arXiv:2603.24833v1 [stat.ME] 25 Mar 2026

具有侧信息的稳健矩阵估计*

AnishAgarwal，Jungjun Choi†，Ming Yuan

哥伦比亚大学工业工程与运筹研究系；罗德岛大学计算机科学与统计学系；哥伦比亚大学统计学系

2026年3月27日

摘要

我们提出了一个灵活的高维矩阵估计框架，将行和列的辅助信息纳入其中。现有方法（如归纳式矩阵补全）往往施加了严格的结构限制，例如要求协变量交互项精确低秩、协变量效应为线性，并且难以利用仅由一侧（行或列）解释或两侧均无法解释的分量；此外，这些方法通常没有明确纳入噪声分量。为克服这些局限，我们将底层矩阵分解为四个互补分量之和：行特征与列特征之间（可能为非线性的）交互项、由行特征驱动的分量、由列特征驱动的分量，以及无法由观测到的特征解释的残差低秩结构。通过结合基于筛法的投影与核范数惩罚，我们可以分别估计各个分量，再将估计出的分量汇总，得到最终估计。我们推导了收敛速率，揭示了该方法在不同模型配置下的稳健性，具体取决于辅助信息所含信息量的多少。我们进一步将该方法扩展到在随机缺失和非随机缺失机制下的部分观测矩阵，其中包括受因果面板数据启发的块状缺失模式。模拟研究和烟草销售真实数据应用表明，利用辅助信息能够提高插补准确度，并且相较于标准低秩方法和基于谱的方法替代方案，还能提升处理效应估计的表现。

关键词：矩阵补全、核范数惩罚、因果推断、非线性估计、协变量信息

*本研究由美国国家科学基金会（NSF）项目 DMS-2052955 资助。

†通信地址：计算机科学与统计学系，泰勒楼，9 Greenhouse Road，罗德岛州金斯顿 02881，电子邮箱：jungjun.choi@uri.edu

1 引言

近年来的技术进步使人们能够收集和处理海量数据，这些数据通常以矩阵形式便于组织，且矩阵的两个维度往往都在快速扩展。因此，矩阵去噪和矩阵补全等高维矩阵估计问题受到了广泛关注，统计和计算领域也都取得了许多令人瞩目的成果。然而，尽管除目标结果数据外通常还可以获得辅助信息，传统方法在进行矩阵估计时通常只使用结果数据。纳入辅助信息能够丰富底层模型，并提高估计和预测准确度。随着我们获取辅助协变量数据的能力不断增强，开发能够有效利用辅助信息的矩阵估计方法，已成为一项重要且及时的研究方向。

近年来，人们提出了许多计算算法，并研究了它们的统计性质。可以说，在矩阵估计中纳入额外信息的最流行模型是归纳式矩阵补全（IMC）模型（例如，Xu 等人，2013；Jain 和 Dhillon，2013；Zhang 等人，2018年）。标准的 IMC 模型形式如下：

$$Y = M = XLZ^\top$$

其中，$Y = [y_{it}]_{i\le N,\ t\le T}$是结果矩阵，$X = [x_1, \ldots, x_N]^\top$ 是 $N \times d_1$ 行特征矩阵，$Z = [z_1, \ldots, z_T]^\top$ 是 $T \times d_2$ 列特征矩阵，而 M是目标矩阵。这里，L 假设为低秩 $d_1 \times d_2$ 矩阵。一种典型的估计方法是求解

$$\min_{L\in\mathbb{R}^{d_1\times d_2}} \left\|\mathcal{P}_\Omega(XLZ^\top - Y)\right\|_F^2 + \lambda \|L\|_*,$$

其中，$||\cdot||_*$表示核范数，$\mathcal{P}_\Omega(A) = \Omega \circ A$，$\Omega$ 是 $N \times T$指示矩阵

用于矩阵补全中的可观测性。

尽管 IMC 是一种流行且实用的利用侧信息进行矩阵估计的方法，但它也存在若干重要局限。首先，它要求两侧都具备特征，且这些特征之间存在线性相互作用。此外，该方法依赖行特征和列特征都具有信息性；如果特征较弱或无关，方法就可能失效。近年来，人们提出了 IMC 的多种扩展，以解决这些不足。 Ledent 等人 (2023) 将噪声分量 E 纳入归纳式矩阵补全模型，并推导了期望 ℓ风险的界。 Zhong 等人 (2019) 允许 (X, Z) 与 M 之间存在非线性关系。 Wang 和 Elhamifar (2018 年) 研究了 L 的秩可能较大的情形。值得注意的是，Chiang 等人(2015) 提出了所谓的“脏”IMC 模型，在标准归纳式矩阵补全形式中加入一个额外的低秩项 R，并通过求解L, R) 来估计

$$\min_{L,R} \left\|\mathcal{P}_\Omega(XLZ^\top + R - Y)\right\|_F^2 + \lambda_1 \|L\|_* + \lambda_2 \|R\|_* .$$

然而，该模型不包含噪声项 E 以及仅由单侧特征解释的分量。此外，它不允许 (X, Z) 与 M 之间存在非线性关系，并且仍要求 L 为低秩。总的来说，每种扩展都只解决了部分局限性，其他问题仍未得到解决。

另一项关于带协变量的矩阵补全的重要研究方向包括Mao 等人 (2019) 和 Ma 等人 (2025)。这些论文考虑如下模型

$$Y = XB^\top + R + E$$

其中， B 是未知系数矩阵， R 是低秩矩阵。由于该方法没有纳入列特征 Z，因此无法捕捉同时涉及 X 和 Z 的交互项（例如 $XLZ^\top$），也无法捕捉仅由 Z解释的分量。此外，该方法不支持非线性效应。因此，该模型仍存在一些

局限性。

最后，另一类相关研究探讨了在不存在缺失数据的情形下进行主成分分析或因子分析（例如，Fan 等（2016）；Chiang 等人（2016）；Niranjan 等人（2017）；Zhu 等人（2016）；Xue 等人（2017））。例如，Fan 等（2016）考虑模型 $Y = (G(X) + \Gamma)F^\top + E = G(X)F^\top + \Gamma F^\top + E$，其中 $G(X)$ 是载荷的一部分，由 X 的未知函数定义，而 F 表示未观测因子。相比之下，Zhu 等人（2016）研究模型 $Y = XB^\top + AZ^\top + R + E$，其中 R 为低秩。前一种框架无法捕捉同时涉及 X 和 Z 的交互项（例如 $XLZ^\top$），也无法捕捉仅由 Z 解释的分量；而后一种框架不包含 X 与 Z 之间的交互项。此外，该模型主要强调线性关系。

为克服现有方法的局限性，我们考虑以下模型：

$$Y = M + E, \qquad M = M_1 + M_2 + M_3 + M_4, \tag{1}$$
$$M_1 = G_1(X)Q_1(Z)^\top, \quad M_2 = G_2(X)V_1^\top, \quad M_3 = W_1Q_2(Z)^\top, \quad M_4 = W_2V_2^\top,$$

其中，G_1、G_2、Q_1 和 Q_2 是未知矩阵值函数，而 W_1、W_2、V_1 和 V_2 是未观测矩阵。该模型更加一般，并涵盖了上述模型。例如，Xu 等人（2013）；Jain 和 Dhillon（2013）；Wang 和 Elhamifar（2018）的模型对应于特殊情形 $M = M_1$，而 Chiang 等人（2015）的模型对应于 $M = M_1 + M_4$。此外，Fan 等（2016）；Mao 等人（2019）；Ma 等人（2025）的模型可视为 $M = M_2 + M_4$ 的特殊情形。例如，Fan 等（2016）的模型可表示为 $M = M_2 + M_4$，其中 $V_1 = V_2$。因此，在该模型下，我们的估计方法不太可能受到模型设定错误的影响。此外，如果数据包含现有模型未考虑的分量，我们的估计量预计会优于基于这些受限模型的估计量。正如第 3 节所述，我们估计量的收敛速率展示了该方法在不同模型下的稳健性，而第 5 节的模拟结果也与

这些理论发现一致。

我们的估计基于筛投影法。利用由筛基构造的投影矩阵来处理 X 和 Z，我们分别估计 M 的各个分量，然后将这些估计相加，得到 M 的估计量。借助筛投影，我们的方法能够处理 X 和 Z 对 M可能产生的非线性效应。此外，分别估计各个分量，使我们能够充分利用（1）中的模型结构。这些特点共同使我们的估计量在数据包含受限模型未能考虑的分量时，更有可能优于基于更严格模型的方法。

我们方法的另一项重要特点是采用核范数惩罚，这对应于一个软阈值程序。因此，如果 M_2、 M_3和 M_4中的某些项恰好为零，那么以高概率计，我们对这些分量的估计也恰好为零。这一性质增强了估计量的稳健性。

相比之下，如果采用谱方法估计各个分量，就必须估计每个部分的秩；当相应分量因信噪比较低而较弱时，现有的秩估计量可能会产生错误的（非零）估计。因此，当 M_2、M_3和 M_4 中的某些项为零或接近零时，谱方法可能表现不佳。相比之下，我们的核范数惩罚估计量无需估计各个分量的秩或信号强度；因此， M_2、 M_3和 M_4 的值较小并不会造成问题。

本文的另一项重要贡献是，我们还考虑了存在缺失条目且缺失机制为非随机缺失的情形。据我们所知，尽管许多论文利用辅助信息在随机缺失（MAR）条件下进行插补，但尚无现有的矩阵补全研究在非随机缺失（MNAR）条件下纳入辅助信息。自Athey 等人（2021）的开创性工作证明矩阵补全技术对因果面板数据模型非常有用以来，矩阵补全已成为一种广受欢迎的工具，用于估计

未处理（对照）条件下的潜在结果。然而，未处理条件下的潜在结果矩阵通常呈现出不符合随机缺失机制的缺失模式。因此，近年来，非随机缺失条件下的矩阵补全及其在因果推断中的应用得到了广泛研究（例如，参见Athey 等人（2021）；Bai 和 Ng（2021）；Agarwal 等人（2023）；Choi 和 Yuan（2024）；Yan 和 Wainwright（2024））。我们提出了一种新颖的矩阵补全方法，可在非随机缺失条件下利用辅助信息。如第5.2节的真实数据实验所示，我们的方法在插补未观测到的潜在结果方面优于标准矩阵补全方法，并展现了其在处理效应估计中的实用性。

本文其余部分安排如下。第2节介绍我们的模型和估计方法，该方法采用带核范数惩罚的筛投影。第3节给出估计量的渐近误差界，并讨论我们的方法在不同模型下的稳健性。第4 节将我们的估计策略扩展到结果矩阵部分观测的情形。值得注意的是，我们同时考虑非随机缺失设定和随机缺失设定。第5 节通过模拟数据和真实数据的数值研究展示我们方法的优势。受篇幅限制，所有证明均放在补充材料中。

2 模型与估计

本文考虑如下面板模型：

$$Y = M + E, \qquad M = M_1 + M_2 + M_3 + M_4,$$

其中，$Y = (y_{it})_{i \le N,\ t \le T}$是结果矩阵，$E = (\epsilon_{it})_{i \le N,\ t \le T}$是噪声矩阵，而$M = (m_{it})_{i \le N,\ t \le T}$ 是我们感兴趣的矩阵。我们将 M 分解为四个部分：(i) M_1，一个能由 X 和 Z共同良好解释的分量；(ii) M_2，一个能由 X 解释、但与 Z无关的分量；(iii) M_3，一个能由 Z 解释、但与 X无关的分量；以及 (iv) M_4，一个

与 X 和 Z均无关的分量，其中 $X = (x_i)_{i \le N}$ 和 $Z = (z_t)_{t \le T}$ 分别是对应于行索引和列索引的可观测特征。

