HAR-Itˆo 模型与高频数据的高维 HAR 建模

Huiling Yuan、Kexin Lu、Yifeng Guo 和
Guodong Li香港大学

2023年3月7日

arXiv:2303.02896v1 [stat.ME] 6 Mar 2023

摘要

对金融和经济学中的高频数据进行已实现波动率建模是一项重要任务，而作为公认最流行的模型，异质自回归（HAR）模型一直主导着该领域的应用。然而，该模型存在三个缺陷：(i.) 其异质波动率分量是具有固定权重的日已实现波动率线性组合，这限制了其对不同类型资产的适应灵活性；(ii.) 对于该模型以及文献中的许多其他 HAR 类模型，其高频概率结构仍然未知；(iii.) 尽管在实际应用中经常会遇到众多资产，但目前尚无用于 HAR 建模的高维推断工具。为克服这些缺陷，本文利用张量技术提出了一种多线性低秩 HAR 模型，并采用数据驱动方法自动选择异质分量。此外，本文引入 HAR-Itˆo 模型，以解释相应的高频动态以及其他 HAR 型模型的高频动态。进一步地，本文建立了高维 HAR 建模的非渐近性质，并提出一种具有理论依据的投影梯度下降算法，用于搜索估计值。模拟研究验证了所提方法的理论和计算性质，真实数据分析则说明了在异质分量中采用数据驱动方法的必要性。

关键词和短语：扩散过程；异质自回归模型；高维分析；高频数据；非渐近性质；张量技术。

1 引言

波动率分析是金融和经济学中最重要的任务之一（Engle，1982年；Bollerslev，1986年），随着高频数据的广泛可用，近年来越来越多的讨论聚焦于已实现波动率建模，这类波动率可由高频日内观测值构造。相关例子包括已实现广义自回归条件异方差（GARCH）模型（Hansen 等，2012年）、基于高频数据的波动率模型（Shephard 和 Sheppard，2010年）、异质自回归（HAR）模型（Corsi，2009年）、乘法误差模型（Engle 和 Gallo，2006年）以及混频数据抽样模型（Ghysels 等，2006年）。HAR模型可谓这些方法中最受欢迎的一种，具有简单的级联结构（Corsi，2009年）。具体而言，首先识别短期、中期和长期波动率分量，它们通常是日已实现波动率的线性组合。随后，我们以这三个异质分量为解释变量，对未来已实现波动率进行回归，从而得到自回归形式。令人惊讶的是，在机器学习领域的一些实际应用中，这种简单的HAR模型甚至能够胜过若干强大的深度神经网络（Bucci，2020）。

自问世以来，HAR模型便受到文献的广泛关注。首先，鉴于跳跃效应对高频数据的重要性（A¨ıt-Sahalia 和 Jacod，2009年；Andersen 等人，2012年），Andersen 等人（2007年）将估计跳跃作为第四个异质性成分纳入模型，以提升预测表现；另见 Corsi 等人（2010年）关于阈值双幂变差的研究，以及 Patton 和 Sheppard（2015年）关于带符号跳跃的研究。Corsi 和 Ren`o（2012年）进一步加入了另一个杠杆效应成分。其次，已有研究致力于推广 HAR 模型。Bollerslev 等人（2016）提出了一种系数随时间变化的 HAR 模型，并采用参数形式；另见 Bekierman 和 Manner（2018）。此外，Chen 等人（2018）研究了一种类似模型，其系数采用非参数形式；McAleer 和 Medeiros（2008）提出了多区制平滑转换 HAR模型；Corsi 等人（2008）则研究了一种更复杂的、带有 GARCH误差的 HAR 模型。跳跃效应也在这些广义HAR模型中得到讨论（Buc- cheri 和 Corsi，2021年；Caporin，2022）。第三，除了普通最小二乘法，文献还探讨了更多高效的估计方法，

其中包括加权最小二乘（Patton 和 Sheppard，2015年）以及稳健回归（Clements 和 Preve，2021年）方法。第四，为了提高预测准确度，通常会在估计之前对已实现波动率进行变换，常用的变换包括对数变换（Corsi，2009年；Chen 等人，2018年；Bekaert 和 Hoerova，2014年）或 Box-Cox变换（Taylor，2017年）。最后，除了股票市场，HAR模型也已应用于比特币（Trucıos，2019年；Hu 等人，2021年）、能源（Lyocsa 和 Todor- ova，2021年）、农产品（Degiannakis 等人，2022年；Luo 等人，2022年），以及许多其他领域。

上述关于 HAR 模型的大多数研究都局限于单变量情形，而在实践中，预测多种资产的已实现波动率十分常见。一种常用方法是将 HAR 模型扩展为多变量版本进行建模（Hong 等人，2020）。例如，Bubak 等人（2011）研究了带有多变量 GARCH 误差项的向量 HAR 模型，以提高预测准确度；Souˇcek 和 Todorova（2013）也采用了类似的模型设定。Bauer 和 Vorkink（2011）对已实现协方差进行了预测，其中向量 HAR 模型在建模中发挥了关键作用；另见 Oh 和 Patton（2016）；Bollerslev 等人（2018）。本文聚焦于用于预测已实现波动率的向量 HAR 模型；当资产数量N 较大时，所得模型的参数数量将大幅增加，其增长率为 $O(N^2)$。因此，需要采用一些降维方法，使预测切实可行。Reinsel（1983）提出了一种向量自回归模型，对系数矩阵的行空间施加低秩假设；Cubadda 等人（2017）将该方法应用于向量 HAR 模型。该方法将参数数量减少至 $O(N)$。不过，人们可能会想知道，是否可以对系数矩阵的列空间，甚至行空间和列空间同时施加低秩假设；详情见备注1。

另一方面，对于上述大多数 HAR 类模型，三个异质波动率分量被设定为日、周和月已实现波动率，其中后两者分别是 5 日和 22 日已实现波动率的简单平均值（Corsi，2009年）。然而，越来越多的实证证据表明，这种设定存在局限性，且

这种局限性越来越明显。Chen 等人（2010年）最先质疑了长期波动率分量（以月已实现波动率衡量）的作用，并指出，结构突变结合低阶自回归模型能够优于 HAR 模型。Audrino 和 Knaus（2016）使用 Lasso 方法（Tibshirani，1996），通过自动选择大阶数自回归模型中的变量，对三个分量设置的合理性进行了实证评估，并在来自美国市场的九只股票中证实了这一负面结论。请注意，这三个分量分别对应于 j 日已实现波动率的平均值，其中 $j = 1$分别为 5 和 22。K¨ohler （2021年）考察了 j 的许多其他组合，并通过实证表明，（1、5、22）的组合表现较差；另见 Hong 等人（2020）。事实上，HAR 模型中的三个分量可解释为滞后阶上的因子，这促使我们考虑采用数据驱动方法来选择这些因子；详情见模型（2.5）和注释3 。本文的首项贡献是在第3.1节提出一种多线性低秩 HAR 模型，利用张量技术预测已实现波动率。该模型对系数矩阵的行空间和列空间均施加低秩假设，并自动选择异质波动率分量。第3.3节讨论了其普通最小二乘估计的渐近正态性。

近年来，人们致力于构建日内资产价格的高频模型，使其对应的低频积分波动率具有 GARCH 型表示。此类模型包括 GARCH-Itˆo 模型（Kimand Wang，2016）、因子 GARCH-Itˆo 模型（Kim 和 Fan，2019）、已实现 GARCH-Itˆo 模型（Song 等人，2021年）、隔夜 GARCH-Itˆo 波动率模型（Kim 等人，2023）、指数型已实现 GARCH-Itˆo 模型（Kim，2023）等。这些 GARCH-Itˆo 型模型搭建了理论桥梁，以协调低频 GARCH 波动率表示与高频波动率过程，从而通过利用 GARCH 模型和已实现波动率，获得更好的解释和更可靠的推断。作为 GARCH-Itˆo 文献自然但并非易事的后续研究，本文的第二项贡献是在第2节提出单变量和多变量 HAR-Itˆo 模型，其低频积分波动率分别具有单变量和多变量 HAR 表示。据我们所知

，这是文献中首次尝试探究 HAR 类模型的高频动态。

与此同时，目前所有使用的 HAR 建模工具都适用于资产数量固定的低维设定。然而，在实际应用中经常会遇到大量金融资产，因此亟须为 HAR 模型设计高维推断工具，此时资产数量 N 可能趋于无穷。本文的第三项贡献是填补这一空白：在第4节对高频数据进行高维 HAR 建模。具体而言，所提出的模型形式类似于含测量误差的向量自回归模型，第4.1节则确立了其高维估计的非渐近性质。第4.2节提出一种投影梯度下降算法来求解估计值，并给出了相应的理论依据。

此外，第5节通过模拟实验评估所提方法的有限样本性能，并在第6节通过实证例子进一步展示其用途。第7节给出简短的结论与讨论，所有技术性证明均置于补充文件中。

2 高频数据的 HAR-Itˆo 模型

2.1 单变量 HAR-Itˆo 模型

本节针对高频资产价格提出单变量和多变量 HAR-Itˆo 模型，使其低频积分波动率分别具有命题 1和2中的单变量和多变量 HAR 表示。

设 X_t 为资产在 $t \in \mathbb{R}_+$ 时刻的对数价格，其中 $\mathbb{R}_+ = [0, \infty)$。记 $\mathcal{F}_t$ 为截至 t 时刻的全部信息集合，则 $\{\mathcal{F}_t, t \geqslant 0\}$ 是一个滤子。为计入金融行业中的跳跃成分，我们考虑如下跳扩散模型：

$$dX_t = \mu_t dt + \sigma_t dB_t + L_t d\Lambda_t, \tag{2.1}$$

其中， μ_t 为漂移项， σ_t 为瞬时波动率过程， B_t 为标准布朗运动， Λ_t 为强度为常数 $\lambda > 0$的标准泊松过程，

跳跃幅度 $\{L_t,\ t \geqslant 0\}$ 独立同分布（$i.i.d.$），且具有有限四阶矩。σ_t、B_t 和 Λ_t 这几个过程均适应于滤子 $\{\mathcal{F}_t\}$，且跳跃幅度与这三个过程相互独立。此外，令 $Z_t = \int_{[t]}^{t} dW_t$，其中$W_t$ 是相对于 $\mathcal{F}_t$ 的另一个标准布朗运动，满足 $dB_t dW_t = \rho dt$，且 $|\rho| < 1$ 几乎必然，而 ρ 可视为对数价格与瞬时波动率之间的杠杆效应。随后，我们定义如下 HAR-Itˆo 模型。

定义 1. 对数价格 $\{X_t,\ t \in \mathbb{R}_+\}$ 在 (2.1) 处称为服从单变量 HAR-Itˆo 模型，若瞬时波动率满足

$$\begin{aligned}\sigma_t^2 = \sigma_{[t]}^2 \quad &+(t-[t])\left(\omega - \sigma_{[t]}^2\right) + \alpha_1 \int_{[t]}^{t} \sigma_s^2 ds + \sum_{l=2}^{P} \alpha_l \int_{[t]-l+1}^{[t]-l+2} \sigma_s^2 ds \\ &+\beta \int_{[t]}^{t} L_s^2 d\Lambda_s + \upsilon([t]+1-t) Z_t^2,\end{aligned} \tag{2.2}$$

其中，$[t]$ 表示 t的整数部分，即对于 $n-1 < t \leqslant n$，$[t] = n-1$；P为正整数，ω、β、υ、α_1 的参数均为正，且 $\alpha_l \geqslant 0$，其中 $2 \leqslant l \leqslant P$。

记 $\boldsymbol{\alpha} = (\alpha_1,\ \alpha_2,\ \cdots,\ \alpha_P)^\top$ 和 $\boldsymbol{\theta} = (\lambda,\ \rho,\ \omega,\ \beta,\ \upsilon,\ \boldsymbol{\alpha}^\top)^\top$，上述定义依赖于参数向量 $\boldsymbol{\theta}$。此外，瞬时波动率 σ_t^2 在 (2.2) 处是一个连续时间过程，并对所有 $t \in \mathbb{R}_+$都有定义。仅考虑整数时间点，则该过程具有如下形式:

$$\sigma_n^2 = \omega + \sum_{l=1}^{P} \alpha_l y_{n-l+1} + \beta J_n,$$

以概率 1 成立，其中

$$y_n = IV_n = \int_{n-1}^{n} \sigma_s^2 ds \quad \text{and} \quad J_n = \int_{n-1}^{n} L_s^2 d\Lambda_s$$

分别为积分波动率和跳跃变差。此外，根据（2.2），通过引入 α_l（其中 $2 \leqslant l \leqslant P$），过去的日波动率（或积分波动率）将对即期波动率（或瞬时波动率）产生影响。本文将重点研究积分波动率，IV_n 或 y_n。

As 在 Song 等人（2021年）中，我们记 $\omega_L = \mathbb{E}(L_t^2)$ 和 $M_t = L_t^2 - \omega_L$。此外，令 t

$\omega^g = \varrho_1\omega + \varrho_2\beta\omega_L\lambda + (\varrho_2 - 2\varrho_3)v$，$\beta^g = (\varrho_1 - \varrho_2)\beta$ 以及对所有 $1 \leqslant l \leqslant P$，其中 $\varrho_1 = \alpha_1^{-1}(e^{\alpha_1} - 1)$、$\varrho_2 = \alpha_1^{-2}(e^{\alpha_1} - 1 - \alpha_1)$ 和 $\varrho_3 = \alpha_1^{-3}(e^{\alpha_1} - 1 - \alpha_1 - \alpha_1^2/2)$。

命题1。 假设 $n \geqslant P+1$ 且 $0<\alpha_1<1$。那么，定义 *1* 中的单变量 *HAR-Itˆo* 模型的积分波动率满足

$$y_n = \omega^g + \sum_{j=1}^{P} \alpha_j^g y_{n-j} + \beta^g J_{n-1} + \varepsilon_n \tag{2.3}$$

以概率*1*，其中 $\varepsilon_n = \varepsilon_n^c + \varepsilon_n^J$,

$$\begin{aligned}
\varepsilon_n^c =& 2v\alpha_1^{-2} \int_{n-1}^{n} \left\{\alpha_1 \left(n-t-\alpha_1^{-1}\right) e^{\alpha_1(n-t)} + 1\right\} Z_t dZ_t, \\
\varepsilon_n^J =& \beta\alpha_1^{-1} \left\{ \int_{n-1}^{n} \left(e^{\alpha_1(n-t)} - 1\right) \left[M_t d\Lambda_t + \omega_L (d\Lambda_t - \lambda dt)\right] \right\},
\end{aligned}$$

且 ε_n^c 和 ε_n^J 分别是来自连续扩散过程和跳跃分量的创新项。此外， $\{\varepsilon_n^c\}$ 和 $\{\varepsilon_n^J\}$均为均值为零且方差有限的 $i.i.d.$ 序列。

根据上述命题，积分波动率 $\{y_n\}$ 可表示为关于 J_{n-1}的含外生变量自回归迭代形式；若对所有满足 $|z|<1$的复数值， $1-\sum_{j=1}^{P} \alpha_j^g z^j \neq 0$ 成立，则它将是平稳的低频时间序列。此外， $Z_t = \int_{[t]}^{t} dW_t$ 这一项在设计式（2.2）中的瞬时波动率 σ_t 时发挥关键作用。我们可以考虑 $Z_t = \int_{[t]}^{t} z_t dW_t$ 的一种更一般形式，其中 z_t 为非随机连续函数，并可建立与命题1类似的结果。然而，当 z_t 是随机函数时，例如 Kim and Wang（2016）中的 $z_t = \sigma_t$，推导出的创新项 $\{\varepsilon_n^c\}$ 只能构成鞅差序列，尽管我们仍可得到类似于（2.3）的自回归形式。事实上，通过为 z_t选择合适的随机函数，推导出的创新项 ε_n^c 甚至可能具有重尾（Shin等人，2021年）。本文将聚焦于定义1中的设定，因为我们的目标是HAR模型。

当不存在跳跃分量时，即 $L_t = 0$ 对于所有 t几乎必然成立时，式（2.3）中的迭代形式可进一步简化为

$$y_n = \omega^g + \sum_{j=1}^{P} \alpha_j^g y_{n-j} + \varepsilon_n, \tag{2.4}$$

并且积分波动率 y_n 可以通过已实现波动率一致地估计，记为 RV_n。令 $IV_n = y_n$、$IV_n^{(w)} = 5^{-1}\sum_{l=1}^{5} y_{n+1-l}$和 $IV_n^{(m)} = 22^{-1}\sum_{l=1}^{22} y_{n+1-l}$分别表示日、周和月积分波动率，我们考虑一种

（2.4）的特例，

$$IV_n = \beta_0 + \beta_d IV_{n-1} + \beta_w IV_{n-1}^{(w)} + \beta_m IV_{n-1}^{(m)} + \varepsilon_n, \tag{2.5}$$

其中 $P = 22$, $\omega^g = \beta_0$, $(\alpha_1^g, \ldots, \alpha_P^g) = (\beta_d, \beta_w, \beta_m)\mathbf{U}_{\mathrm{C}}^{\top}$，且

$$\mathbf{U}_{\mathrm{C}}^{\top} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 & 0 & \cdots & 0 \\ 1/22 & 1/22 & 1/22 & 1/22 & 1/22 & 1/22 & \cdots & 1/22 \end{pmatrix} \in \mathbb{R}^{3\times 22}.$$

Corsi （2009年）利用模型（2.5）提出了用于预测已实现波动率的 HAR模型，而定义1和命题1中提出的 HAR-Itˆo 模型则为这一广泛应用的模型提供了概率论支持。另一方面，（2.5）中的级联结构揭示，参数 α_j 和 α_j^g 通常沿滞后阶方向呈现低维结构。

对于存在跳跃的情形，已实现波动率 RV_n 是$IV_n + J_n$的一致估计量，而积分波动率 IV_n 可通过双幂变差进行一致估计，记为 BV_n。根据命题 1，可使用以下 HAR型模型预测双幂变差：

$$BV_n = \beta_0 + \beta_d BV_{n-1} + \beta_w BV_{n-1}^{(w)} + \beta_m BV_{n-1}^{(m)} + \beta_J \widehat{J}_{n-1} + \epsilon_n^J,$$

其中 $BV_{n-1}^{(w)}$ 和 $BV_{n-1}^{(m)}$ 分别是日双幂变差的周平均值和月平均值$\frown BV_n$，而 $J_n = \max\{RV_n - BV_n, 0\}$ 是估计跳跃；参见 Cubadda 等人 （2017年）。误差项 ϵ_n^J 指命题 ε_n 中的模型误差 以及使用$\widehat{}$和 BV_n 估计 J_n 和 IV_n 时产生的 J_n估计误差。在文献中，对于其他包含跳跃成分的 HAR 型模型，例如 Andersen等人 （2007年），其高频模型可以构造成与定义1类似的形式。

2.2 多元 HAR-Itˆo 模型

本小节定义一种适用于 N 金融资产的多元 HAR-Itˆo 模型。分别用$X_{i,t}$ 和 $\sigma_{i,t}$ 表示第 i 个资产在 $t \in \mathbb{R}_+$ 时的对数价格及其瞬时波动率，其中每个 $1 \leqslant i \leqslant N$ 均成立，并令 $\mathbf{X}_t = (X_{1,t}, \ldots, X_{N,t})^{\top}$ 和 $\boldsymbol{\sigma}_t = (\sigma_{1,t}, \ldots, \sigma_{N,t})^{\top}$。

定义两个 N 维布朗运动，$\mathbf{B}_t = (B_{1,t}, \ldots, B_{N,t})^{\top}$ 和 $\mathbf{W}_t = (W_{1,t}, \ldots, W_{N,t})^{\top}$，其中 $dB_{i,t} \cdot dB_{j,t} = \rho_{i,j}^B dt$，$dW_{i,t} \cdot dW_{j,t} = \rho_{i,j}^W dt$，$dB_{i,t} \cdot dW_{j,t} = \rho_{i,j} dt$，

ρ_i^B，$_i = 1$和ρ_i^W，$_i = 1$对所有 $1 \leqslant i$，$j \leqslant N$ 均成立。令 $\boldsymbol{\rho}^B = (\rho_i^B$，$_j) \in \mathbb{R}^{N\times N}$，$\boldsymbol{\rho}^W = (\rho_i^W$，$_j) \in \mathbb{R}^{N\times N}$ 和 $\boldsymbol{\rho} = (\rho_i$，$_j) \in \mathbb{R}^{N\times N}$，其中 $\boldsymbol{\rho}^B$ 和 $\boldsymbol{\rho}^W$ 为对称矩阵。此外，令 Z_i，$_t = \int_{[t]}^t dW_i$，$_t$，其中 $1 \leqslant i \leqslant N$，并记 $\mathbf{Z}_t = (Z_1$，$_t$，...，Z_N，$_t)^\top$。对于跳跃分量，令 $\boldsymbol{\Lambda}_t = (\Lambda_1$，$_t$，...，$\Lambda_N$，$_t)^\top$ 为一个 N 维泊松过程，$\mathbf{L}_t = (L_1$，$_t$，...，L_N，$_t)^\top$ 为 N 维跳跃幅度。我们假设不同资产的跳跃成分彼此独立，$\{\mathbf{L}_t$，$t \in \mathbb{R}_+\}$ 为具有有限四阶矩的 $i.i.d.$，$\mathbf{L}_t$ 独立于 $\boldsymbol{\Lambda}_t$，$\mathbf{B}_t$，$\mathbf{W}_t$ 和 $\boldsymbol{\sigma}_t$，且上述所有扩散过程均适应于滤子$\{\mathcal{F}_t$，$t \in \mathbb{R}_+\}$。因此，跳扩散模型可如下引入，

$$dX_{i,t} = \mu_{i,t}dt + \sigma_{i,t}dB_{i,t} + L_{i,t}d\Lambda_{i,t}, \quad 1 \leqslant i \leqslant N. \tag{2.6}$$

定义2。资产 N 的对数价格 $\{\mathbf{X}_t$，$t \in \mathbb{R}_+\}$ 在（2.6）中称为服从多元 HAR-Itˆo 模型，如果对每个 $1 \leqslant i \leqslant N$，第$i$个资产的瞬时波动率满足

$$\begin{aligned}\sigma_{i,t}^2 = \sigma_{i,[t]}^2 &+(t-[t])\left(\omega_i - \sigma_{i,[t]}^2\right) + \sum_{j=1}^N \alpha_{i,j}^{(1)} \int_{[t]}^t \sigma_{j,s}^2 ds + \sum_{l=2}^P \sum_{j=1}^N \alpha_{i,j}^{(l)} \int_{[t]-l+1}^{[t]-l+2} \sigma_{j,s}^2 ds \\ &+\beta_i \int_{[t]}^t L_{i,s}^2 d\Lambda_{i,s} + \upsilon_i([t]+1-t)Z_{i,t}^2,\end{aligned} \tag{2.7}$$

其中，$[t]$ 表示 t的整数部分，即对于 $n-1 < t \leqslant n$，$[t] = n-1$；P是正整数，ω_i、β_i、υ_i、$\alpha_i^{(1)}$，$_j$的参数均为正数，并且对所有 $2 \leqslant l \leqslant P$，$\alpha_i^{(l)}$，$_j \geqslant 0$。

令 $\boldsymbol{\omega} = (\omega_1$，$\cdots$，$\omega_N)^\top \in \mathbb{R}^N$，$\boldsymbol{\alpha}_i^{(l)} = (\alpha_{i\,1}^{(l)}$，$\cdots$，$\alpha_i^{(l)}$，$_N)^\top \in \mathbb{R}^N$，$\boldsymbol{\alpha}^{(l)} = (\boldsymbol{\alpha}_1^{(l)}$，$\cdots$，$\boldsymbol{\alpha}_N^{(l)})^\top \in \mathbb{R}^{N\times N}$，以及 $\boldsymbol{\beta}$ =对角$\{\beta_1$，.....，$\beta_N\} \in \mathbb{R}^{N\times N}$。在（$\boldsymbol{\sigma}_t$ 2.7）中定义的瞬时波动率是一个 N维连续时间过程；将其限制在整数时间点时，其形式为

$$\boldsymbol{\sigma}_n^2 = \boldsymbol{\omega} + \sum_{l=1}^P \boldsymbol{\alpha}^{(l)}\mathbf{y}_{n-l+1} + \boldsymbol{\beta}\mathbf{J}_n,$$

以概率 1 成立，其中第 i 个资产的积分波动率和跳跃变差定义为

$$y_{i,n} = \int_{n-1}^n \sigma_{i,t}^2 dt \quad \text{and} \quad J_{i,n} = \int_{n-1}^n L_{i,s}^2 d\Lambda_{i,s},$$

分别为$\mathbf{y}_n = (y_{1,n},\cdots,y_{N,n})^\top$和$\mathbf{J}_n = (J_{1,n},\cdots,J_{N,n})^\top$。此外，根据（2.7），$\alpha_i^{(l)}$，$_j$ 的取值，其中 $2 \leqslant l \leqslant P$ 测度了第 j 只股票的日波动率对第 i 只股票即期波动率的贡献。

对于 $1 \leqslant i \leqslant N$，令 $\omega_{i,L} = E(L_{i,t}^2)$ 和 $M_{i,t} = L_{i,t}^2 - \omega_{i,L}$，并记 $\boldsymbol{\omega}_L = \text{diag}\{\omega_{1,L}, \cdots, \omega_{N,L}\} \in \mathbb{R}^{N\times N}$，$\mathbf{V} = \text{diag}\{v_1, \cdots, v_N\} \in \mathbb{R}^{N\times N}$，以及 $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_N)^\top \in \mathbb{R}^N$。此外，令 $\boldsymbol{\varrho}_1 = \boldsymbol{\alpha}^{(1)^{-1}}(\mathrm{e}^{\boldsymbol{\alpha}^{(1)}} - \mathbf{I}_N) \in \mathbb{R}^{N\times N}$，$\boldsymbol{\varrho}_2 = \boldsymbol{\alpha}^{(1)^{-2}}(\mathrm{e}^{\boldsymbol{\alpha}^{(1)}} - \mathbf{I}_N - \boldsymbol{\alpha}^{(1)}) \in \mathbb{R}^{N\times N}$，且 $\boldsymbol{\varrho}_3 = \boldsymbol{\alpha}^{(1)^{-3}}(\mathrm{e}^{\boldsymbol{\alpha}^{(1)}} - \mathbf{I}_N - \boldsymbol{\alpha}^{(1)} - \frac{\boldsymbol{\alpha}^{(1)^2}}{2}) \in \mathbb{R}^{N\times N}$，其中对于矩阵 $\mathbf{A} \in \mathbb{R}^{N\times N}$，$\mathrm{e}^{\mathbf{A}} = \sum_{k=0}^{\infty} \frac{\mathbf{A}^k}{k!} \in \mathbb{R}^{N\times N}$。记 $\overline{\boldsymbol{\omega}^g} = \boldsymbol{\varrho}_1\boldsymbol{\omega} + \boldsymbol{\varrho}_2\boldsymbol{\beta}\boldsymbol{\omega}_L\boldsymbol{\lambda} + (\boldsymbol{\varrho}_2 - 2\boldsymbol{\varrho}_3)\mathbf{V}\mathbf{1}_N \in \mathbb{R}^N$，

$$\boldsymbol{\beta}^g = (\boldsymbol{\varrho}_1 - \boldsymbol{\varrho}_2)\,\boldsymbol{\beta} \in \mathbb{R}^{N\times N} \quad \text{and} \quad \mathbf{A}_j = (\boldsymbol{\varrho}_1 - \boldsymbol{\varrho}_2)\,\boldsymbol{\alpha}^{(j)} \in \mathbb{R}^{N\times N} \quad \text{with } 1 \leqslant j \leqslant P,$$

其中 $\mathbf{1}_N = (1, \ldots, 1)^\top \in \mathbb{R}^N$。为简单起见，我们进一步定义 $d\boldsymbol{\Lambda}_t = (d\Lambda_{1,t}, \ldots, d\Lambda_{N,t})$以及 $\mathbf{Z}_t d\mathbf{Z}_t = (Z_{1,t}dZ_{1,t}, \cdots, Z_{N,t}dZ_{N,t})^\top$。

