_t^\top\theta_1=x_t^\top\theta_1^0+x_t^\top(\theta_1-\theta_1^0)=x_t^\top\theta_1^0+x_{n,t}^\top\alpha_1,\quad \text{其中 } x_{n,t}\equiv D_n^{-1}P^\top x_t,\\&z_t^\top\theta_{2j}=z_t^\top\theta_{2j}^0+z_t^\top(\theta_{2j}-\theta_{2j}^0)=z_t^\top\theta_{2j}^0+p_n^{-1}z_t^\top\alpha_{2j}.\end{aligned}$$

从而,

$$\begin{aligned}L_n(\theta,\gamma)&=\sum_{t=1}^{n}\left[\rho\left(y_t-\sum_{j=1}^{p_1}\gamma_{1j}g_{1j}(x_t^\top\theta_1)-\sum_{j=1}^{p_2}\gamma_{2j}g_{2j}(z_t^\top\theta_{2j})\right)-\rho(e_t)\right]\\&=\sum_{t=1}^{n}\left[\rho\left(e_t+\sum_{j=1}^{p_1}[\gamma_{1j}^0g_{1j}(x_t^\top\theta_1^0)-\gamma_{1j}g_{1j}(x_t^\top\theta_1)]+\sum_{j=1}^{p_2}[\gamma_{2j}^0g_{2j}(z_t^\top\theta_{2j}^0)-\gamma_{2j}g_{2j}(z_t^\top\theta_{2j})]\right)-\rho(e_t)\right]\\&=\sum_{t=1}^{n}\left[\rho\left(e_t+\sum_{j=1}^{p_1}[\gamma_{1j}^0g_{1j}(x_t^\top\theta_1^0)-\gamma_{1j}g_{1j}(x_t^\top\theta_1^0+x_{n,t}^\top\alpha_1)]\right.\right.\\&\qquad\left.\left.+\sum_{j=1}^{p_2}[\gamma_{2j}^0g_{2j}(z_t^\top\theta_{2j}^0)-\gamma_{2j}g_{2j}(z_t^\top\theta_{2j}^0+z_t^\top p_n^{-1}\alpha_{2j})]\right)-\rho(e_t)\right]\\&=\sum_{t=1}^{n}\left[\rho\left(e_t+\sum_{j=1}^{p_1}[\gamma_{1j}^0(g_{1j}(x_t^\top\theta_1^0)-g_{1j}(x_t^\top\theta_1^0+x_{n,t}^\top\alpha_1))+(\gamma_{1j}^0-\gamma_{1j})g_{1j}(x_t^\top\theta_1^0+x_{n,t}^\top\alpha_1)]\right.\right.\\&\qquad\left.\left.+\sum_{j=1}^{p_2}[\gamma_{2j}^0(g_{2j}(z_t^\top\theta_{2j}^0)-g_{2j}(z_t^\top\theta_{2j}^0+z_t^\top p_n^{-1}\alpha_{2j}))+(\gamma_{2j}^0-\gamma_{2j})g_{2j}(z_t^\top\theta_{2j}^0+z_t^\top p_n^{-1}\alpha_{2j})]\right)-\rho(e_t)\right]\\&=\sum_{t=1}^{n}\left[\rho\left(e_t+\sum_{j=1}^{p_1}[\gamma_{1j}^0(g_{1j}(x_t^\top\theta_1^0)-g_{1j}(x_t^\top\theta_1^0+x_{n,t}^\top\alpha_1))+q_n^{-1}\beta_{1j}g_{1j}(x_t^\top\theta_1^0+x_{n,t}^\top\alpha_1)]\right.\right.\end{aligned}$$

$$
\begin{aligned}
&\quad + \sum_{j=1}^{p_2} [\gamma_{2j}^0 (g_{2j}(z_t^\top \theta_{2j}^0) - g_{2j}(z_t^\top \theta_{2j}^0 + z_t^\top p_n^{-1} \alpha_{2j})) + p_n^{-1} \beta_{2j} g_{2j}(z_t^\top \theta_{2j}^0 + z_t^\top p_n^{-1} \alpha_{2j})]\Big) - \rho(e_t)\Big] \\
&\equiv \tilde{L}_n(\alpha, \beta).
\end{aligned}
$$

$\tilde{L}_n(\alpha, \beta)$ 重新参数化的原因和优点与定理 3.1 中所述相同.