更具体地说，每个部分都可以表示为

$$M_1 = G_1(X)Q_1(Z)^\top, \quad M_2 = G_2(X)V_1^\top, \quad M_3 = W_1Q_2(Z)^\top, \quad M_4 = W_2V_2^\top, \tag{2}$$

其中，$G_1(X) = (g_{1,k}(x_i))_{i \le N, k \le K_1}$, $G_2(X) = (g_{2,k}(x_i))_{i \le N, k \le K_2}$, $Q_1(Z) = (q_{1,k}(z_t))_{t \le T, k \le K_1}$, 以及 $Q_2(Z) = (q_{2,k}(z_t))_{t \le T, k \le K_3}$ 均对应某些未知函数 $g_{1,k}(\cdot)$、$g_{2,k}(\cdot)$、$q_{1,k}(\cdot)$ 和 $q_{2,k}(\cdot)$。这里，$W_1 = (w_{1,ik})_{i \le N, k \le K_3}$ 和 $W_2 = (w_{2,ik})_{i \le N, k \le K_4}$ 表示无法由 X 解释的分量，而 $V_1 = (v_{1,tk})_{t \le T, k \le K_2}$ 和 $V_2 = (v_{2,tk})_{t \le T, k \le K_4}$ 表示无法由 Z 解释的分量。该模型具有一般性，涵盖了许多现有模型。

估计。为恰当纳入并利用模型（2）中的结构，我们提出使用筛投影法来估计 M。对于两组基函数 $\{\phi_1(x) \ldots, \phi_J(x)\}$和 $\{\psi_1(z) \ldots, \psi_J(z)\}$（例如，B样条、傅里叶级数、小波或多项式级数），定义

$$\boldsymbol{\phi}(x_i) = \left[\phi_1(x_{i1}), \ldots, \phi_J(x_{i1}), \ldots, \phi_1(x_{id_1}), \ldots, \phi_J(x_{id_1})\right]^\top \in \mathbb{R}^{Jd_1},$$
$$\boldsymbol{\psi}(z_t) = \left[\psi_1(z_{t1}), \ldots, \psi_J(z_{t1}), \ldots, \psi_1(z_{td_2}), \ldots, \psi_J(z_{td_2})\right]^\top \in \mathbb{R}^{Jd_2},$$

其中，d_1 和 d_2 分别是 x_i 和 z_t的维数。相应的投影矩阵为

$$P_X = \Phi(X)\big(\Phi(X)^\top\Phi(X)\big)^{-1}\Phi(X)^\top, \qquad P_Z = \Psi(Z)\big(\Psi(Z)^\top\Psi(Z)\big)^{-1}\Psi(Z)^\top,$$

其中 $\Phi(X) = \left[\boldsymbol{\phi}(x_1) \ldots, \boldsymbol{\phi}(x_N)\right]^\top$ 以及 $\Psi(Z) = \left[\boldsymbol{\psi}(z_1) \ldots, \boldsymbol{\psi}(z_T)\right]^\top$。注意，只要 $g_{1,k}(\cdot)$、$g_{2,k}(\cdot)$、$q_{1,k}(\cdot)$ 和 $q_{2,k}(\cdot)$ 足够光滑，对于任意 $\iota \in (1,2)$，我们都有

$$P_X G_\iota(X) \approx G_\iota(X), \qquad P_Z Q_\iota(Z) \approx Q_\iota(Z). \tag{3}$$

此外，由于 (X, Z) 与 E 之间的正交性，$\|P_X E P_Z\|_F$ 可能远小于 $\|E\|_F$。利用这一性质，我们提出如下估计 M。

算法 1估计过程

步骤 1： 计算 $\widehat{M}_1 = P_X Y P_Z$。

步骤 2： 计算以下核范数惩罚估计量：

$$\widehat{M}_2 := \arg\min_{A \in \mathbb{R}^{N\times T}} \left\|P_X Y (I_T - P_Z) - A\right\|_F^2 + \nu_2 \|A\|_*,$$

$$\widehat{M}_3 := \arg\min_{A \in \mathbb{R}^{N\times T}} \left\|(I_N - P_X) Y P_Z - A\right\|_F^2 + \nu_3 \|A\|_*,$$

$$\widehat{M}_4 := \arg\min_{A \in \mathbb{R}^{N\times T}} \left\|(I_N - P_X) Y (I_T - P_Z) - A\right\|_F^2 + \nu_4 \|A\|_*,$$

其中 $\nu_2 = C_2\sqrt{T}$、$\nu_3 = C_3\sqrt{N}$ 和 $\nu_4 = C_4\sqrt{N+T}$ 为某些足够大的常数 C_2，C_3，$C_4 > 0$。

步骤 3： 构造最终估计量 $\widehat{M} = \widehat{M}_1 + \widehat{M}_2 + \widehat{M}_3 + \widehat{M}_4$。

为理解该估计量的工作原理，注意到根据（3）和核范数惩罚的基本性质，我们有

$$\widehat{M}_1 \approx G_1 Q_1^\top + G_2 V_1^\top P_Z + P_X W_1 Q_2^\top + P_X W_2 V_2^\top P_Z,$$

$$\widehat{M}_2 \approx G_2 V_1^\top (I_T - P_Z) + P_X W_2 V_2^\top (I_T - P_Z),$$

$$\widehat{M}_3 \approx (I_N - P_X) W_1 Q_2^\top + (I_N - P_X) W_2 V_2^\top P_Z,$$

$$\widehat{M}_4 \approx (I_N - P_X) W_2 V_2^\top (I_T - P_Z),$$

在关于噪声以及$g_{1,k}(\cdot)$、$g_{2,k}(\cdot)$、$q_{1,k}(\cdot)$和$q_{2,k}(\cdot)$的光滑性满足适当条件时。重要的是，将四个估计量相加时，涉及 $P_X W_\iota$或 $P_Z V_\iota$的项会相互抵消。因此，无需对 X 和 W（或$\widehat{}$与 Z 和 V之间）施加任何正交条件，我们的最终估计量 M 仍能很好地估计 M。

此外，由于核范数惩罚相当于阈值估计量，当 M_2、M_3 和 M_4 中的某些（或全部）为零或足够小时，它有助于我们获得更紧的界。

3 渐近结果

本节将给出我们估计量的收敛速率。首先，我们提出以下条件。

假设 3.1（噪声）。随机变量 $(\epsilon_{it})_{i\leq N,\ t\leq T}$ 相互独立、均值为零、服从次高斯分布，并满足 $\mathbb{E}[\epsilon_{it}^2] \leq \sigma^2 \leq C_1$ 和 $\mathbb{E}[\exp(s\epsilon_{it})] \leq \exp(C_2 s^2\sigma^2)$，对所有 $s \in \mathbb{R}$均成立，其中常数为 C_1，$C_2 > 0$。此外，$(\epsilon_{it})_{i\leq N,\ t\leq T}$与 X和 Z相互独立。

独立性和次高斯性假设用于推导 $\|P_X E P_Z\|$，$\|P_X E\|$和 $\|E P_Z\|$的紧确界。我们可以将此条件推广到具有适当矩条件的弱相关噪声，但需要在定理 3.1 的界中增加依赖于 J 的项。

假设 3.2（基函数）。(i) 存在常数 c，$C > 0$ 使得，以趋于一的概率，

$$c < \lambda_{\min}\big(N^{-1}\Phi(X)^\top\Phi(X)\big) \leq \lambda_{\max}\big(N^{-1}\Phi(X)^\top\Phi(X)\big) < C,$$
$$c < \lambda_{\min}\big(T^{-1}\Psi(Z)^\top\Psi(Z)\big) \leq \lambda_{\max}\big(T^{-1}\Psi(Z)^\top\Psi(Z)\big) < C,$$

其中，$\lambda_{\max}(A)$ 和 $\lambda_{\min}(A)$ 分别表示 A 的最大和最小奇异值。(ii) $\max_{j\leq J,\ i\leq N,\ l\leq d_1}\mathbb{E}[\phi_j(x_{il})^4] < \infty$ 以及 $\max_{j\leq J,\ t\leq T,\ l\leq d_2}\mathbb{E}[\psi_j(z_{tl})^4] < \infty$。

这一条件在筛估计文献中很常见（例如，Fan 等人(2016)；Chen 等人(2023)）。由于我们关注的是 $Jd_1 \ll N$和 $Jd_2 \ll T$ 的情形，因此根据大数定律，该条件成立，因而限制并不苛刻。

假设 3.3（筛逼近）。(i) 存在常数 $\gamma_1^G, \gamma_2^G, \gamma_1^Q, \gamma_2^Q \geq 2$，使得对于某些筛系数向量 $b_{1,k}, b_{2,k} \in \mathbb{R}^{Jd_1}$ 和 $a_{1,k}, a_{2,k} \in \mathbb{R}^{Jd_2}$，筛逼近满足

$$\sup_{x\in\mathcal{X}}\left|g_{\iota,k}(x) - b_{\iota,k}^\top\boldsymbol{\phi}(x)\right| = O\left(J^{-\gamma_\iota^G}\right), \qquad \sup_{z\in\mathcal{Z}}\left|q_{\iota,k}(z) - a_{\iota,k}^\top\boldsymbol{\psi}(z)\right| = O\left(J^{-\gamma_\iota^Q}\right),$$

其中 $\mathcal{X}$ 和 $\mathcal{Z}$ 分别是 x_i 和 z_t的支撑集。

(ii) 筛维数 J 满足

$$\sqrt{T}\,K_1/J^{\gamma_1^G}\to 0,\qquad \sqrt{N}\,K_1/J^{\gamma_1^Q}\to 0,\qquad \max\{\sqrt{N},\sqrt{T}\}\,K_2/J^{\gamma_2^G}\to 0,$$
$$\max\{\sqrt{N},\sqrt{T}\}\,K_3/J^{\gamma_2^Q}\to 0.$$

条件 (i) 是筛估计中的标准假设。例如，若 $g_{\iota,k}(\cdot)$ 具有加性形式 $g_{\iota,k}(x_i)=\sum_{l=1}^{d_1} g_{\iota,kl}(x_{il})$，且每个 $g_{\iota,kl}(\cdot)$ 都属于 H¨older 类$\mathcal{H}(\rho_\iota^G,\tau_\iota^G)$，其中

$$\mathcal{H}(\rho,\tau)=\left\{h:\left|h^{(\rho)}(s)-h^{(\rho)}(t)\right|\le C|s-t|^{\tau}\right\}$$

对于某个 $C>0$，则 $\gamma_\iota^G=\rho_\iota^G+\tau_\iota^G$对于典型的基函数选择（例如参见 Chen(2007）)。另一方面，条件 (ii) 要求函数$g_{\iota,k}(\cdot)$ 和 $q_{\iota,k}(\cdot)$ 具有足够的光滑性。注意，γ_ι^G 和 γ_ι^Q 可分别视为 $g_{\iota,k}(\cdot)$ 和$q_{\iota,k}(\cdot)$ 的光滑度参数。因此，若 $g_{\iota,k}(\cdot)$ 和 $q_{\iota,k}(\cdot)$ 足够光滑，则 γ_ι^G 和γ_ι^Q 会较大，即使 J 增长缓慢，条件 (ii) 也能得到满足。

最后，我们施加以下矩条件。

假设 3.4（矩）。（i）对于所有 i和 t，$\mathbb{E}[m_{it}^4]$ 有界。

（ii）存在一个常数 $C_1>0$，使得对于所有 i，t，k，

$$\mathbb{E}\big[g_{1,k}^2(x_i)\big],\ \mathbb{E}\big[g_{2,k}^2(x_i)\big],\ \mathbb{E}\big[q_{1,k}^2(z_t)\big],\ \mathbb{E}\big[q_{2,k}^2(z_t)\big]\le C_1.$$