命题 2。假设 $n \geqslant P+1$，且 $\boldsymbol{\alpha}^{(1)}$ 的谱半径小于一。那么定义 2 中的多变量 *HAR-Itˆo* 模型的积分波动率满足

$$\mathbf{y}_n = \boldsymbol{\omega}^g + \sum_{j=1}^{P} \mathbf{A}_j \mathbf{y}_{n-j} + \boldsymbol{\beta}^g \mathbf{J}_{n-1} + \boldsymbol{\varepsilon}_n \tag{2.8}$$

以概率 *1* 成立，其中 $\boldsymbol{\varepsilon}_n = \boldsymbol{\varepsilon}_n^c + \boldsymbol{\varepsilon}_n^J \in \mathbb{R}^N$，

$$\begin{aligned}
\boldsymbol{\varepsilon}_n^c =& 2\mathbf{V}\boldsymbol{\alpha}^{(1)^{-2}} \int_{n-1}^{n} \left\{ \left((n-t)\boldsymbol{\alpha}^{(1)} - \mathbf{I}_N\right) \mathrm{e}^{(n-t)\boldsymbol{\alpha}^{(1)}} + \mathbf{I}_N \right\} \mathbf{Z}_t d\mathbf{Z}_t \in \mathbb{R}^N, \\
\boldsymbol{\varepsilon}_n^J =& \boldsymbol{\beta}\boldsymbol{\alpha}^{(1)^{-1}} \left\{ \int_{n-1}^{n} \left(\mathrm{e}^{(n-t)\boldsymbol{\alpha}^{(1)}} - \mathbf{I}_N \right) \left[\mathbf{M}_t d\boldsymbol{\Lambda}_t + \boldsymbol{\omega}_L (d\boldsymbol{\Lambda}_t - \boldsymbol{\lambda} dt) \right] \right\} \in \mathbb{R}^N,
\end{aligned}$$

且 $\boldsymbol{\varepsilon}_n^c$ 和 $\boldsymbol{\varepsilon}_n^J$ 分别是来自连续扩散过程和跳跃分量的创新项。此外，$\{\boldsymbol{\varepsilon}_n^c\}$ 和 $\{\boldsymbol{\varepsilon}_n^J\}$ 是零均值且具有有限方差矩阵的序列。

与单变量情形类似，积分波动率 $\{\mathbf{y}_n\}$ 可表示为带外生变量 $\mathbf{J}_{n-1}$的向量自回归。在不存在跳跃分量的情况下，式（2.8）中的表示将简化为以下向量 HAR 模型，

$$\mathbf{y}_n = \boldsymbol{\omega}^g + \sum_{j=1}^{P} \mathbf{A}_j \mathbf{y}_{n-j} + \boldsymbol{\varepsilon}_n, \tag{2.9}$$

而积分波动率 $\{\mathbf{y}_n\}$ 可以通过已实现波动率进行一致估计。令 $\mathbf{y}_n^{(j)} = j^{-1}\sum_{l=1}^{j} \mathbf{y}_{n+1-l}$ 表示 j日积分波动率的平均值，则模型（2.9）可改写为 $\mathbf{y}_n = \boldsymbol{\omega}^g + \sum_{j=1}^{P} \boldsymbol{\beta}_j \mathbf{y}_{n-1}^{(j)} + \boldsymbol{\varepsilon}_n$，Hong 等人（Hong 等人 （2020）用它来$_{j=1}$预测已实现波动率。因此，定义 2和命题 2 中提出的 HAR-Itˆo 模型为该向量 HAR 模型提供了概率论支持。

备注 1。 当资产数量 N 较大时，系数矩阵 $\mathbf{A}_j$或 $\boldsymbol{\alpha}^{(j)}$可能具有低维结构，因为由大量资产构成的市场通常由少数综合因素驱动（Bai 和 Ng，2008年；Lam 和 Yao，2012年）。为此，借鉴自回归指数模型（Reinsel，1983年）的思路，Cubadda 等人（2017年）对以下矩阵施加了低秩假设

$$(\mathbf{A}_1^\top, \cdots, \mathbf{A}_P^\top) = (\boldsymbol{\alpha}^{(1)\top}, \cdots, \boldsymbol{\alpha}^{(P)\top})(\boldsymbol{\varrho}_1^\top - \boldsymbol{\varrho}_2^\top) \in \mathbb{R}^{N \times NP},$$

对应于$\mathbf{A}_j$'或 $\boldsymbol{\alpha}^{(j)}$'的行空间，并且可以观察到已实现波动率的预测精度显著提高。再考虑另一个矩阵$(\mathbf{A}_1, \cdots, \mathbf{A}_P) = (\boldsymbol{\varrho}_1 - \boldsymbol{\varrho}_2)(\boldsymbol{\alpha}^{(1)}, \cdots, \boldsymbol{\alpha}^{(P)}) \in \mathbb{R}^{N \times NP}$，它对应于$\mathbf{A}_j$'或 $\boldsymbol{\alpha}^{(j)}$'的列空间。一般而言，这两个矩阵的秩并不相等。这促使我们考虑$\mathbf{A}_j$'或 $\boldsymbol{\alpha}^{(j)}$'的列空间的低秩性，甚至同时考虑行空间和列空间的低秩性。

3 多线性低秩 HAR 模型

3.1 张量记法与分解

本小节介绍一些张量记法和 Tucker分解，这些内容将用于阐述新模型，并在本节及下一节中推导推断工具。

张量是矩阵针对高阶数据的自然推广，可提供实用的数据表示格式。张量的阶数是其维数，也称为维或模。多维数组 $\mathcal{X} \in \mathbb{R}^{p_1 \times \cdots \times p_d}$ 称为 d阶张量，本文将重点讨论三阶张量。关于张量记法与运算的详细综述，请参阅 Kolda 和 Bader (2009年)。

本文中，我们用花体大写字母表示张量 $\mathcal{X}$，$\mathcal{Y}$，.....。对于矩阵 $\mathbf{X}$，其 Frobenius 范数、算子范数、核范数、向量化、转置、谱半径和第 j 大奇异值分别记为 $\|\mathbf{X}\|_{\mathrm{F}}$、$\|\mathbf{X}\|_{\mathrm{op}}$、$\|\mathbf{X}\|_*$、$\mathrm{vec}(\mathbf{X})$、$\mathbf{X}^\top$、$\rho(\mathbf{X})$ 和 $\sigma_j(\mathbf{X})$。对于方阵 $\mathbf{X}$，最大和最小特征值分别记为 $\lambda_{\max}(\mathbf{X})$ 和 $\lambda_{\min}(\mathbf{X})$。此外，对于张量 $\mathcal{X} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$，令 $\|\mathcal{X}\|_{\mathrm{F}} = \left(\sum_{i=1}^{p_1}\sum_{j=1}^{p_2}\sum_{k=1}^{p_3}\mathcal{X}_{ijk}^2\right)^{1/2}$ 表示其 Frobenius 范数。

矩阵化（也称为展开或扁平化）是将高阶张量的元素重新排列为矩阵的过程。对于任意三阶张量 $\mathcal{X} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$，其模-1矩阵化 $\mathcal{X}_{(1)}$ 是一个 $p_1 \times (p_2 p_3)$ 矩阵，其中将第一个张量模作为行，并将其余各模合并为列。具体而言，$\mathcal{X}$ 的第 (i_1, i_2, i_3) 个元素映射为 $\mathcal{X}_{(1)}$ 的第 (i_1, j) 个元素，其中

$$j = 1 + \sum_{k=2}^{3} (i_k - 1) J_k \quad \text{with} \quad J_k = \prod_{\ell=2}^{k-1} p_\ell.$$

模-2和模-3矩阵化可以类似地定义。张量的矩阵化有助于关联矩阵与张量的概念和性质。张量 $\mathcal{X} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$ 与矩阵 $\mathbf{Y} \in \mathbb{R}^{q_1 \times p_1}$ 的模-1乘法 $\times_1$ 定义为

$$\mathcal{X} \times_1 \mathbf{Y} = \left(\sum_{i=1}^{p_1} \mathcal{X}_{ijk} \mathbf{Y}_{si} \right)_{1 \leqslant s \leqslant q_1, 1 \leqslant j \leqslant p_2, 1 \leqslant k \leqslant p_3} \in \mathbb{R}^{q_1 \times p_2 \times p_3},$$

我们也可以类似地定义模-2和模-3乘法。

对于任意三阶张量 $\mathcal{X} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$，其多线性秩 (r_1, r_2, r_3) 定义为其各模矩阵化的矩阵秩，即 $r_i = \text{秩}\left(\mathcal{X}_{(i)}\right)$ 其中 $1 \leqslant i \leqslant 3$；对于三阶及更高阶张量，这些秩不一定相等。因此，存在如下 *Tucker*分解 (Tucker,1966;De Lathauwer 等人,2000):

$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3 = \mathcal{G} \times_{i=1}^{3} \mathbf{U}_i, \tag{3.1}$$

其中 $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ 是核心张量，$\mathbf{U}_i \in \mathbb{R}^{p_i \times r_i}$ 是 $i = 1,2,3$对应的因子矩-阵。上述分解可简记为 $\mathcal{X} = [\![\mathcal{G}; \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3]\!]$。

注意，对于任意满足 $1 \leqslant i \leqslant 3$ 的非奇异矩阵 $\mathbf{O}_i \in \mathbb{R}^{r_i \times r_i}$，都有$[\![\mathcal{G}; \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3]\!] = [\![\mathcal{G} \times_1 \mathbf{O}_1 \times_2 \mathbf{O}_2 \times_3 \mathbf{O}_3; \mathbf{U}_1 \mathbf{O}_1^{-1}, \mathbf{U}_2 \mathbf{O}_2^{-1}, \mathbf{U}_3 \mathbf{O}_3^{-1}]\!]$，也就是说，该分解并非唯一。我们可以考虑 $\mathcal{X}$的高阶奇异值分解（HOSVD），即一种特殊的 Tucker分解：选取 $\mathbf{U}_i$ 为由 $\mathcal{X}_{(i)}$ 的前 r_i 个左奇异向量组成的高矩阵，并令 $\mathcal{G} = \mathcal{X} \times_1 \mathbf{U}_1^\top \times_2 \mathbf{U}_2^\top \times_3 \mathbf{U}_3^\top$，从而唯一确定该分解。此时，因子矩阵 $\mathbf{U}_i$ 均为正交归一的，而 $\mathcal{G}$ 具有全正交性质：对于每个 $1 \leqslant j \leqslant 3$，$\mathcal{G}_{(j)}$ 的各行两两正交。

3.2 多线性低秩HAR模型

考虑 N 金融资产及其对数价格 $\mathbf{X}_t = (X_{1,t}, \ldots\ldots, X_{N,t})^\top$ 其中 $t \in \mathbb{R}_+$ 由无跳跃分量的多变量 HAR-Itˆo 模型（2.6）和（2.7）生成。令 $\mathbf{y}_n = (y_{1,n}, \cdots, y_{N,n})^\top \in \mathbb{R}^N$ 表示中心化的 N 维积分波动率，且 $\mathbb{E}(\mathbf{y}_n) = 0$，接下来我们将为其构造一个已实现测度。

对于资产 i，令 $t_{n,k} \in [n-1, n]$（其中 $0 \leqslant k \leqslant m$）为具有日内观测值的时间点，其中 m 是第 n 天内的观测值数量，且 $n-1 = t_{n,0} < t_{n,1} < \cdots < t_{n,m} = n$。已实现波动率可定义为 $RV_{i,n} = \sum_{k=1}^{m}(X_{i,t_{n,k}} - X_{i,t_{n,k-1}})^2$，我们关注其中心化版本，即 $\widetilde{y}_{i,n} = RV_{i,n} - T^{-1}\sum_{l=1}^{T} RV_{i,l}$，其中 T 为总天数。记 $\widetilde{\mathbf{y}}_n = (\widetilde{y}_{1,n}, \cdots, \widetilde{y}_{N,n})^\top \in \mathbb{R}^N$，于是有

$$\widetilde{\mathbf{y}}_n = \mathbf{y}_n + \boldsymbol{\eta}_n, \qquad 1 \leqslant n \leqslant T, \tag{3.2}$$

其中 $\boldsymbol{\eta}_n = (\eta_{1,n}, \cdots, \eta_{N,n})^\top \in \mathbb{R}^N$ 为估计误差。为简洁起见，省略记号 $\widetilde{\mathbf{y}}_n$ 和 $\boldsymbol{\eta}_n$ 对 m 的依赖关系。假设所有 $1 \leqslant n \leqslant T$ 和 $1 \leqslant i \leqslant N$ 对应的观测价格数量 m 相同，且 $\{t_{n,k}\}$ 同步且等间距，即 $t_{n,k} - t_{n,k-1} = m^{-1}$。采用与 Kim and Wang（2016）、Kim 等人（2023）和 Song等（2021）类似的方法，可以证明

$$\max_{1\leqslant n\leqslant T}\max_{1\leqslant i\leqslant N} E(|\eta_{i,n}|^2) \leqslant O(m^{-1/2}). \tag{3.3}$$

注2。同步和等间距的假设仅为简化表述，且可放宽为广义采样时间（A¨ıt-Sahalia 等人，2010年）、刷新时间（Barndorff-Nielsen 等人，2011年）或前一时点（Zhang，2011年）。此外，除了已实现波动率，我们还可以考虑其他已实现度量，例如多尺度已实现波动率（Zhang，2006年）、核已实现波动率（Barndorff-Nielsen 等人，2008年）和预平均已实现波动率（Jacod 等人，2009年）。

根据命题2，低频积分波动率满足

$$\mathbf{y}_n = \sum_{j=1}^{P} \mathbf{A}_j \mathbf{y}_{n-j} + \boldsymbol{\varepsilon}_n, \qquad P+1 \leqslant n \leqslant T, \tag{3.4}$$

其中，截距 $\boldsymbol{\omega}^g$ 被中心化处理，因为 $\mathbb{E}(\mathbf{y}_n) = 0$、$\mathbf{A}_j$均为 $N \times N$ 系数矩阵，而创新项 $\{\boldsymbol{\varepsilon}_n\}$ 为 $i.i.d$，其中$\boldsymbol{\varepsilon}_n = (\varepsilon_{1,n}, \cdots, \varepsilon_{N,n})^\top \in \mathbb{R}^N$、$\mathbb{E}(\boldsymbol{\varepsilon}_n) = 0$ 和

$\text{var}(\boldsymbol{\varepsilon}_n) < \infty$。模型（3.4）的矩阵多项式定义为 $\boldsymbol{\mathcal{A}}(z) = \mathbf{I}_N - \mathbf{A}_1 z - \cdots - \mathbf{A}_P z^P$，其中 $z \in \mathbb{C}$ 其中 $\mathbb{C}$ 为复空间。

假设1. 对于所有 $|z| < 1$， $\boldsymbol{\mathcal{A}}(z)$ 的行列式均不等于零。

上述假设是向量自回归严格平稳的充要条件，因此确保 $\{\mathbf{y}_n\}$ 严格平稳。同时，积分波动率 $\{\mathbf{y}_n\}$ 由式（2.6）和（2.7）中的 Itˆo 过程生成，而非由式（3.4）生成；也就是说，序列 $\{\mathbf{y}_n\}$ 不是向量自回归过程。因此，假设 1 并非必要，而 $\{\mathbf{y}_n\}$ 的平稳性则能极大简化表述和技术细节。假设 1 不成立时，由于积分波动率不再平稳，我们无法对其进行中心化。这个问题可以通过使用式（3.2）中的原始积分波动率，并在式（3.4）中加入截距来解决。相关理论讨论可参见 Zheng 和 Cheng (2021年)。

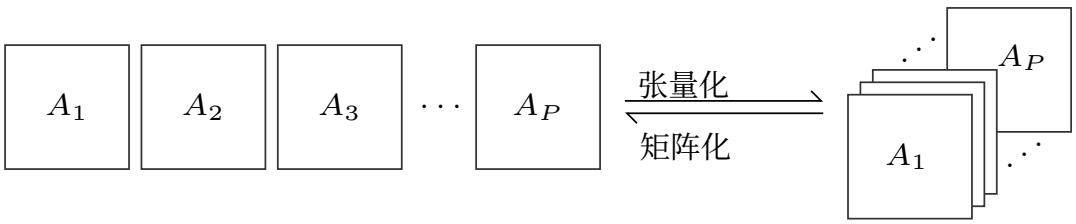

图1：将$\mathbf{A}_j$重新排列为三阶张量 $\boldsymbol{\mathcal{A}}$。

对于模型（3.4），参数的数量为 N^2P，可能非常庞大，因此本文采用张量技术对参数空间进行降维。具体而言，首先将系数矩阵重新排列为三阶张量$\boldsymbol{\mathcal{A}} \in \mathbb{R}^{N\times N\times P}$，使得 $\boldsymbol{\mathcal{A}}_{(1)} = (\mathbf{A}_1, \ldots, \mathbf{A}_P)$；示意见图1。随后，我们假设系数张量 (r_1, r_2, r_3) 具有多线性低秩 $\boldsymbol{\mathcal{A}}$，因此存在 Tucker 分解（De Lathauwer 等人，2000）。

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{U}_1 \times {}_2\mathbf{U}_2 \times_3 \mathbf{U}_3, \tag{3.5}$$

其中 $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{r_1\times r_2\times r_3}$ 是核心张量，而 $\mathbf{U}_1 \in \mathbb{R}^{N\times r_1}$、$\mathbf{U}_2 \in \mathbb{R}^{N\times r_2}$ 和 $\mathbf{U}_3 \in \mathbb{R}^{P\times r_3}$ 是因子矩阵。为简便起见，我们将公式（3.2）和（3.4），以及式（3.5）中的低秩结构，称为多线性低秩 *HAR*（MLR-HAR）模型。

注意，$\boldsymbol{\mathcal{A}}_{(2)} = (\mathbf{A}_1^\top、\cdots、\mathbf{A}_P^\top)$ 和 $\boldsymbol{\mathcal{A}}_{(3)} = (\text{vec}(\mathbf{A}_1)、.....、\text{vec}(\mathbf{A}_P))^\top$，以及由 $\boldsymbol{\mathcal{A}}_{(1)}$、$\boldsymbol{\mathcal{A}}_{(2)}$ 和 $\boldsymbol{\mathcal{A}}_{(3)}$ 所张成的空间，分别是系数矩阵的列空间、行空间和时间空间。因此，式（3.5）中的低秩假设同时从三个方向限制了参数空间，参数数量减少至 $r_1 r_2 r_3 + (N - r_1)\, r_1 + (N - r_2)\, r_2 + (P - r_3)\, r_3$。

所提出的 MLR-HAR 模型与含测量误差的向量自回归模型（Staudenmayer and Buonaccorsi，2005年）形式相似，将（3.4）代入（3.2）可得

$$\widetilde{\mathbf{y}}_n = \sum_{j=1}^{P} \mathbf{A}_j \widetilde{\mathbf{y}}_{n-j} + \boldsymbol{\epsilon}_n \quad \text{with} \quad \boldsymbol{\epsilon}_n = \underbrace{\boldsymbol{\eta}_n - \sum_{j=1}^{P} \mathbf{A}_j \boldsymbol{\eta}_{n-j}}_{estimation \quad error} + \underbrace{\boldsymbol{\varepsilon}_n}_{model \quad error}, \tag{3.6}$$

其中 $P + 1 \leqslant n \leqslant T$。假设系数张量具有 HOSVD 分解，即 $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3,$，也就是说 $\boldsymbol{\mathcal{G}}$ 是全正交的，且 $\mathbf{U}_j$ 均为正交归一。令 $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{G}} \times_3 \mathbf{U}_3,$ 和 $\mathbf{H}_j \in \mathbb{R}^{r_1 \times r_2}$ 分别表示其第 j 个正面切片，其中 $1 \leqslant j \leqslant P$，即 $\boldsymbol{\mathcal{H}}_{(1)} = (\mathbf{H}_1, \mathbf{H}2, \cdots, \mathbf{H}_P)$。因此，$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{H}} \times_1 \mathbf{U}_1 \times {}_2\mathbf{U}_2,$，我们可以将模型 (3.6) 重写为

$$\widetilde{\mathbf{y}}_n = \mathbf{U}_1 \sum_{j=1}^{P} \mathbf{H}_j \mathbf{U}_2^\top \widetilde{\mathbf{y}}_{n-j} + \boldsymbol{\epsilon}_n \quad \text{or} \quad \mathbf{U}_1^\top \widetilde{\mathbf{y}}_n = \sum_{j=1}^{P} \mathbf{H}_j \mathbf{U}_2^\top \widetilde{\mathbf{y}}_{n-j} + \mathbf{U}_1^\top \boldsymbol{\epsilon}_n,$$

其中 $\mathbf{U}_1^\top \widetilde{\mathbf{y}}_n$ 和 $\mathbf{U}_2^\top \widetilde{\mathbf{y}}_{n-j}$ 分别是响应变量和预测变量的汇总因子。Cubadda 等人 (2017年) 中的 HAR 模型对应于 $r_1 = N$和$\mathbf{U}_1$ 均为单位矩阵的情形，因此其参数数量多于所提出的 MLR-HAR 模型；更多讨论请参见备注 1 。

注释 3. 设 $\boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \in \mathbb{R}^{N \times N \times r_3}$，且 $\mathbf{S}_j \in \mathbb{R}^{N \times N}$ 为其 j 阶前向切片，其中 $1 \leqslant j \leqslant r_3$，即 $\boldsymbol{\mathcal{S}}_{(1)} = (\mathbf{S}_1, \mathbf{S}2, \cdots, \mathbf{S}_{r_3})$。记 $\mathbf{U}_3 = (\mathbf{u}^{(1)}、.....、\mathbf{u}^{(r_3)}) \in \mathbb{R}^{P \times r_3}$ 以及 $\mathbf{u}^{(k)} = (u_1^{(k)}、.....、u_P^{(k)})^\top \in \mathbb{R}^P$，其中 $1 \leqslant k \leqslant r_3$。因此，$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{S}} \times_3 \mathbf{U}_3,$，模型（3.6）可改写为

$$\widetilde{\mathbf{y}}_n = \mathbf{S}_1 \widetilde{\mathbf{x}}_n^{(1)} + \cdots + \mathbf{S}_{r_3} \widetilde{\mathbf{x}}_n^{(r_3)} + \boldsymbol{\epsilon}_n \quad \text{with} \quad \widetilde{\mathbf{x}}_n^{(k)} = \sum_{j=1}^{P} u_j^{(k)} \widetilde{\mathbf{y}}_{n-j}. \tag{3.7}$$

与日、周和月已实现波动率类似，$\widetilde{\mathbf{x}}_n^{(k)}$ 是沿时间方向汇总的因子，可视为 r_3 异质波动率成分，并由估计方法自动选出。此外，

$\mathbf{U}_3$ 是相应的载荷矩阵，当$\mathbf{U}_3$ =$\mathbf{U}_C$ 满足（2.5）时，模型（3.7）将退化为向量 HAR 模型（Bub´ak 等，2011年；Sou˘cek 和 Todorova，2013年）。

3.3 普通最小二乘法估计

假设系数张量的多线性秩 $(r_1$、r_2、$r_3)$ 已知。根据（3.5）和（3.6），MLR-HAR模型的普通最小二乘法（OLS）估计量可定义为

$$\widehat{\mathcal{A}}_{\mathrm{MLR}} \equiv [\![\widehat{\mathcal{G}}; \widehat{\mathbf{U}}_1, \widehat{\mathbf{U}}_2, \widehat{\mathbf{U}}_3]\!] = \operatorname{argmin} L(\mathcal{G}, \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3),$$

其中，$\tilde{}\mathbf{x}_n = (\tilde{}\mathbf{y}_{n-1}^{\top}$、……、$\tilde{}\mathbf{y}_{n-P}^{\top})^{\top}$，且

$$L(\mathcal{G}, \mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3) = \frac{1}{T} \sum_{n=P+1}^{T} ||\tilde{\mathbf{y}}_n - (\mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3)_{(1)} \tilde{\mathbf{x}}_n||_2^2.$$

尽管Tucker分解的各个分量 $\mathcal{G}$、$\mathbf{U}_1$, $\mathbf{U}_2$ 和$\mathbf{U}_3$, 无法识别， $\mathcal{A}$ 却可以被唯一识别。对于 N 和 $P\hat{}$均固定的低维情形，本小节通过调整 $\mathcal{A}_{\mathrm{MLR}}$ Shapiro （1986年）中针对过参数化模型的技术，确立其渐近正态性。

令 $\boldsymbol{\varphi} = \left(\mathrm{vec}(\mathcal{G}_{(1)})^{\top}、\mathrm{vec}(\mathbf{U}_1)^{\top}、\mathrm{vec}(\mathbf{U}_2)^{\top}、\mathrm{vec}(\mathbf{U}_3)^{\top}\right)^{\top}$以及 $\boldsymbol{h}(\boldsymbol{\varphi})$ =vec $(\mathcal{A}_{(1)})$ =vec$(\mathbf{U}_1\mathcal{G}_{(1)}(\mathbf{U}_3 \otimes \mathbf{U}_2)^{\top})$ 为 $\boldsymbol{\varphi}$的函数。记 $\boldsymbol{\Sigma}_{\varepsilon} = \mathrm{var}(\boldsymbol{\varepsilon}_{\mathrm{n}})$，且

$$\boldsymbol{\Gamma}^* = \begin{pmatrix} \boldsymbol{\Gamma}_0 & \boldsymbol{\Gamma}_1 & \cdots & \boldsymbol{\Gamma}_{P-1} \\ \boldsymbol{\Gamma}_1^{\top} & \boldsymbol{\Gamma}_0 & \cdots & \boldsymbol{\Gamma}_{P-2} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}_{P-1}^{\top} & \boldsymbol{\Gamma}_{P-2}^{\top} & \cdots & \boldsymbol{\Gamma}_0 \end{pmatrix} \in \mathbb{R}^{NP \times NP},$$

其中， $\boldsymbol{\Gamma}_j = \mathrm{cov}(\mathbf{y}^{n+j}, \mathbf{y}_n) \in \mathbb{R}^{N \times N}$，且满足 $j \geqslant 0$。因此，雅可比矩阵 $\mathbf{H}$ ： = $\partial \boldsymbol{h}(\boldsymbol{\varphi})/\partial \boldsymbol{\varphi}$ 具有如下形式

$$\mathbf{H} = \big((\mathbf{U}_3 \otimes \mathbf{U}_2 \otimes \mathbf{U}_1), [(\mathbf{U}_3 \otimes \mathbf{U}_2)\mathcal{G}_{(1)}^{\top}] \otimes \mathbf{I}_N, \mathbf{T}_{21}\left\{\left[(\mathbf{U}_3 \otimes \mathbf{U}_1)\mathcal{G}_{(2)}^{\top}\right] \otimes \mathbf{I}_N\right\},$$
$$\mathbf{T}_{31}\left\{\left[(\mathbf{U}_2 \otimes \mathbf{U}_1)\mathcal{G}_{(3)}^{\top}\right] \otimes \mathbf{I}_P\right\}\big) \in \mathbb{R}^{N^2P \times (r_1r_2r_3 + Nr_1 + Nr_2 + Pr_3)},$$

其中，$\mathbf{I}_{\ell} \in \mathbb{R}^{\ell \times \ell}$ 是单位矩阵，$\mathbf{T}_{ij} \in \mathbb{R}^{(N^2P) \times (N^2P)}$ 是置换矩阵，使得对于 $1 \leqslant i, j \leqslant 3$, $\mathrm{vec}(\mathcal{A}_{(j)}) = \mathbf{T}_{ij}\mathrm{vec}(\mathcal{A}_{(i)})$。此外，令 $\mathbf{J} = \boldsymbol{\Gamma}^* \otimes \boldsymbol{\Sigma}_{\varepsilon}^{-1}$，并令$\boldsymbol{\Sigma}_{\mathrm{MLR}}$ =$\mathbf{H}(\mathbf{H}^{\top}\mathbf{J}\mathbf{H})^{\dagger}\mathbf{H}^{\top}$, 其中 † 表示 Moore-Penrose 逆。