考虑 $\tilde{L}_n(\alpha, \beta)$ 对 (α, β) 的优化. 显然, 如果 $(\widehat{\alpha}, \widehat{\beta})$ 使得 $\tilde{L}_n(\alpha, \beta)$ 达到最小, 则 $(\widehat{\theta}, \widehat{\gamma})$ 使得 $L_n(\theta, \gamma)$ 达到最小, 其中 $\widehat{\beta}_{1j} = q_n(\widehat{\gamma}_{1j} - \gamma_{1j}^0)$、$\widehat{\beta}_{2j} = p_n(\widehat{\gamma}_{2j} - \gamma_{2j}^0)$, $\widehat{\alpha}_1 = D_n P^\top (\widehat{\theta}_1 - \theta_1^0)$, $\widehat{\alpha}_{2j} = p_n(\widehat{\theta}_{2j} - \theta_{2j}^0)$.

观察到

$$
\begin{aligned}
\Delta_{1t}(\alpha_1, \beta_{1j}) &\equiv \gamma_{1j}^0 (g_{1j}(x_t^\top \theta_1^0) - g_{1j}(x_t^\top \theta_1^0 + x_{n,t}^\top \alpha_1)) + q_n^{-1} \beta_{1j} g_{1j}(x_t^\top \theta_1^0 + x_{n,t}^\top \alpha_1) \\
&= \gamma_{1j}^0 (g_{1j}(x_t^\top \theta_1^0) - g_{1j}(x_t^\top \theta_1^0 + x_{n,t}^\top \alpha_1)) + q_n^{-1} \beta_{1j} g_{1j}(x_t^\top \theta_1^0) \\
&\quad + q_n^{-1} \beta_{1j} [g_{1j}(x_t^\top \theta_1^0 + x_{n,t}^\top \alpha_1) - g_{1j}(x_t^\top \theta_1^0)], \\
\Delta_{2t}(\alpha_{2j}, \beta_{2j}) &\equiv \gamma_{2j}^0 (g_{2j}(z_t^\top \theta_{2j}^0) - g_{2j}(z_t^\top \theta_{2j}^0 + z_t^\top p_n^{-1} \alpha_{2j})) + p_n^{-1} \beta_{2j} g_{2j}(z_t^\top \theta_{2j}^0 + z_t^\top p_n^{-1} \alpha_{2j}) \\
&= \gamma_{2j}^0 (g_{2j}(z_t^\top \theta_{2j}^0) - g_{2j}(z_t^\top \theta_{2j}^0 + z_t^\top p_n^{-1} \alpha_{2j})) + p_n^{-1} \beta_{2j} g_{2j}(z_t^\top \theta_{2j}^0) \\
&\quad + p_n^{-1} \beta_{2j} [g_{2j}(z_t^\top \theta_{2j}^0 + z_t^\top p_n^{-1} \alpha_{2j}) - g_{2j}(z_t^\top \theta_{2j}^0)].
\end{aligned}
$$

进一步地, 使用 Taylor 展开, 可得

$$
\begin{aligned}
\Delta_{1t}(\alpha_1, \beta_{1j}) &= -\gamma_{1j}^0 \dot{g}_{1j}(x_t^\top \theta_1^0) x_{n,t}^\top \alpha_1 + \frac{1}{2} \ddot{g}_{1j}(x_t^\top \theta_1^*)(x_{n,t}^\top \alpha_1)^2 + q_n^{-1} \beta_{1j} g_{1j}(x_t^\top \theta_1^0) \\
&\quad + q_n^{-1} \beta_{1j} [\dot{g}_{1j}(x_t^\top \theta_1^0) x_{n,t}^\top \alpha_1 + \frac{1}{2} \ddot{g}_{1j}(x_t^\top \theta_1^*)(x_{n,t}^\top \alpha_1)^2],
\end{aligned}
$$

其中 θ_1^* 在连接 θ_1^0 和 α_1 的线段上.

注意 $x_{n,t} = D_n^{-1} P^\top x_t = [n^{-1/4} x_t^\top \theta_1^0, n^{-3/4} x_t^\top P_2]^\top$, 因此有

$$
q_n^{-1} \dot{g}_{1j}(x_t^\top \theta_1^0) x_{n,t}^\top = [n^{-1/2} \dot{g}_{1j}(x_t^\top \theta_1^0) x_t^\top \theta_1^0, n^{-1} \dot{g}_{1j}(x_t^\top \theta_1^0) x_t^\top P_2]
$$