（iii）存在一个常数 $C_2>0$，使得对于所有 i，t，k，

$$\mathbb{E}[w_{1,ik}^2],\ \mathbb{E}[w_{2,ik}^2],\ \mathbb{E}[v_{1,tk}^2],\ \mathbb{E}[v_{2,tk}^2]\le C_2.$$

现在我们可以陈述估计量的统计性质。下面的定理给出了所提出估计量的收敛速率。

定理 3.1 (收敛速率)。 假设 *3.1–3.4* 成立。那么，

$$\|\widehat{M} - M\|_F = O_p\Bigg(J + \sqrt{K_2 + K_4}\,\min\Big\{\sqrt{T},\, \|M_2\|_F + \|P_X M_4\|_F\Big\} + \sqrt{K_3 + K_4}\,\min\Big\{\sqrt{N},\, \|M_3\|_F + \|M_4 P_Z\|_F\Big\} + \sqrt{K_4}\,\min\Big\{\sqrt{N+T},\, \|M_4\|_F\Big\} + \sqrt{NT}\left[\frac{K_1}{J^{\gamma_1^G}} + \frac{K_1}{J^{\gamma_1^Q}} + \frac{K_2}{J^{\gamma_2^G}} + \frac{K_3}{J^{\gamma_2^Q}}\right]\Bigg).$$

这里有几点显而易见的说明。首先，注意到我们估计量的误差界中起主导作用的部分不依赖于 K_1。因此，只要$g_{1,k}(\cdot)$和$q_{1,k}(\cdot)$足够光滑，我们就允许 K_1 较大。最后一项，

$$\sqrt{NT}\left[\frac{K_1}{J^{\gamma_1^G}} + \frac{K_1}{J^{\gamma_1^Q}} + \frac{K_2}{J^{\gamma_2^G}} + \frac{K_3}{J^{\gamma_2^Q}}\right],$$

源自筛逼近（平滑）误差。当函数$g_{\iota,k}(\cdot)$和$q_{\iota,k}(\cdot)$足够光滑时（即 γ_ι^G和 γ_ι^Q较大时），这一项较小，并且被其他项主导。

定理3.1 说明了我们估计量的稳健性。首先，考虑最有利的情形 $M = M_1$，此时 M 可由 X 和 Z很好地解释。在这种情况下，只要$g_{\iota,k}(\cdot)$和$q_{\iota,k}(\cdot)$足够光滑，我们估计量的收敛速率为 $O_p(J)$。这与“双重投影”估计量 $P_X Y P_Z$的速率相匹配；当我们事先知道$a\ priori\ \ M = M_1$时，该估计量最为适用。相比之下，在这种情形下，若使用核范数惩罚等标准低秩方法估计 M，其收敛速率将为 $O_p(\sqrt{K_1(N+T)})$，远大于本文方法。此外，即使 M 除了 M_1之外还包含一个额外分量 $M_2 + M_3 + M_4$，只要该分量足够小（即 $\|M_2 + M_3 + M_4\|_F \ll \sqrt{N+T}$），我们的速率$O_p\big(J + \sqrt{K^*}\,\|M_2 + M_3 + M_4\|_F\big)$ 仍小于通常的低秩估计量$O_p\big(\sqrt{K(N+T)}\big)$的速率，其中 $K^* = \max\{K_2, K_3, K_4\}$，而 K是 M 的秩。

接下来，考虑最不利情形 $M = M_4$，此时 M 与 X 无关，且

Z 辅助信息没有提供有效信息。在这种情况下，只要 $g_{\iota,\ k}(\cdot)$ 和 $q_{\iota,\ k}(\cdot)$ 足够平滑，且 $J \ll \sqrt{K_4(N+T)}$，我们的估计量的收敛速率为 $O_p(\sqrt{K_4(N+T)})$。请注意，这一速率与标准低秩估计量的速率一致。因此，即使在最不利情形下，我们的估计量的误差界也与典型低秩方法相当。相比之下，“双重投影”估计量$P_X Y P_Z$ 在这种情况下是不一致的。此外，如果 M 除了 M_4之外还包含一个较小的分量$M_1+M_2+M_3$，那么我们的估计量的收敛速率将变为 $O_p(\sqrt{K^*(N+T)})$，而典型低秩估计量的收敛速率仍为 $O_p(\sqrt{K(N+T)})$。因此，当 K_1 较大时，我们的方法可以得到更紧的误差界。

最后，考虑情形 $M = M_2$。在此情形下，只要 $g_{\iota,\ k}(\cdot)$ 和 $q_{\iota,\ k}(\cdot)$ 足够光滑且满足 $J \ll \sqrt{K_2 T}$，我们的估计量的收敛速率为 $O_p(\sqrt{K_2 T})$。相比之下，典型低秩估计量的收敛速率为 $O_p(\sqrt{K_2(N+T)})$。因此，当 $N \gg T$ 时，我们的方法能得到更紧的界。此外，如果 M 除了 M_2 之外还包含一个较小的分量 $M_1 + M_3 + M_4$，那么我们的估计量的收敛速率将变为 $O_p(\sqrt{(K_2+K_4)T})$，而典型低秩估计量的收敛速率为$O_p(\sqrt{K(N+T)})$。因此，当 K_1 较大或 $N \gg T$ 时，我们的方法能得到更好的界。对情形 $M = M_3$ 也有类似的讨论。

表 1 总结了上述情形下估计量的收敛速率。可以看出，在每种情形下，只要 $g_{\iota,\ k}(\cdot)$ 和 $q_{\iota,\ k}(\cdot)$ 足够光滑，我们的估计量的收敛速率至少不逊于其他估计量。

	$M = M_1$	$M = M_2$	$M = M_3$	$M = M_4$
我们的估计量	J	$\sqrt{K_2 T}$	$\sqrt{K_3 N}$	$\sqrt{K_4(N+T)}$
双重投影	J	$\sqrt{NT}$	$\sqrt{NT}$	$\sqrt{NT}$
低秩估计	$\sqrt{K_1(N+T)}$	$\sqrt{K_2(N+T)}$	$\sqrt{K_3(N+T)}$	$\sqrt{K_4(N+T)}$

表1：矩阵估计量的收敛速率

最后，作为一项推论，我们给出估计奇异向量的收敛速率，因为因子和载荷往往是研究者感兴趣的对象（参见例如 Bai 等人 (2008)）。

令 $\widehat{U} \in \mathbb{R}^{N\times K}$和 $\widehat{V} \in \mathbb{R}^{T\times K}$分别为 $\widehat{M}$的左右奇异向量。类似地，令 U 和 V 分别表示 M的左右奇异向量。对于$\widehat{\ }$记号上的方便起见，将 $\|M - M\|_F$在定理 3.1 中的上界记为

$$\mathcal{R} = J + \sqrt{K_2 + K_4}\,\min\left\{\sqrt{T}, \|M_2\|_F + \|P_X M_4\|_F\right\} + \sqrt{K_3 + K_4}\,\min\left\{\sqrt{N}, \|M_3\|_F + \|M_4 P_Z\|_F\right\}$$
$$+ \sqrt{K_4}\,\min\left\{\sqrt{N+T}, \|M_4\|_F\right\} + \sqrt{NT}\left(\frac{K_1}{J^{\gamma_1^G}} + \frac{K_1}{J^{\gamma_1^Q}} + \frac{K_2}{J^{\gamma_2^G}} + \frac{K_3}{J^{\gamma_2^Q}}\right).$$

于是，我们得到以下收敛速率。

推论 3.2. 假设条件 *3.1–3.4* 成立。此外，假设 $\mathcal{R}/\lambda_{\min} \to_p 0$，其中 $\lambda_{\min}$ 表示 M的最小非零奇异值。

$$\max\left\{\min_{R\in\mathcal{O}_{K\times K}} \|\widehat{U} - RU\|_F,\quad \min_{R\in\mathcal{O}_{K\times K}} \|\widehat{V} - RV\|_F\right\} = O_p\left(\frac{\mathcal{R}}{\lambda_{\min}}\right).$$

4 扩展至缺失情形

接下来，我们将估计策略扩展到结果矩阵仅被部分观测到的情形。基础模型与第2节相同，此外还假设研究人员观测到 $\Omega \circ Y$而非 Y，其中 $\Omega = (\omega_{it})_{i\le N,\ t\le T} \in \{0,1\}^{N\times T}$。

4.1 随机缺失情形

本节考虑结果条目随机缺失的情形。具体而言，我们假设$(\omega_{it})_{i\le N,\ t\le T}$ 是均值为 p 的独立同分布的伯努利随机变量，这在矩阵补全文献中很常见。

估计。 与上述完全观测情形类似，我们使用投影方法来利用模型（2）的结构。然而，存在缺失条目时，一个关键难点是，当我们试图通过核范数惩罚来估计 $\circ\big(P_X Y(I_T - P_Z)\big)$和 $\circ\big((I_N - P_X)Y P_Z\big)$时，无法直接观测到 Ω $P_X M(I_T - P_Z)$ 或 $(I_N - P_X)MP_Z$。

另一方面，我们仍然可以使用投影估计量$p^{-1}P_X(\Omega \circ Y)P_Z$准确地估计 $P_X M P_Z$。

因此，存在缺失条目的情况下，我们按以下步骤进行。

算法 2随机缺失情形的估计过程

步骤 1： 推导 $\widehat{M_1} = p^{-1}P_X(\Omega \circ Y)P_Z$。

步骤 2： 对 $\Omega \circ (Y - M_1)$ 采用核范数惩罚：

$$\widehat{M}_{rest} := \underset{A:\|A\|_\infty \le M_{\max}}{\arg\min} \left\| \Omega \circ (Y - \widehat{M_1} - A) \right\|_F^2 + \nu \|A\|_*,$$

其中 $M_{\max} > 0$ 是某个较大的常数，且 $\nu = Cp^{1/2}\sqrt{N + T}$，其中常数为 $C > 0$.

步骤 3： 得到最终估计量 $M = M_1 + M_{rest}$。

注意，根据投影关系（3）以及矩阵补全中核范数惩罚的常见性质，我们有

$$\widehat{M_1} \approx G_1 Q_1^\top + G_2 V_1^\top P_Z + P_X W_1 Q_2^\top + P_X W_2 V_2^\top P_Z,$$
$$\widehat{M}_{\text{rest}} \approx G_2 V_1^\top (I_T - P_Z) + (I_N - P_X) W_1 Q_2^\top + W_2 V_2^\top - P_X W_2 V_2^\top P_Z,$$

在噪声以及 $g_{1,k}(\cdot)$、$g_{2,k}(\cdot)$、$q_{1,k}(\cdot)$ 和 $q_{2,k}(\cdot)$ 的光滑性方面满足与上一节类似的条件时。如上所述，将两个估计量相加时，涉及 $P_X W_\iota$ 或 $P_Z V_\iota$ 的项会相互抵消。因此，我们的最终估计量 M 可以很好地估计 M。具体而言，由于我们使用投影方法而非低秩估计量来估计 M_1 分量，因此当 K_1 较大，或当 M_1 相对于 M_2、M_3 和 M_4 较大时，我们的方法可能具有优势。

渐近结果。现在我们给出估计量的收敛速率。首先介绍几个额外的假设。

假设 4.1（随机缺失）。随机变量 $(\omega_{it})_{i \le N,\ t \le T}$ 服从参数为 $\mathbb{E}[\omega_{it}] = p$ 的独立同分布伯努利分布。此外，Ω 与 E、X、Z 和 M 相互独立。

除了假设 4.1之外，我们还要求筛逼近误差满足一个略有不同的条件，该条件不同于完全观测情形下的条件。

假设 4.2 (筛逼近)。 (i) 假设 3.3 (i) 成立。

(ii) 筛逼近满足

$$\frac{\min\{\sqrt{N},\sqrt{T}\}}{\sqrt{p}}\left(\frac{K_1}{J^{\gamma_1^G}}+\frac{K_1}{J^{\gamma_1^Q}}+\frac{K_2}{J^{\gamma_2^G}}+\frac{K_3}{J^{\gamma_2^Q}}\right)\to 0.$$