定理1. 假设 (3.3) 和假设 *1* 成立， $\mathbb{E}\|\boldsymbol{\varepsilon}_n\|^4 < \infty$，且 N 和P均为固定值。如果 $m \to \infty$，$T \to \infty$ 且 $T^{4+\delta}m^{-1} \to 0$ 对于某个 $\delta > 0$，则

$$\sqrt{T}\{\mathrm{vec}((\widehat{\boldsymbol{\mathcal{A}}}_{\mathrm{MLR}})_{(1)}) - \mathrm{vec}(\boldsymbol{\mathcal{A}}_{(1)})\} \to \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathrm{MLR}})$$

在分布意义下，其中$\mathbf{0}$是零向量。

上述定理中的渐近方差矩阵 $\boldsymbol{\Sigma}_{\mathrm{MLR}}$ 是退化的，因为$\widetilde{\boldsymbol{\mathcal{A}}}_{\mathrm{MLR}}$ 具有低 Tucker 秩。此外，日内观测值的数量 m要求以快于 T^4的速率发散；如果我们能在（3.3）中获得更高阶的估计误差矩，则可以放宽这一条件。

作为比较，我们还考虑没有低秩约束以及仅对 $\boldsymbol{\mathcal{A}}_{(2)}$ 施加低秩性约束的情形，其普通最小二乘估计量如下：

$$\widehat{\boldsymbol{\mathcal{A}}}_{\mathrm{OLS}} = \underset{}{\text{极小化}} \frac{1}{T}\sum_{n=P+1}^{T} ||\widetilde{\mathbf{y}}_n - \boldsymbol{\mathcal{A}}_{(1)}\widetilde{\mathbf{x}}_n||_2^2 \text{以及} \quad \widehat{\boldsymbol{\mathcal{A}}}_{\mathrm{MRI}} = \underset{(\boldsymbol{\mathcal{A}}_{(2)}) \leqslant r_2}{\text{极小化}} \text{秩} \frac{1}{T}\sum_{n=P+1}^{T} ||\widetilde{\mathbf{y}}_n - \boldsymbol{\mathcal{A}}_{(1)}\widetilde{\mathbf{x}}_n||_2^2,$$

其中 $\widehat{\boldsymbol{\mathcal{A}}}_{\mathrm{MRI}}$ 对应于Reinsel（1983年）中的多元自回归指数模型（MRI）。对于两个半正定矩阵 $\boldsymbol{\Sigma}_1$ 和 $\boldsymbol{\Sigma}_2$，记作 $\boldsymbol{\Sigma}_1 \leqslant \boldsymbol{\Sigma}_2$，若 $\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1$是半正定的。下一个推论从理论上验证了 $\boldsymbol{\mathcal{A}}_{\mathrm{MLR}}$ 是最高效的，而的表现 $\boldsymbol{\mathcal{A}}_{\mathrm{OLS}}$ 最差。

推论 1. 若定理 *1* 的条件成立，则 $\sqrt{T}\{\mathrm{vec}((\widehat{\boldsymbol{\mathcal{A}}}_{\mathrm{OLS}})_{(1)}) - \mathrm{vec}(\boldsymbol{\mathcal{A}}_{(1)})\} \to N(\mathbf{0}, \boldsymbol{\Sigma}_{\mathrm{OLS}})$ 且 $\sqrt{T}\{\mathrm{vec}((\widehat{\boldsymbol{\mathcal{A}}}_{\mathrm{MRI}})_{(1)}) - \mathrm{vec}(\boldsymbol{\mathcal{A}}_{(1)})\} \to N(\mathbf{0}, \boldsymbol{\Sigma}_{\mathrm{MRI}})$ 当 $m \to \infty$ 且 $T \to \infty$，其中 $\boldsymbol{\Sigma}_{\mathrm{OLS}} = \mathbf{J}^{-1}$ 且 $\boldsymbol{\Sigma}_{\mathrm{MRI}}$ 在证明中定义。此外，有 $\boldsymbol{\Sigma}_{\mathrm{MLR}} \leqslant \boldsymbol{\Sigma}_{\mathrm{MRI}} \leqslant \boldsymbol{\Sigma}_{\mathrm{OLS}}$。

4 高维 HAR 建模

4.1 高维HAR建模

在实际应用中，经常会遇到许多资产，资产数量 N可能非常庞大，例如以任意速率随样本量 T 增长。这对应于高维情形，此时上一节推导出的渐近性质不再成立。本节转而建立高维情形下 OLS 估计的非渐近性质。

假设2。 模型误差 $\varepsilon_n = \mathbf{\Sigma}_\varepsilon^{1/2}\boldsymbol{\xi}_n$, ，其中 $\{\boldsymbol{\xi}_n\}$ 是 $i.i.d.$ 随机向量，满足$\mathbb{E}(\boldsymbol{\xi}_n) = 0$, 方差$(\boldsymbol{\xi}_n) = \mathbf{I}_N$，且 $\mathbf{\Sigma}_\varepsilon =$ 方差(ε_n) 是一个正定矩阵。此外， $\boldsymbol{\xi}_n$ 的元素 $(\boldsymbol{\xi}_{in})_{1\leqslant i\leqslant N}$ 彼此独立且服从 κ^2-次高斯分布，模型误差 $\{\varepsilon_n\}$与估计误差 $\{\boldsymbol{\eta}_n\}$相互独立。

上述假设中的次高斯性在文献中常用于高维情形（Wainwright，2019年）。模型误差与估计误差相互独立这一假设，主要用于简化本节定理的技术性证明；通过冗长的证明过程，也可以放宽这一假设。

接下来，我们将推导非渐近误差界，这些界将基于 $\{\tilde{\mathbf{y}}_n\}$ （Basu 和 Michailidis，2015年）的时间依赖性和横截面依赖性。为此，下面首先定义两个依赖性测度，

$$\mu_{\min}(\boldsymbol{\mathcal{A}}) := \min_{|z|=1} \lambda_{\min}(\boldsymbol{\mathcal{A}}^*(z)\boldsymbol{\mathcal{A}}(z)) \quad \text{and} \quad \mu_{\max}(\boldsymbol{\mathcal{A}}) := \max_{|z|=1} \lambda_{\max}(\boldsymbol{\mathcal{A}}^*(z)\boldsymbol{\mathcal{A}}(z)),$$

其中，矩阵多项式 $\boldsymbol{\mathcal{A}}(z) = \mathbf{I}_N - \mathbf{A}_1 z - \cdots - \mathbf{A}_P z^P$, $\boldsymbol{\mathcal{A}}^*(z)$ 是 $\boldsymbol{\mathcal{A}}(z)$ 的共轭转置。对于任意两个序列 $\{a_n\}$ 和 $\{b_n\}$，若存在常数 C ，使得对所有 n 均有 $a_n \leqslant Cb_n$ （或 $a_n \geqslant Cb_n$），则记为 $a_n \lesssim b_n$ （或 $a_n \gtrsim b_n$）。设 $\kappa_L = \lambda_{\min}(\mathbf{\Sigma}_\varepsilon)/\mu_{\max}(\boldsymbol{\mathcal{A}})$, $\kappa_U = \lambda_{\max}(\mathbf{\Sigma}_\varepsilon)/\mu_{\min}(\boldsymbol{\mathcal{A}})$, 和 $d_{\mathcal{M}} = r_1 r_2 r_3 + N r_1 + N r_2 + P r_3$ 为 MLR-HAR 模型的参数个数。

定理 2. 假设 (3.3) 以及假设条件 *1* 和 *2* 成立。若样本量$T \gtrsim$最大值$(\kappa^2, \kappa^4)(\kappa_U/\kappa_L)^2 d_{\mathcal{M}}$, $m^{1/4}T^{2\delta} \gtrsim N^2 P$指数$(d_{\mathcal{M}})$且 $m^{1/4} \gtrsim T^{1+2\delta}$对于某个$\delta > 0$，则

$$\|\widehat{\boldsymbol{\mathcal{A}}}_{\text{MLR}} - \boldsymbol{\mathcal{A}}\|_{\text{F}} \leqslant \frac{C}{\kappa_L}\left[(\kappa^2\sqrt{\lambda_{\max}(\mathbf{\Sigma}_\varepsilon)\kappa_U} + \kappa\sqrt{\kappa_U})\sqrt{\frac{d_{\mathcal{M}}}{T}} + \frac{T^{1+2\delta}}{m^{1/4}}\right] \quad \textit{and}$$

$$\frac{1}{T}\sum_{n=P+1}^{T} \|(\widehat{\boldsymbol{\mathcal{A}}}_{\text{MLR}})_{(1)}\tilde{\mathbf{x}}_n - \boldsymbol{\mathcal{A}}_{(1)}\tilde{\mathbf{x}}_n\|_2^2 \leqslant \frac{C}{\kappa_L}\left[(\kappa^2\sqrt{\lambda_{\max}(\mathbf{\Sigma}_\varepsilon)\kappa_U} + \kappa\sqrt{\kappa_U})\sqrt{\frac{d_{\mathcal{M}}}{T}} + \frac{T^{1+2\delta}}{m^{1/4}}\right]^2,$$

with probability at least $1 - \exp(-Cd_{\mathcal{M}}) - 2\exp(-CT(\kappa_L/\kappa_U)^2 \min\{\kappa^{-2}, \kappa^{-4}\})$, *where* C *is a positive constant given in the proof.*

上述定理给出了估计误差和预测误差的上界，它们由两项构成：第一项源于对

模型 (3.4) 的 OLS 估计，第二项则是由使用已实现波动率来估计 (3.2) 中的积分波动率所导致的。当 κ_L 和 κ_U 有界且远离零和无穷大时，估计误差 $\|\widehat{\boldsymbol{\mathcal{A}}}_{\mathrm{MLR}}-\boldsymbol{\mathcal{A}}\|_{\mathrm{F}}=O_P(\sqrt{d_{\mathcal{M}}/T+T^{1+2\delta}/m^{1/4}})$，其中 $d_{\mathcal{M}}$ 衡量 MLR-HAR 模型的复杂度，而 $T^{1+2\delta}/m^{1/4}$ 项在 m 以快于 T^4 的速率发散时趋于零，这与定理 1中的低维情形相同。类似地，预测误差 $T^{-1}\sum_{n=P+1}^{T}\|(\widehat{\boldsymbol{\mathcal{A}}}_{\mathrm{MLR}})_{(1)}\widetilde{\mathbf{x}}_n-\boldsymbol{\mathcal{A}}_{(1)}\widetilde{\mathbf{x}}_n\|_2^2=O_P(d_{\mathcal{M}}/T+T^{2+4\delta}/m^{1/2})$，并且若 $m\to\infty, T\to\infty, d_{\mathcal{M}}/T\to 0$ 和 $T^{4+8\delta}m^{-1}\to 0$。

4.2 投影梯度下降算法

第3.3节指出，寻找普通最小二乘法估计量是一个非凸问题 $\boldsymbol{\mathcal{A}}_{\mathrm{MLR}}$，这使得参数估计在数值上和理论上都颇具挑战性。本小节采用Chen 等人(2019)的方法，引入一种投影梯度下降（PGD）算法，并给出其理论保证。

考虑位于3.2、3.4和3.5处的 MLR-HAR 模型参数空间 ），

$$\boldsymbol{\Theta}(r_1,r_2,r_3)=\{\boldsymbol{\mathcal{A}}\in\mathbb{R}^{N\times N\times P}:\operatorname{rank}(\boldsymbol{\mathcal{A}}_{(i)})\leqslant r_i \quad \text{for} \quad 1\leqslant i\leqslant 3\}.$$

我们首先介绍任意张量 $\boldsymbol{\mathcal{B}}\in\mathbb{R}^{N\times N\times P}$ 到 $\boldsymbol{\Theta}(r_1$、r_2、$r_3)$的一个投影。对于 $1\leqslant i\leqslant 3$，令 $\mathcal{M}_i$ 为矩阵化算子，它将张量映射为其模-i 矩阵化；令 $\mathcal{M}_i^{-1}$ 为逆算子，即 $\mathcal{M}_i(\boldsymbol{\mathcal{B}})=\boldsymbol{\mathcal{B}}_{(i)}$ 和 $\mathcal{M}_i^{-1}(\boldsymbol{\mathcal{B}}_{(i)})=\boldsymbol{\mathcal{B}}$。此外，记 P_r 为投影算子，它将矩阵映射为其最佳秩 r 近似。具体而言，对于一个矩阵，P_r 首先对其进行 SVD，然后保留最大的 r 个奇异值，并将其余奇异值压缩为零。因此，对于张量 $\boldsymbol{\mathcal{B}}\in\mathbb{R}^{N\times N\times P}$，我们可以定义其在 $\boldsymbol{\Theta}(r_1$、r_2、$r_3)$ 上的投影如下：

$$\widehat{P}_{\boldsymbol{\Theta}(r_1,r_2,r_3)}(\boldsymbol{\mathcal{B}}):=(\mathcal{M}_3^{-1}\circ P_{r_3}\circ\mathcal{M}_3)\circ(\mathcal{M}_2^{-1}\circ P_{r_2}\circ\mathcal{M}_2)\circ(\mathcal{M}_1^{-1}\circ P_{r_1}\circ\mathcal{M}_1)(\boldsymbol{\mathcal{B}}).$$

具体而言，我们首先对 $\boldsymbol{\mathcal{B}}$进行模-1矩阵化，然后通过奇异值分解找出最佳秩为 r_1的近似，最后将其折叠回三阶张量。随后依次对第二模和第三模执行相同操作；详情请参见算法 1 。矩阵化的执行顺序并不重要，也不会影响后续的收敛性分析。此外，$P_{\boldsymbol{\Theta}(r_1,\ r_2,\ r_3)}(\cdot)$ 是一种近似

投影到 $\boldsymbol{\Theta}(r_1、\ r_2、\ r_3)$ 中的投影，而精确投影众所周知是一个 NP难问题（Hillar 和 Lim，2013年）。

算法1 HAR建模的投影梯度下降算法 输入：数据 $\{\tilde{\mathbf{y}}_n\}$、参数空间 $\boldsymbol{\Theta} = \boldsymbol{\Theta}(r_1, r_2, r_3)$、迭代次数 K、步长 η 初始化： $k = 0$ 和 $\boldsymbol{\mathcal{A}}_0 \in \boldsymbol{\Theta}$。对于 $k = 1, 2, \ldots, K$，执行 $\tilde{\boldsymbol{\mathcal{A}}}_k = \boldsymbol{\mathcal{A}}_{k-1} - \eta \nabla L(\boldsymbol{\mathcal{A}}_{k-1})$（梯度下降）对于 $j = 1, 2, 3$，执行 $\tilde{\boldsymbol{B}}_j = \mathcal{M}_j(\boldsymbol{\mathcal{A}}_k)$（模式-$j$ 矩阵化）$\tilde{\boldsymbol{B}}_j = P_{r_j}(\boldsymbol{B}_j)$（通过奇异值分解求最佳秩 r_j 近似）$\widehat{\boldsymbol{\mathcal{A}}}_k = \mathcal{M}_j^{-1}(\widehat{\boldsymbol{B}}_j)$（通过逆向执行模式-$j$ 矩阵化，将其折叠为张量）结束循环 结束循环 输出： $\boldsymbol{\mathcal{A}}_K$

现在可以介绍 PGD方法来搜索最小二乘估计量 $\boldsymbol{\mathcal{A}}_{\mathrm{MLR}}$；见算法 1。具体而言，我们首先采用常用的梯度下降方法更新估计值，然后将更新后的张量投影到 $\boldsymbol{\Theta}(r_1, r_2, r_3)$，因为它可能不具有低 Tucker 秩。令 (r_1', r_2', r_3') 表示算法1中使用的运行 Tucker 秩，并记

$d'_{\mathcal{M}} = (r_1 + r_1')(r_2 + r_2')(r_3 + r_3') + N(r_1 + r_1') + N(r_2 + r_2') + P(r_3 + r_3')$。

定理3。 假设 (3.3) 以及假设 *1* 和 *2* 均成立，步长 $\eta = 2/(3\kappa_U)$,和运行中的 *Tucker* 秩 $r_i' \geqslant \left(\sqrt[3]{1 + \kappa_L/(24\kappa_U)} - 1\right)^{-2} r_i$ 满足 ~~$1 \leqslant i \leqslant 3$~~。若$T \gtrsim$最大值$(\kappa^2, \kappa^4)\ (\kappa_U/\kappa_L)^2 d'_{\mathcal{M}}$, $m^{1/4}T^{2\delta} \gtrsim N^2 P$指数$(d'_{\mathcal{M}})$且 $m^{1/4} \gtrsim T^{1+2\delta}$对某个$\delta > 0$成立，则

$$\|\widehat{\boldsymbol{\mathcal{A}}}_K - \boldsymbol{\mathcal{A}}\|_{\mathrm{F}} \leqslant (1 - \frac{\kappa_L}{24\kappa_U})^K \|\widehat{\boldsymbol{\mathcal{A}}}_0 - \boldsymbol{\mathcal{A}}\|_{\mathrm{F}} + \frac{C}{\kappa_L}\left[(\kappa^2\sqrt{\lambda_{\max}(\boldsymbol{\Sigma}_\varepsilon)}\kappa_U + \kappa\sqrt{\kappa_U})\sqrt{\frac{d'_{\mathcal{M}}}{T} + \frac{T^{1+2\delta}}{m^{1/4}}}\right],$$

以至少 $1 -$指数$(-Cd'_{\mathcal{M}}) - 2$指数$(-CT(\kappa_L/\kappa_U)^2$最小值$\{\kappa^{-2}, \kappa^{-4}\})$ 的概率成立，其中 C 是证明中给出的正常数。

上述定理中上界的两项分别对应优化误差和统计误差，而统计误差的形式与

定理2中的形式类似。注意到 $\kappa_L < \kappa_U$，因此优化误差满足线性收敛速率。具体而言，对于任意 $\epsilon > 0$，我们可以选择ˆ迭代次数 $K = [\log(\epsilon) - \log \|\mathcal{A}_0 - \mathcal{A}\|_{\mathrm{F}}]/\log [1 - \kappa_L/(24\kappa_U)]$，使得优化ˆ误差小于 ϵ。对于初始值 $\mathcal{A}_0$，实际操作中可以简单地将其设为零。最后，实际应用中 $\mathcal{A}$ 的 Tucker 秩通常未知，可以通过经验方法或高维贝叶斯信息准则（BIC）来选择，

$$\mathrm{BIC}(\mathbf{r}) = \log\left\{\frac{1}{T}\sum_{n=P+1}^{T}\|\tilde{\mathbf{y}}_n - \widehat{\mathcal{A}}(\mathbf{r})\tilde{\mathbf{x}}_n\|_2^2\right\} + \frac{\lambda d_{\mathcal{M}}(\mathbf{r})\log(T)}{T}, \tag{4.1}$$

ˆ ˆ其中 $\mathcal{A}(\mathbf{r}) = \mathcal{A}_{\mathrm{MLR}}$的 Tucker 秩为 $\mathbf{r} = (r_1, r_2, r_3)$, $d_{\mathcal{M}}(\mathbf{r}) = r_1r_2r_3 + Nr_1 + Nr_2 + Pr_3$是参数的数量，而 λ 是一个调优参数。

5 模拟研究

本节开展了三项模拟实验：前两项分别用于评估第3.3节和第4.1节中提出的 MLR-HAR 模型在低维和高维情形下的 OLS 估计的有限样本性能，第三项则用于验证第4.2节中所提出算法的收敛性。

在第一个实验中，使用（2.6）和（2.7）中的多元 HAR-Itˆo 模型生成高频数据，并假设不存在跳跃成分和漂移项，即 $L_{i,t} = 0$ 和 $\mu_{i,t} = 0$，其中 $1 \leqslant i \leqslant N$。维度固定为 $N = 5$，离散化时将时间间隔设为 $\Delta = 1/780$。对于两个布朗运动 B_t 和 W_t，增量 $\mathrm{B}_{t+\Delta} - \mathrm{B}_t$ 和 $\mathrm{W}_{t+\Delta} - \mathrm{W}_t$ 服从零均值、方差矩阵为 $\Delta \cdot \mathrm{I}_N$ 的多元正态分布，且其相关系数矩阵为 $-0.6 \cdot \mathrm{I}_N$。对于所有 $1 \leqslant i \leqslant N$，初始对数价格固定为 $X_{i,0} = \log(50)$，初始瞬时波动率固定为 $\sigma_{i,0} = 0.1$。模型参数设为 $(\omega_i, v_i) = (0.2, 0.4)$，其中 $1 \leqslant i \leqslant N$，且 $(\boldsymbol{\alpha}^{(1)}, \ldots, \boldsymbol{\alpha}^{(P)}) = \boldsymbol{\alpha}^*(\mathrm{U}_{\mathrm{C}}^\top \otimes \mathrm{I}_N) \in \mathbb{R}^{N\times N\times 22}$；这里，$\mathrm{U}_{\mathrm{C}} \in \mathbb{R}^{22\times 3}$ 定义见（2.5），而 $\boldsymbol{\alpha}^* \in \mathbb{R}^{N\times 3N}$ 是一个随机生成的秩为二且满足假设1的矩阵。根据备注1，我们可以计算其低频表示的系数张量 $\mathcal{A} \in \mathbb{R}^{N\times N\times 22}$，并可验证其具有低 Tucker 秩 $(r_1, r_2, r_3) = (2, 2, 3)$。

我们考虑五种不同的样本量，$T = 500(1 + j)$ 且 $1 \leqslant j \leqslant 5$，每种设置重复实验2000次。首先根据日内观测值数量为 $m = 78$ 或780计算已实现波动率，然后采用算法1，以步长 5×10^{-4}、容差 10^{-7} 和初始值$\mathcal{A}_0 = 0$搜索估计量 $\widehat{\mathcal{A}}_{\mathrm{MLR}}$。作为比较，我们还计算了不施加低秩约束的估计量$\widehat{\quad}$，以及仅对 $\mathcal{A}_{(2)}$ 施加低秩约束的估计量，即第3.3节中的 $\mathcal{A}_{\mathrm{OLS}}$ 和 $\mathcal{A}_{\mathrm{MRI}}$。此外，根据定理1和推论 1，可以得到三个估计量的渐近方差矩阵。图 2 展示了经验方差（EVar）矩阵的最大奇异值，以及估计渐近方差（AVar）矩阵的最大奇异值的平均值。偏差也以相对于真实系数张量的平均绝对偏差表示，并将其平方，以便与 EVar 和 AVar 进行比较。我们有以下四点发现。（i.）随着样本量 T 增加，所有偏差和方差都趋于零，因此可以得出这三个估计量具有一致性。（ii.）$\widehat{\quad}$与 $\mathcal{A}_{\mathrm{OLS}}$ 和 $\mathcal{A}_{\mathrm{MRI}}$相比，$\mathcal{A}_{\mathrm{MLR}}$ 的偏差和方差最小，这$\widehat{\quad}$与我们的直觉一致：$\mathcal{A}_{\mathrm{MLR}}$ 利用了更多低秩结构。（iii.）EVar 通常与相应的 AVar 很吻合，且随着 T 增加，两者差异逐渐缩小，不过所有情况下 AVar 都倾向于低估方差。（iv.）最后，日内观测值数量较大时，即 $m = 780$，这三个估计量的偏差、EVar 和 AVar 均略小。

第二个实验旨在验证定理 2 中的非渐近估计误差界，该误差界由两部分组成：$O_p(\sqrt{d_{\mathcal{M}}/T})$ 和 $O_p(T^{1+2\delta}/m^{1/4})$。这两部分分别源于低频表示中的模型误差和已实现波动率带来的高频误差，二者难以拆分并单独研究。因此，我们考虑了两种不同的数据生成过程，以分别对二者进行评估。第一种数据生成过程直接根据模型（$\{y_{i,\ n}\}$）生成低频数据3.4，其中不涉及高频误差，即 $m = \infty$，因此误差界简化为 $O_p(\sqrt{d_{\mathcal{M}}/T})$。系数张量的形式为 $\mathcal{A} = \mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3 \in \mathbb{R}^{N \times N \times 22}$，其中核心张量 $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ 的元素首先从标准正态分布中独立生成，然后重新缩放，使得 $\|\mathcal{G}\|_{\mathrm{F}} = 0.5$，而因子矩阵

U$_i$'通过提取高斯随机矩阵的前 r_i 个左奇异向量生成，同时确保满足假设1中的平稳性条件。误差项$\{\boldsymbol{\varepsilon}_n\}$ 服从多元标准正态分布，Tucker秩设为 $(r_1、\ r_2、\ r_3) = (2、2、2)$ 或 $(3、3、3)$。我们考虑三个维度，分别为 $N = 10、20$ 和 25，以及五个样本量，$T = 50(4 + 3j + j^2)$ 其中 $1 \leqslant j \leqslant 5$，并再次使用算法1 $\widehat{}$ $\widehat{}$搜索估计值 $\boldsymbol{\mathcal{A}}_{\mathrm{MLR}}$。图 3 给出了估计误差 $\|\boldsymbol{\mathcal{A}}_{\mathrm{MLR}} - \boldsymbol{\mathcal{A}}\|_{\mathrm{F}}$在500次重复实验中的平均值，并可以观察到其关于 $\sqrt{d_{\mathcal{M}}/T}$ 的线性关系。事实上，随着样本量 T 不断增加，估计误差将趋近于零。因此，误差界的第一部分得到了验证。还可以看出，这些直线具有不同的斜率，这是因为定理 2 中误差界的常数项（例如 κ、 κ_L和 κ_U）可能随 N的不同而变化。

对于第二个实验中的第二个数据生成过程，我们生成一个$i.i.d.$ 高频估计误差序列 $\{\boldsymbol{\eta}_n\}$ 来模拟（3.2），其中 $\boldsymbol{\eta}_n$ 服从均值为零、方差为 $m^{-1/2} \cdot \mathbf{I}_N$ 的多元正态分布，而 $\{\mathbf{y}_n\}$ 则由第一个数据生成过程生成。我们设置了五个样本量 $T = 50(1 + j)$，其 $1 \leqslant j \leqslant 5$，其他所有设置均与第一个数据生成过程相同。计算已实现波动率时，日内观测值的数量设为 $m = T^4$。注意，根据定理 2， $\sqrt{T/d_{\mathcal{M}}}\|\widehat{\boldsymbol{\mathcal{A}}}_{\mathrm{MLR}} - \boldsymbol{\mathcal{A}}\|_{\mathrm{F}} = O_p(1) + O_p(T^{3/2+2\delta}/m^{1/4})$,因此，我们在图 $\sqrt{T/d_{\mathcal{M}}}\|\widehat{\boldsymbol{\mathcal{A}}}_{\mathrm{MLR}} - \boldsymbol{\mathcal{A}}\|_{\mathrm{F}}$ 中绘制调整后的估计误差与 $T^{3/2}/m^{1/4}$ 的关系3。明显的线性关系证实了定理 2的后半部分。此外，由于 $O_p(1)$ 项的存在，这些直线具有不同的截距。

第三个实验旨在评估第4.2节中算法1的收敛性能。使用第一次实验中的数据生成过程生成样本，其中 $(N、\ T) = (30、\ 1000)$，并使用 $m = 78、390$或780计算已实现波动率。需要注意的是，真实的Tucker秩为 $(r_1、\ r_2、\ r_3) = (2、2、3)$，我们在算法中考虑了三组不同的运行秩：$(r_1'、\ r_2'、\ r_3') = (2、2、3)、\ (5、5、5)$ 和 $(10、10、10)$。 $\widehat{}$图 4给出了前20次迭代的标准化均方根误差 $\|\boldsymbol{\mathcal{A}}_K - \boldsymbol{\mathcal{A}}\|_{\mathrm{F}}/\|\boldsymbol{\mathcal{A}}\|_{\mathrm{F}}$，可以看出所有情形都具有相似的衰减模式。同时，更准确的预先指定的秩和/或更多的日内观测值意味着更低的估计误差。我们还尝试了更多的重复实验次数

以及不同的数据生成过程，也观察到了类似的现象。

6 真实数据分析

本节分析2009年4月1日至2013年12月30日期间标准普尔500指数成分股的高频交易数据。为降低结构突变的影响，剔除了2009年第一季度的数据。具体而言，我们考察 $N = 31$只股票，即2013年1月2日交易量最大的42只或90只股票，并使用2013年的数据评估样本外表现。因此，估计期有 $T = 937$ 天，预测期有 $M = 249$ 天。