和

$$
\begin{aligned}
\ddot{g}_{1j}(x_t^\top \theta_1^*)(x_{n,t}^\top \alpha_1)^2 &= \ddot{g}_{1j}(x_t^\top \theta_1^*)(x_t^\top (\theta_1 - \theta_1^0))^2 \\
&= \ddot{g}_{1j}(x_t^\top \theta_1^*)(x_t^\top P^* D_n^{-1} D_n P^{*\top} (\theta_{1j} - \theta_{1j}^0))^2 = \ddot{g}_{1j}(x_t^\top \theta_1^*)(x_t^\top P^* D_n^{-1} \alpha_1^*)^2 \\
&= n^{-1/2} \ddot{g}_{1j}(x_t^\top \theta_1^*)(x_t^\top \theta_1^*)^2 + n^{-3/2} \ddot{g}_{1j}(x_t^\top \theta_1^*)(x_t^\top P^* \alpha_{1,2}^*)^2 \\
&\quad + 2n^{-1} \ddot{g}_{1j}(x_t^\top \theta_1^*)(x_t^\top \theta_1^*)(x_t^\top P_2^* \alpha_{1,2}^*),
\end{aligned}
$$

其中 $P^* = (\theta_1^*/\|\theta_1^*\|, P_2^*)$ 是一个正交矩阵, $\alpha_1^* = D_n P^{*\top} (\theta_{1j} - \theta_{1j}^0)$, $\alpha_{1,2}^* = \sqrt[4]{n}^3 P_2^{*\top} (\theta_{1j} - \theta_{1j}^0)$. 因此, 根据文献 [34, 第 1256 页, 引理 2 和 3], 以及 $u\dot{g}_{1j}(u)$, $u\ddot{g}_{1j}(u)$ 和 $u^2 \ddot{g}_{1j}(u)$ 的 I- 正则性质, 可知 $\Delta_{1t}(\alpha_{1j}, \beta_{1j})$ 有两个主项, 即

$$
\Delta_{1t}(\alpha_1, \beta_{1j}) = [-\gamma_{1j}^0 \dot{g}_{1j}(x_t^\top \theta_1^0) x_{n,t}^\top \alpha_1 + q_n^{-1} \beta_{1j} g_{1j}(x_t^\top \theta_{1j}^0)](1 + o_P(1)). \tag{C.9}
$$

同样地,

$$
\Delta_{2t}(\alpha_{2j}, \beta_{2j}) = -\gamma_{2j}^0 \dot{g}_{2j}(z_t^\top \theta_{2j}^0) z_t^\top p_n^{-1} \alpha_{2j} + \frac{1}{2} \gamma_{2j}^0 \ddot{g}_{2j}(z_t^\top \theta_{2j}^*)(z_t^\top p_n^{-1} \alpha_{2j})^2 + p_n^{-1} \beta_{2j} g_{2j}(z_t^\top \theta_{2j}^0)
$$

$$
\begin{aligned}
&+ p_n^{-1}\beta_{2j}\left[\dot{g}_{2j}(z_t^\top\theta_{2j}^0)z_t^\top p_n^{-1}\alpha_{2j} + \frac{1}{2}\ddot{g}_{2j}(z_t^\top\theta_{2j}^*)(z_t^\top p_n^{-1}\alpha_{2j})^2\right] \\
&= p_n^{-1}[-\gamma_{2j}^0\dot{g}_{2j}(z_t^\top\theta_{2j}^0)z_t^\top\alpha_{2j} + \beta_{2j}g_{2j}(z_t^\top\theta_{2j}^0)] \\
&\quad + \frac{1}{2}p_n^{-2}\gamma_{2j}^0\ddot{g}_{2j}(z_t^\top\theta_{2j}^*)(z_t^\top\alpha_{2j})^2 + p_n^{-2}\beta_{2j}\left[\dot{g}_{2j}(z_t^\top\theta_{2j}^0)z_t^\top\alpha_{2j} + \frac{1}{2}p_n^{-1}\ddot{g}_{2j}(z_t^\top\theta_{2j}^*)(z_t^\top\alpha_{2j})^2\right] \\
&= p_n^{-1}[-\gamma_{2j}^0\dot{g}_{2j}(z_t^\top\theta_{2j}^0)z_t^\top\alpha_{2j} + \beta_{2j}g_{2j}(z_t^\top\theta_{2j}^0)](1+o_P(1)).
\end{aligned}
$$

为方便起见, 定义

$$
\Delta_t(\alpha,\beta) = \sum_{j=1}^{p_1}\Delta_{1t}(\alpha_1,\beta_{1j}) + \sum_{j=1}^{p_2}\Delta_{2t}(\alpha_{2j},\beta_{2j}), \tag{C.10}
$$

并且由于所涉及的函数的 I- 正则性和一阶单整 x_t, 利用引理 B.1, 可得 $\Delta_{1t}(\alpha,\beta) \sim O_P(n^{-3/4})$ 对于充分大的 t 成立, 而 $\Delta_{2t}(\alpha,\beta) \sim O_P(n^{-1/2})$. 于是, $\Delta_t(\alpha,\beta) \sim O_P(n^{-1/2})$.