以下定理给出了我们估计量的收敛速率。

定理 4.1 (随机缺失情形下的收敛速率)。 假设 *3.1*、 *3.2*、*3.4*、 *4.1*和 *4.2* 均成立。此外，假设 $J \ll p\sqrt{N+T}$。那么，若

$$M_{\max} \geq \left\|M_{\text{rest}} - M_2 P_Z - P_X M_3 - P_X M_4 P_Z\right\|_\infty,$$

我们有

$$\|\widehat{M}-M\|_F = O_p\left(\frac{J}{\sqrt{p}}+\sqrt{K^*}\,\min\left\{\frac{\sqrt{N+T}\,(1+M_{\max})}{\sqrt{p}},\,\|M_2\|_F+\|M_3\|_F+\|M_4\|_F\right\}\right.$$
$$\left.+\sqrt{NT}\left[\frac{K_1}{J^{\gamma_1^G}}+\frac{K_1}{J^{\gamma_1^Q}}+\frac{K_2}{J^{\gamma_2^G}}+\frac{K_3}{J^{\gamma_2^Q}}\right]\right),$$

其中 $K^* =$ 最大值$\{K_2, K_3, K_4\}$。

与完全观测情形类似的讨论同样适用于此处。只要 K_1 as long as$g_{1,k}(\cdot)$ 和 $q_{1,k}(\cdot)$ 足够平滑，我们的估计量的误差界就不依赖于 M_1 ，因为我们使用投影方法而非低秩估计量进行估计。因此，我们可以允许 K_1 tobelarge。

此外，该估计量具有稳健性。为简单起见，假设$M_{\max}$ 和 $\|M\|_\infty$ 有界。首先，考虑最理想的情况，即 $M = M_1$。在这种情况下，只要 $g_{\iota,k}(\cdot)$ 和$q_{\iota,k}(\cdot)$ 足够光滑，我们的估计量的收敛速率为 $O_p(J/\sqrt{p})$。相比之下，如果我们使用标准低秩补全

方法（例如，核范数惩罚）来估计 O_p，收敛速率将为 $O_p(\sqrt{K_1(N+T)/p})$，远高于本文方法。此外，即使 M 包含一个超出 M_1的额外分量$M_2+M_3+M_4$，只要该分量足够小（即 $\|M_2+M_3+M_4\|_F \ll \sqrt{(N+T)/p}$），我们的速率

$$O_p\Big(J+\sqrt{K^*}\,\|M_2+M_3+M_4\|_F\Big)$$

小于通常低秩补全方法的 $O_p(\sqrt{K(N+T)/p})$，其中K 是 M 的秩，且 $K^*=\max\{K_2,K_3,K_4\}$。

接下来，考虑最不利情形 $M=M_4$，其中 M 与 X 和Z无关。在这种情况下，只要$g_{\iota,k}(\cdot)$和$q_{\iota,k}(\cdot)$足够光滑，我们的估计量的收敛速率为 $O_p(\sqrt{K_4(N+T)/p})$。注意，该速率与标准低秩补全方法的收敛速率一致。因此，即使在最不利情形下，我们的估计量也与典型的低秩补全方法相当。此外，如果 M 除 M_4之外还包含一个较小的分量 $M_1+M_2+M_3$，那么我们的估计量的收敛速率将变为 $O_p(\sqrt{K^*(N+T)/p})$，而典型低秩补全方法的收敛速率为 $O_p(\sqrt{K(N+T)/p})$。因此，当 K_1 较大时，我们的方法能够给出更紧的界。对于 $M=M_2$ 和 $M=M_3$的情况，也有类似的讨论：我们的估计量达到与标准低秩补全方法相同的收敛速率；当 M 还包含一个 K_1较大的 M_1 分量时，则能给出更好的界。

4.2 非随机缺失情形

尽管随机缺失假设在矩阵补全文献中很常见，但对于某些重要应用而言，这一假设可能并不恰当，例如在因果面板模型中填补对照组潜在结果。在这类模型中，处理会从某个时点开始分配给部分个体（也可能以交错方式分配）。在这种情形下，将缺失模式视为固定的（即非随机的）可能更为恰当。

遵循非随机缺失（MNAR）条件下矩阵补全领域的相关文献，例如，Bai 和 Ng(2021)；Choi 和 Yuan (2024)；Yan 和 Wainwright (2024)，我们假设

缺失模式如图 1 所示。在这种情形下，“miss”子矩阵中的所有（或部分）条目均未观测到，而“tall”和“wide”子矩阵中的所有条目均已观测到。这种模式在因果面板数据中十分常见：“wide”子矩阵对应于对照组在所有时间段内的观测值，“tall”子矩阵对应于所有个体在处理前时期的观测值，此时结果为对照条件下的潜在结果。

图1：非随机缺失情形下的缺失模式

$$Y = \begin{pmatrix} \checkmark & \checkmark & \checkmark & ? & ? & ? & \checkmark & ? \\ \checkmark & \checkmark & \checkmark & ? & ? & ? & ? & ? \\ \checkmark & \checkmark & \checkmark & ? & \checkmark & \text{miss} & ? & ? \\ \checkmark & \text{tall} & \checkmark & ? & ? & ? & ? & \checkmark \\ \checkmark & \checkmark & \checkmark & \vdots & \iddots & ? & ? & ? \\ \checkmark & \checkmark & \checkmark & ? & \cdots & ? & \checkmark & ? \\ \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark \\ \vdots & \iddots & \checkmark & \checkmark & \text{wide} & \checkmark & \checkmark & \checkmark \\ \checkmark & \cdots & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark & \checkmark \end{pmatrix}$$

估计。注意，“高”子矩阵和“宽”子矩阵中的条目均被完整观测。因此，我们可以对“高”子矩阵和“宽”子矩阵应用算法1 来估计$M_{\text{tall}} = (m_{it})_{i \le N,\ t \le T_0}$以及 $M_{\text{wide}} = (m_{it})_{i \le N_0,\ t \le T}$。然后，如推论3.2所述，我们可以分别估计 M_{tall} 和 M_{wide}的左右奇异向量。重要的是，M_{tall} 和 M 的左奇异向量张成相同的空间。同样，M_{wide} 和 M 的右奇异向量张成相同的空间。因此，通过结合 M_{tall} 的左奇异向量估计量与 M_{wide}的右奇异向量估计量，并进行适当的旋转调整，我们可以估计 M。

具体而言，我们按如下方式估计 M 。

算法 3非随机缺失情形的估计过程

步骤 1： 从“高”的子矩阵 $Y_{\text{tall}} = (y_{it})_{i \le N,\ t \le T_0}$，使用算法1求得并计算其左奇异向量 $\widehat{U}_{\text{tall}} \in \mathbb{R}^{N \times K}$。

步骤 2： 从“宽”子矩阵 $Y_{\text{wide}} = (y_{it})_{i \le N_0,\ t \le T}$，使用 M_{wide} 算法进行估计1并计算其左右奇异向量 $\widehat{U}_{\text{wide}} \in \mathbb{R}^{N_0 \times K}$ 和 $\widehat{V}_{\text{wide}} \in \mathbb{R}^{T \times K}$，以及对应的奇异值 $\widehat{D}_{\text{wide}} \in \mathbb{R}^{K \times K}$。

步骤 3： 通过对 U_{tall} 中对应于 $i \le N_0$ 的子矩阵回归 $\widehat{U}_{\text{wide}}$，求得旋转矩阵 $\widehat{H}_{\text{adj}} \in \mathbb{R}^{K \times K}$。

步骤 4： 构造最终估计量 $\widehat{M} = \widehat{U}_{\text{tall}} \widehat{H}_{\text{adj}} \widehat{D}_{\text{wide}} \widehat{V}^{\top}$宽。

由于该估计量是基于完全观测情形的算法 1 构建的，我们预期它具有与第 3 节讨论的相似优势。

渐近结果。为了更准确地说明这一点，我们给出估计量的收敛速率。我们先提出一个额外的假设。

假设 4.3（分块非相干性）。分别用 u_i 和 v_t 表示 $U^{\top}$ 的第 i 列和 $V^{\top}$ 的第 t 列。那么存在常数 c_1，$c_2 > 0$，使得当样本量趋于无穷时，以趋近于 1 的概率，

$$c_1 \le \lambda_{\min}\left(\frac{N}{N_0}\sum_{i \le N_0} u_i u_i^{\top}\right) \le \lambda_{\max}\left(\frac{N}{N_0}\sum_{i \le N_0} u_i u_i^{\top}\right) \le c_2,$$
$$c_1 \le \lambda_{\min}\left(\frac{T}{T_0}\sum_{t \le T_0} v_t v_t^{\top}\right) \le \lambda_{\max}\left(\frac{T}{T_0}\sum_{t \le T_0} v_t v_t^{\top}\right) \le c_2.$$

这一假设可视为一种非相干条件，它确保 M 的左奇异向量不会被处理组单位或未处理组单位中的任何一方所主导，且右奇异向量不会被处理前时期或处理后时期中的任何一方所主导。这在非随机缺失矩阵补全文献中很常见（例如 Bai 和 Ng(2021) 中的假设 D，以及 Choi 和 Yuan (2024) 中的定理 3.1(v)），并使我们能够将子矩阵 M_{tall} 和 M_{wide} 的性质与 M的性质联系起来。例如，若 $\{u_i\}_{i \le N}$ 是平稳的，

then

$$\frac{N}{N_0}\sum_{i\le N_0} u_i u_i^\top \approx \frac{1}{N}\sum_{i\le N} u_i u_i^\top = I_K,$$

且该条件得到满足。

现在我们可以陈述估计量的收敛速率。令 $\mathcal{R}_{\text{tall}}$ 和 $\mathcal{R}_{\text{wide}}$ 分别表示 $\|M_{\text{tall}} - M_{\text{tall}}\|_F$ 和 $\|M_{\text{wide}} - M_{\text{wide}}\|_F$ 的上界，如定理 3.1所示：

$$\begin{aligned}
\mathcal{R}_{\text{tall}} &= J + \sqrt{K_2+K_4}\min\left\{\sqrt{T_0}, \|M_{2,\text{tall}}\|_F + \|P_X M_{4,\text{tall}}\|_F\right\} \\
&\quad + \sqrt{K_3+K_4}\min\left\{\sqrt{N}, \|M_{3,\text{tall}}\|_F + \|M_{4,\text{tall}} P_{Z,\text{sub}}\|_F\right\} \\
&\quad + \sqrt{K_4}\min\left\{\sqrt{N+T_0}, \|M_{4,\text{tall}}\|_F\right\} + \sqrt{NT_0}\left(\frac{K_1}{J^{\gamma_1^G}} + \frac{K_1}{J^{\gamma_1^Q}} + \frac{K_2}{J^{\gamma_2^G}} + \frac{K_3}{J^{\gamma_2^Q}}\right), \\
\mathcal{R}_{\text{wide}} &= J + \sqrt{K_2+K_4}\min\left\{\sqrt{T}, \|M_{2,\text{wide}}\|_F + \|P_{X,\text{sub}} M_{4,\text{wide}}\|_F\right\} \\
&\quad + \sqrt{K_3+K_4}\min\left\{\sqrt{N_0}, \|M_{3,\text{wide}}\|_F + \|M_{4,\text{wide}} P_Z\|_F\right\} \\
&\quad + \sqrt{K_4}\min\left\{\sqrt{N_0+T}, \|M_{4,\text{wide}}\|_F\right\} + \sqrt{N_0 T}\left(\frac{K_1}{J^{\gamma_1^G}} + \frac{K_1}{J^{\gamma_1^Q}} + \frac{K_2}{J^{\gamma_2^G}} + \frac{K_3}{J^{\gamma_2^Q}}\right),
\end{aligned}$$