从沃顿研究数据服务下载上午9:30至下午4:00的日交易数据，并采用文献中最常用的五分钟收益率，其对应的日内观测值数量为 $m = 78$（参见Zhang，2011年；Andersen和Bollerslev，1998；Liu等人，2015年）。由于隔夜收益率通常包含受外部因素影响的跳跃，因此将其排除在外。我们使用两种已实现度量——已实现方差（RV）和中位数已实现方差（medRV）——来估计积分波动率，其中medRV能够降低跳跃效应；参见Andersen等人(2012年)。两种度量均先转换为对数形式，再进行中心化处理，使其均值为零。

提出的阶数为 $P = 22$ 的 MLR-HAR 模型与两个竞争模型进行了比较：向量 HAR (VHAR) 模型和向量 HAR 指数 (VHARI) 模型。两个竞争模型中的波动率分量均固定为日、周和月波动率，而我们模型中的波动率分量则由模型自动选择。我们还考虑 MLR-HAR 模型的阶数 $P = 66$，以探索可能存在的更长期波动率分量。VHAR 模型不涉及进一步降维，而 VHARI 模型假设系数矩阵的行空间秩为 $r < N$。因此，VHAR、VHARI 和 MLR-HAR 模型分别有 $3N^2$、$4Nr - r^2$ 和 $r_1r_2r_3 + (N - r_1)r_1 + (N - r_2)r_2 + (P - r_3)r_3$ 个参数，我们的模型参数少得多。使用 (4.1) 中的贝叶斯信息准则，为调优参数为 $\lambda = 10^{-4}$ 和 $1 \leqslant r_1$、r_2、$r_3 \leqslant 10$ 的 MLR-HAR 模型搜索 Tucker 秩，得到 $(r_1$、r_2、$r_3) = (3$、3、$2)$；对于阶数 $P = 22$ 和 66，以及股票数为 $N = 31$、42 和 90 的所有情况，均使用

2009年到2012年的训练数据。该方法也经过调整，用于选择 VHARI 模型的秩；对于股票数为 $N = 31$、42 和 90 的情况，所选秩分别为 $r = 4$、4 和 5。

我们采用滚动预测程序评估这四个模型的样本外表现：在 2013 年的样本外期间，历史数据的结束点逐步向前滚动，窗口大小固定为 $T = 937$，并在每次迭代中进行一步预测。预测期间，MLR-HAR 和 VHARI 模型的秩与上述设定相同；我们采用文献中最常用的经验准似然（QLIKE）来评估预测准确度，

$$\text{QLIKE}_i = \frac{1}{M}\sum_{n=1}^{M}\left(\frac{\widetilde{y}_{i,n}}{\widehat{y}_{i,n}} - \log\left(\frac{\widetilde{y}_{i,n}}{\widehat{y}_{i,n}}\right) - 1\right) \quad \text{with } 1 \leqslant i \leqslant N,$$

其中$\widehat{y}_{i,\ n}$ 和$\widetilde{y}_{i,\ n}$ 分别表示 2013 年第 i个资产在第 n个交易日的预测值和计算得到的已实现测度， $M = 249$表示 2013 年的交易日数；参见Patton 和 Sheppard (2009年); Patton (2011a,b)。图5 展示了 VHAR、VHARI 和两个 MLR-HAR 模型在 $N = 31$、42 或 90 只股票下的 QLIKE值箱线图，得出以下四点发现。(i.) MLR-HAR 模型的预测准确度远高于 VHAR 和 VHARI 模型，说明从两个方向探索 N 只股票之间的低秩结构十分重要。(ii.) 两个 MLR-HAR 模型的表现相近，且 $P = 66$模型的表现甚至略差。这证实了文献中的常见做法，即使用一个月内的波动率分量来预测已实现测度，而较长期的波动率分量可能没有显著贡献。(iii.) 股票数量越多，预测误差通常越大，这可能是因为更大的 N 会使模型更复杂。(iv.) 最后，medRV 的 QLIKE值远小于 RV 的 QLIKE值，说明数据中可能存在跳跃。

为了更好地理解高维性对 MLR-HAR 模型的影响，我们考虑了 2011年1月2日 至 2012年12月30日 的较短周期训练数据，总共只有 $T = 497$ 天。所选 VHARI模型的秩分别为 $r = 3$、4 和 5，对应于 $N = 31$、42 和 90 只股票，而 MLR-HAR 模型的 Tucker 秩与之前相同。相应的 QLIKE值 箱线图见图 5，可以看出，所有预测误差都变大了。尤其是拥有 $N = 90$ 只股票的 VHAR 模型表现明显最差，

这可能是由于历史数据较少，但参数数量却十分庞大。作为比较，两个 MLR-HAR 模型受到的影响都不大，因此我们可以认为，我们的模型具备同时处理大量资产的能力。

文献中广泛采用包含三个异质波动率分量的级联结构，而上述 MLR-HAR建模在所有情况下都只采用两个数据驱动分量。这促使我们进一步研究：文献中常用的三个固定形式分量是否最适合预测波动率。为此，我们首先使用阶数为 $P = 22$ 的 MLR-HAR 模型重新拟合 2009年4月1日 至 2013年12月30日 的数据，并将 Tucker 秩设为 $(r_1、\ r_2、\ r_3) = (3、3、3)$，即将异质波动率分量的数量固定为三个，以便比较估计得到的载荷矩阵 $\widehat{\mathrm{U}}_3 \in \mathbb{R}^{22\times 3}$ 与式（2.5）中的 $\mathrm{U}_\mathrm{C} \in \mathbb{R}^{22\times 3}$ 。请注意，根据式（3.5）的 Tucker分解，$\widehat{\mathrm{U}_3}$ 并不唯一，但其列空间 $\mathrm{colspace}(\mathrm{U}_3)$ 可以唯一确定。接下来，我们采用 Pan和Yao（2008）中的差异度量来评估 $\widehat{\mathrm{U}}_3$ 与 U_C 之间的距离：$\mathcal{D}\{\mathrm{colspace}(\widehat{\mathrm{U}}_3),\ \mathrm{colspace}(\mathrm{U}_\mathrm{C})\} = \{1 - \mathrm{tr}(\mathrm{D}_1\mathrm{D}_1^\top\mathrm{D}_2\mathrm{D}_2^\top)/3\}^{1/2},$，其中 $\widehat{\mathrm{D}}_1$ 和 D_2 分别是 $\mathrm{colspace}(\mathrm{U}_3)$ 和 $\mathrm{colspace}(\mathrm{U}_\mathrm{C})$ 的正交归一基；该度量的取值范围为 $[0、\ 1]$，数值越大表示两个空间之间的差异越大。对于股票数量分别为 $N = 31$、42 和 90 的情况，该度量的值分别为 0.717、0.736 和 0.652；也就是说，U_3 与 U_C 之间存在显著差异。这证实了所提出的数据驱动方法的必要性，该方法能够在 HAR 模型中自动选择异质波动率分量。

7 结论与讨论

本文对广受欢迎的 HAR模型进行了重要拓展，并解决了其在高频域和低频域中的主要缺陷。具体而言，在高频情形下，我们建立了单变量和多变量 HAR-Itˆo 模型，将 HAR 类模型与其对应的 Itˆo 扩散过程联系起来，并首次探究了 HAR 类模型的高频动态。另一方面，在低频情形下，我们提出了多线性低秩HAR模型，该模型

同时对响应、预测变量和滞后方向施加低秩假设。这种低秩结构不仅大幅缩小了参数空间，使模型能够处理资产数量多得多的情形，还以数据驱动的波动率分量取代了HAR模型中固定的异质波动率分量。因此，模型的灵活性大大增强，真实数据分析也证实了这种做法的必要性。最后，我们推导了高维 HAR 建模的理论性质，并提出了投影梯度下降算法，其线性收敛性得到了理论证明。

下面简要讨论本研究可能的几项扩展。首先，HAR 模型（Corsi，2009年）的思想源于将不同的时间跨度分组为不同类型的波动率分量。沿着这一思路，我们可以将 $\tilde{\mathbf{y}}_{n-1}$，…，$\tilde{\mathbf{y}}_{n-P}$在（3.4）中重新排列为一个三阶张量，使其三个模分别对应资产、周和月，随后便可利用张量技术自动选择周度（中期）和月度（长期）波动率分量。这将形成一个四阶系数张量 $\mathcal{A}$，本文中的高维建模工具也可以据此进行调整。其次，第4节中推导的非渐近性质依赖于次高斯假设，而金融和经济数据通常具有重尾特征（Shin等人，2021年）。因此，在更重尾的假设下讨论所提出高维建模方法的理论性质具有重要的实际意义。最后，已实现协方差（Bauer和Vorkink，2011年）近年来受到越来越多的关注，并且可以将其视为矩阵值时间序列（Chen 等人，2021年）。与本文类似，我们也可以将HAR协方差表示与Itˆo扩散过程联系起来，进而提出低秩HAR协方差模型，以预测已实现协方差。

参考文献

A¨ıt-Sahalia, Y.、Fan, J. 和 Xiu, D.（2010年）。带噪声和异步金融数据的高频协方差估计。《美国统计学会杂志》，105卷：1504–1517页。

A¨ıt-Saha lia, Y. 和 Jacod, J. (2009年). 对离散观测过程中的跳跃进行检验 .

《统计学年鉴》，37:184–222。

Andersen, T. G. 和 Bollerslev, T.（1998）。回应怀疑者：是的，标准波动率模型确实能提供准确的预测。《国际经济评论》，39:885–905。

Andersen, T. G.、Bollerslev, T. 和 Diebold, F.（2007）。纳入跳跃成分：收益率波动率测度、建模与预测中的跳跃成分。《经济学与统计学评论》，89:701–720。

Andersen, T. G.、Dobrev, D. 和 Schaumburg, E.（2012）。使用最近邻截断进行稳健于跳跃的波动率估计。《计量经济学杂志》，169:75–93。

Audrino, F. 和 Knaus, S. D.（2016）。HAR模型的Lasso化：从模型选择视角看已实现波动率动态。《计量经济学评论》，35:1485–1521。

Bai, J. 和 Ng, S.（2008年）。大维因子分析。《计量经济学基础与趋势》，3：89–163。

Barndorff-Nielsen, O. E.、Hansen, P. R.、Lunde, A. 和 Shephard, N.（2008年）。在存在噪声的情况下设计已实现核，以测度股票价格的事后变动。《计量经济学》，76：1481–1536。

Barndorff-Nielsen, O. E.、Hansen, P. R.、Lunde, A. 和 Shephard, N.（2011年）。多元已实现核：在存在噪声和非同步交易的情况下，对股票价格协变动的一致半正定估计量。《计量经济学杂志》，162：149–169。

Basu, S. 和 Michailidis, G.（2015年）。稀疏高维时间序列模型中的正则化估计。《统计学年鉴》，43：1535–1567。

Bauer, G. H. 和 Vorkink, K.（2011年）。预测多元已实现股票市场波动率。《计量经济学杂志》，160：93–101。

Bekaert, G. 和 Hoerova, M. (2014年). VIX指数、方差溢价与股票市场波动率.《计量经济学杂志》, 183:181–192.

Bekierman, J. 和 Manner, H. (2018). 使用时变系数模型预测已实现方差测度. 《国际预测学杂志》, 34:276–287. Bollerslev, T. (1986年). 广义自回归条件异方差. 《计量经济学杂志》, 31:307–327. Bollerslev, T., Patton, A. J. 和 Quaedvlieg, R. (2016). 利用误差：改进波动率预测的一种简单方法. 《计量经济学杂志》, 192:1–18.

Bollerslev, T., Patton, A. J. 和 Quaedvlieg, R.（2018）。对（不）可靠的已实现协方差进行建模和预测，以支持更可靠的金融决策。《计量经济学杂志》，207：71–91。

Bub´ak, V., Koˇcenda, E. 和 ˇZikeˇs, F.（2011）。新兴欧洲外汇市场中的波动率传导。《银行与金融杂志》，35：2829–2841。

Buccheri, G. 和 Corsi, F.（2021）。HARK 和 SHARK：考虑测量误差和非线性依赖关系的已实现波动率建模。《金融计量经济学杂志》，19：614–649。

Bucci, A.（2020）。利用神经网络预测已实现波动率。《金融计量经济学杂志》，18：502–531。

Caporin, M.（2022）。跳跃在已实现波动率建模与预测中的作用。《金融计量经济学杂志》，nbab030。

Chen, H.、Raskutti, G. 和 Yuan, M.（2019）。广义低秩张量回归的非凸投影梯度下降。《机器学习研究杂志》，20：172–208。

Chen, R.、Xiao, H. 和 Yang, D.（2021年）。矩阵值时间序列的自回归模型。《计量经济学杂志》，222：539–560。

Chen, X. B., Gao, J., Li, D., 和 Silvapulle, P. (2018)。时变系数已实现波动率模型的非参数估计与预测。《商业与经济统计杂志》，36：88–100。

Chen, Y., H¨ardle, W.K., 和 Pigorsch, U.(2010)。局部化已实现波动率建模。《美国统计学会杂志》，105卷：1376–1393。

Clements, A. 和 Preve, D. P. A.（2021年）。HAR 波动率模型应用实务指南。《银行与金融杂志》，133:106285。

Corsi, F.（2009年）。已实现波动率的简单近似长记忆模型。《金融计量经济学杂志》，7:174–196。

Corsi, F.、Mittnik, S.、Pigorsch, C. 和 Pigorsch, U.（2008年）。实现波动率的波动率。《计量经济学评论》，27:46–78。

Corsi, F.、Pirino, D. 和 Ren`o, R.（2010年）。阈值双幂变差及跳跃对波动率预测的影响。《计量经济学杂志》，159:276–288。

Corsi, F. 和 Ren`o, R.（2012年）。具有持久杠杆效应的离散时间波动率预测及其与连续时间波动率建模的联系。《商业与经济统计杂志》，30:368–380。

Cubadda, G., Guardabascio, B., 和 Hecq, A.（2017年）。一种用于已实现波动率度量的向量异质自回归指数模型。《国际预测学杂志》，33：337–344。

De Lathauwer, L., De Moor, B., 和 Vandewalle, J.（2000）。一种多线性奇异值分解。《SIAM 矩阵分析与应用杂志》，21：1253–1278。

Degiannakis, S., Filis, G., Klein, T., 和 Walther, T.（2022年）。农产品已实现波动率预测。《国际预测学杂志》，38：74–96。

Engle, R. F.（1982年）。基于英国通货膨胀方差估计的自回归条件异方差。《计量经济学》，50：987–1007。

Engle, R. F. 和 Gallo, G. M.（2006年）。利用日内数据建立波动率多指标模型。《计量经济学杂志》，131：3–27。

Ghysels, E.、Santa-Clara, P. 和 Valkanov, R.（2006年）。预测波动率：充分利用不同频率采样的收益数据。《计量经济学杂志》，131：59–95。

Hansen, P. R.、Huang, Z. 和 Shek, H. H.（2012年）。已实现 GARCH：收益与已实现波动率度量的联合模型。应用计量经济学杂志，27：877–906。

Hillar, C. J. 和 Lim, L.-H.（2013年）。大多数张量问题都是 NP 困难的。美国计算机协会会刊（*JACM*），60（6）：1–39。

Hong, W. T.、Lee, J. 和 Hwang, E.（2020）。关于多元 HAR-RV 模型中最小二乘估计渐近正态性理论的注记。数学，8：2083。

Hu, J.、Kuo, W. Y. 和 H¨ardle, W. K.（2021年）。比特币市场风险：波动率、跳跃与预测。*arXiv* 预印本 *arXiv:1912.05228*。

Jacod, J.、Li, Y.、Mykland, P. A.、Podolskij, M. 和 Vetter, M.（2009年）。连续情形下的微观结构噪声：预平均方法。随机过程及其应用，119：2249–2276。

Kim, D.（2023）。指数 GARCH-Itˆo 波动率模型。计量经济学理论，即将发表。

Kim, D. 和 Fan, J.（2019）。用于高频数据的因子 GARCH-Itˆo 模型及其在大规模波动率矩阵预测中的应用。《计量经济学杂志》，208：395–417。

Kim, D.、M.、S. 和Wang, Y.（2023）。隔夜 GARCH-Itˆo 波动率模型。《商业与经济统计杂志》，即将发表。

Kim, D. 和 Wang, Y.（2016）。统一的离散时间与连续时间模型，以及针对合并的低频与高频金融数据的统计推断。《计量经济学杂志》，194：220–230。

K¨ohler, S.（2021）。级联自回归（CAR）模型的模型阶数选择。工作论文。
Kolda, T. G. 和 Bader, B. W.（2009）。张量分解及其应用。《*SIAM*评论》，51：455–500。Lam, C. 和 Yao, Q.（2012）。高维时间序列的因子建模：因子数量的推断。《统计学年鉴》，40：694–726。

Liu, L. Y., Patton, A. J., and Sheppard, K.（2015年）。有没有什么指标胜过5分钟已实现波动率？多个资产类别的已实现度量比较。《计量经济学杂志》，187：293–311。

Luo, J., Demirer, R., Gupta, R., and Ji, Q.（2022a）。结构突变下利用情绪指标预测石油和黄金波动率。《能源经济学》，105：105751。

Luo, J., Klein, T., Ji, Q., and Hou, C.（2022b）。利用无限隐马尔可夫HAR模型预测农产品期货已实现波动率。《国际预测学杂志》，38：51–73。

Ly´ocsa, ˇS. 和 Todorova, N.（2021年）。是什么驱动了美国石油和天然气公司的波动率？《能源经济学》，100:105367。McAleer, M. 和 Medeiros, M. C.（2008年）。用于长记忆和非对称性的多状态平滑转换异质自回归模型。《计量经济学杂志》，147:104–119。Oh, D. H. 和 Patton, A. J.（2016）。适用于混频数据的高维基于Copula的分布。《计量经济学杂志》，193:349–366。Pan, J. 和 Yao, Q.（2008年）。基于共同因子的多元时间序列建模。《生物统计学》，95:365–379。Patton, A. J.（2011a）。已实现波动率估计量的基于数据的排序。《计量经济学杂志》，161:284–303。

Patton, A. J.（2011b）。使用不完美波动率代理变量进行波动率预测比较。《计量经济学杂志》，160:246–256。

Patton, A. J. 和 Sheppard, K.（2009年）。评估波动率和相关性预测。载于：Andersen, T. G.、Davis, R. A.、Kreiss, J-P. 和 Mikosch, T.（编），《金融时间序列手册》。施普林格出版社。

Patton, A. J. 和 Sheppard, K.（2015年）。好波动率，坏波动率：带符号跳跃与波动率的持续性。《经济学与统计学评论》，97：683–697。

Reinsel, G.（1983年）。关于多元自回归指数模型的一些结果。《生物统计学》，70：145–156。

Shapiro, A.（1986年）。过度参数化结构模型的渐近理论。《美国统计学会杂志》，81：142–149。

Shephard, N. 和 Sheppard, K.（2010年）。实现未来：使用基于高频数据的波动率（HEAVY）模型进行预测。应用计量经济学杂志，25：197–231。

Shin, M., Kim, D., Wang, Y., 和 Fan, J.（2021年）。适用于厚尾高频金融数据的因子与特质 VAR-Itˆo 波动率模型。*arXiv* 预印本 *arXiv:2109.05227*。

Song, X., Kim, D., Yuan, H., Cui, X., Lu, Z., Zhou, Y., 和 Wang, Y.（2021年）。基于已实现 GARCH-Itˆo 模型的波动率分析。《计量经济学杂志》，222:393–410。

Souˇcek, M. 和 Todorova, N.（2013年）。原油与股票期货市场之间的已实现波动率传导：一种多元HAR方法。《能源经济学》，40:586–597。

Staudenmayer, J. 和 Buonaccorsi, J. P.（2005年）。线性自回归模型中的测量误差。《美国统计学会杂志》，100:841–852。

Taylor，N.（2017年）。Box-Cox变换下的已实现方差预测。《国际预测学杂志》，33：770–785。

Tibshirani，R.（1996）。通过套索进行回归收缩与选择。《皇家统计学会杂志：*B*辑》，58：267–288。Truc´ıos，C.（2019）。比特币风险度量预测：一种稳健方法。《国际预测学杂志》，35：836–847。

Tucker，L. R.（1966）。关于三模因子分析的一些数学札记。《心理计量学》，31：279–311。

Wainwright，M. J.（2019）。《高维统计学：非渐近视角》。剑桥大学出版社，剑桥。

Wang，D.、Huang，F.、Zhao，J.、Li，G. 和 Tian，G.（2020）。紧凑型自回归网络。载于*AAAI*人工智能会议论文集，第34卷，第6145–6152页。

Zhang，L.（2006）。利用含噪观测高效估计随机波动率：一种多尺度方法。《伯努利》，12：1019–1043。

Zhang，L.（2011）。协变动估计：Epps效应与微观结构噪声。《计量经济学杂志》，160：33–47。

Zheng，Y. 和 Cheng，G.（2021年）。线性约束下向量自回归模型的有限时间分析。《生物统计学》，108：469–489。

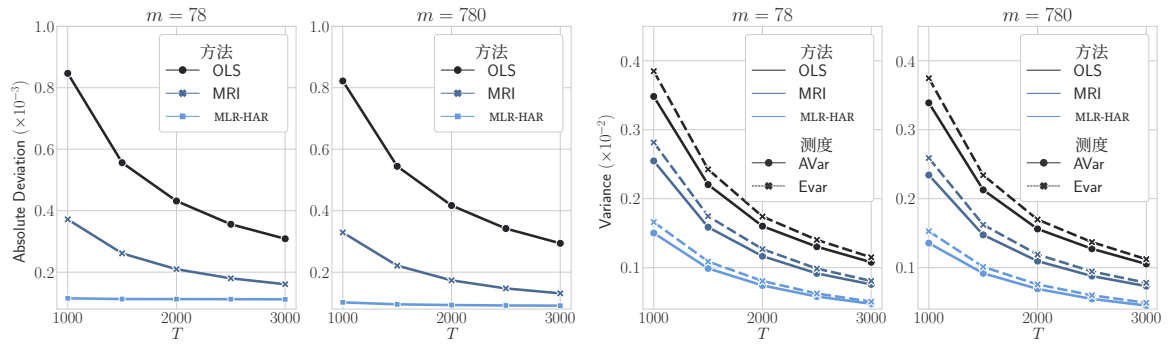

图 2：绝对偏差平方（左侧两个面板），以及估计量$\hat{\mathcal{A}}_{\mathrm{MLR}}$、$\hat{\mathcal{A}}_{\mathrm{MRI}}$和$\hat{\mathcal{A}}_{\mathrm{OLS}}$的经验方差（EVar）与估计渐近方差（AVar）矩阵（右侧两个面板）；日内观测值数量为 $m = 78$ 或 780。

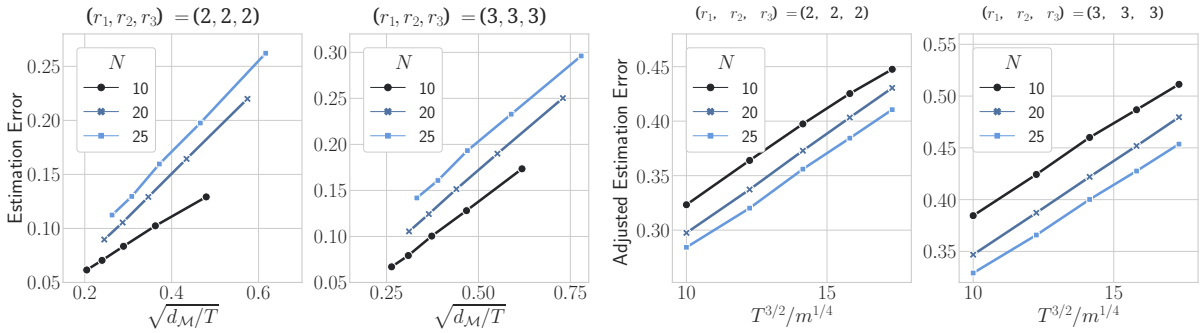

图 3：第一个数据生成过程下相对于$\sqrt{d_{\mathcal{M}}/T}$的估计误差$\|\hat{\mathcal{A}}_{\mathrm{MLR}} - \mathcal{A}\|_{\mathrm{F}}$（左侧两个面板），以及第二个数据生成过程下相对于$T^{3/2}/m^{1/4}$的调整后的估计误差$\sqrt{T/d_{\mathcal{M}}}\|\hat{\mathcal{A}}_{\mathrm{MLR}} - \mathcal{A}\|_{\mathrm{F}}$（右侧两个面板）。

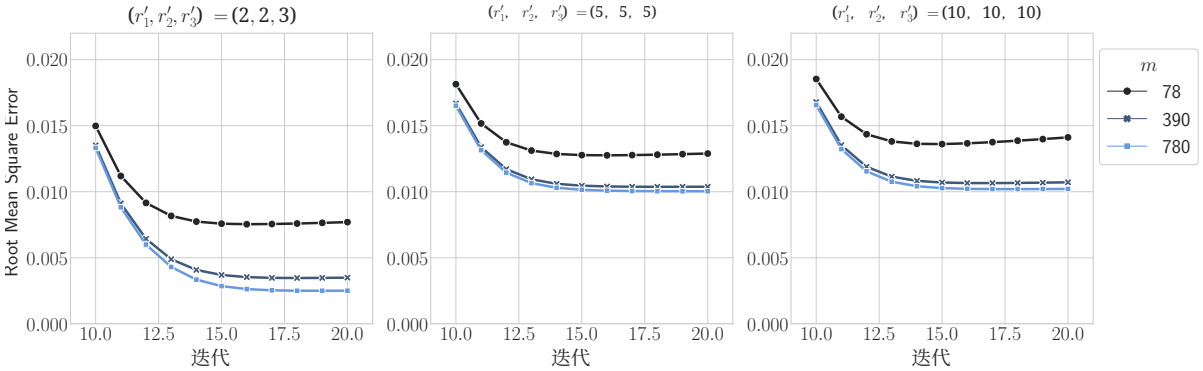

图 4：运行秩$(r'_1, r'_2, r'_3) = (2, 2, 3)$、$(5, 5, 5)$和$(10, 10, 10)$下前 20 次迭代的标准化均方根误差。

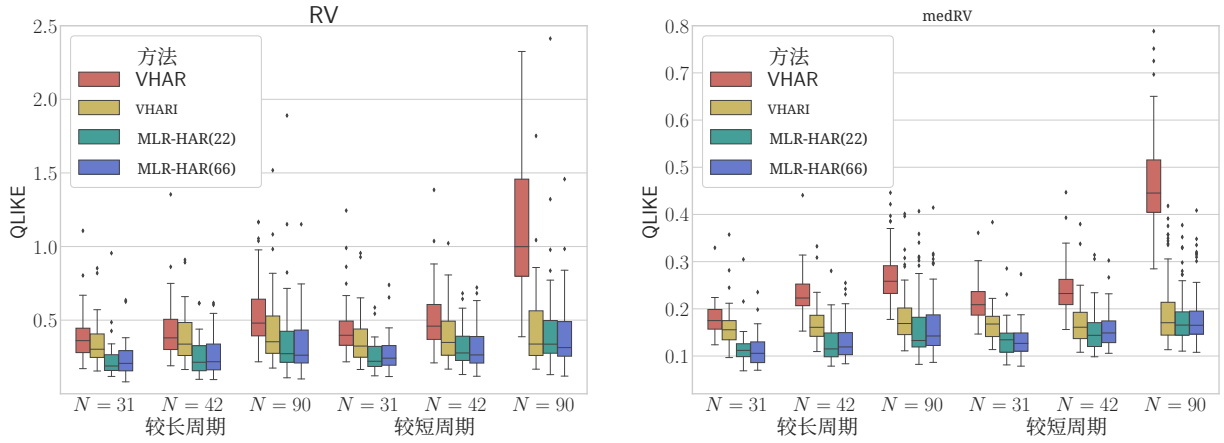

图5：较长周期（2009.4–2013.12）和较短周期（2011.1–2013.12）的数据中，VHAR模型、VHARI模型和阶数为 $P = 22$ 和66的MLR-HAR 模型在 $N = 31$、42和90只股票上的QLIKE值箱线图。使用了两种已实现测度：RV（左面板）和medRV（右面板）。

附录A：技术细节

本附　　录给出了第2–4节理论结果的技术性证明　　.