所以,

$$
\tilde{L}_n(\alpha,\beta) = \sum_{t=1}^{n}[\rho(e_t+\Delta_t(\alpha,\beta)) - \rho(e_t)],
$$

并定义

$$
\tilde{L}_{mn}(\alpha,\beta) \equiv \sum_{t=1}^{n}[\rho_m(e_t+\Delta_t(\alpha,\beta)) - \rho_m(e_t)],
$$

其中 $\rho_m(\cdot)$ 是引理 A.1 中定义的正则序列. 因此

$$
\sup_{\alpha,\beta}|\tilde{L}_n(\alpha,\beta) - \tilde{L}_{mn}(\alpha,\beta)| \leqslant Cm^{-1/2}n, \tag{C.11}
$$

其中 C 是与 m 和 n 无关的绝对常数.

选择足够大的 m, 如 $m=[n^{2+\varepsilon}]$ 且 $\varepsilon>0$, 从而 $\tilde{L}_n(\alpha,\beta)$ 和 $\tilde{L}_{mn}(\alpha,\beta)$ 之间的差异可以忽略不计. 现在关注 $\tilde{L}_{mn}(\alpha,\beta)$, 它对所有参数都是无限次可微的. 利用 Taylor 展开式, 有

$$
\begin{aligned}
\tilde{L}_{mn}(\alpha,\beta) &= \sum_{t=1}^{n}[\rho_m(e_t+\Delta_t(\alpha,\beta)) - \rho_m(e_t)] \\
&= \sum_{t=1}^{n}\left[\rho_m'(e_t)\Delta_t(\alpha,\beta) + \frac{1}{2}\rho_m''(e_t)\Delta_t(\alpha,\beta)^2 + O_P(n^{-3/2})\right].
\end{aligned} \tag{C.12}
$$

为了将 $\tilde{L}_{mn}(\alpha,\beta)$ 写成向量形式, 注意到

$$
\Lambda = (\Lambda_1^\top,\Lambda_2^\top)^\top, \quad \Lambda_1 = (\alpha_1^\top,\beta_{11},\ldots,\beta_{1p_1})^\top, \quad \Lambda_2 = (\alpha_{21}^\top,\ldots,\alpha_{2p_2}^\top,\beta_{21},\ldots,\beta_{2p_2})^\top,
$$

以及 $p_n=\sqrt{n}$, $q_n=\sqrt[4]{n}$, $\dot{g}_1(u)=\sum_{j=1}^{p_1}\gamma_{1j}^0\dot{g}_{1j}(u)$ 和 (3.11) 中给出的 $Z_t=(Z_1(x_{nt})^\top,Z_2(z_t)^\top)^\top$. 于是有

$$
\Delta_t(\alpha,\beta) = Z_1(x_{nt})^\top\Lambda_1 + Z_2(z_t)^\top\Lambda_2 = Z_t^\top\Lambda, \tag{C.13}
$$

从而 (C.12) 中的 $\tilde{L}_{mn}(\alpha,\beta)$ 可以写成

$$\begin{aligned}\tilde{L}_{mn}(\alpha,\beta) &= \sum_{t=1}^{n}\left[\rho_m'(e_t)Z_t^\top\Lambda + \frac{1}{2}\rho_m''(e_t)(Z_t^\top\Lambda)^2\right]\\ &= \left(\sum_{t=1}^{n}\rho_m'(e_t)Z_t\right)^\top\Lambda + \frac{1}{2}\Lambda^\top\left(\sum_{t=1}^{n}\rho_m''(e_t)Z_tZ_t^\top\right)\Lambda\\ &= \left(\sum_{t=1}^{n}\psi(e_t)Z_t\right)^\top\Lambda + \frac{1}{2}a_2\Lambda^\top\left(\sum_{t=1}^{n}Z_tZ_t^\top\right)\Lambda\\ &\quad + \left(\sum_{t=1}^{n}[\rho_m'(e_t)-\psi(e_t)]Z_t\right)^\top\Lambda + \frac{1}{2}\Lambda^\top\left(\sum_{t=1}^{n}[\mathbb{E}(\rho_m''(e_t)\mid\mathcal{F}_{t-1})-a_2]Z_tZ_t^\top\right)\Lambda\\ &\quad + \frac{1}{2}\Lambda^\top\left(\sum_{t=1}^{n}[\rho_m''(e_t)-\mathbb{E}(\rho_m''(e_t)\mid\mathcal{F}_{t-1})]Z_tZ_t^\top\right)\Lambda. \end{aligned} \tag{C.14}$$