其中

$$P_{X,\text{sub}} = \Phi_{\text{sub}}(X)\big(\Phi_{\text{sub}}(X)^\top\Phi_{\text{sub}}(X)\big)^{-1}\Phi_{\text{sub}}(X)^\top, \quad P_{Z,\text{sub}} = \Psi_{\text{sub}}(Z)\big(\Psi_{\text{sub}}(Z)^\top\Psi_{\text{sub}}(Z)\big)^{-1}\Psi_{\text{sub}}(Z)^\top,$$

$\Phi_{\text{sub}}(X) = \big[\boldsymbol{\phi}(x_1)\ldots,\boldsymbol{\phi}(x_{N_0})\big]^\top$，以及 $\Psi_{\text{sub}}(Z) = \big[\boldsymbol{\psi}(z_1)\ldots,\boldsymbol{\psi}(z_{T_0})\big]^\top$。此外，令 $\delta_N = N_0/N$ 和 $\delta_T = T_0/T$。下面的定理给出了我们的估计量的收敛速率。

定理 4.2（非随机缺失情形下的收敛速率）。假设条件 *3.1–3.4*对子矩阵 Y_{tall} 和 Y_{wide}成立，且假设条件 *4.3* 成立。此外，假设

$$\frac{\max\{\mathcal{R}_{\text{wide}}, \mathcal{R}_{\text{tall}}\}}{\lambda_{\min}\sqrt{\delta_N\delta_T}} \to_p 0,$$

其中， $\lambda_{\min}$ 是 M的最小非零奇异值。于是，

$$\|\widehat{M} - M\|_F = O_p\left(\frac{\kappa \max\{\mathcal{R}_{\text{wide}}, \mathcal{R}_{\text{tall}}\}}{\sqrt{\delta_N \delta_T}}\right),$$

其中 $\kappa = \lambda_{\max}/\lambda_{\min}$。

定理 4.2 突显了本文估计量的优势。适用与第 3 节类似的讨论。为简单起见，考虑一个典型情形，其中 κ 有界，且$N_0 \geq cN$ 和 $T_0 \geq cT$ 满足某个 $c > 0$。首先，考虑最有利情形，即 $M = M_1$。在此情形下，只要 g_ι、$_k$(·) 和 q_ι、$_k$(·) 足够光滑，本文估计量的收敛速率为 O_p(J)。相比之下，如果使用非随机缺失低秩方法（例如 Bai 和 Ng（2021））估计子矩阵 M_{tall} 和 M_{wide}，收敛速率将为O_p($\sqrt{K_1}$($N + T$))，远高于本文方法。

接下来，考虑最不利情形，即 $M = M_4$。在此情形下，只要 $g_{\iota,\ k}$(·) 和 $q_{\iota,\ k}$(·) 足够光滑，且满足 $J \ll \sqrt{K_4}$~~($N + T$),~~ 本文估计量的收敛速率为 O_p($\sqrt{K_4}$($N + T$))。这一速率与对子矩阵 M_{tall} 和 M_{wide}应用非随机缺失低秩估计量所得的速率一致。

此外，若 $M = M_2$，在 $g_{\iota,\ k}$(·) 和 $q_{\iota,\ k}$(·) 足够光滑且 $J \ll \sqrt{K_2 T}$ 的条件下，我们的估计量的收敛速率为 O_p($\sqrt{K_2 T}$)；而低秩方法得到的速率为 O_p($\sqrt{K_2}$($N + T$))。类似地，若 $M = M_3$，我们的估计量的收敛速率为 O_p($\sqrt{K_3 N}$)，而低秩方法得到的速率为 O_p($\sqrt{K_3}$($N + T$))。因此，当 M 至少部分由可观测特征 X 和 Z 解释时，利用这些信息可以显著提高估计精度。

5 模拟实验

为展示我们方法的实际优势和有限样本表现，我们开展了几组模拟实验。

5.1 各部分相对规模的变化

首先，为研究我们的估计量相较于现有方法的相对优势如何随各分量的贡献而变化，我们调整组成部分权重，并比较不同估计量的估计性能。具体而言，我们考虑如下模型

$$M = \alpha_1 M_1 + \alpha_2 M_2 + \alpha_3 M_3 + \alpha_4 M_4,$$

其中，对于所有 $1 \le r \le 4$， $\sum_{r=1}^{4} \alpha_r = 1$ 和 $\|M_r\|_F = 2\sqrt{NT}$。我们改变 α_r 的取值，并评估这些估计量的均方误差。

数据生成过程。我们考虑八个特征，其中 $x_i = [x_{1,i}, x_{2,i}, x_{3,i}, x_{4,i}]^\top$和 $z_t = [z_{1,t}, z_{2,t}, z_{3,t}, z_{4,t}]^\top$。我们抽取 $x_{1、\ i}$，服从 i.i.d. Unif$[-1,1]$、 $x_{2；\ i}$，服从 i.i.d. Unif$[-0.5,0.5]$、 $x_{3；\ i}$，服从 i.i.d. $\mathcal{N}$(0, 0.22)；以及 $x_{4、\ i}$，服从 i.i.d. (0, 0.32)。我们以相同的方式生成 $z_{1、\ t}$、 $z_{2、\ t}$、 $z_{3、\ t}$、 $z_{4、\ t}$。

对于矩阵 M_1，我们设定

$$g_{1,k}(x_i) = b_0^{(1,k)} + \sum_{d=1}^{4} \left(b_{d,1}^{(1,k)} x_{d,i} + b_{d,2}^{(1,k)} x_{d,i}^2 + b_{d,3}^{(1,k)} x_{d,i}^3 + b_{d,4}^{(1,k)} x_{d,i}^4 \right), \qquad k \le K_1 = 17,$$

并从标准正态分布中抽取系数 $b_0^{(1,k)}$ 和 $b_{d\ ,\ j}^{(1,k)}$。类似地，我们设定

$$q_{1,k}(z_t) = a_0^{(1,k)} + \sum_{d=1}^{4} \left(a_{d,1}^{(1,k)} z_{d,t} + a_{d,2}^{(1,k)} z_{d,t}^2 + a_{d,3}^{(1,k)} z_{d,t}^3 + a_{d,4}^{(1,k)} z_{d,t}^4 \right), \qquad k \le K_1 = 17,$$

并抽 取系数 $a_0^{(1,k)}$ 和 $a_{d\ ,\ j}^{(1,k)}$，服从标准正态分布。

对于矩阵 M_2，我们设定 $K_2 = 3$，并使用与上述相同的设定生成 $G_2(X)$。此外，我们生成 $v_{1,\ t} \in \mathbb{R}^3$，使其独立同分布于 $\mathcal{N}\big(0,$ diag(0.5，1，1.5)$\big)$，并将它们堆叠成 $V_1 = [v_{1,1}, \ldots, v_{1,T}]^\top$。

对于矩阵 M_3，我们设定 $K_3 = 3$，并使用与上述相同的设定生成 $Q_2(Z)$。我们从 $\mathcal{N}\big(0、$ diag(0.5,1、1.5)$\big)$ 中独立同分布地生成 $w_{1,\ i} \in \mathbb{R}^3$，并将它们堆叠成

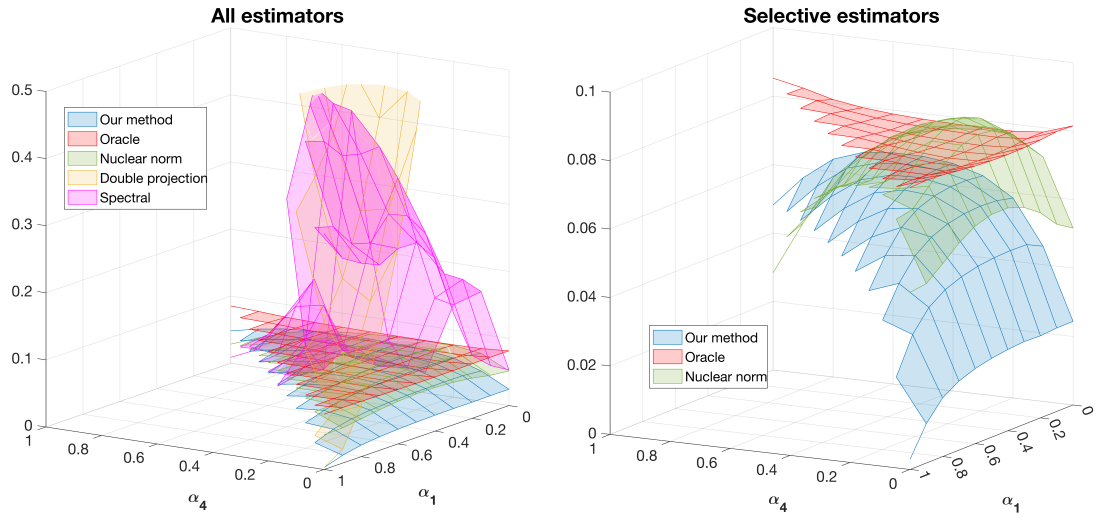
$W_1 = [w_{1,1}, \ldots, w_{1,N}]^\top$。

对于矩阵 M_4，我们从 $\mathcal{N}\big(0, \text{diag}(0.5,1,1.5)\big)$ 中独立同分布地抽取 $w_{2,i} \in \mathbb{R}^3$，并从 $\mathcal{N}\big(0, 1.5^2 I_3\big)$ 中独立同分布地抽取 $v_{2,t} \in \mathbb{R}^3$。最后，我们对所有四个矩阵进行归一化，使得 $\|M_r\|_F = 2\sqrt{NT}$对$r = 1, 2, 3, 4$ 成立。我们从 $\mathcal{N}(0, 0.5_2)$ 中独立同分布地生成噪声项。

对于完全观测情形，我们设置 $N = T = 200$。对于随机缺失（MAR）情形，我们设置 $N = T = 400$，并设置观测概率 $p = 0.6$。对于非随机缺失（MNAR）情形，我们设置 $N = T = 400$ 和 $N_0 = T_0 = 200$。

结果。这里，我们采用 $J = 5$的多项式筛，并将迭代次数设为100。我们在 $\sum_{r=1}^4 \alpha_r = 1$ 和 $\alpha_2 = \alpha_3$的限制下改变 α_r。此外，为了保持 M的秩不变，我们仅考虑对所有 $r = 1$、2、3、4均满足 $\alpha_r \geq 0.01$ 的情形。

图2：不同 α_r取值下的平均均方误差（AMSE）



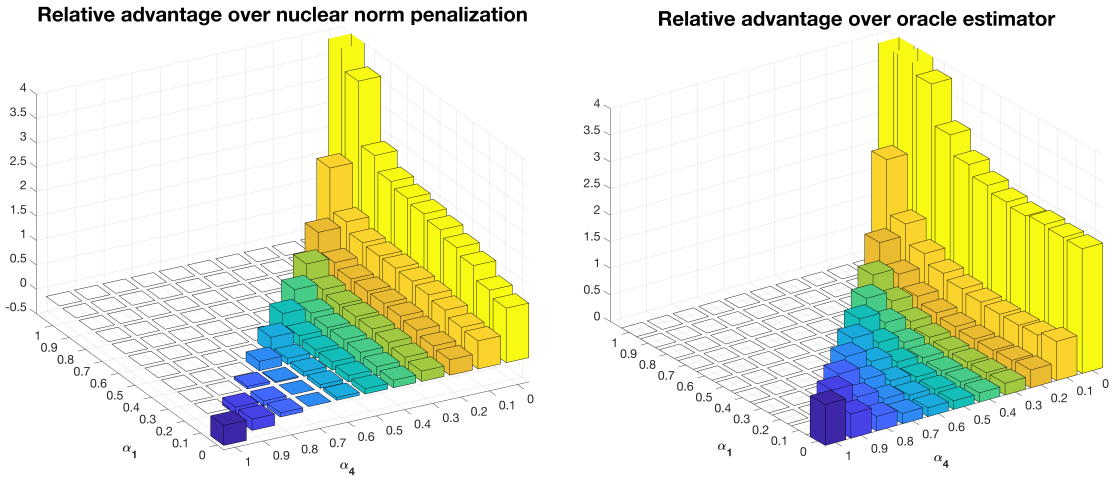
脚注：我们在$\sum_{r=1}^4 \alpha_r = 1$ 和 $\alpha_2 = \alpha_3$的限制下改变 α_r。