A.1 第1 个命题和第2 节中的证明2

命题*1*的证明。定义

$$R(k)=\int_{n-1}^{n}\frac{(n-t)^k}{k!}\sigma_t^2dt=\frac{\sigma_{n-1}^2}{(k+1)!}+\int_{n-1}^{n}\frac{(n-t)^{k+1}}{(k+1)!}d\sigma_t^2$$

对所有 $k\geqslant 0$ 和 $n\geqslant P+1$。根据定义1 以及 Itˆo 引理，有

$$d\sigma_t^2=[\omega+v(n-t)-\sigma_{n-1}^2-vZ_t^2+\alpha_1\sigma_t^2]dt+2v(n-t)Z_tdZ_t+\beta L_t^2d\Lambda_t$$

对于 $n-1<t\leqslant n$，并且

$$v\int_{n-1}^{n}\frac{(n-t)^{k+1}}{(k+1)!}Z_t^2dt=\frac{v}{(k+3)!}+2v\int_{n-1}^{n}\frac{(n-t)^{k+2}}{(k+2)!}Z_tdZ_t.$$

因此，对于任意 $k\geqslant 0$,

$$\begin{aligned}R(k)&=\frac{w}{(k+2)!}+\frac{(k+2)v}{(k+3)!}+\frac{(k+1)\sigma_{n-1}^2}{(k+2)!}+2v\int_{n-1}^{n}\frac{(n-t)^{k+2}}{(k+1)!}Z_tdZ_t\\&\quad-v\int_{n-1}^{n}\frac{(n-t)^{k+1}}{(k+1)!}Z_t^2dt+\beta\int_{n-1}^{n}\frac{(n-t)^{k+1}}{(k+1)!}L_t^2d\Lambda_t+\alpha_1R(k+1)\\&:=R_1(k)+R_2(k)+R_3(k)+R_4(k)+\alpha_1R(k+1),\end{aligned}\tag{A.1}$$

其中

$$\begin{aligned}&R_1(k)=2v\int_{n-1}^{n}\frac{(k+1)(n-t)^{k+2}}{(k+2)!}Z_tdZ_t,\quad R_2(k)=\beta\int_{n-1}^{n}\frac{(n-t)^{k+1}}{(k+1)!}L_t^2d\Lambda_t,\\&R_3(k)=\frac{(k+3)w+(k+1)v}{(k+3)!},\qquad\text{and}\qquad R_4(k)=\frac{(k+1)\sigma_{n-1}^2}{(k+2)!}.\end{aligned}$$

由于 $R(k)\leqslant\int_{n-1}^{n}\sigma_t^2dt$ 且 $0<\alpha_1<1$，因此 $\alpha_1^kR(k)\to 0$ 几乎必然成立，因为 $k\to\infty$ 并且对公式（A.1）进行迭代可得

$$R(0)=\int_{n-1}^{n}\sigma_t^2dt=\sum_{k=0}^{\infty}\alpha_1^kR_1(k)+\sum_{k=0}^{\infty}\alpha_1^kR_2(k)+\sum_{k=0}^{\infty}\alpha_1^kR_3(k)+\sum_{k=0}^{\infty}\alpha_1^kR_4(k)\tag{A.2}$$

以概率 1 成立。经过一些代数运算，可以验证

$$\begin{aligned}
\sum_{k=0}^{\infty}\alpha_1^k R_1(k) &= \sum_{k=0}^{\infty} 2v\alpha_1^{-1}\int_{n-1}^{n}\frac{[\alpha_1(n-t)]^{k+1}(n-t)}{(k+1)!}Z_t dZ_t \\
&\quad - \sum_{k=0}^{\infty} 2v\alpha_1^{-2}\int_{n-1}^{n}\frac{[\alpha_1(n-t)]^{k+2}}{(k+2)!}Z_t dZ_t \\
&= 2v\alpha_1^{-2}\int_{n-1}^{n}\left\{\alpha_1\left(n-t-\alpha_1^{-1}\right)e^{\alpha_1(n-t)}+1\right\}Z_t dZ_t := \varepsilon_n^c, \\
\sum_{k=0}^{\infty}\alpha_1^k R_2(k) &= \sum_{k=0}^{\infty}\beta\alpha_1^{-1}\int_{n-1}^{n}\frac{[\alpha_1(n-t)]^{k+1}}{(k+1)!}L_t^2 d\Lambda_t = \beta\alpha_1^{-1}\int_{n-1}^{n}\left(e^{\alpha_1(n-t)}-1\right)L_t^2 d\Lambda_t \\
&= \varrho_2\beta\omega_L\lambda + \varepsilon_n^J, \\
\sum_{k=0}^{\infty}\alpha_1^k R_3(k) &= \varrho_2\omega + (\varrho_2 - 2\varrho_3)v \quad \text{and} \quad \sum_{k=0}^{\infty}\alpha_1^k R_4(k) = (\varrho_1-\varrho_2)\sigma_{n-1}^2,
\end{aligned} \tag{A.3}$$

其中 $\varrho_1 = \alpha_1^{-1}(e^{\alpha_1}-1)$, $\varrho_2 = \alpha_1^{-2}(e^{\alpha_1}-1-\alpha_1)$, $\varrho_3 = \alpha_1^{-3}(e^{\alpha_1}-1-\alpha_1-\alpha_1^2/2)$, 个 d

$$\varepsilon_n^J = \beta\alpha_1^{-1}\left\{\int_{n-1}^{n}\left(e^{\alpha_1(n-t)}-1\right)\left[M_t d\Lambda_t + \omega_L\left(d\Lambda_t - \lambda dt\right)\right]\right\}.$$

此外，整数时间点 n 处的瞬时波动率 σ_t^2 具有如下形式： $\sigma_n^2 = \omega + \sum_{l=1}^{P}\alpha_l y_{n-l+1} + \beta J_n$，其中

$$y_n = \int_{n-1}^{n}\sigma_s^2 ds \quad \text{and} \quad J_n = \int_{n-1}^{n} L_s^2 d\Lambda_s$$

分别为积分波动率和跳跃变差。结合式（A.2）和（A.3），可得式（2.3）。

注意，随机变量 ε_n^c 和 ε_n^J 均独立于 σ-域 $\mathcal{F}_{n-1}$，因此 $\{\varepsilon_n^c\}$ 和 $\{\varepsilon_n^J\}$ 是两个 $i.i.d.$ 序列。此外，容易验证 $\{\varepsilon_n^c\}$ 和 $\{\varepsilon_n^J\}$ 均为鞅差序列（Song 等人，2021年），即 $E(\varepsilon_n^c) = E(\varepsilon_n^J) = 0$。接下来，我们证明 $\mathrm{var}(\varepsilon_n^c)$ 和 $\mathrm{var}(\varepsilon_n^J)$ 均有限。

令 $\varrho_4(t) = \alpha_1\left(n-t-\alpha_1^{-1}\right)e^{\alpha_1(n-t)}+1$，则对所有 $n-1<t\leqslant n$，都有 $|\varrho_4(t)| < 2e+1$。此外，$Z_t = \int_{[t]}^{t} dW_t = W_t - W_{n-1}$、$dZ_t = dW_t$ 和 $(dZ_t)^2 = dt$。因此，由 Itˆo 等距性，

$$\begin{aligned}
E(\varepsilon_n^c)^2 &= E\left(2v\alpha_1^{-2}\int_{n-1}^{n}\varrho_4(t)Z_t dZ_t\right)^2 = 4v^2\alpha_1^{-4}E\int_{n-1}^{n}[\varrho_4(t)]^2 Z_t^2 (dZ_t)^2 \\
&< 4v^2\alpha_1^{-4}(2e+1)^2 E\int_{n-1}^{n} Z_t^2 dt.
\end{aligned} \tag{A.4}$$

根据 Itˆo 引理和 Z_t 的定义，有 $dZ_t^2 = 2Z_t dZ_t + dt$、 $Z_{n-1} = 0$，且 $\{\int_{n-1}^t Z_s dZ_s\}$ 是一个鞅差序列。因此， $Z_t^2 = Z_t^2 - Z_{n-1}^2 = \int_{n-1}^t (2Z_s dZ_s + ds)$，并且

$$E\int_{n-1}^n Z_t^2 dt = 2E\left(\int_{n-1}^n \int_{n-1}^t Z_s dZ_s dt\right) + 0.5 = 0.5,$$

这与（A.4）一起意味着 $\mathrm{var}(\varepsilon_n^c) < 2v^2\alpha_1^{-4}(2e+1)^2 < \infty$。

接下来考虑 $\mathrm{var}(\varepsilon_n^J)$。令 $\varrho_5(t) = e^{\alpha_1(n-t)} - 1$，则 $|\varrho_5(t)| < e+1$ 对所有 $n-1 < t \leqslant n$均成立。由（A.3）以及 $E(\varepsilon_n^J) = 0$这一事实可得

$$\mathrm{var}(\varepsilon_n^J) = \beta^2\alpha_1^{-2}E\left(\int_{n-1}^n \varrho_5(t)L_t^2 d\Lambda_t\right)^2 - (\varrho_2\beta\omega_L\lambda)^2. \tag{A.5}$$

注意到， $\{L_t\}$ 与泊松过程 $\{\Lambda_t\}$ 相互独立，并且根据划分方法，我们有

$$\begin{aligned} E\left(\int_{n-1}^n \varrho_5(t)L_t^2 d\Lambda_t\right)^2 &= E\left(\sum_{n-1<t\leqslant n} \varrho_5(t)L_t^2\Delta\Lambda_t\right)^2 < (e+1)^2 E\left(\sum_{n-1<t\leqslant n} L_t^2\Delta\Lambda_t\right)^2 \\ &= (e+1)^2 E(L_t^4)E\left(\sum_{n-1<t\leqslant n}\Delta\Lambda_t^2\right) = (e+1)^2(\lambda^2+\lambda)E(L_t^4), \end{aligned}$$

这与（A.5）以及 $E(L_t^4) < \infty$这一事实一起意味着 $\mathrm{var}(\varepsilon_n^J) < \infty$。

□

命题 2 的证明。 定义

$$R_i(k) = \int_{n-1}^n \frac{(n-t)^k}{k!}\sigma_{i,t}^2 dt = \frac{\sigma_{i,n-1}^2}{(k+1)!} + \int_{n-1}^n \frac{(n-t)^{k+1}}{(k+1)!} d\sigma_{i,t}^2, \quad 1 \leqslant i \leqslant N,$$

对所有 $k \geqslant 0$ 和 $n \geqslant P+1$均成立。根据定义 2 和伊藤引理，我们有

$$\begin{aligned} d\sigma_{i,t}^2 = &\left(\omega_i + \upsilon_i(n-t) - \sigma_{i,n-1}^2 - \upsilon_i Z_{i,t}^2 + \sum_{j=1}^N \alpha_{i,j}^{(1)}\sigma_{j,t}^2\right) dt + 2\upsilon_i(n-t)Z_{i,t}dZ_{i,t} \\ &+ \beta_i L_{i,t}^2 d\Lambda_{i,t} \end{aligned}$$

对于 $n-1 < t \leqslant n$，以及

$$\upsilon_i\int_{n-1}^n \frac{(n-t)^{k+1}}{(k+1)!} Z_{i,t}^2 dt = \frac{\upsilon_i}{(k+3)!} + 2\upsilon_i\int_{n-1}^n \frac{(n-t)^{k+2}}{(k+2)!} Z_{i,t}dZ_{i,t}.$$

因此，对于任意 $k \geqslant 0$，令

$$R_i^{(1)}(k) = 2v_i \int_{n-1}^{n} \frac{(k+1)(n-t)^{k+2}}{(k+2)!} Z_{i,t} dZ_{i,t}, \quad R_i^{(2)}(k) = \beta_i \int_{n-1}^{n} \frac{(n-t)^{k+1}}{(k+1)!} L_{i,t}^2 d\Lambda_{i,t}$$
$$R_i^{(3)}(k) = \frac{(k+3)w_i + (k+1)v_i}{(k+3)!}, \qquad \text{and} \qquad R_i^{(4)}(k) = \frac{(k+1)\sigma_{i,n-1}^2}{(k+2)!},$$

于是有

$$\begin{aligned} R_i(k) =& \frac{\omega_i}{(k+2)!} + \frac{(k+2)v_i}{(k+3)!} + \frac{(k+1)\sigma_{i,n-1}^2}{(k+2)!} \\ &+ 2v_i \int_{n-1}^{n} \frac{(n-t)^{k+2}}{(k+1)!} Z_{i,t} dZ_{i,t} - v_i \int_{n-1}^{n} \frac{(n-t)^{k+1}}{(k+1)!} Z_{i,t}^2 dt \\ &+ \beta_i \int_{n-1}^{n} \frac{(n-t)^{k+1}}{(k+1)!} L_{i,t}^2 d\Lambda_{i,t} + \sum_{j=1}^{N} \alpha_{i,j}^{(1)} R_j(k+1) \\ =& R_i^{(1)}(k) + R_i^{(2)}(k) + R_i^{(3)}(k) + R_i^{(4)}(k) + \sum_{j=1}^{N} \alpha_{i,j}^{(1)} R_j(k+1), \end{aligned}$$

可改写为

$$\mathbf{R}(k) = \mathbf{R}^{(1)}(k) + \mathbf{R}^{(2)}(k) + \mathbf{R}^{(3)}(k) + \mathbf{R}^{(4)}(k) + \boldsymbol{\alpha}^{(1)} \mathbf{R}(k+1), \tag{A.6}$$

其中，$\mathbf{R}(k) = (R_1(k), R_2(k), \cdots, R_N(k))^\top$ 和 $\mathbf{R}^{(j)}(k) = (R_1^{(j)}(k), R_2^{(j)}(k), \cdots, R_N^{(j)}(k))^\top$，且 $1 \leqslant j \leqslant 4$。

由于以下事实 $R_i(k) \leqslant \int_{n-1}^{n} \sigma_{i,\,t}^2 dt$ 且 $\boldsymbol{\alpha}^{(1)}$ 的谱半径小于 1，因此 ${\boldsymbol{\alpha}^{(1)}}^k \mathbf{R}(k) \to 0$ 几乎必然成立，当 $k \to \infty$时。此外，通过迭代公式（A.6），可得

$$\begin{aligned} \mathbf{R}(0) = \int_{n-1}^{n} \boldsymbol{\sigma}_t^2 dt = \sum_{k=0}^{\infty} {\boldsymbol{\alpha}^{(1)}}^k \mathbf{R}^{(1)}(k) + \sum_{k=0}^{\infty} {\boldsymbol{\alpha}^{(1)}}^k \mathbf{R}^{(2)}(k) + \sum_{k=0}^{\infty} {\boldsymbol{\alpha}^{(1)}}^k \mathbf{R}^{(3)}(k) \\ + \sum_{k=0}^{\infty} {\boldsymbol{\alpha}^{(1)}}^k \mathbf{R}^{(4)}(k) \end{aligned} \tag{A.7}$$

以概率 1 成立。与命题1类似，可以验证

$$\begin{aligned}
\sum_{k=0}^{\infty} \boldsymbol{\alpha}^{(1)^k} \mathbf{R}_1(k) &= \sum_{k=0}^{\infty} \boldsymbol{\alpha}^{(1)^k} \left[2v_i \int_{n-1}^{n} \frac{(k+1)(n-t)^{k+2}}{(k+2)!} Z_{i,t} dZ_{i,t} \right]_{i=1,\cdots,N}^{\top} \\
&= 2\mathbf{V}\boldsymbol{\alpha}^{(1)^{-2}} \int_{n-1}^{n} \left\{ \left((n-t)\boldsymbol{\alpha}^{(1)} - \mathbf{I}_N \right) \mathbf{e}^{(n-t)\boldsymbol{\alpha}^{(1)}} + \mathbf{I}_N \right\} \mathbf{Z}_t d\mathbf{Z}_t \\
&:= \boldsymbol{\varepsilon}_n^c, \\
\sum_{k=0}^{\infty} \boldsymbol{\alpha}^{(1)^k} \mathbf{R}_2(k) &= \sum_{k=0}^{\infty} \boldsymbol{\alpha}^{(1)^k} \int_{n-1}^{n} \left[\beta_i \frac{(n-t)^{k+1}}{(k+1)!} L_{i,t}^2 d\Lambda_{i,t} \right]_{i=1,\cdots,N}^{\top} \\
&= \boldsymbol{\varrho}_2 \boldsymbol{\beta} \boldsymbol{\omega}_L \boldsymbol{\lambda} + \boldsymbol{\varepsilon}_n^J, \\
\sum_{k=0}^{\infty} \boldsymbol{\alpha}^{(1)^k} \mathbf{R}_3(k) &= \boldsymbol{\varrho}_2 \boldsymbol{\omega} + (\boldsymbol{\varrho}_2 - 2\boldsymbol{\varrho}_3)\mathbf{V}\mathbf{1}_N \quad \text{and} \\
\sum_{k=0}^{\infty} \boldsymbol{\alpha}^{(1)^k} \mathbf{R}_4(k) &= (\boldsymbol{\varrho}_1 - \boldsymbol{\varrho}_2)\boldsymbol{\sigma}_{n-1}^2,
\end{aligned} \tag{A.8}$$

其中

$$\boldsymbol{\varepsilon}_n^J = \boldsymbol{\beta}\boldsymbol{\alpha}^{(1)^{-1}} \left\{ \int_{n-1}^{n} \left(\mathbf{e}^{\boldsymbol{\alpha}^{(1)}(n-t)} - \mathbf{I}_N \right) \left[\mathbf{M}_t d\boldsymbol{\Lambda}_t + \boldsymbol{\omega}_L (d\boldsymbol{\Lambda}_t - \boldsymbol{\lambda} dt) \right] \right\}.$$

注意，在整数时间点 n ， $\boldsymbol{\sigma}_t^2$ 具有如下形式

$$\boldsymbol{\sigma}_n^2 = \boldsymbol{\omega} + \sum_{l=1}^{P} \boldsymbol{\alpha}^{(l)} \mathbf{y}_{n-l+1} + \boldsymbol{\beta}\mathbf{J}_n,$$

于是我们得到式（2.8）中的自回归形式。采用与命题1类似的方法，我们进一步证明，$\{\boldsymbol{\varepsilon}_n^c\}$ 和 $\{\boldsymbol{\varepsilon}_n^J\}$ 均为 *i.i.d.* 序列，且具有零均值和有限方差矩阵。

□

A.2 定理1 和推论1在第3节中的证明

定理*1*的证明。该定理的证明分为两步：首先，我们推导不施加低秩约束时最小二乘估计量的渐近正态性；然后，借鉴^中Shapiro （1986年）的证明技巧，以建立 $\boldsymbol{\mathcal{A}}_{\text{MLR}}$ 的渐近正态性。

第一步考虑模型（3.6）在不施加任何低秩约束时的最小二乘估计量， $\widehat{\boldsymbol{\mathcal{A}}}_{\text{OLS}} = \operatorname{argmin} T^{-1} \sum_{n=P+1}^{T} ||\tilde{}\mathbf{y}_n - \boldsymbol{\mathcal{A}}_{(1)} \tilde{}\mathbf{x}_n||_2^2$，并且可以验证

$$(\widehat{\boldsymbol{\mathcal{A}}}_{\text{OLS}})_{(1)} = \sum_{n=P+1}^{T} \tilde{\mathbf{y}}_n \tilde{\mathbf{x}}_n^{\top} \left(\sum_{n=P+1}^{T} \tilde{\mathbf{x}}_n \tilde{\mathbf{x}}_n^{\top} \right)^{-1} = \boldsymbol{\mathcal{A}}_{(1)} + \sum_{n=P+1}^{T} \boldsymbol{\epsilon}_n \tilde{\mathbf{x}}_n^{\top} \left(\sum_{n=P+1}^{T} \tilde{\mathbf{x}}_n \tilde{\mathbf{x}}_n^{\top} \right)^{-1}, \tag{A.9}$$

其中 $\tilde{\mathbf{x}}_n = (\tilde{\mathbf{y}}_{n-1}^\top, \ldots\ldots, \tilde{\mathbf{y}}_{n-P}^\top)^\top$，$\boldsymbol{\epsilon}_n = \boldsymbol{\eta}_n - \sum_{j=1}^P \mathbf{A}^j \boldsymbol{\eta}_{n-j} + \boldsymbol{\varepsilon}_n$，且 $\boldsymbol{\eta}_n = \tilde{\mathbf{y}}_n - \mathbf{y}_n$。此外，由（3.3），可得

$$\mathbb{P}\left\{\max_{1\leqslant n\leqslant T}\max_{1\leqslant i\leqslant N}|\eta_{i,n}| \geqslant \frac{(NT)^{1/2+\delta/4}}{m^{1/4}}\right\} \leqslant \frac{C_1}{(NT)^{\delta/2}}, \tag{A.10}$$

其中 C_1 是一个正常数，$\delta > 0$ 在定理中给出。为简单起见，令 $\eta_{\max} = \max_{1\leqslant n\leqslant T}\max_{1\leqslant i\leqslant N}|\eta_{i,n}|$，则 $\eta_{\max} = O_p(T^{1/2+\delta/4}m^{-1/4})$。因此，
$T^{1/2}\eta_{\max} = O_p(T^{1+\delta/4}m^{-1/4}) = o_p(1)$，因为 $T^{4+\delta}m^{-1} \to 0$。

我们首先处理 $T^{-1}\sum_{n=P+1}^T \tilde{\mathbf{x}}_n\tilde{\mathbf{x}}_n^\top$在（A.9）右侧的项，可以得到

$$\begin{aligned}\frac{1}{T}\sum_{n=P+1}^T \widetilde{\mathbf{x}}_n\widetilde{\mathbf{x}}_n^\top =& \frac{1}{T}\sum_{n=P+1}^T \mathbf{x}_n\mathbf{x}_n^\top + \frac{1}{T}\sum_{n=P+1}^T (\widetilde{\mathbf{x}}_n - \mathbf{x}_n)(\widetilde{\mathbf{x}}_n - \mathbf{x}_n)^\top \\ &+ \frac{1}{T}\sum_{n=P+1}^T \mathbf{x}_n(\widetilde{\mathbf{x}}_n - \mathbf{x}_n)^\top + \frac{1}{T}\sum_{n=P+1}^T (\widetilde{\mathbf{x}}_n - \mathbf{x}_n)\mathbf{x}_n^\top,\end{aligned} \tag{A.11}$$

其中$\mathbf{x}_n = (\mathbf{y}_{n-1}^\top, \ldots, \mathbf{y}_{n-P}^\top)^\top$。根据假设1，由 $\mathbb{E}\|\boldsymbol{\varepsilon}_n\|^4 < \infty$这一事实以及遍历定理，可以得到

$$\frac{1}{T}\sum_{n=P+1}^T \mathbf{x}_n\mathbf{x}_n^\top = \boldsymbol{\Gamma}^* + o_p(1) \quad \text{and} \quad \frac{1}{T}\sum_{n=P+1}^T |\mathbf{x}_n| = \mathbb{E}(|\mathbf{x}_n|) + o_p(1), \tag{A.12}$$

其中，$|\cdot|$表示逐元素取绝对值。此外，可以验证 $|\tilde{\mathbf{x}}_n - \mathbf{x}_n| \leqslant \eta_{\max}\mathbf{1}_{NP}$，其中 $\mathbf{1}_{NP}$是一个 (NP)维全1向量。因此，

$$\begin{aligned}&\left|\frac{1}{T}\sum_{n=P+1}^T (\widetilde{\mathbf{x}}_n - \mathbf{x}_n)(\widetilde{\mathbf{x}}_n - \mathbf{x}_n)^\top\right| \leqslant \eta_{\max}^2\mathbf{1}_{NP}\mathbf{1}_{NP}^\top \quad \text{and} \\ &\left|\frac{1}{T}\sum_{n=P+1}^T \mathbf{x}_n(\widetilde{\mathbf{x}}_n - \mathbf{x}_n)^\top\right| \leqslant \eta_{\max}\left(\frac{1}{T}\sum_{n=P+1}^T |\mathbf{x}_n|\right)\mathbf{1}_{NP}^\top,\end{aligned}$$

结合（A.11）、（A.12）以及 $\eta_{\max} = o_p(1)$这一事实，可得

$$\frac{1}{T}\sum_{n=P+1}^T \widetilde{\mathbf{x}}_n\widetilde{\mathbf{x}}_n^\top = \boldsymbol{\Gamma}^* + o_p(1). \tag{A.13}$$

接下来，我们处理 $T^{-1/2}\sum_{n=P+1}^T \boldsymbol{\epsilon}_n\tilde{\mathbf{x}}_n^\top$在（A.9）右侧的项，可以得到

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{n=P+1}^T \boldsymbol{\epsilon}_n\widetilde{\mathbf{x}}_n^\top =& \frac{1}{\sqrt{T}}\sum_{n=P+1}^T \boldsymbol{\varepsilon}_n\mathbf{x}_n^\top + \frac{1}{\sqrt{T}}\sum_{n=P+1}^T (\boldsymbol{\epsilon}_n - \boldsymbol{\varepsilon}_n)(\widetilde{\mathbf{x}}_n - \mathbf{x}_n)^\top \\ &+ \frac{1}{\sqrt{T}}\sum_{n=P+1}^T \boldsymbol{\varepsilon}_n(\widetilde{\mathbf{x}}_n - \mathbf{x}_n)^\top + \frac{1}{\sqrt{T}}\sum_{n=P+1}^T (\boldsymbol{\epsilon}_n - \boldsymbol{\varepsilon}_n)\mathbf{x}_n^\top.\end{aligned} \tag{A.14}$$

根据假设1，存在一个正常数 C 使得 $(\mathbf{I}_N + \sum_{j=1}^{P} |\mathbf{A}_j|)\mathbf{1}_N \leqslant C \cdot \mathbf{1}_N$ 进而 $|\boldsymbol{\epsilon}_n - \boldsymbol{\varepsilon}_n| = |\boldsymbol{\eta}_n - \sum_{j=1}^{P} \mathbf{A}_j \boldsymbol{\eta}_{n-j}| \leqslant C\eta_{\max}\mathbf{1}_N$。因此，

$$\left|\frac{1}{\sqrt{T}} \sum_{n=P+1}^{T} (\boldsymbol{\epsilon}_n - \boldsymbol{\varepsilon}_n)(\widetilde{\mathbf{x}}_n - \mathbf{x}_n)^\top\right| \leqslant T^{1/2}\eta_{\max}^2 \mathbf{1}_{NP}\mathbf{1}_{NP}^\top,$$

$$\left|\frac{1}{\sqrt{T}} \sum_{n=P+1}^{T} \boldsymbol{\varepsilon}_n(\widetilde{\mathbf{x}}_n - \mathbf{x}_n)^\top\right| \leqslant T^{1/2}\eta_{\max} \left(\frac{1}{T} \sum_{n=P+1}^{T} |\boldsymbol{\varepsilon}_n|\right) \mathbf{1}_{NP}^\top, \text{ and}$$

$$\left|\frac{1}{\sqrt{T}} \sum_{n=P+1}^{T} (\boldsymbol{\epsilon}_n - \boldsymbol{\varepsilon}_n)\mathbf{x}_n^\top\right| \leqslant CT^{1/2}\eta_{\max}\mathbf{1}_{NP} \left(\frac{1}{T} \sum_{n=P+1}^{T} |\mathbf{x}_n^\top|\right).$$

它们与（A.14）、（A.12）以及以下事实：$T^{-1}\sum_{n=P+1}^{T} |\boldsymbol{\varepsilon}_n| = \mathbb{E}(|\boldsymbol{\varepsilon}_n|) + o_p(1)$和 $T^{1/2}\eta_{\max} = o_p(1)$, 共同得到

$$\frac{1}{\sqrt{T}} \sum_{n=P+1}^{T} \boldsymbol{\epsilon}_n \widetilde{\mathbf{x}}_n^\top = \frac{1}{\sqrt{T}} \sum_{n=P+1}^{T} \boldsymbol{\varepsilon}_n \mathbf{x}_n^\top + o_p(1). \tag{A.15}$$