令

$$Q_n(\alpha,\beta) = \left(\sum_{t=1}^{n}\psi(e_t)Z_t\right)^\top\Lambda + \frac{1}{2}a_2\Lambda^\top\left(\sum_{t=1}^{n}Z_tZ_t^\top\right)\Lambda. \tag{C.15}$$

于是对于任意的 (α,β), 根据引理 B.4, 取 $m=[n^{2+\varepsilon}]$, 有

$$\begin{aligned}|\tilde{L}_{mn}(\alpha,\beta)-Q_n(\alpha,\beta)| &\leqslant \left\|\sum_{t=1}^{n}[\rho_m'(e_t)-\psi(e_t)]Z_t\right\|\|\Lambda\| + \frac{1}{2}\|\Lambda\|^2\left\|\sum_{t=1}^{n}[\mathbb{E}(\rho_m''(e_t)|\mathcal{F}_{t-1})-a_2]Z_tZ_t^\top\right\|\\ &\quad + \frac{1}{2}\|\Lambda\|^2\left\|\sum_{t=1}^{n}[\rho_m''(e_t)-\mathbb{E}\rho_m''(e_t)]Z_tZ_t^\top\right\|\\ &= O_P(n^{1/2}m^{-1/2}) + O_P(m^{-1/2}) + O_P(n^{-1/4})\\ &= O_P(n^{-1/4}),\end{aligned}$$

这在 (α,β) 的任意紧集上也是一致地成立. 注意, $Q_n(\alpha,\beta)$ 有唯一的最小值

$$\widehat{\Lambda}_Q = -\left(a_2\sum_{t=1}^{n}Z_tZ_t^\top\right)^{-1}\sum_{t=1}^{n}\psi(e_t)Z_t,$$

并且, 类似于文献 [13], 可以证明 $\widehat{\Lambda}=\widehat{\Lambda}_Q+o_P(1)$.

利用引理 B.5 中的联合收敛性, 由 $\widehat{\Lambda}_Q$ 的联合收敛性得到 $\widehat{\Lambda}=O_P(1)$, 并且当 $n\to\infty$ 时,

$$\widehat{\Lambda} \to_D \frac{\sqrt{a_1}}{a_2}\mathcal{R}^{-1/2}N(0,I),$$

其中 I 是 $d_1+p_1+p_2d_2+p_2$ 维单位矩阵. □

推论 3.2 的证明　考虑 $\widehat{\theta}_1-\theta_1^0$ 的收敛性. 注意到

$$\begin{aligned}\sqrt[4]{n}(\widehat{\theta}_1-\theta_1^0) &= \sqrt[4]{n}(D_nP^\top)^{-1}(D_nP^\top)(\widehat{\theta}-\theta_0)\\ &\to_D \frac{\sqrt{a_1}}{a_2}P\mathrm{diag}(1,\mathbf{0}_{d_1-1})N(0,r_{11})\\ &=_D \frac{\sqrt{a_1}}{a_2}N(0,r_{11}^{11}\theta_1^0(\theta_1^0)^\top).\end{aligned}$$

结论成立. □

Robust M estimation for additive single-index cointegrating time series models

Chaohua Dong, Chen Zhou & Yundong Tu

Abstract In this paper, we study robust M estimation for an additive single-index cointegrating time series parametric model, where the usual smooth squared loss function of the error is replaced by a class of loss functions that may have kinks; examples include but not limited to least absolute deviation (LAD) estimator, quantile estimator, and Huber's estimator as special cases. Because of the nonstationarity of the time series, the existing approach could not deliver the limit distribution for the extremum estimators. Nonsmoothness of the loss functions has been a genesis of technical difficulties in developing a unified method for deriving the asymptotic properties of the estimators. The family of generalized functions offers an elegant way of approximating nonsmooth functions by infinitely smooth functions, and thereby overcoming many of the technical difficulties. We derive the asymptotic properties of the M estimators corresponding to nonsmooth loss functions for such an important additive single-index model, and evaluate the finite-sample performance of the proposed estimation method and theory by both simulated data and an empirical analysis of predictive regression of stock returns.

Keywords **additive single index model, generalized function, nonstationary time series, quadratic approximation, regular function sequence, robust M estimation**

MSC(2020) **62E20, 62F35, 62J02**

doi: 10.1360/SSM-2024-0187