我们首先研究完全观测情形。图2 报告了各估计量的平均均方误差（AMSE）。这里，“Oracle”表示已知K的谱估计量；“核范数”表示普通的核范数惩罚估计量；“双重投影”表示 $P_X Y P_Z$；“谱方法”表示估计 K后的谱估计量。对于秩估计，我们采用Ahn和Horenstein（2013）提出的特征值比率法。

从左侧　　从左侧面板可以看出，总体而言，双重投影估计量和

估计秩的谱估计量相较于其他估计量表现较差。从右侧面板可以看出，我们的方法优于秩已知的谱估计量。预言估计量的平均均方误差相当稳定，不受 α_r变化的影响。相比之下，我们的方法和核范数惩罚估计量的平均均方误差会受到 α_r的显著影响。总体而言，除非 α_1 非常小且 α_4 较大，否则我们的方法优于核范数惩罚。当 α_1 较大且 α_4 较小时，我们的估计量表现尤为出色。

图3：$(AMSE_{\text{other}} - AMSE_{\text{our}})/AMSE_{\text{our}}$在 α_r取不同值时

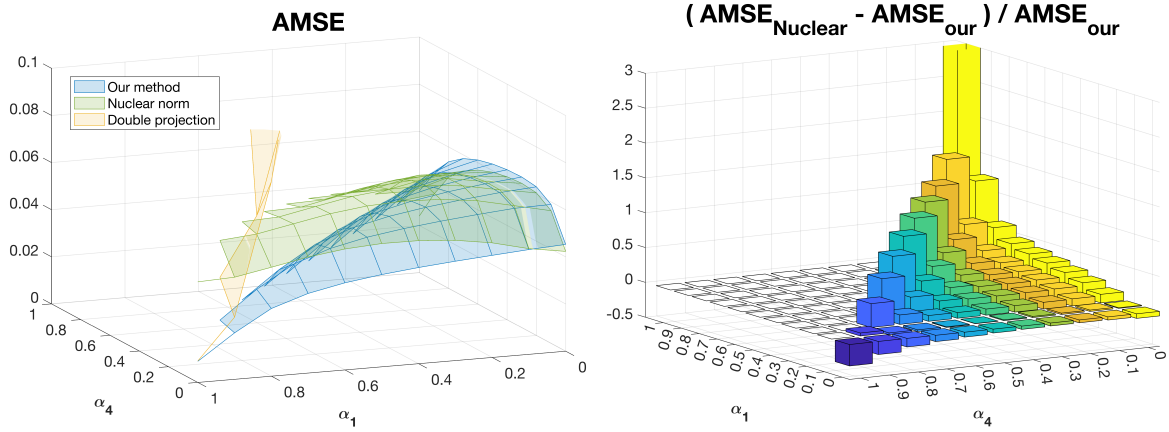


脚注：左图中， $\alpha_1 = 1$ 和 $\alpha_4 = 0.01$ 处的值为16.66。右图中， $\alpha_1 = 1$ 和 $\alpha_4 = 0.01$ 处的值为27.12。

为了评估我们的估计量相较于其他方法的相对优势，图3 绘制了 $(AMSE_{\text{other}} - AMSE_{\text{our}})/AMSE_{\text{our}}$。与核范数惩罚相比，我们的估计量的优势会随着 α_1 增大而增大。粗略来说，随着 α_4减小，优势也会增大。特别是当 α_1 接近零时（例如， $\alpha_1 = 0.01$），相对表现会随着 $\alpha_2 = \alpha_3$ 增大且 α_4 减小而改善。右图比较了我们的估计量与预言估计量，其中两者对 α_4 的依赖关系不太明确；不过，我们仍然观察到相对优势会随着 α_1增大。

接下来，我们研究随机缺失（MAR）情形。图4 报告了各估计量的平均均方误差（AMSE）及其相对表现。这里，我们

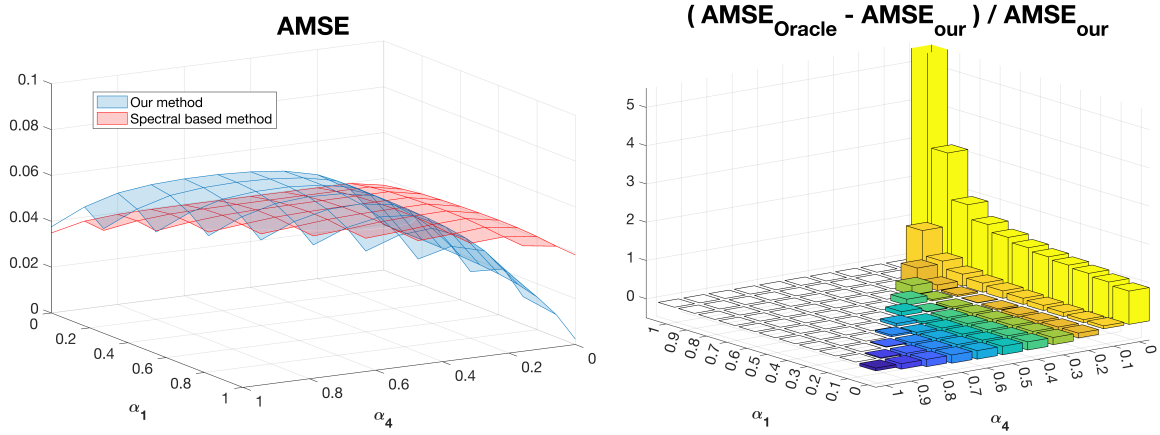
图4：随机缺失情形下的性能比较



脚注：右侧面板中， $\alpha_1 = 1$ 和 $\alpha_4 = 0.01$ 处的值为 3。77。

纳入了双重投影方法 $\left(p^{-1}P_X(\Omega \circ Y)P_Z\right)$ 以及核范数惩罚估计量，后者是随机缺失（MAR）情形下的一种标准方法。我们发现，除非 $\alpha_1 = 1$，双重投影方法的表现都很差。与核范数惩罚相比，随着 α_1 增大，我们的方法的优势也随之增大。不过，相较于完全观测情形，关于 α_4 的变化趋势并不那么明显。这可能是因为在随机缺失（MAR）情形下，我们无法分别估计 M_2、 M_3和 M_4。当 α_1 非常小且α_4 相对较大时，核范数惩罚的表现优于我们的方法。

图5：非随机缺失情形下的性能比较



脚注：右侧面板中， $\alpha_1 = 1$ 和 $\alpha_4 = 0.01$ 处的值为 15。58。

最后，我们 st 究非随机缺失（MNAR）情形。图5 报告了平均均方误差（AMSE）

以及我们估计量的相对优势。此处，我们假设已知秩，并将我们的方法与 Bai 和 Ng (2021) 针对非随机缺失设定提出的标准基于谱的估计量进行比较；Yan 和 Wainwright (2024)。我们发现，当 α_1 较大和/或 α_4 较小时，我们的方法通常优于基于谱的估计量。相反，当 α_1 较小和/或 α_4 较大时，基于谱的估计量通常表现更好。然而，当 α_1 较大（或 α_4 较小）时的相对优势，远大于当 α_1 较小（或 α_4 较大）时的相对劣势。此外，即使 α_1 接近于零，只要 α_2 和α_3 较大，我们的方法仍可能表现更好。

总而言之，在所有设置下，当 α_1 较大和/或 α_4 较小时，我们的方法明显优于其他竞争估计量。当 α_1 较小和/或 α_4较大时，与 α_1 较大和/或 α_4 较小时所取得的优势相比，我们方法的劣势相对有限。

此外，为了考察我们估计量相对于现有方法的相对优势如何随各分量的秩而变化，我们改变秩并比较估计表现。总体而言，当K_1 较大时，我们的方法明显优于其他竞争估计量。相反，当 K_1 较小且 K_4 较大时，优势则较为有限，表现与其他估计量相当。详情请参阅附录中的B 节。

5.2 烟草销售模拟实验

在本节中，我们使用Abadie 等人(2010) 的烟草销售数据开展一项真实数据实验，该数据在相关文献中被广泛使用。1988年，加利福尼亚州在美国推出了首部重大反烟草法案（第99号提案）。为研究这项法案对烟草销售的影响，Abadie 等人(2010) 使用了1970年至2000年间从38个在2000年前未实施反烟草法案的美国各州收集的人均香烟销售数据（$N = 38$，$T = 31$）。我们将这些数据编码为一个 38×31 矩阵 Y，其中的条目

y_{it} 表示在“控制”条件下，即不存在任何干预时，州 i在年份 t 的人均香烟销量（以包为单位）“潜在”结果。为生成缺失条目，我们人为地对部分州引入干预（即缺失机制）：每次迭代中，我们随机选择8个州，使其从时期 $T_0 + 1$开始接受干预（例如烟草控制计划）。重新排列矩阵后，便得到图1所示的块状缺失模式，其中包含一个 $8 \times (T - T_0)$缺失子矩阵。

对于州级特征，我们使用香烟零售价格的时间平均值、州人均个人收入的对数、18至24岁人口的比例、完成四年或以上大学教育的成年人的比例，以及人均啤酒消费量。这些变量大多取1970–2000年期间的平均值。此外，为了将其作为州总体烟草偏好的代理变量，我们使用2001年的人均香烟销量。对于年度特征，我们使用人均实际国内生产总值的对数，以及各年度各州香烟零售价格的平均值。我们还使用佛罗里达州和密歇根州的人均香烟销量平均值，作为各年度总体烟草偏好的代理变量。尽管佛罗里达州和密歇根州因在2000年之前接受过干预而未纳入上述38个州，但与其他接受干预的州相比，这些干预的影响相对较小。附录第A 节提供了这些特征构建方式的更多细节。

我们将本估计量的表现与Bai 和 Ng (2021)；以及Yan 和 Wainwright (2024)提出的基于谱的估计量进行比较，后者是处理块状缺失模式的标准方法。该方法使用谱估计量来估计“高”和“宽”子矩阵。进行秩估计时，我们采用Ahn和Horenstein (2013)的特征值比率法。在我们的方法中，投影步骤采用二阶多项式筛（$J = 2$）。我们将迭代次数设为100。

我们首先比较缺失条目上的平均均方误差（AMSE）。具体而言，在每次迭代中，我们将全部 8($T - T_0$) 个缺失条目的估计误差平方求和，再除以 8($T - T_0$)，然后对各次迭代中的这一数值取平均。我们将估计误差计算为估计值与

每个缺失条目对应的观测人均香烟销量之差。表 2的前两行报告了不同采用时间下的结果 T_0。我们发现，在所有情况下，我们的方法都优于基于谱的估计量。尤其是，当观测期数量相对较少时（即 T_0 较小时），性能差距更大。

表2：平均均方误差

目标参数	方法	$T_0 = 10$	15	20	25
每个缺失元素	Ours	239.28	219.69	217.05	181.76
	谱方法	268.28	238.57	233.12	194.19
每年缺失元素的平均值	Ours	43.56	43.67	32.19	27.01
	谱方法	50.45	48.19	34.91	27.91
所有缺失元素的平均值	Ours	32.89	35.77	25.28	23.75
	谱方法	41.35	40.41	28.69	25.47