结合（A.9）、（A.13）、（A.15）以及鞅差序列的中心极限定理，可得

$$\sqrt{T}\{\mathrm{vec}((\widehat{\boldsymbol{\mathcal{A}}}_{\mathrm{OLS}})_{(1)}) - \mathrm{vec}(\boldsymbol{\mathcal{A}}_{(1)})\} \to \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathrm{OLS}})$$

依分布收敛，其中 $m \to \infty$ 和 $T \to \infty$，且 $\boldsymbol{\Sigma}_{\mathrm{OLS}} = \boldsymbol{\Gamma}^{*-1} \otimes \boldsymbol{\Sigma}_{\varepsilon}$。

第二步主要依据 Shapiro (1986) 中针对过度参数化模型的命题 4.1。设 $\boldsymbol{\varphi} = \Big(\mathrm{vec}(\mathcal{G}_{(1)})^\top$、$\mathrm{vec}(\mathrm{U}_1)^\top$、$\mathrm{vec}(\mathrm{U}_2)^\top$、$\mathrm{vec}(\mathrm{U}_3)^\top\Big)^\top$ 为 Tucker分解中的分量参数，且 $\boldsymbol{h}(\boldsymbol{\varphi}) = \mathrm{vec}(\boldsymbol{\mathcal{A}}_{(1)}) = \mathrm{vec}(\mathrm{U}_1\mathcal{G}_{(1)}(\mathrm{U}_3 \otimes \mathrm{U}_2)^\top)\widehat{\ }$ $\widehat{\ }$是 $\boldsymbol{\varphi}$的函数。记 $\boldsymbol{h}_{\mathrm{OLS}} = \mathrm{vec}((\boldsymbol{\mathcal{A}}_{\mathrm{OLS}})_{(1)})$，其渐近正态性已在第一步中得到证明。与 Shapiro (1986) 中的方法相同，对任意 $\boldsymbol{h}$，我们构造如下差异函数，

$$F(\boldsymbol{h}, \widehat{\boldsymbol{h}}_{\mathrm{OLS}}) = \sum_{n=P+1}^{T} ||\widetilde{\mathbf{y}}_n - (\widetilde{\mathbf{x}}_n^\top \otimes \mathbf{I}_N)\boldsymbol{h}||_2^2 - \sum_{n=P+1}^{T} ||\widetilde{\mathbf{y}}_n - (\widetilde{\mathbf{x}}_n^\top \otimes \mathbf{I}_N)\widehat{\boldsymbol{h}}_{\mathrm{OLS}}||_2^2,$$

这是一个非负且二阶连续可微的函数。此外，当且仅当 $\boldsymbol{h} = \boldsymbol{h}_{\mathrm{OLS}}$ 时，它等于零。

令$\mathbf{T}_{ij}(N,\ N,\ P) \in \mathbb{R}^{N^2P \times N^2P}$为张量矩阵化变换算子，该算子满足$\mathrm{vec}(\boldsymbol{\mathcal{A}}_{(j)}) = \mathbf{T}_{ij}(N,\ N,\ P)\mathrm{vec}(\boldsymbol{\mathcal{A}}_{(\mathrm{i})})$，对任意张量 $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{N \times N \times P}$均成立。注意，$\mathbf{T}_{ij}(N,\ N,\ P)$是一个满秩矩阵，表示张量 $\boldsymbol{\mathcal{A}}$的

vec($\mathcal{A}_{(j)}$)中每个元素的对应位置，并且可以视为矩阵转置的置换矩阵的自然扩展。尽管$\mathbf{T}_{ij}(N,\ N,\ P)$仅取决于 N和 P的值，但由于 N 和 P 在该定理中均为固定值，我们可以将其简写为$\mathbf{T}^{ij}$ 。因此，

$$\begin{aligned}\operatorname{vec}(\mathcal{A}_{(1)}) =&\operatorname{vec}(\mathbf{U}_1\mathcal{G}_{(1)}(\mathbf{U}_3\otimes\mathbf{U}_2)^\top) = \mathbf{T}_{21}\operatorname{vec}(\mathbf{U}_2\mathcal{G}_{(2)}(\mathbf{U}_3\otimes\mathbf{U}_1)^\top)\\ =&\mathbf{T}_{31}\operatorname{vec}(\mathbf{U}_3\mathcal{G}_{(3)}(\mathbf{U}_2\otimes\mathbf{U}_1)^\top),\end{aligned}$$

以及 $\boldsymbol{h}$ 的雅可比矩阵具有以下形式

$$\begin{aligned}\mathbf{H} = \partial\boldsymbol{h}(\boldsymbol{\varphi})/\partial\boldsymbol{\varphi} = \big((\mathbf{U}_3\otimes\mathbf{U}_2\otimes\mathbf{U}_1), [(\mathbf{U}_3\otimes\mathbf{U}_2)\mathcal{G}_{(1)}^\top]\otimes\mathbf{I}_N,\\ \mathbf{T}_{21}\left\{\left[(\mathbf{U}_3\otimes\mathbf{U}_1)\mathcal{G}_{(2)}^\top\right]\otimes\mathbf{I}_N\right\}, \mathbf{T}_{31}\left\{\left[(\mathbf{U}_2\otimes\mathbf{U}_1)\mathcal{G}_{(3)}^\top\right]\otimes\mathbf{I}_P\right\}\big).\end{aligned}$$

记 $\boldsymbol{h}(\widehat{\boldsymbol{\varphi}}_{\mathrm{MLR}})$为 $F(\cdot$的极小化解$\boldsymbol{h}_{\mathrm{OLS}})$，它对应于 MLR-HAR 估计量，即 $\boldsymbol{h}(\widehat{\boldsymbol{\varphi}}_{\mathrm{MLR}}) = \operatorname{vec}((\mathcal{A}_{\mathrm{MLR}})_{(1)})$。根据 Shapiro(1986年）中的命题 4.1，我们可以得到如下渐近正态性结果:

$$\sqrt{T}\{\boldsymbol{h}(\widehat{\boldsymbol{\varphi}}_{\mathrm{MLR}}) - \boldsymbol{h}\} \to N(\mathbf{0}, \boldsymbol{\Sigma}_{\mathrm{MLR}})$$

依分布收敛于 $m\to\infty$ 和 $T\to\infty$时，其中 $\boldsymbol{\Sigma}_{\mathrm{MLR}} = \mathrm{P}\boldsymbol{\Sigma}_{\mathrm{OLS}}\mathrm{P}^\top$，$\mathrm{P} = \mathrm{H}(\mathrm{H}^\top\mathrm{J}\mathrm{H})^\dagger\mathrm{H}^\top\mathrm{J}$ 为投影矩阵，$\mathrm{J} = \boldsymbol{\Gamma}^*\otimes\boldsymbol{\Sigma}_\varepsilon^{-1}$ 是 $\boldsymbol{h}$的费 Fisher 信息矩阵，$\dagger$ 表示 Moore-Penrose 逆。由于 $\boldsymbol{\Sigma}_{\mathrm{OLS}} = \mathrm{J}^{-1}$，我们可以容易地得到$\boldsymbol{\Sigma}_{\mathrm{MLR}}$ =$\mathrm{H}(\mathrm{H}^\top\mathrm{J}\mathrm{H})^\dagger\mathrm{H}^\top$。至此，定理1的证明完成。

□

推论 *1* 的证明。关于 $\boldsymbol{\mathcal{A}}_{\mathrm{OLS}}$ 的渐近正态性已在定理 1证明的第一步中得到证明。注意，$\boldsymbol{\Sigma}_{\mathrm{OLS}} = \mathbf{J}^{-1}$ 和 $\boldsymbol{\Sigma}_{\mathrm{MLR}}$ =$\mathbf{P}\mathbf{J}^{-1}\mathbf{P}^\top$，其中$\mathbf{P}$ =$\mathbf{H}(\mathbf{H}^\top\mathbf{J}\mathbf{H})^\dagger\mathbf{H}^\top\mathbf{J}$ 是投影矩阵。因此，

$$\boldsymbol{\Sigma}_{\mathrm{OLS}} - \boldsymbol{\Sigma}_{\mathrm{MLR}} = \mathbf{J}^{-1} - \mathbf{H}(\mathbf{H}^\top\mathbf{J}\mathbf{H})^\dagger\mathbf{H}^\top = \mathbf{J}^{-1/2}\mathbf{Q}_{\mathbf{J}^{1/2}\mathbf{H}}\mathbf{J}^{-1/2}, \tag{A.16}$$

其中 $\mathbf{Q}_{\mathbf{J}^{1/2}\mathbf{H}}$ =$\mathbf{I}$ −$\mathbf{J}^{1/2}\mathbf{H}(\mathbf{H}^\top\mathbf{J}\mathbf{H})^\dagger\mathbf{H}^\top\mathbf{J}^{1/2}$ 是投影到 span $(\mathbf{J}^{1/2}\mathbf{H})$的正交补上的投影矩阵，因此是半正定的。于是，$\boldsymbol{\Sigma}_{\mathrm{OLS}} \geqslant \boldsymbol{\Sigma}_{\mathrm{MLR}}$。

对于 MRI 估计量，等价于限制系数张量具有如下形式的 Tucker分解：$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{H}}\times_2\mathbf{U}_2$，其中秩 r_2 已知，并采用定理1 中的证明技术来完成相应的证明。

令 $\boldsymbol{\theta} = \left(\text{vec}(\mathcal{H}_{(1)})^\top、\text{vec}(\mathbf{U}_2)^\top\right)^\top$ 和 $\boldsymbol{h}(\boldsymbol{\theta}) = \text{vec}(\boldsymbol{\mathcal{A}}_{(1)}) = \text{vec}(\mathcal{H}_{(1)}(\mathbf{I}_P \otimes \mathbf{U}_2)^\top)\widehat{}$ 为 $\boldsymbol{\theta}$ 的函数。类似于 $\boldsymbol{\mathcal{A}}_{\text{MLR}}$ 的情形，我们定义如下雅可比矩阵$\mathbf{R}:= \partial \boldsymbol{h}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$

$$\mathbf{R} = \left(\mathbf{I}_P \otimes \mathbf{U}_2 \otimes \mathbf{I}_N, \mathbf{T}_{21}\left\{\left[(\mathbf{I}_P \otimes \mathbf{I}_N)\mathcal{H}_{(2)}^\top\right] \otimes \mathbf{I}_N\right\}\right) \in \mathbb{R}^{N^2P \times (NPr_2 + Nr_2)}.$$

记 $\boldsymbol{\Sigma}_{\text{MRI}} = \mathbf{R}\left(\mathbf{R}^\top \mathbf{J} \mathbf{R}\right)^\dagger \mathbf{R}^\top$。与 $\widehat{\boldsymbol{\mathcal{A}}}_{\text{MLR}}$ 的证明类似，我们可以证明$\sqrt{T}\{\text{vec}((\widehat{\boldsymbol{\mathcal{A}}}_{\text{MRI}})_{(1)}) - \text{vec}(\boldsymbol{\mathcal{A}}_{(1)})\} \to N(\mathbf{0}, \boldsymbol{\Sigma}_{\text{MRI}})$ 依分布收敛于 $m \to \infty$ 和 $T \to \infty$，且 $\boldsymbol{\Sigma}_{\text{MRI}} \leqslant \mathbf{J}^{-1} = \boldsymbol{\Sigma}_{\text{OLS}}$。

最后，由于 $\mathbf{U}_2$ 在 Tucker 分解中与 $\boldsymbol{\mathcal{A}}_{(2)}$的奇异值分解中的左奇异向量完全相同，因此我们可以将 Tucker 分解视为对矩阵 $\mathcal{H}_{(1)}$的进一步分解，即 $\boldsymbol{\mathcal{A}} = \mathcal{H} \times_2 \mathbf{U}_2$ 和 $\mathcal{H} = \mathcal{G} \times_1 \mathbf{U}_1 \times_3 \mathbf{U}_3$。因此，$\mathbf{H} = \partial \boldsymbol{h}(\boldsymbol{\theta})/\partial \boldsymbol{\varphi} = \partial \boldsymbol{h}(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \cdot \partial \boldsymbol{\theta}(\boldsymbol{\varphi})/\partial \boldsymbol{\varphi} = \mathbf{R} \cdot \partial \boldsymbol{\theta}(\boldsymbol{\varphi})/\partial \boldsymbol{\varphi}$。通过类似于（A.16）的方法，我们可以证明 $\boldsymbol{\Sigma}_{\text{MLR}} \leqslant \boldsymbol{\Sigma}_{\text{MIR}}$，因为 $\text{span}\left(\mathbf{J}^{1/2}\mathbf{H}\right) \subset \text{span}\left(\mathbf{J}^{1/2}\mathbf{R}\right)$。

□

A.3 定理2 和3 第4节的证明

本小节给出第2 和3 4节中定理的技术性证明。同时，我们还给出四个辅助引理。引理A.1 建立了低多线性秩张量的覆盖数与离散化，引理A.2 推导出受限强凸性（RSC）和受限光滑性（RSM），引理A.3推导出偏差界；这些引理将在定理2 和3的证明中使用。引理A.4 推导出压缩投影性质（CPP），该性质用于定理3的证明中。在本小节中，我们将用 C 表示任意正数，其值可能随行变化。

$\widehat{}$定理2的证明。为简便起见，将多线性低秩估计量 $\boldsymbol{\mathcal{A}}_{\text{MLR}}$ 记为$\widehat{\boldsymbol{\mathcal{A}}}$，并令 $\boldsymbol{\Delta} = \widehat{\boldsymbol{\mathcal{A}}} - \boldsymbol{\mathcal{A}}$，其中 $\boldsymbol{\mathcal{A}}$ 为真实参数张量。损失函数的形式为

$$L(\boldsymbol{\mathcal{A}}) = \frac{1}{T}\sum_{n=P+1}^{T} ||\widetilde{\mathbf{y}}_n - \boldsymbol{\mathcal{A}}_{(1)}\widetilde{\mathbf{x}}_n||_2^2,$$

其中 $\tilde{\mathbf{x}}_n = (\tilde{\mathbf{y}}_{n-1}^\top, \ldots\ldots, \tilde{\mathbf{y}}_{n-P}^\top)^\top \in \mathbb{R}^{NP\times 1}$。由于 $\widehat{\boldsymbol{\mathcal{A}}}$ 的最优性，有

$$\frac{1}{T}\sum_{n=P+1}^{T}\|\tilde{\mathbf{y}}_n - \widehat{\boldsymbol{\mathcal{A}}}_{(1)}\tilde{\mathbf{x}}_n\|_2^2 \leqslant \frac{1}{T}\sum_{n=P+1}^{T}\|\tilde{\mathbf{y}}_n - \boldsymbol{\mathcal{A}}_{(1)}\tilde{\mathbf{x}}_n\|_2^2,$$

这意味着

$$\frac{1}{T}\sum_{n=P+1}^{T}\|\boldsymbol{\Delta}_{(1)}\tilde{\mathbf{x}}_n\|_2^2 \leqslant \frac{2}{T}\sum_{n=P+1}^{T}\langle\boldsymbol{\epsilon}_n, \boldsymbol{\Delta}_{(1)}\tilde{\mathbf{x}}_n\rangle \leqslant 2\langle\frac{1}{T}\sum_{n=P+1}^{T}\boldsymbol{\epsilon}_n\circ\tilde{\mathbf{X}}_n, \boldsymbol{\Delta}\rangle, \tag{A.17}$$

其中 $\tilde{\mathbf{X}}_n = (\tilde{\mathbf{y}}_{n-1}, \ldots\ldots, \tilde{\mathbf{y}}_{n-P}) \in \mathbb{R}^{N\times P}$、$\sum_{n=P+1}^{T}\langle\boldsymbol{\epsilon}_n, \boldsymbol{\Delta}_{(1)}\tilde{\mathbf{x}}_n\rangle = \langle\sum_{n=P+1}^{T}\boldsymbol{\epsilon}_n\circ\tilde{\mathbf{X}}_n, \boldsymbol{\Delta}\rangle$，而$\circ$表示外积。

记张量集合为

$$\mathcal{S}(r_1, r_2, r_3) = \{\boldsymbol{\mathcal{A}} \in \mathbb{R}^{N\times N\times P} : \|\boldsymbol{\mathcal{A}}\|_{\mathrm{F}} = 1, \mathrm{rank}_i(\boldsymbol{\mathcal{A}}_i) \leqslant r_i, 1 \leqslant i \leqslant 3\}.$$

注意到，$\boldsymbol{\mathcal{A}}$ 和 $\widehat{\boldsymbol{\mathcal{A}}}$ 的 Tucker 秩均为 (r_1, r_2, r_3)，因此 $\boldsymbol{\Delta}$ 的 Tucker 秩至多为 $(2r_1, 2r_2, 2r_3)$。因此，由式（A.17），

$$\frac{1}{T}\sum_{n=P+1}^{T}\|\boldsymbol{\Delta}_{(1)}\tilde{\mathbf{x}}_n\|_2^2 \leqslant 2\|\boldsymbol{\Delta}\|_{\mathrm{F}}\sup_{\boldsymbol{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)}\langle\frac{1}{T}\sum_{n=P+1}^{T}\boldsymbol{\epsilon}_n\circ\tilde{\mathbf{X}}_n, \boldsymbol{\Delta}\rangle,$$

因此，我们可以应用引理A.2 和引理A.3 推导出估计误差界，其中 δ 为 $\delta+1/2$，以简化记号。利用上述不等式、估计误差界和引理A.3也可以得到预测误差界。

□

定理的证明 3。对于固定的 $1 \leqslant k \leqslant K$，定义一个线性空间，

$$\mathcal{A} = \{\alpha_1\widehat{\boldsymbol{\mathcal{A}}}_k + \alpha_2\boldsymbol{\mathcal{A}}, \ \alpha_1, \alpha_2 \in \mathbb{R}\},$$

并将 $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{N\times N\times P}$ 在空间 $\mathcal{A}$ 上的投影记为 $(\boldsymbol{\mathcal{B}})_{\mathcal{A}}$。为简洁起见，省略 $\mathcal{A}$ 对 k 的依赖关系。由于 $\boldsymbol{\mathcal{A}} \in \boldsymbol{\Theta}(r_1, r_2, r_3)$ 和 $\boldsymbol{\mathcal{A}}_k \in \boldsymbol{\Theta}(r_1', r_1', r_3')$，因此有 $\mathcal{A} \subset \boldsymbol{\Theta}(r_1 + r_1', r_2 + r_2', r_3 + r_3')$。

注意，$\widetilde{\boldsymbol{\mathcal{A}}}_k = \widehat{\boldsymbol{\mathcal{A}}}_{k-1} - \eta\nabla L(\widehat{\boldsymbol{\mathcal{A}}}_{k-1})$，且 $r_i' \geqslant \left(\sqrt[3]{1 + \frac{\kappa_L}{24\kappa_U}} - 1\right)^{-2} r_i$ 其中 $1 \leqslant i \leqslant 3$。由引理A.4，可得

$$\|\widehat{\boldsymbol{\mathcal{A}}}_k - (\widetilde{\boldsymbol{\mathcal{A}}}_k)_{\mathcal{A}}\|_{\mathrm{F}} \leqslant [\prod_{i=1}^{3}(\sqrt{\frac{r_i}{r_i'}} + 1) - 1]\|\boldsymbol{\mathcal{A}} - (\widetilde{\boldsymbol{\mathcal{A}}}_k)_{\mathcal{A}}\|_{\mathrm{F}} \leqslant \frac{\kappa_L}{24\kappa_U}\|\boldsymbol{\mathcal{A}} - (\widetilde{\boldsymbol{\mathcal{A}}}_k)_{\mathcal{A}}\|_{\mathrm{F}},$$

这与 $1<1+\frac{\kappa_L}{24\kappa_U}<2$这一事实相结合，意味着

$$\begin{aligned}\|\widehat{\boldsymbol{\mathcal{A}}}_k-\boldsymbol{\mathcal{A}}\|_{\mathrm{F}} \leqslant & \|\widehat{\boldsymbol{\mathcal{A}}}_k-(\widetilde{\boldsymbol{\mathcal{A}}}_k)_{\mathcal{A}}\|_{\mathrm{F}}+\|\boldsymbol{\mathcal{A}}-(\widetilde{\boldsymbol{\mathcal{A}}}_k)_{\mathcal{A}}\|_{\mathrm{F}} \leqslant(1+\frac{\kappa_L}{24\kappa_U})\|\boldsymbol{\mathcal{A}}-(\widetilde{\boldsymbol{\mathcal{A}}}_k)_{\mathcal{A}}\|_{\mathrm{F}} \\ \leqslant & (1+\frac{\kappa_L}{24\kappa_U})\|(\boldsymbol{\mathcal{A}}-\widehat{\boldsymbol{\mathcal{A}}}_{k-1}-\eta[\nabla L(\boldsymbol{\mathcal{A}})-\nabla L(\widehat{\boldsymbol{\mathcal{A}}}_{k-1})])_{\mathcal{A}}\|_{\mathrm{F}}+2\eta\|(\nabla L(\boldsymbol{\mathcal{A}}))_{\mathcal{A}}\|_{\mathrm{F}} \\ := & A_1+A_2. \end{aligned} \tag{A.18}$$

我们首先处理 A_1这一项。令$\mathbf{H}=T^{-1}\sum_{n=P+1}^{T}(\tilde{}\mathbf{x}_n\tilde{}\mathbf{x}_n^{\top}\otimes\mathbf{I}_N)$，并且有 $T^{-1}\sum_{n=P+1}^{T}\|\boldsymbol{\Delta}_{(1)}\tilde{}\mathbf{x}_n\|_2^2=T^{-1}\sum_{n=P+1}^{T}\|(\tilde{}\mathbf{x}_n^{\top}\otimes\mathbf{I}_N)\mathrm{vec}(\boldsymbol{\Delta})\|_2^2=\mathrm{vec}(\boldsymbol{\Delta})^{\top}\mathbf{H}\mathrm{vec}(\boldsymbol{\Delta})$。于是，根据引理A.2，对所有 $\boldsymbol{\Delta}\in\boldsymbol{\Theta}(r_1+r_1',\ r_2+r_2',\ r_3+r_3')$,

$$\frac{1}{8}\kappa_L\|\boldsymbol{\Delta}\|_{\mathrm{F}}^2\leqslant\mathrm{vec}(\boldsymbol{\Delta})^{\top}\mathbf{H}\mathrm{vec}(\boldsymbol{\Delta})\leqslant\frac{8}{3}\kappa_{\mathrm{U}}\|\boldsymbol{\Delta}\|_{\mathrm{F}}^2$$

其概率至少为

$$1-2\exp(-CT(\kappa_L/\kappa_U)^2\min\{\kappa^{-2},\kappa^{-4}\})-\exp(-Cd'_{\mathcal{M}}). \tag{A.19}$$

注意，$\eta=2/(3\kappa_U)$, 并且$\mathbf{H}$ 是损失函数 $L(\mathcal{B})$ 关于$\hat{}$ $\hat{}\mathrm{vec}(\mathcal{B})$的海森矩阵。有 $\mathrm{vec}(\nabla L(\boldsymbol{\mathcal{A}})-\nabla L(\boldsymbol{\mathcal{A}}_{k-1}))=\mathbf{H}\mathrm{vec}(\boldsymbol{\mathcal{A}}-\boldsymbol{\mathcal{A}}_{k-1})$，并且

$$\begin{aligned}A_1 &= (1+\frac{\kappa_L}{24\kappa_U})\|((\mathbf{I}-\eta\mathbf{H})\mathrm{vec}(\boldsymbol{\mathcal{A}}-\widehat{\boldsymbol{\mathcal{A}}}_{k-1}))_{\mathcal{A}}\|_2 \\ &\leqslant (1+\frac{\kappa_L}{24\kappa_U})(1-\frac{\kappa_L}{12\kappa_U})\|\widehat{\boldsymbol{\mathcal{A}}}_{k-1}-\boldsymbol{\mathcal{A}}\|_{\mathrm{F}} \\ &\leqslant (1-\frac{\kappa_L}{24\kappa_U})\|\widehat{\boldsymbol{\mathcal{A}}}_{k-1}-\boldsymbol{\mathcal{A}}\|_{\mathrm{F}}, \end{aligned} \tag{A.20}$$

以（A.19）处的概率成立，其中 $(\mathrm{vec}(\mathcal{B}))_{\mathcal{A}}=(\mathcal{B})_{\mathcal{A}}$，第一个不等式由Chen 等人（2019）中的引理 4 得出。

接下来处理 A_2 这一项，并根据Chen 等人（2019）中的引理 5,

$$A_2\leqslant\frac{32}{\kappa_L}\|(\nabla L(\boldsymbol{\mathcal{A}}))_{\mathcal{A}}\|_{\mathrm{F}}=\frac{32}{\kappa_L}\sup_{\boldsymbol{\mathcal{S}}\in\mathcal{A},\|\boldsymbol{\mathcal{S}}\|_{\mathrm{F}}=1}\langle\nabla L(\boldsymbol{\mathcal{A}}),\boldsymbol{\mathcal{S}}\rangle\leqslant\frac{32}{\kappa_L}\xi,$$

其中 $\mathcal{S}(r_1+r_1',\ r_2+r_2',\ r_3+r_3')=\boldsymbol{\Theta}(r_1+r_1',\ r_2+r_2',\ r_3+r_3')\bigcap\{\|\boldsymbol{\mathcal{S}}\|_{\mathrm{F}}=1\}$ 和

$$\begin{aligned}\xi &= \sup_{\boldsymbol{\mathcal{S}}\in\mathcal{S}(r_1+r_1',r_2+r_2',r_3+r_3')}\langle\nabla L(\boldsymbol{\mathcal{A}}),\boldsymbol{\mathcal{S}}\rangle=\sup_{\boldsymbol{\mathcal{S}}\in\mathcal{S}(r_1+r_1',r_2+r_2',r_3+r_3')}\left\langle\frac{2}{T}\sum_{n=P+1}^{T}\boldsymbol{\epsilon}_n\circ\tilde{\mathbf{X}}_n,\boldsymbol{\mathcal{S}}\right\rangle \\ &\leqslant C\left[(\kappa^2\sqrt{\lambda_{\max}(\boldsymbol{\Sigma}_\varepsilon)\kappa_U}+\kappa\sqrt{\kappa_U})\sqrt{\frac{d'_{\mathcal{M}}}{T}}+\frac{T^{1+2\delta}}{m^{1/4}}\right]\end{aligned}$$

以至少 $1-\exp(-Cd'_{\mathcal{M}})-2\exp(-CT(\kappa_L/\kappa_U)^2\min\{\kappa^{-2},\ \kappa^{-4}\})$。结合式（A.18）和（A.20），即可完成证明。

□

引理 A.1。（低多线性秩张量的覆盖数与离散化）。设 $\bar{\mathcal{S}}(r_1,r_2,r_3)$ 是集合 $S(r_1,r_2,r_3):=\{\boldsymbol{\mathcal{A}}\in\mathbb{R}^{N\times N\times P}:\ \|\boldsymbol{\mathcal{A}}\|_{\mathrm{F}}=1,\text{秩}_i(\boldsymbol{\mathcal{A}}_{(i)})\leqslant r_i,1\leqslant i\leqslant 3\}$ 的一个 ϵ-网。

(i) $\mathcal{S}(r_1,r_2,r_3)$ 的基数满足

$$\left|\bar{\mathcal{S}}(r_1,r_2,r_3)\right|\leqslant(12/\epsilon)^{(r_1r_2r_3+Nr_1+Nr_2+Pr_3)}.$$

(ii) 对任意张量 $\boldsymbol{\mathcal{N}}\in\mathbb{R}^{N\times N\times P}$ 和矩阵 $\mathbf{Z}\in\mathbb{R}^{NP\times T}$，有，

$$\sup_{\boldsymbol{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)}\langle\boldsymbol{\mathcal{N}},\boldsymbol{\Delta}\rangle\leqslant(1-2\sqrt{2}\epsilon)^{-1}\max_{\bar{\boldsymbol{\Delta}}\in\bar{\mathcal{S}}(2r_1,2r_2,2r_3)}\left\langle\boldsymbol{\mathcal{N}},\bar{\boldsymbol{\Delta}}\right\rangle,\quad and$$

$$\sup_{\boldsymbol{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)}\left\|\boldsymbol{\Delta}_{(1)}\mathbf{Z}\right\|_{\mathrm{F}}\leqslant(1-2\sqrt{2}\epsilon)^{-1}\max_{\bar{\boldsymbol{\Delta}}\in\bar{\mathcal{S}}(2r_1,2r_2,2r_3)}\left\|\bar{\boldsymbol{\Delta}}_{(1)}\mathbf{Z}\right\|_{\mathrm{F}}.$$

引理 *A.1* 的证明。(i) 中的结果来自 Wang 等人（2020）的引理 A.1，本文仅证明 (ii)。

考虑 $\mathcal{S}(2r_1,2r_2,2r_3)$ 的一个 ϵ-网 $\bar{\mathcal{S}}(2r_1,2r_2,2r_3)$。则对任意张量 $\boldsymbol{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)$，存在一个 $\bar{\boldsymbol{\Delta}}\in\bar{\mathcal{S}}(2r_1,2r_2,2r_3)$，使得 $\|\boldsymbol{\Delta}-\bar{\boldsymbol{\Delta}}\|_{\mathrm{F}}\leqslant\epsilon$。由于 $\boldsymbol{\mathcal{W}}=\boldsymbol{\Delta}-\bar{\boldsymbol{\Delta}}$ 的秩至多为 $(4r_1,4r_2,4r_3)$，我们可以将 $\boldsymbol{\mathcal{W}}$ 的 HOSVD 分为 8 个部分，使得 $\overline{\boldsymbol{\mathcal{W}}}=\sum_{i=1}^{8}\overline{\boldsymbol{\mathcal{W}}}_i$，其中对于 $1\leqslant i\leqslant 8$ 和 $1\leqslant j\leqslant 3$，秩$_j(\overline{\boldsymbol{\mathcal{W}}}_i)\leqslant 2r_j$，并且对任意 $j\neq k$，有 $\langle\overline{\boldsymbol{\mathcal{W}}}_j,\overline{\boldsymbol{\mathcal{W}}}_k\rangle=0$。于是对任意 $\boldsymbol{\mathcal{N}}\in\mathbb{R}^{N\times N\times P}$，我们有

$$\langle\boldsymbol{\mathcal{N}},\boldsymbol{\Delta}\rangle=\left\langle\boldsymbol{\mathcal{N}},\bar{\boldsymbol{\Delta}}\right\rangle+\sum_{i=1}^{8}\langle\boldsymbol{\mathcal{N}},\overline{\boldsymbol{\mathcal{W}}}_i\rangle=\left\langle\boldsymbol{\mathcal{N}},\bar{\boldsymbol{\Delta}}\right\rangle+\sum_{i=1}^{8}\left\langle\boldsymbol{\mathcal{N}},\overline{\boldsymbol{\mathcal{W}}}_i/\|\overline{\boldsymbol{\mathcal{W}}}_i\|_{\mathrm{F}}\right\rangle\left\|\overline{\boldsymbol{\mathcal{W}}}_i\right\|_{\mathrm{F}},\quad(A.21)$$

其中 $\overline{\boldsymbol{\mathcal{W}}}_i/\left\|\overline{\boldsymbol{\mathcal{W}}}_i\right\|_{\mathrm{F}}\in S(2r_1,2r_2,2r_3)$，和 $\left\langle\boldsymbol{\mathcal{N}},\ \overline{\boldsymbol{\mathcal{W}}}_i/\|\overline{\boldsymbol{\mathcal{W}}}_i\|_{\mathrm{F}}\right\rangle\leqslant$上确界$_{\boldsymbol{\Delta}\in S(2r_1,\ 2r_2,\ 2r_3)}\langle\boldsymbol{\mathcal{N}},\ \boldsymbol{\Delta}\rangle$。

注意，$\left\|\overline{\boldsymbol{\mathcal{W}}}\right\|^2{}_{\mathrm{F}}=\sum_{i=1}^{8}\left\|\overline{\boldsymbol{\mathcal{W}}}_i\right\|^2{}_{\mathrm{F}}$，并且有 $\sum_{i=1}^{8}\left\|\overline{\boldsymbol{\mathcal{W}}}_i\right\|_{\mathrm{F}}\leqslant 2\sqrt{2}\|\overline{\boldsymbol{\mathcal{W}}}\|_{\mathrm{F}}\leqslant 2\sqrt{2}\epsilon$，结合（A.21），可得

$$\gamma:=\sup_{\boldsymbol{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)}\langle\boldsymbol{\mathcal{N}},\boldsymbol{\Delta}\rangle\leqslant\max_{\bar{\boldsymbol{\Delta}}\in\bar{\mathcal{S}}(2r_1,2r_2,2r_3)}\left\langle\boldsymbol{\mathcal{N}},\bar{\boldsymbol{\Delta}}\right\rangle+2\sqrt{2}\gamma\epsilon,\quad\text{or}$$

$$\gamma=\sup_{\boldsymbol{\Delta}\in S(2r_1,2r_2,2r_3)}\langle\boldsymbol{\mathcal{N}},\boldsymbol{\Delta}\rangle\leqslant(1-2\sqrt{2}\epsilon)^{-1}\max_{\bar{\boldsymbol{\Delta}}\in\bar{\mathcal{S}}(2r_1,2r_2,2r_3)}\left\langle\boldsymbol{\mathcal{N}},\bar{\boldsymbol{\Delta}}\right\rangle.$$

对于矩阵 $\mathbf{Z} \in \mathbb{R}^{NP\times T}$，有

$$\begin{aligned}\left\|\mathbf{\Delta}_{(1)}\mathbf{Z}\right\|_{\mathrm{F}} &\leqslant \left\|\bar{\mathbf{\Delta}}_{(1)}\mathbf{Z}\right\|_{\mathrm{F}} + \sum_{i=1}^{8}\left\|(\overline{\mathcal{W}}_i)_{(1)}\mathbf{Z}\right\|_{\mathrm{F}} \\ &= \left\|\bar{\mathbf{\Delta}}_{(1)}\mathbf{Z}\right\|_{\mathrm{F}} + \sum_{i=1}^{8}\left\|\overline{\mathcal{W}}_i\right\|_{\mathrm{F}}\left\|(\overline{\mathcal{W}}_i)_{(1)}/\|\overline{\mathcal{W}}_i\|_{\mathrm{F}}\mathbf{Z}\right\|_{\mathrm{F}} \\ &\leqslant \left\|\bar{\mathbf{\Delta}}_{(1)}\mathbf{Z}\right\|_{\mathrm{F}} + \sum_{i=1}^{8}\left\|\overline{\mathcal{W}}_i\right\|_{\mathrm{F}} \sup_{\mathbf{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)}\left\|\mathbf{\Delta}_{(1)}\mathbf{Z}\right\|_{\mathrm{F}} \\ &\leqslant \left\|\bar{\mathbf{\Delta}}_{(1)}\mathbf{Z}\right\|_{\mathrm{F}} + 2\sqrt{2}\epsilon \sup_{\mathbf{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)}\left\|\mathbf{\Delta}_{(1)}\mathbf{Z}\right\|_{\mathrm{F}},\end{aligned}$$

通过对等式两边取上确界，我们完成证明。 □

引理 A.2. （受限强凸性与光滑性）。假设 (3.3) 和假设 *1* 与 *2* 成立。若 $T \gtrsim \max(\kappa^2, \kappa^4)(\kappa_U/\kappa_L)^2 d_{\mathcal{M}}$, $m^{1/4} \gtrsim T^{2\delta}$, $m^{1/4}T^{2\delta-1} \gtrsim N^2 P\exp(d_{\mathcal{M}})$ 对于某个 $\delta > 1/2$, 则

$$\frac{1}{8}\kappa_L\left\|\mathbf{\Delta}\right\|_{\mathrm{F}}^2 \leqslant \frac{1}{T}\sum_{n=P+1}^{T}\left\|\mathbf{\Delta}_{(1)}\tilde{\mathbf{x}}_n\right\|_2^2 \leqslant \frac{8}{3}\kappa_U\left\|\mathbf{\Delta}\right\|_{\mathrm{F}}^2,$$

对所有 $\mathbf{\Delta} \in \mathcal{S}(2r_1, 2r_2, 2r_3)$，以至少.....的概率

$$1 - 2\exp(-CT(\kappa_L/\kappa_U)^2\min\{\kappa^{-2}, \kappa^{-4}\}) - \exp(-Cd_{\mathcal{M}}),$$

其中 κ, κ_U, κ_U 和 $d_{\mathcal{M}}$ 在定理 *2*中定义。

引理 *A.2* 的证明。记 $R_T(\mathbf{\Delta}) = \sum_{n=P+1}^{T}\|\mathbf{\Delta}_{(1)}\tilde{}\mathbf{x}_n\|_2^2$，并且有

$$\begin{aligned}R_T(\mathbf{\Delta}) &= \sum_{n=P+1}^{T}\left\|\mathbf{\Delta}_{(1)}(\mathbf{x}_{\mathrm{n}} + \boldsymbol{\vartheta}_{\mathrm{n}})\right\|_2^2 \\ &= \sum_{n=P+1}^{T}\mathbf{x}_n^\top\mathbf{\Delta}_{(1)}^\top\mathbf{\Delta}_{(1)}\mathbf{x}_n + 2\sum_{n=P+1}^{T}\mathbf{x}_n^\top\mathbf{\Delta}_{(1)}^\top\mathbf{\Delta}_{(1)}\boldsymbol{\vartheta}_n + \sum_{n=P+1}^{T}\boldsymbol{\vartheta}_n^\top\mathbf{\Delta}_{(1)}^\top\mathbf{\Delta}_{(1)}\boldsymbol{\vartheta}_n \\ &:= \mathrm{R}_1 + \mathrm{R}_2 + \mathrm{R}_3, \qquad (A.22)\end{aligned}$$

其中 $\boldsymbol{\vartheta}_n = \tilde{}\mathbf{x}_n - \mathbf{x}_n = (\boldsymbol{\eta}_{n-1}^\top, \cdots, \boldsymbol{\eta}_{n-P}^\top)^\top \in \mathbb{R}^{NP\times 1}$。根据 Basu 和 Michailidis (2015年) 关于 ARMA 过程的谱测度，我们有 $\lambda_{\min}\{\mathbb{E}(\mathbf{x}_n\mathbf{x}_n^\top)\} \geqslant \lambda_{\min}(\mathbf{\Sigma}_\varepsilon)/\mu_{\max}(\boldsymbol{\mathcal{A}}) = \kappa_L$ 并且 $\lambda_{\max}\{\mathbb{E}(\mathbf{x}_n\mathbf{x}_n^\top)\} \leqslant \lambda_{\max}(\mathbf{\Sigma}_\varepsilon)/\mu_{\min}(\boldsymbol{\mathcal{A}}) = \kappa_U$, 因此有

$$(T-P)\kappa_L \leqslant \mathbb{E}(\mathrm{R}_1) = \mathbb{E}\left(\sum_{n=P+1}^{T}\mathrm{vec}(\mathbf{\Delta})^\top(\mathbf{I}_N \otimes \mathbf{x}_n\mathbf{x}_n^\top)\mathrm{vec}(\mathbf{\Delta})\right) \leqslant (T-P)\kappa_U \leqslant T\kappa_U,$$

(A.23)

当 $\|\mathbf{\Delta}\|_{\mathrm{F}}=1$ 时。此外， $R_T(\mathbf{\Delta})\geqslant \mathbb{E}(\mathrm{R}_1)-\sup_{\mathbf{\Delta}\in\mathcal{S}}\{|\mathrm{R}_1-\mathbb{E}(\mathrm{R}_1)|+|\mathrm{R}_2|+|\mathrm{R}_3|\}$，且 $R_T(\mathbf{\Delta})\leqslant \mathbb{E}(\mathrm{R}_1)+\sup_{\mathbf{\Delta}\in\mathcal{S}}\{|\mathrm{R}_1-\mathbb{E}(\mathrm{R}_1)|+|\mathrm{R}_2|+|\mathrm{R}_3|\}$，其中 $\mathcal{S}=\mathcal{S}(2r_1,2r_2,2r_3)$。接下来，我们先对每个固定的 $\mathbf{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)$，给出 $|\mathrm{R}_1-\mathbb{E}(\mathrm{R}_1)|+|\mathrm{R}_2|+|\mathrm{R}_3|$ 的上界。

考虑 $\mathrm{R}_1-\mathbb{E}(\mathrm{R}_1)$ 项。注意，$\mathrm{x}_n=(\mathrm{y}_{n-1}^\top、\ldots、\mathrm{y}_{n-P}^\top)^\top\in\mathbb{R}^{NP\times 1}$ 以及 $\boldsymbol{\mathcal{A}}_{(1)}=(\mathrm{A}1、\ldots、\mathrm{A}_P)$，其中每个 A_i 都是一个 $N\times N$ 矩阵。由 (3.4) 可得 $\mathrm{y}_n=\boldsymbol{\mathcal{A}}_{(1)}\mathrm{x}_n+\boldsymbol{\varepsilon}_n$。它还可以进一步改写为 VAR(1) 形式 $\mathrm{x}_n=\mathrm{B}\mathrm{x}_{n-1}+\boldsymbol{e}_n$，因此 $\mathrm{x}_n=\sum_{j=0}^{\infty}\mathrm{B}^j\boldsymbol{e}_{n-j}$ 的 VMA 表示为 $\mathrm{z}=\mathrm{P}\boldsymbol{e}$，其中

$$\mathbf{B}=\begin{pmatrix}\mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{P-1} & \mathbf{A}_P\\ \mathbf{I}_N & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_N & \cdots & \mathbf{0} & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_N & \mathbf{0}\end{pmatrix},\quad \mathbf{P}=\begin{pmatrix}\mathbf{I}_{NP} & \mathbf{B} & \mathbf{B}^2 & \cdots & \mathbf{B}^{T-1} & \cdots\\ \mathbf{0} & \mathbf{I}_{NP} & \mathbf{B} & \cdots & \mathbf{B}^{T-2} & \cdots\\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{NP} & \cdots & \mathbf{B}^{T-3} & \cdots\\ \vdots & \vdots & \vdots & \ddots & \vdots & \cdots\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_{NP} & \cdots\end{pmatrix},$$

$\boldsymbol{e}_n=(\boldsymbol{\varepsilon}_n^\top、\ldots、\mathbf{0})^\top\in\mathbb{R}^{NP\times 1}$、$\boldsymbol{e}=(\boldsymbol{e}_{T-1}^\top、\ldots、\boldsymbol{e}_P^\top、\ldots)^\top$，以及 $\mathbf{z}=(\mathbf{x}_T^\top、\ldots、\mathbf{x}_{P+1}^\top)^\top\in\mathbb{R}^{NP(T-P)\times 1}$。此外，根据假设 2，误差项的形式为 $\boldsymbol{e}=\bar{\mathbf{\Sigma}}\bar{\boldsymbol{\xi}}$，其中 $\bar{\boldsymbol{\xi}}_n=(\boldsymbol{\xi}_n^\top、\ldots、\mathbf{0})^\top\in\mathbb{R}^{NP\times 1}$、$\bar{\boldsymbol{\xi}}=(\bar{\boldsymbol{\xi}}_{T-1}^\top、\ldots、\bar{\boldsymbol{\xi}}_P^\top、\ldots)^\top$、

$$\bar{\mathbf{\Sigma}}_\varepsilon=\begin{pmatrix}\mathbf{\Sigma}_\varepsilon & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{\Sigma}_\varepsilon & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Sigma}_\varepsilon\end{pmatrix}\in\mathbb{R}^{NP\times NP}\quad\text{and}\quad \bar{\mathbf{\Sigma}}=\begin{pmatrix}\bar{\mathbf{\Sigma}}_\varepsilon^{1/2} & \mathbf{0} & \mathbf{0} & \cdots\\ \mathbf{0} & \bar{\mathbf{\Sigma}}_\varepsilon^{1/2} & \mathbf{0} & \cdots\\ \mathbf{0} & \mathbf{0} & \bar{\mathbf{\Sigma}}_\varepsilon^{1/2} & \cdots\\ \vdots & \vdots & \vdots & \ddots\end{pmatrix}.$$

记 $\mathbf{\Sigma}_{\mathbf{\Delta}}=\bar{\mathbf{\Sigma}}\mathbf{P}^\top(\mathbf{I}_{T-P}\otimes\mathbf{\Delta}_{(1)}^\top\mathbf{\Delta}_{(1)})\mathbf{P}\,\bar{\mathbf{\Sigma}}$，则

$$\mathrm{R}_1=\sum_{n=P+1}^{T}\mathbf{x}_n^\top\mathbf{\Delta}_{(1)}^\top\mathbf{\Delta}_{(1)}\mathbf{x}_n=\bar{\boldsymbol{\xi}}^\top\bar{\mathbf{\Sigma}}\mathbf{P}^\top(\mathbf{I}_{T-P}\otimes\mathbf{\Delta}_{(1)}^\top\mathbf{\Delta}_{(1)})\mathbf{P}\bar{\mathbf{\Sigma}}\bar{\boldsymbol{\xi}}=\bar{\boldsymbol{\xi}}^\top\mathbf{\Sigma}_{\mathbf{\Delta}}\bar{\boldsymbol{\xi}}.$$

注意， $\lambda_{\max}(\mathbf{P}\mathbf{P}^\top)=1/\mu_{\min}(\boldsymbol{\mathcal{A}})$, $\left\|\mathbf{\Sigma}_{\mathbf{\Delta}}\right\|_{\mathrm{op}}\leqslant\kappa_U$ 且

$$\left\|\mathbf{\Sigma}_{\mathbf{\Delta}}\right\|_{\mathrm{F}}\leqslant\|\bar{\mathbf{\Sigma}}\|_{\mathrm{op}}^2\|\mathbf{P}\|_{\mathrm{op}}\|\mathbf{P}^\top\|_{\mathrm{op}}\|\mathbf{I}_{T-P}\otimes\mathbf{\Delta}_{(1)}^\top\mathbf{\Delta}_{(1)}\|_{\mathrm{F}}=\sqrt{T-P}\kappa_U\leqslant\sqrt{T}\kappa_U.$$

对于任意 $t>0$，根据 Hanson–Wright不等式，我们可以将 $\mathrm{R}_1-\mathbb{E}(\mathrm{R}_1)$ 限制在以下范围内 $_{\mathrm{W}}$,

$$\begin{aligned}\mathbb{P}\left[|\mathrm{R}_1-\mathbb{E}(\mathrm{R}_1)|\geqslant t\right]&\leqslant 2\exp\left(-C\min\left(\frac{t}{\kappa^2\left\|\mathbf{\Sigma}_{\mathbf{\Delta}}\right\|_{\mathrm{op}}},\frac{t^2}{\kappa^4\left\|\mathbf{\Sigma}_{\mathbf{\Delta}}\right\|_{\mathrm{F}}^2}\right)\right)\\&\leqslant 2\exp\left(-C\min\left(\frac{t}{\kappa^2\kappa_U},\frac{t^2}{\kappa^4T\kappa_U^2}\right)\right).\end{aligned}\tag{A.24}$$

接下来，我们考虑 R_2，这一项，其形式为

$$\mathrm{R}_2 = 2\mathbf{z}^\top(\mathbf{I}_{T-P}\otimes\boldsymbol{\Delta}_{(1)}^\top\boldsymbol{\Delta}_{(1)})\mathbf{c} = 2\bar{\boldsymbol{\xi}}^\top\bar{\boldsymbol{\Sigma}}\boldsymbol{P}^\top(\mathbf{I}_{T-P}\otimes\boldsymbol{\Delta}_{(1)}^\top\boldsymbol{\Delta}_{(1)})\mathbf{c} = 2\boldsymbol{m}_1^\top\bar{\boldsymbol{\xi}},$$

其中 $\boldsymbol{m}_1 = \bar{\boldsymbol{\Sigma}}\boldsymbol{P}^\top(\mathbf{I}_{T-P}\otimes\boldsymbol{\Delta}_{(1)}^\top\boldsymbol{\Delta}_{(1)})\mathbf{c}$，$\mathbf{c} = (\boldsymbol{\vartheta}_T^\top, \cdots, \boldsymbol{\vartheta}_{P+1}^\top)^\top \in \mathbb{R}^{NP(T-P)\times 1}$。令 $\eta_{\max}$ =最大值$_{1\leqslant n\leqslant T}$最大值$_{1\leqslant i\leqslant N}|\eta_{i,\ n}|$，且有 $\|\mathbf{c}\|_2^2 \leqslant NPT\eta_{\max}^2$ 以及 $\|\boldsymbol{m}_1\|_2^2 \leqslant NPT\eta_{\max}^2\kappa_U$。此外，由不等式（3.3）和马尔可夫不等式可得

$$\mathbb{P}\left\{\eta_{\max} \geqslant \frac{x}{m^{1/4}}\right\} \leqslant \frac{CNT}{x^2}, \tag{A.25}$$

通过改变 $\eta_{\max}$ 的取值，可以用来控制 x 的速率。根据假设 $\bar{\boldsymbol{\xi}}2$ 中的次高斯条件，并在（2A.25）中令 $x = m^{1/8}T^\delta/N^{1/2}P^{1/2}$，

$$\mathbb{P}\left[2\left|\boldsymbol{m}_1{}^\top\bar{\boldsymbol{\xi}}\right| \geqslant t\right] \leqslant 2\exp\left(-\frac{Cm^{1/4}t^2}{\kappa^2T^{1+2\delta}\kappa_U}\right) + \frac{CN^2P}{m^{1/4}T^{2\delta-1}} \tag{A.26}$$

对任意 $t>0$。

最后，对于（$_3$ A.22）中 R 的第三项，有

$$\sup_{\boldsymbol{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)}|\mathrm{R}_3| = \sup_{\boldsymbol{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)}\left|\mathbf{c}^\top(\mathbf{I}_{T-P}\otimes\boldsymbol{\Delta}_{(1)}^\top\boldsymbol{\Delta}_{(1)})\mathbf{c}\right| \leqslant NPT\eta_{\max}^2,$$

并令 $x = m^{1/8}T^\delta/N^{1/2}P^{1/2},$，我们有

$$\mathbb{P}\left[\sup_{\boldsymbol{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)}|\mathrm{R}_3| \geqslant \frac{T^{1+2\delta}}{m^{1/4}}\right] \leqslant \frac{CN^2P}{m^{1/4}T^{2\delta-1}}. \tag{A.27}$$

因此，结合（A.24）、（A.26）和（A.27），我们有

$$\begin{aligned}
&\mathbb{P}\left[|\mathrm{R}_1+\mathrm{R}_2+\mathrm{R}_3-\mathbb{E}(\mathrm{R}_1)| \geqslant t_1+t_2+\frac{T^{1+2\delta}}{m^{1/4}}\right] \\
&\quad\leqslant \mathbb{P}\left[|\mathrm{R}_1-\mathbb{E}(\mathrm{R}_1)| \geqslant t_1\right] + \mathbb{P}\left[|\mathrm{R}_2| \geqslant t_2\right] + \mathbb{P}\left[|\mathrm{R}_3| \geqslant \frac{T^{1+2\delta}}{m^{1/4}}\right] \\
&\quad\leqslant 2\exp\left(-C\min\left(\frac{t_1}{\kappa^2\kappa_U}, \frac{t_1^2}{\kappa^4T\kappa_U^2}\right)\right) + 2\exp\left(-\frac{Cm^{1/4}t^2}{\kappa^2T^{1+2\delta}\kappa_U}\right) + \frac{CN^2P}{m^{1/4}T^{2\delta-1}}.
\end{aligned}$$

令 $t_1 = t_2 = T\kappa_L/6$ 并由（A.23）可得

$$\begin{aligned}
&\mathbb{P}\left[0.5\kappa_L \leqslant T^{-1}R_T(\boldsymbol{\Delta}) \leqslant 1.5\kappa_U\right] \\
&\quad\geqslant 1-2\exp\left(-C\min\left(\frac{T\kappa_L}{\kappa^2\kappa_U}, \frac{T\kappa_L^2}{\kappa^4\kappa_U^2}\right)\right) - 2\exp\left(-\frac{C\kappa_L^2m^{1/4}}{T^{2\delta-1}\kappa^2\kappa_U}\right) - \frac{CN^2P}{m^{1/4}T^{2\delta-1}},
\end{aligned}$$

(A.28)

如 $m^{1/4} \geqslant 6T^{2\delta}/\kappa_L$。令 $\bar{\mathcal{S}}$ 为 $S(2r_1, 2r_2, 2r_3)$ 的 ϵ-覆盖网。为构造并合界，我们将 $R_T(\mathbf{\Delta})$ 改写为 $R_T(\mathbf{\Delta}) = \|\mathbf{\Delta}_{(1)}\tilde{\mathbf{X}}\|_{\mathrm{F}}^2$，其中 $\tilde{\mathbf{X}} = (\tilde{\mathbf{x}}_{P+1}, \cdots, \tilde{\mathbf{x}}_T) \in \mathbb{R}^{NP\times(T-P)}$。定义事件

$$\mathcal{E}(\epsilon) = \left\{\forall \mathbf{\Delta} \in \bar{\mathcal{S}}(2r_1, 2r_2, 2r_3) : \sqrt{0.5\kappa_L} \leqslant \frac{1}{\sqrt{T}}\|\mathbf{\Delta}_{(1)}\tilde{\mathbf{X}}\|_{\mathrm{F}} \leqslant \sqrt{1.5\kappa_U}\right\}.$$

然后，根据式（A.28）中的逐点界以及引理A.1(i)中的覆盖数，

$$\begin{aligned}\mathbb{P}\left[\mathcal{E}^c(\epsilon)\right] \leqslant\ & 2\exp\left(Cd_{\mathcal{M}} - C\min\left(\frac{T\kappa_L}{\kappa^2\kappa_U}, \frac{T\kappa_L^2}{\kappa^4\kappa_U^2}\right)\right) \\ & + 2\exp\left(Cd_{\mathcal{M}} - \frac{C\kappa_L^2 m^{1/4}}{T^{2\delta-1}\kappa^2\kappa_U}\right) + \exp(Cd_{\mathcal{M}})\frac{CN^2P}{m^{1/4}T^{2\delta-1}}.\end{aligned}$$

注意，根据引理A.1(ii)，

$$\begin{aligned}\mathcal{E}(\epsilon) &\subset \left\{\max_{\mathbf{\Delta}\in\bar{\mathcal{S}}(2r_1,2r_2,2r_3)} \frac{1}{\sqrt{T}}\|\mathbf{\Delta}_{(1)}\tilde{\mathbf{X}}\|_{\mathrm{F}} \leqslant \sqrt{1.5\kappa_U}\right\} \\ &\subset \left\{\sup_{\mathbf{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)} \frac{1}{\sqrt{T}}\|\mathbf{\Delta}_{(1)}\tilde{\mathbf{X}}\|_{\mathrm{F}} \leqslant \frac{\sqrt{1.5\kappa_U}}{1-2\sqrt{2}\epsilon}\right\}.\end{aligned}$$

此外，类似于引理A.1(ii)，我们可以证明

$$\begin{aligned}\|\mathbf{\Delta}_{(1)}\mathbf{Z}\|_{\mathrm{F}} &\geqslant \|\bar{\mathbf{\Delta}}_{(1)}\mathbf{Z}\|_{\mathrm{F}} - \sum_{i=1}^{8}\|(\overline{\mathbf{W}}_i)_{(1)}\mathbf{Z}\|_{\mathrm{F}} \\ &\geqslant \|\bar{\mathbf{\Delta}}_{(1)}\mathbf{Z}\|_{\mathrm{F}} - \sum_{i=1}^{8}\|\overline{\mathbf{W}}_i\|_{\mathrm{F}} \sup_{\mathbf{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)}\|\mathbf{\Delta}_{(1)}\mathbf{Z}\|_{\mathrm{F}} \\ &\geqslant \|\bar{\mathbf{\Delta}}_{(1)}\mathbf{Z}\|_{\mathrm{F}} - 2\sqrt{2}\epsilon \sup_{\mathbf{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)}\|\mathbf{\Delta}_{(1)}\mathbf{Z}\|_{\mathrm{F}},\end{aligned}$$