此外，我们考察了 (i) 每年缺失条目的平均值，以及 (ii) 所有缺失条目的平均值的平均均方误差。对于逐年平均值的平均均方误差，在每次迭代中，我们计算每个干预后年份缺失条目的平均值，计算每个此类平均值的估计误差平方，将这些平方误差求和，再除以 $T - T_0$。随后，我们对各次迭代中的这一数值取平均。对于总体平均值的平均均方误差，在每次迭代中，我们计算所有缺失条目的平均值，计算其估计误差平方，然后对各次迭代的结果取平均。这类平均值目标的优势在于，结果变量中的噪声会相互抵消；因此， y_{it}的平均值接近于 m_{it}的平均值。表 2 最后四行报告了不同采用时间 T_0下的结果。我们发现，我们的方法优于基于谱的估计量，并且当观测期数量较少时（即，当 T_0 较小时），性能差距会扩大。

作为替代方案，我们还考虑烟草偏好的不同代理变量。对于各州总体烟草偏好的代理变量，在估计“宽型”子矩阵时，我们使用整个样本期内人均香烟销量的时间平均值；在估计“高型”,$\ldots, T_0$时，我们使用干预前时期 1 的平均值

子矩阵。类似地，对于各年份总体烟草偏好的代理变量，在估计“宽型”子矩阵时，我们使用每年 30 个对照州的人均香烟销量平均值；在估计“高型”子矩阵时，我们使用每年全部 38 个州的人均香烟销量平均值。

表3：平均均方误差

目标参数	方法	$T_0 = 10$	15	20	25
每个缺失元素	Ours	218.32	211.43	212.12	175.56
	谱方法	268.28	238.57	233.12	194.19
每年缺失元素的平均值	Ours	40.26	42.67	32.02	25.35
	谱方法	50.45	48.19	34.91	27.91
所有缺失元素的平均值	Ours	30.78	34.83	25.82	22.84
	谱方法	41.35	40.41	28.69	25.47

表 3 报告了将这些替代代理变量用作特征时的结果。我们发现，在大多数情况下，我们的方法表现有所提升，其相对优势也更加明显。这类代理变量与理论并不完全一致，因为它可能违反外生性条件；然而，由于该代理变量能够平均抵消结果噪声，实际中由此产生的内生性可能微不足道。在我们的实验中，使用这些代理变量得到的结果确实更为理想。

总之，实证结果表明，与典型的低秩方法相比，在因果面板设定中纳入辅助信息可以改善对控制潜在结果的估计。

6 结论

本文提出了一种灵活的高维矩阵估计框架，系统地纳入有关行和列的丰富辅助信息。该方法将信号分解为由 $(X,\ Z)$ 共同解释的分量、仅由 X 解释的分量、仅由 Z 解释的分量，以及两者都无法解释的分量，并结合筛投影与核范数惩罚来估计这些分量。我们的方法能够处理非线性协变量效应，避免对每个分量进行显式

秩选择，并自动对较弱或可忽略的信号进行阈值处理。我们推导出收敛速率，证明该方法在多种模型配置下具有稳健性：在有利情形下，其表现可与专门设计的方法相当；而在辅助信息缺乏效用时，仍具有竞争力。我们还将该方法扩展到随机缺失和非随机缺失机制下的部分观测矩阵，包括受因果面板数据启发的块缺失模式。通过模拟和烟草销售应用，我们进一步表明，利用辅助信息能够显著提高插补准确度，并改善处理效应估计。

参考文献

Abadie, A., Diamond, A., 和 Hainmueller, J.（2010）。用于比较案例研究的合成控制方法：估计加利福尼亚州烟草控制计划的效果。《美国统计学会杂志》，105(490)：493–505。

Agarwal, A., Dahleh, M., Shah, D., 和 Shen, D.（2023）。因果矩阵补全。载于第三十六届学习理论年会，第 3821–3826 页。PMLR。

Ahn, S. C. 和 Horenstein, A. R.（2013）。用于确定因子数量的特征值比检验。《计量经济学》，81(3)：1203–1227。

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., 和 Khosravi, K.（2021）。用于因果面板数据模型的矩阵补全方法。《美国统计学会杂志》，116(536)：1716–1730。

Bai, J. 和Ng, S.（2021）。矩阵补全、反事实与缺失数据的因子分析。《美国统计学会杂志》，116(536)：1746–1763。

Bai, J.、Ng, S. 等（2008）。大维因子分析。《基础与趋势》®计量经济学，3(2)：89–163。

B¨uhlmann, P. 和 Van De Geer, S.（2011）。高维数据统计：方法、理论与应用。施普林格科学与商业媒体。

Chen, Q.、Roussanov, N. 和 Wang, X.（2023）。半参数条件因子模型：估计与推断。技术报告，美国国家经济研究局。

Chen, X.（2007）。半非参数模型的大样本筛估计。《计量经济学手册》，6：5549–5632。

Chen, Y., Chi, Y., Fan, J., Ma, C., 等（2021）。数据科学的谱方法：统计学视角。机器学习基础与趋势®，14(5):566–806。

Chernozhukov, V., Hansen, C., Liao, Y., 和Zhu, Y.（2023）。低秩模型的推断。《统计年鉴》，51(3):1309–1330。

Chiang, K.-Y., Hsieh, C.-J. 和 Dhillon, I. (2016). 利用辅助信息进行稳健主成分分析。载于国际机器学习会议，第 2291–2299 页。PMLR。

Chiang, K.-Y., Hsieh, C.-J. 和 Dhillon, I. S. (2015). 利用含噪侧信息进行矩阵补全。神经信息处理系统进展，28。

Choi, J. 和 Yuan, M. (2024). 当缺失并非随机时的矩阵补全及其在因果面板数据模型中的应用。《美国统计学会杂志》，（刚刚接收）：1–24。

Fan, J., Liao, Y. 和 Wang, W. (2016). 因子模型中的投影主成分分析。《统计年鉴》，44(1):219。

Jain, P. 和 Dhillon, I. S. (2013). 可证明的归纳式矩阵补全。*arXiv* 预印本 *arXiv:1306.0626*。

Koltchinskii, V.（2011）。经验风险最小化和稀疏恢-´ ´复问题中的预言机不等式：圣弗卢尔概率论暑期学校´*XXXVIII-2008*´，第2033卷。施普林格科学与商业媒体。

Ledent, A.、Alves, R.、Lei, Y.、Guermeur, Y. 和 Kloft, M.（2023）。低噪声情形下归纳式矩阵补全的泛化界。载于《*AAAI*人工智能会议论文集》，第37卷，第8447–8455页。

Ma, S., Niu, P.-Y., Zhang, Y., and Zhu, Y.（2025）。融合辅助信息的含噪矩阵补全的统计推断。《美国统计学会杂志》，120(549):343–355。

Mao, X., Chen, S. X., and Wong, R. K.（2019）。利用协变量信息的矩阵补全。《美国统计学会杂志》，114(525):198–210。

Niranjan, U., Rajkumar, A., 和 Tulabandhula, T.（2017）。通过迭代硬阈值实现可证明的归纳式鲁棒主成分分析。*arXiv* 预印本 *arXiv:1704.00367*。

Vershynin, R.（2018年）。高维概率：一本包含数据科学应用的入门读物，第 47 卷。剑桥大学出版社。

Wang, Y. 和 Elhamifar, E.（2018年）。利用辅助信息进行高秩矩阵补全。载于人工智能促进协会人工智能会议论文集，第 32 卷。

Xu, M., Jin, R. 和 Zhou, Z.-H.（2013）。利用辅助信息加速矩阵补全：多标签学习中的应用。神经信息处理系统进展，第 26 卷。

Xue, N., Panagakis, Y. 和 Zafeiriou, S.（2017）。鲁棒主成分分析中的辅助信息：算法与应用。载于*IEEE*国际计算机视觉会议，第 4317–4325 页。

Yan, Y. 和 Wainwright, M. J.（2024）。因果面板数据的逐元素推断：一种简单且实例最优的方法。*arXiv* 预印本 *arXiv:2401.13665*。

Zhang, X.、Du, S. 和 Gu, Q.（2018）。通过多阶段普罗克鲁斯流实现快速且样本高效的归纳式矩阵补全。载于国际机器学习会议，第 5756–5765 页。PMLR。

Zhong, K.、Song, Z.、Jain, P. 和 Dhillon, I. S.（2019）。可证明的非线性归纳式矩阵补全。神经信息处理系统进展，32。

Zhu, Y., Shen, X., and Ye, C. (2016). 潜在因子模型中的个性化预测与稀疏性追求。《美国统计学会杂志》，111(513):241–252。

附录

A 数据描述

本节介绍实验中使用的数据及其来源。

- 人均香烟销量（以包计）。来源：美国疾病控制与预防中心（CDC）提供的 Orzechowski 和 Walker 所著《烟草税负》。

- 香烟平均零售价格（美元）：对于每个州，我们计算了 1970—2000 年期间香烟（年度）零售价格的平均值。这里的零售价格包括《烟草税负》（Orzechowski 和 Walker）数据中的平均成本和销售税。我们还使用消费者价格指数将其换算为 2000 年美元。来源：美国疾病控制与预防中心（CDC）提供的 Orzechowski 和 Walker 所著《烟草税负》。

- 州人均个人收入（取对数）：对于每个州，我们计算了 1970—2000 年期间州人均个人收入（年度对数值）的平均值。我们使用消费者价格指数将美国经济分析局的数据换算为 2000 年美元，并将其转换为对数值。来源：美国经济分析局（BEA）。

- 18至24岁人口比例：对于每个州，我们计算了1970年、1980年、1990年和2000年美国人口普查中18至24岁人口比例的平均值。来源：综合公共使用微观数据系列（IPUMS USA）。

- 完成四年大学学业或更高学历的成年人比例：对于每个州，我们计算了1970、1980、1990和2000年美国人口普查中完成四年大学学业或更高学历的成年人比例的平均值。来源：美国农业部经济研究局。

- 人均啤酒消费量（加仑）：由于数据始于1977年，我们计算了1977—2000年期间各州（年度）人均啤酒消费量的平均值。来源：美国国立卫生研究院下属美国国家酒精滥用与酒精中毒研究所的《监测报告第121号：表观人均酒精消费量：1977—2022年全国、州及地区趋势》。

- 人均实际国内生产总值（取对数）：我们将世界银行开放数据中的数据转换为对数值。来源：世界银行开放数据。

- 各州香烟平均零售价格（美元）：对于每一年，我们计算了38个州香烟零售价格的平均值。此处的零售价格包含《烟草税负》（Orzechowski 和 Walker）数据中的平均成本和销售税。我们还使用消费者价格指数将其换算为2000年美元。来源：美国疾病控制与预防中心（CDC）提供的 Orzechowski 和 Walker 所著《烟草税负》。

B 附加模拟：各部分秩大小的变化

为了研究我们的方法相较于现有方法的相对优势如何随各分量的秩变化，我们改变各秩并比较估计性能。具体而言，我们在约束条件 $\sum_{r=1}^{4} K_r = 15$、 $K_2 = K_3$ 和 $K_r \geq 1$ 下改变 K_r，其中对所有 r 均成立（如果 $15 -$ （$K_1 + K_4$）为奇数，则令 $K_2 = K_3 + 1$）。此外，我们固定 α_r，使得$\alpha_1 = \cdots = \alpha_4 = 0.25$。

数据生成过程。我们按照第5.1节中的相同方式生成八个特征。对于矩阵 M_1，我们设定

$$g_{1,k}(x_i) = b_0^{(1,k)} + \sum_{d=1}^{4} \left(b_{d,1}^{(1,k)} x_{d,i} + b_{d,2}^{(1,k)} x_{d,i}^2 + b_{d,3}^{(1,k)} x_{d,i}^3 \right), \qquad \text{for } k \leq K_1,$$