其中，最后一个不等式由 $\sum_{i=1}^{8}\|\overline{\mathbf{W}}_i\|_{\mathrm{F}} \leqslant 2\sqrt{2}\epsilon$ 得出。对两边取下确界，若 $0 \leqslant \epsilon \leqslant \frac{1}{4\sqrt{2}}$，则有

$$\begin{aligned}\inf_{\mathbf{\Delta}\in S(2r_1,2r_2,2r_3)} \frac{1}{\sqrt{T}}\|\mathbf{\Delta}_{(1)}\mathbf{Z}\|_{\mathrm{F}} \geqslant & \min_{\mathbf{\Delta}\in\bar{\mathcal{S}}(2r_1,2r_2,2r_3)} \frac{1}{\sqrt{T}}\|\mathbf{\Delta}_{(1)}\mathbf{Z}\|_{\mathrm{F}} \\ & - 2\sqrt{2}\epsilon\frac{1}{\sqrt{T}} \sup_{\mathbf{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)}\|\mathbf{\Delta}_{(1)}\mathbf{Z}\|_{\mathrm{F}} \\ \geqslant & \sqrt{0.5\kappa_L} - 2\sqrt{2}\epsilon\frac{\sqrt{1.5\kappa_U}}{1-2\sqrt{2}\epsilon} \geqslant \sqrt{0.5\kappa_L} - 4\epsilon\sqrt{3\kappa_U}.\end{aligned}$$

当 ϵ 取为 $\frac{1}{8}\sqrt{\frac{\kappa_L}{6\kappa_U}}$ 时，

$$\mathcal{E}(\epsilon) \subset \left\{ \inf_{\mathbf{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)} \frac{1}{\sqrt{T}}\|\mathbf{\Delta}_{(1)}\widetilde{\mathbf{X}}\|_{\mathrm{F}} \geqslant \frac{\sqrt{0.5\kappa_L}}{2} \right\},$$

因此，采用上述 ϵ 的选择，

$$\mathcal{E}(\epsilon) \subset \left\{ \frac{\kappa_L}{8} \leqslant \inf_{\mathbf{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)} \frac{1}{T}\|\mathbf{\Delta}_{(1)}\widetilde{\mathbf{X}}\|_{\mathrm{F}}^2 \leqslant \sup_{\mathbf{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)} \frac{1}{T}\|\mathbf{\Delta}_{(1)}\widetilde{\mathbf{X}}\|_{\mathrm{F}}^2 \leqslant \frac{8\kappa_U}{3} \right\}.$$

根据条件 $T \gtrsim (\kappa_U/\kappa_L)^2 \max(\kappa^2, \kappa^4) d_{\mathcal{M}}$, $m^{1/4} \gtrsim T^{2\delta}$, $m^{1/4}T^{2\delta-1} \gtrsim N^2 P \exp(d_{\mathcal{M}})$（其中某个 $\delta > 1/2,$），可知对于所有 $\mathbf{\Delta} \in \mathcal{S}(2r_1, 2r_2, 2r_3)$,

$$\mathbb{P}\left[\frac{\kappa_L}{8} \leqslant \frac{1}{T}\sum_{n=P+1}^{T} \|\mathbf{\Delta}_{(1)}\widetilde{\mathbf{x}}_n\|_2^2 \leqslant \frac{8\kappa_U}{3} \right] \geqslant 1 - 2\exp\left(-CT(\kappa_L/\kappa_U)^2 \min(\kappa^{-2}, \kappa^{-4})\right) - \exp\left(-Cd_{\mathcal{M}}\right).$$

至此，证明完成。 □

引理 A.3。 假设 (3.3) 以及假设条件*1* 和 *2* 均成立。若样本量 $T \gtrsim \max(\kappa^2, \kappa^4)(\kappa_U/\kappa_L)^2 d_{\mathcal{M}}$, $m^{1/4} \gtrsim T^{2\delta}$ 且 $m^{1/4}T^{2\delta-1} \gtrsim N^2 P^{1/2}\exp(d_{\mathcal{M}})$ 对于某个 $\delta > 1/2,$ 则

$$\sup_{\mathbf{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)} \langle \frac{1}{T}\sum_{n=P+1}^{T} \boldsymbol{\epsilon}_n \circ \widetilde{\mathbf{X}}_n, \mathbf{\Delta} \rangle \leqslant C(\kappa^2\sqrt{\lambda_{\max}(\mathbf{\Sigma}_\varepsilon)\kappa_U} + \kappa\sqrt{\kappa_U})\sqrt{\frac{d_{\mathcal{M}}}{T}} + \frac{T^{2\delta}}{m^{1/4}}$$

以至少.....的概率

$$1 - \exp(-Cd_{\mathcal{M}}) - 2\exp(-CT(\kappa_L/\kappa_U)^2 \min\{\kappa^{-2}, \kappa^{-4}\}),$$

其中 κ、 κ_U以及 κ_U 和 $d_{\mathcal{M}}$ 定义于定理*2*。

引理*A.3*的证明。 为区分模型误差和估计误差的影响，令 $\boldsymbol{\upsilon}_n = \boldsymbol{\eta}_n - \sum_{j=1}^{P} \mathbf{A}_j \boldsymbol{\eta}_{n-j} \in \mathbb{R}^{N\times 1}$, $\mathbf{X}_n = (\mathbf{y}_{n-1}, \ldots, \mathbf{y}_{n-P}) \in \mathbb{R}^{N\times P}$, 以及 $\mathbf{\Upsilon}_n = (\boldsymbol{\eta}_{n-1}, \ldots, \boldsymbol{\eta}_{n-P}) \in \mathbb{R}^{N\times P}$。注意， $\boldsymbol{\epsilon}_n = \boldsymbol{\varepsilon}_n + \boldsymbol{\upsilon}_n$ 以及 $\widetilde{\mathbf{X}}_n = \mathbf{X}_n + \mathbf{\Upsilon}_n$。因此，

$$\begin{aligned}
&\sup_{\mathbf{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)} \langle \frac{1}{T}\sum_{n=P+1}^{T} \boldsymbol{\epsilon}_n \circ \widetilde{\mathbf{X}}_n, \mathbf{\Delta} \rangle \\
&\quad\leqslant \sup_{\mathbf{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)} \langle \frac{1}{T}\sum_{n=P+1}^{T} \boldsymbol{\varepsilon}_n \circ \widetilde{\mathbf{X}}_n, \mathbf{\Delta} \rangle \\
&\quad\quad + \sup_{\mathbf{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)} \langle \frac{1}{T}\sum_{t=P+1}^{T} \boldsymbol{\upsilon}_n \circ \mathbf{X}_n, \mathbf{\Delta} \rangle + \sup_{\mathbf{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)} \langle \frac{1}{T}\sum_{n=P+1}^{T} \boldsymbol{\upsilon}_n \circ \mathbf{\Upsilon}_n, \mathbf{\Delta} \rangle.
\end{aligned}$$

(A.29)

我们将逐一给出这三项的上界。

~对于（A.29）的第一项，很容易验证 $\langle \boldsymbol{\varepsilon}_n \circ \mathbf{X}_n,\ \boldsymbol{\Delta} \rangle = \langle \boldsymbol{\varepsilon}_n,\ \boldsymbol{\Delta}_{(1)} \tilde{\mathbf{x}}_n \rangle$。记 $S_t(\boldsymbol{\Delta}) = \sum_{n=P+1}^{t} \langle \boldsymbol{\varepsilon}_n,\ \boldsymbol{\Delta}_{(1)} \tilde{\mathbf{x}}_n \rangle$ 以及 $R_t(\boldsymbol{\Delta}) = \sum_{n=P+1}^{t} \|\boldsymbol{\Delta}_{(1)} \tilde{\mathbf{x}}_n\|_2^2$，其中 $P+1 \leqslant t \leqslant T$。根据误差的 Chernoff界，对任意 $\alpha > 0,\ \beta > 0$ 以及 $c > 0$，存在 $\eta > 0$,

$$
\begin{aligned}
&\mathbb{P}\left[\{S_T(\boldsymbol{\Delta}) \geqslant \alpha\} \bigcap \{R_T(\boldsymbol{\Delta}) \leqslant \beta\}\right] \\
=&\inf_{\eta>0} \mathbb{P}\left[\{\exp\left(\eta S_T(\boldsymbol{\Delta})\right) \geqslant \exp\left(\eta\alpha\right)\} \bigcap \{R_T(\boldsymbol{\Delta}) \leqslant \beta\}\right] \\
=&\inf_{\eta>0} \mathbb{P}\left[\exp\left(\eta S_T\left(\boldsymbol{\Delta}\right)\right) \mathbb{I}\left(R_T(\boldsymbol{\Delta}) \leqslant \beta\right) \geqslant \exp(\eta\alpha)\right] \\
\leqslant&\inf_{\eta>0} \exp\left(-\eta\alpha\right) \mathbb{E}\left[\exp(\eta S_T(\boldsymbol{\Delta})) \mathbb{I}(R_T(\boldsymbol{\Delta}) \leqslant \beta)\right] \\
=&\inf_{\eta>0} \exp(-\eta\alpha + c\eta^2\beta) \mathbb{E}\left[\exp(\eta S_T(\boldsymbol{\Delta}) - c\eta^2\beta) \mathbb{I}(R_T(\boldsymbol{\Delta}) \leqslant \beta)\right] \\
\leqslant&\inf_{\eta>0} \exp(-\eta\alpha + c\eta^2\beta) \mathbb{E}\left[\exp(\eta S_T(\boldsymbol{\Delta}) - c\eta^2 R_T(\boldsymbol{\Delta}))\right].
\end{aligned} \tag{A.30}
$$

根据条件期望的塔式性质，我们有

$$
\begin{aligned}
&\mathbb{E}\left[\exp\left(\eta S_T(\boldsymbol{\Delta}) - c\eta^2 R_T(\boldsymbol{\Delta})\right)\right] \\
=&\mathbb{E}\left[\mathbb{E}\left[\exp\left(\eta S_T(\boldsymbol{\Delta}) - c\eta^2 R_T(\boldsymbol{\Delta})\right) | \mathcal{F}_{T-1}\right]\right] \\
=&\mathbb{E}\left[\exp\left(\eta S_{T-1}(\boldsymbol{\Delta}) - c\eta^2 R_{T-1}(\boldsymbol{\Delta})\right) \mathbb{E}\left[\exp\left(\eta\langle \boldsymbol{\varepsilon}_T, \boldsymbol{\Delta}_{(1)} \tilde{\mathbf{x}}_T \rangle - c\eta^2 \|\boldsymbol{\Delta}_{(1)} \tilde{\mathbf{x}}_T\|_2^2\right) | \mathcal{F}_{T-1}\right]\right].
\end{aligned}
$$

根据假设 2中的次高斯性条件，则 $\langle \boldsymbol{\varepsilon}_T,\ \boldsymbol{\Delta}_{(1)} \tilde{\mathbf{x}}_T \rangle = \langle \boldsymbol{\xi}_T,\ \boldsymbol{\Sigma}_\varepsilon^{1/2} \boldsymbol{\Delta}_{(1)} \tilde{\mathbf{x}}_T \rangle$，且 $\mathbb{E}\left[\exp(\eta\langle \boldsymbol{\varepsilon}_T, \boldsymbol{\Delta}_{(1)} \tilde{\mathbf{x}}_T \rangle)\right] \leqslant \exp\left(\eta^2\kappa^2\lambda_{\max}(\boldsymbol{\Sigma}_\varepsilon)\|\boldsymbol{\Delta}_{(1)} \tilde{\mathbf{x}}_T\|_2^2/2\right)$。由于 $\tilde{\mathbf{x}}_n$ 是 $\mathcal{F}_{n-1}$-可测的，$\boldsymbol{\varepsilon}_n$ 是 $\mathcal{F}_n$-可测的，且 $\boldsymbol{\varepsilon}_n | \mathcal{F}_{n-1}$ 均值为零，令 $c = \kappa^2\lambda_{\max}(\boldsymbol{\Sigma}_\varepsilon)/2$, 即可轻易推出以下不等式,

$$
\begin{aligned}
&\mathbb{E}[\exp(\eta S_T(\boldsymbol{\Delta}) - \eta^2\kappa^2\lambda_{\max}(\boldsymbol{\Sigma}_\varepsilon) R_T(\boldsymbol{\Delta})/2)] \\
\leqslant&\mathbb{E}[\exp(\eta S_{T-1}(\boldsymbol{\Delta}) - \eta^2\kappa^2\lambda_{\max}(\boldsymbol{\Sigma}_\varepsilon) R_{T-1}(\boldsymbol{\Delta})/2)] \\
\leqslant&\cdots \leqslant \mathbb{E}[\exp(\eta S_{P+1}(\boldsymbol{\Delta}) - \eta^2\kappa^2\lambda_{\max}(\boldsymbol{\Sigma}_\varepsilon) R_{P+1}(\boldsymbol{\Delta})/2)] \leqslant 1.
\end{aligned}
$$

因此，对于任意 $\alpha>0$ 和 $\beta>0$，我们可以得到以下不等式（A.30），

$$\begin{aligned}
&\mathbb{P}[\{S_T(\boldsymbol{\Delta})\geqslant\alpha\}\bigcap\{R_T(\boldsymbol{\Delta})\leqslant\beta\}]\\
\leqslant&\inf_{\eta>0}\exp(-\eta\alpha+\eta^2\kappa^2\lambda_{\max}(\boldsymbol{\Sigma}_\varepsilon)\beta/2)\\
=&\exp\left(-\frac{\alpha^2}{2\kappa^2\lambda_{\max}(\boldsymbol{\Sigma}_\varepsilon)\beta}\right).
\end{aligned}\tag{A.31}$$

此外，根据引理 A.2，以下关于 $R_T(\boldsymbol{\Delta})$ 的界满足

$$\frac{T}{8}\kappa_L\leqslant R_T(\boldsymbol{\Delta})\leqslant\frac{8T}{3}\kappa_U\tag{A.32}$$

w概率至少为 $1-2\exp(-CT(\kappa_L/\kappa_U)^2\min\{\kappa^{-2},\ \kappa^{-4}\})-\exp(-Cd_{\mathcal{M}}\qquad)$.

根据引理 A.1 （ii），对于任意 $x>0$,

$$\begin{aligned}
&\mathbb{P}\left[\sup_{\boldsymbol{\Delta}\in S(2r_1,2r_2,2r_3)}\left\langle\frac{1}{T}\sum_{n=P+1}^{T}\boldsymbol{\varepsilon}_n\circ\widetilde{\mathbf{X}}_n,\boldsymbol{\Delta}\right\rangle\geqslant x\right]\\
\leqslant&\mathbb{P}\left[\max_{\boldsymbol{\Delta}\in\bar{\mathcal{S}}(2r_1,2r_2,2r_3)}\left\langle\frac{1}{T}\sum_{n=P+1}^{T}\boldsymbol{\varepsilon}_n\circ\widetilde{\mathbf{X}}_n,\boldsymbol{\Delta}\right\rangle\geqslant(1-2\sqrt{2}\epsilon)x\right]\\
\leqslant&|\bar{\mathcal{S}}(2r_1,2r_2,2r_3)|\cdot\mathbb{P}\left[\left\langle\frac{1}{T}\sum_{n=P+1}^{T}\boldsymbol{\varepsilon}_n\circ\widetilde{\mathbf{X}}_n,\boldsymbol{\Delta}\right\rangle\geqslant(1-2\sqrt{2}\epsilon)x\right],
\end{aligned}\tag{A.33}$$

结合（A.31）和（A.32），可得

$$\begin{aligned}
&\mathbb{P}\left[\left\langle\frac{1}{T}\sum_{n=P+1}^{T}\boldsymbol{\varepsilon}_n\circ\widetilde{\mathbf{X}}_n,\boldsymbol{\Delta}\right\rangle\geqslant(1-2\sqrt{2}\epsilon)x\right]\\
\leqslant&\mathbb{P}[\{S_T(\boldsymbol{\Delta})\geqslant T(1-2\sqrt{2}\epsilon)x\}\bigcap\{R_T(\boldsymbol{\Delta})\leqslant CT\kappa^2\kappa_U\}]\\
&+\mathbb{P}[R_T(\boldsymbol{\Delta})\geqslant CT\kappa^2\kappa_U]\\
\leqslant&\exp\left[-\frac{(1-2\sqrt{2}\epsilon)^2Tx^2}{2C\kappa^4\lambda_{\max}(\boldsymbol{\Sigma}_\varepsilon)\kappa_U}\right]+2\exp(-CT(\kappa_L/\kappa_U)^2\min\{\kappa^{-2},\kappa^{-4}\})+\exp(-Cd_{\mathcal{M}})
\end{aligned}$$

对于任意 $x>0$。注意，根据引理 A.1，$|\bar{\mathcal{S}}(r_1,\ r_2,\ r_3)|\leqslant(12/\epsilon)^{d_{\mathcal{M}}}$。令$\epsilon=1/10$, 且 $x=C\sqrt{d_{\mathcal{M}}\kappa^4\lambda_{\max}(\boldsymbol{\Sigma}_\varepsilon)\kappa_U/T}$,，则有

$$\begin{aligned}
&\mathbb{P}\left[\sup_{\boldsymbol{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)}\left\langle\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_n\circ\widetilde{\mathbf{X}}_n,\boldsymbol{\Delta}\right\rangle\geqslant C\kappa^2\sqrt{\lambda_{\max}(\boldsymbol{\Sigma}_\varepsilon)\kappa_U}\sqrt{\frac{d_{\mathcal{M}}}{T}}\right]\\
&\leqslant\exp(-Cd_{\mathcal{M}})+2\exp(-CT(\kappa_L/\kappa_U)^2\min\{\kappa^{-2},\kappa^{-4}\}).
\end{aligned}\tag{A.34}$$

对于（A.29）中的第二项，由于 $\mathbf{x}_n=(\mathbf{y}_{n-1}^\top,\ldots,\mathbf{y}_{n-P}^\top)^\top\in\mathbb{R}^{NP\times1}$，可以验证 $\langle\boldsymbol{\upsilon}_n\circ\mathbf{X}_n,\boldsymbol{\Delta}\rangle=\langle\boldsymbol{\upsilon}_n,\boldsymbol{\Delta}_{(1)}\mathbf{x}_n\rangle$。记 $\mathbf{z}=(\mathbf{x}_T^\top,\ldots,\mathbf{x}_{P+1}^\top)^\top\in\mathbb{R}^{NP(T-P)\times1}$,

以及 $\boldsymbol{\upsilon} = (\boldsymbol{\upsilon}_T^\top, \ldots, \boldsymbol{\upsilon}_{P+1}^\top)^\top \in \mathbb{R}^{N(T-P)\times 1}$，并且有

$$\sum_{n=P+1}^{T} \langle \boldsymbol{\upsilon}_n, \boldsymbol{\Delta}_{(1)} \mathbf{x}_n \rangle = \boldsymbol{\upsilon}^\top (\mathbf{I}_{T-P} \otimes \boldsymbol{\Delta}_{(1)}) \mathbf{z} := \boldsymbol{m}_2^\top \bar{\boldsymbol{\xi}},$$

其中 $\boldsymbol{m}_2 = \bar{\boldsymbol{\Sigma}} \mathbf{P}^\top (\mathbf{I}_{T-P} \otimes \boldsymbol{\Delta}_{(1)}) \boldsymbol{\upsilon}$，$\mathbf{z} = \mathbf{P}\boldsymbol{e}$，且 $\boldsymbol{e} = \bar{\boldsymbol{\Sigma}} \bar{\boldsymbol{\xi}}$。注意，$\|\boldsymbol{\upsilon}\|_2^2 \leqslant CNT\eta_{\max}^2$且 $\|\boldsymbol{m}_2\|_2^2 \leqslant CNT\eta_{\max}^2 \kappa_U$，其中 $\eta_{\max}$ =最大值$_{1\leqslant n\leqslant T}$最大值$_{1\leqslant i\leqslant N} |\eta_{n,\,i}|$。更多细节请参见引理 A.2的证明。

根据假设 $\boldsymbol{\xi}$ 2中的次高斯性条件，与引理A.2的证明类似，我们可以得到

$$\mathbb{P}[|\boldsymbol{m}_2^\top \bar{\boldsymbol{\xi}}| \geqslant t | \eta_{\max}] \leqslant \exp\left(-\frac{Ct^2}{\kappa^2 \kappa_U NT \eta_{\max}^2}\right) \quad \text{for all } t > 0$$

并令 $x = m^{1/8} T^\delta / N^{1/2}$代入（A.25），

$$\mathbb{P}\left[|\boldsymbol{m}_2^\top \bar{\boldsymbol{\xi}}| \geqslant t\right] \leqslant \exp\left(-\frac{Cm^{1/4} t^2}{\kappa^2 \kappa_U T^{1+2\delta}}\right) + \frac{CN^2}{m^{1/4} T^{2\delta-1}}.$$

因此，采用与（A.33）类似的方法，

$$\begin{aligned}
&\mathbb{P}\left[\sup_{\boldsymbol{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)} \langle \frac{1}{T} \sum_{n=P+1}^{T} \boldsymbol{\upsilon}_n \circ \mathbf{X}_t, \boldsymbol{\Delta} \rangle \geqslant t\right] \\
&\leqslant \mathbb{P}\left[\max_{\boldsymbol{\Delta}\in\bar{\mathcal{S}}(2r_1,2r_2,2r_3)} \langle \frac{1}{T} \sum_{n=P+1}^{T} \boldsymbol{\upsilon}_n \circ \mathbf{X}_t, \boldsymbol{\Delta} \rangle \geqslant (1-2\sqrt{2}\epsilon)t\right] \\
&\leqslant |\bar{\mathcal{S}}(2r_1,2r_2,2r_3)| \mathbb{P}\left[\langle \sum_{n=P+1}^{T} \boldsymbol{\upsilon}_n \circ \mathbf{X}_t, \boldsymbol{\Delta} \rangle \geqslant (1-2\sqrt{2}\epsilon)Tt\right] \\
&\leqslant \exp(Cd_{\mathcal{M}}) \frac{CN^2}{m^{1/4} T^{2\delta-1}} + \exp\left(Cd_{\mathcal{M}} - \frac{Cm^{1/4} t^2}{\kappa^2 \kappa_U T^{2\delta-1}}\right).
\end{aligned}$$

令 $t = C\sqrt{\kappa^2 \kappa_U T^{2\delta-1} d_{\mathcal{M}} / m^{1/4}}$，并且在条件 $m^{1/4} \gtrsim T^{2\delta}$ 和 $m^{1/4} T^{2\delta-1} \gtrsim N^2 \exp(d_{\mathcal{M}})$ 对某个 $\delta > 1/2$, 成立时，我们有

$$\mathbb{P}\left[\sup_{\boldsymbol{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)} \langle \frac{1}{T} \sum_{n=P+1}^{T} \boldsymbol{\upsilon}_n \circ \mathbf{X}_t, \boldsymbol{\Delta} \rangle \geqslant C\kappa\sqrt{\kappa_U}\sqrt{\frac{d_{\mathcal{M}}}{T}}\right] \leqslant \exp\left(-Cd_{\mathcal{M}}\right). \quad \text{(A.35)}$$

最后，对于 (A.29) 处的第三项，记 $\boldsymbol{\vartheta}_n = (\boldsymbol{\eta}_{n-1}^\top, \ldots, \boldsymbol{\eta}_{n-P}^\top)^\top \in \mathbb{R}^{NP\times 1}$, $\boldsymbol{\upsilon}_n = \boldsymbol{\eta}_n - \sum \mathbf{A}_j \boldsymbol{\eta}_{n-j} \in \mathbb{R}^{N\times 1}$, $\boldsymbol{\Upsilon}_n = (\boldsymbol{\eta}_{n-1}, \ldots, \boldsymbol{\eta}_{n-P}) \in \mathbb{R}^{N\times P}$，并且有 $\langle \boldsymbol{\upsilon}_n \circ_{j=1} \boldsymbol{\Upsilon}_n, \boldsymbol{\Delta} \rangle = \langle \boldsymbol{\upsilon}_n, \boldsymbol{\Delta}_{(1)} \boldsymbol{\vartheta}_n \rangle$。令 $\mathbf{c} = (\boldsymbol{\vartheta}_T^\top, \cdots, \boldsymbol{\vartheta}_{P+1}^\top)^\top$，并令

$$\begin{aligned}
\sup_{\boldsymbol{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)} \sum_{n=P+1}^{T} \langle \boldsymbol{\upsilon}_n, \boldsymbol{\Delta}_{(1)} \boldsymbol{\vartheta}_n \rangle &= \sup_{\boldsymbol{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)} \boldsymbol{\upsilon}^\top (\mathbf{I}_{T-P} \otimes \boldsymbol{\Delta}_{(1)}) \mathbf{c} \\
&\leqslant \sup_{\boldsymbol{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)} CNP^{1/2} T \eta_{\max}^2
\end{aligned}$$

为 $\|\boldsymbol{\upsilon}\|_2 \leqslant C\sqrt{NT}\eta_{\max}$和 $\|\mathbf{c}\|_2 \leqslant \sqrt{NTP}\eta_{\max}$。因此，令 $x = m^{1/8}T^{\delta}/N^{1/2}P^{1/4}$位于 (A.25) 处，我们有

$$\mathbb{P}\left[\sup_{\boldsymbol{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)}\sum_{n=P+1}^{T}\langle\boldsymbol{\upsilon}_n,\boldsymbol{\Delta}_{(1)}\boldsymbol{\vartheta}_n\rangle \geqslant \frac{T^{1+2\delta}}{m^{1/4}}\right] \leqslant \frac{CN^2P^{1/2}}{m^{1/4}T^{2\delta-1}}. \tag{A.36}$$

结合（A.34）、（A.35）和（A.36），即可得到完整的偏差界，

$$\mathbb{P}\left[\sup_{\boldsymbol{\Delta}\in\mathcal{S}(2r_1,2r_2,2r_3)}\frac{1}{T}\sum_{n=P+1}^{T}\langle\boldsymbol{\epsilon}_n,\boldsymbol{\Delta}_{(1)}\widetilde{\mathbf{x}}_n\rangle \geqslant C(\kappa^2\sqrt{\lambda_{\max}(\boldsymbol{\Sigma}_{\varepsilon})\kappa_U}+\kappa\sqrt{\kappa_U})\sqrt{\frac{d_{\mathcal{M}}}{T}}+\frac{T^{2\delta}}{m^{1/4}}\right]$$
$$\leqslant \exp(-Cd_{\mathcal{M}})+2\exp(-CT(\kappa_L/\kappa_U)^2\min\{\kappa^{-2},\kappa^{-4}\}).$$

至此，证明完成。 □

引理 A.4。设三个 *Tucker* 秩为 $(r_1^{(j)}$、$r_2^{(j)}$、$r_3^{(j)})$，其中 $1 \leqslant j \leqslant 3$，并假设 $r_i^{(1)} < r_i^{(2)} < r_i^{(0)}$ 满足 $1 \leqslant i \leqslant 3$。对于 $\boldsymbol{\mathcal{X}} \in \boldsymbol{\Theta}(r_1^{(0)}$、$r_2^{(0)}$、$r_3^{(0)})$，有

$$\|P_{\boldsymbol{\Theta}(r_1^{(2)},r_2^{(2)},r_3^{(2)})}(\boldsymbol{\mathcal{X}})-\boldsymbol{\mathcal{X}}\|_{\mathrm{F}} \leqslant [\Pi_{i=1}^{3}(\beta_i+1)-1]\|\boldsymbol{\mathcal{Y}}-\boldsymbol{\mathcal{X}}\|_{\mathrm{F}} \tag{A.37}$$

对于任意 $\boldsymbol{\mathcal{Y}} \in \boldsymbol{\Theta}(r_1^{(1)}$、$r_2^{(1)}$、$r_3^{(1)})$，其中 $\beta_i = \sqrt{(r_i^{(0)}-r_i^{(2)})/(r_i^{(0)}-r_i^{(1)})}$。

引理 *A.4* 的证明。技术性证明请参见 Chen 等人（2019）。 □