并从标准正态分布中抽取系数 $b_0^{(1,k)}$ 和 $b_d^{(1,k)}$，j

$$q_{1,k}(z_t) = a_0^{(1,k)} + \sum_{d=1}^{4} \left(a_{d,1}^{(1,k)} z_{d,t} + a_{d,2}^{(1,k)} z_{d,t}^2 + a_{d,3}^{(1,k)} z_{d,t}^3 \right), \qquad \text{for } k \le K_1,$$

并从标准正态分布中抽取系数 $a_0^{(1,k)}$ 和 $a_d^{(1,k)}$，j。

对于矩阵 M_2，我们按照上述相同的设定生成 $G_2(X)$。对于$V_1 = [v_{1,1}, \ldots, v_{1,T}]^\top$，我们从$v_{1,\ t} \in \mathbb{R}^{K_2}$中独立同分布地抽取前 K_2 个坐标

$$\mathcal{N}\big(0, \operatorname{diag}(1, 0.75^2, 1.25^2, 0.5^2, 1.5^2, 0.25^2, 1.75^2)\big). \tag{4}$$

类似地，对于矩阵 M_3，我们按照上述相同的设定生成 $Q_2(Z)$，并从$w_{1,\ i} \in \mathbb{R}^{K_3}$ 中独立同分布地抽取前 K_3 个坐标（4）。

For 对于矩阵 M_4，我们从 $w_{2,\ i} \in \mathbb{R}^{K_4}$ 中独立同分布地抽取前 K_4 个坐标

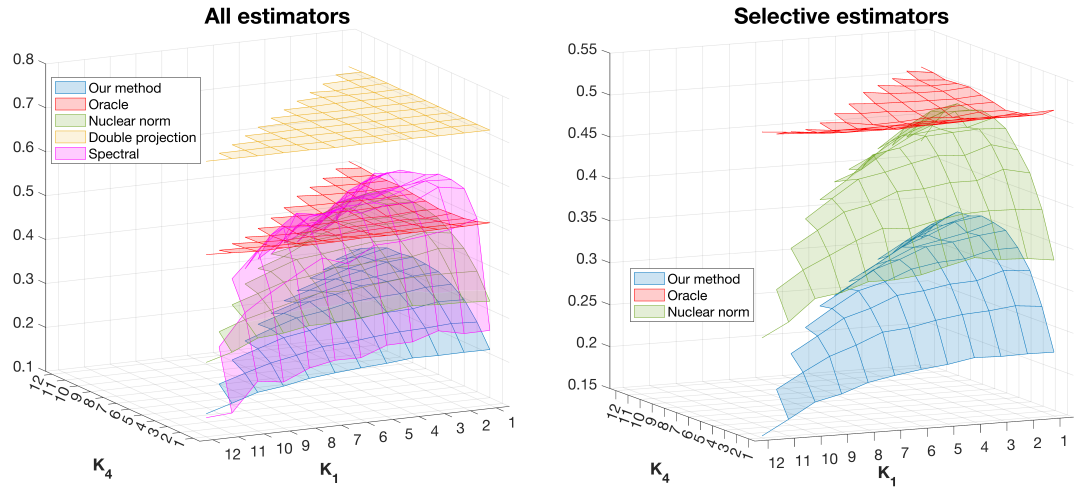
$$\mathcal{N}\Big(0, \operatorname{diag}(1, 0.75^2, 1.25^2, 0.75^2, 1.25^2, 0.5^2, 1.5^2, 0.5^2, 1.5^2, 0.25^2, 1.75^2, 0.25^2)\Big),$$

并从 $v_{2,\ t} \in \mathbb{R}^{K_4}$中独立同分布地抽取 $\mathcal{N}(0,\ 1.5^2 I_{K_4})$。最后，我们对所有矩阵进行归一化，使得 $\|M_r\|_F = 2\sqrt{NT}$对于 $r = 1$、2、3、4 均成立。我们从 $\mathcal{N}(0,\ 1.5_2)$中独立同分布地抽取噪声项。

结果。 我们采用多项式筛，参数为 $J = 4$。迭代次数和样本量与第 5.1 节相同。我们首先研究完全观测情形。图 6 报告了各估计量的平均均方误差（AMSE）。我们考虑与第 5.1 节相同的一组估计量。预言估计量是已知 K 时的谱估计量。

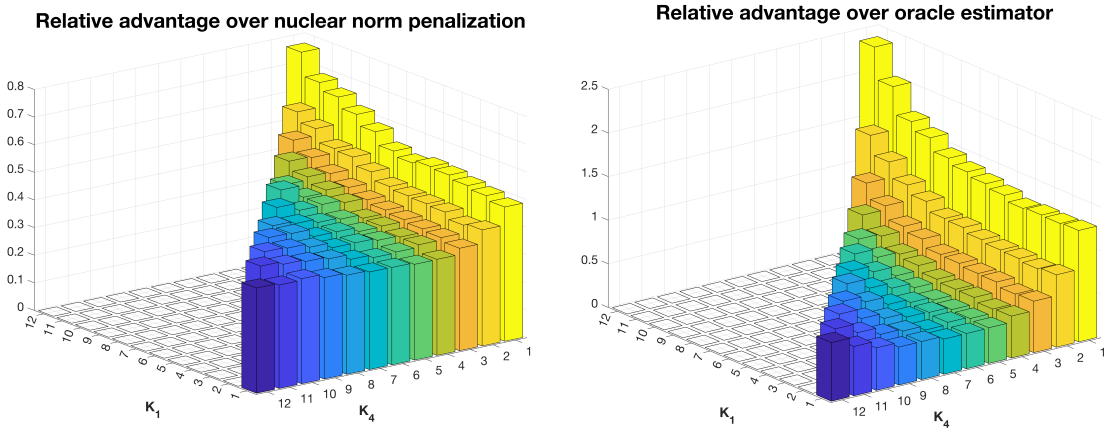
从左图可以看出，双重投影估计量表现不佳，而采用估计秩的谱估计量表现不稳定。从右图可以看出，我们的方法优于其他估计量。此外，Oracle 估计量的平均均方误差相当稳定，不受 K_r 变化的影响，而平均均方误差

图6：不同 K_r 取值下的平均均方误差



脚注：我们改变 K_r，同时满足 $\sum_{r=1}^{4} K_r = 15$ 和 $K_2 = K_3$。

图7：$(AMSE_{\text{other}} - AMSE_{\text{our}})/AMSE_{\text{our}}$ 在不同 K_r 取值下



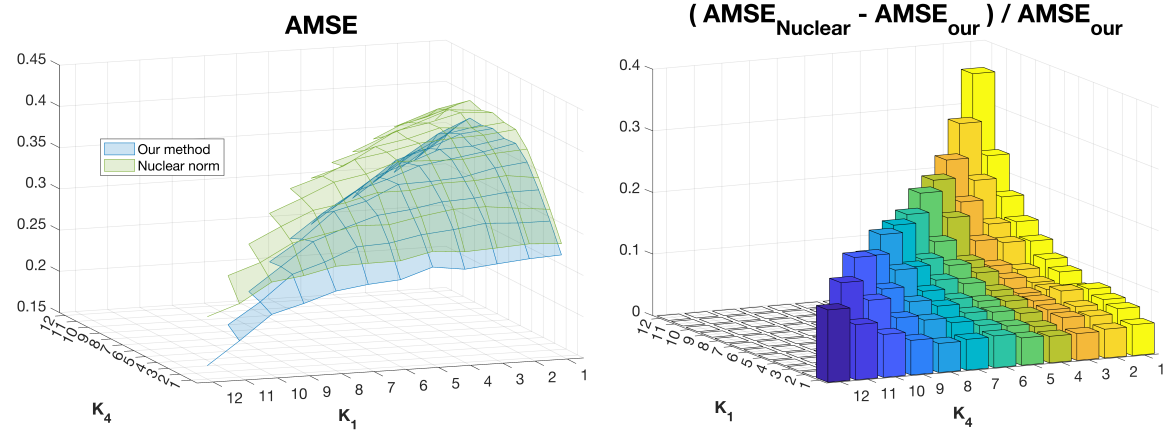
脚注：我们改变 K_r，同时满足 $\sum_{r=1}^{4} K_r = 15$ 和 $K_2 = K_3$。

我们的方法和核范数惩罚的表现随 K_r 变化。总体而言，随着 K_1 增大，我们的方法往往表现更好。

为了进一步评估我们估计量的相对优势，图7 绘制了 $(AMSE_{\text{other}} - AMSE_{\text{our}})/AMSE_{\text{our}}$。与核范数惩罚相比，我们估计量的优势会随着 K_1 的增大而增大。粗略来说，该优势也会随着 K_4的减小而增大。右图比较了我们的估计量与预言估计量，其中的

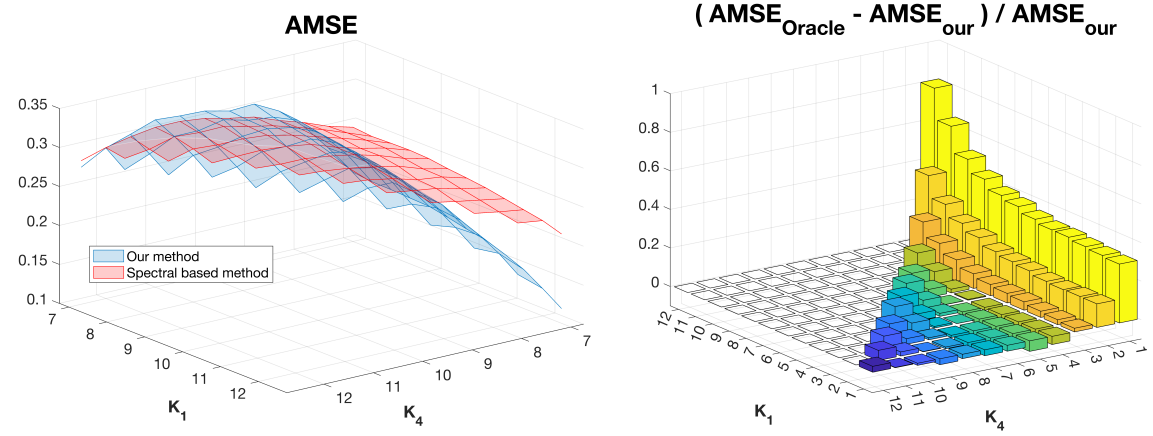
对 K_4 的依赖关系不太明确；不过，我们仍然观察到相对优势会随着 K_1的增大而增大。

图8：随机缺失情形下的性能比较



接下来，我们研究随机缺失（MAR）情形。图8 报告了平均均方误差（AMSE），以及我们的方法与核范数惩罚的平均均方误差之比；核范数惩罚是在随机缺失设定下的标准估计量。与核范数惩罚相比，我们的方法表现更好，且随着 K_1 增加，其优势也随之增大。然而，关于 K_4 的变化规律则不那么明晰。

图9：非随机缺失情形下的性能比较



脚注：在右侧面板中， $\alpha_1 = 1$ 和 $\alpha_4 = 0.01$ 处的值为15。58。

最后，我们研 研究非随机缺失（MNAR）情形。图9 报告了平均均方误差

以及我们估计量的相对优势。这里，我们考虑与第5.1节相同的估计量。基于谱的方法指的是使用预言估计量来估计 M_{tall} 和 M_{wide}。

当 K_1 较大和/或 K_4 较小时，我们的方法通常优于基于谱的方法。反之，当 K_1 较小和/或 K_4 较大时，基于谱的方法表现更好。不过，前一种情况下的相对优势远大于后一种情况下的相对劣势。总之，在所有观测模式下，当 K_1 较大时，我们的方法明显优于竞争估计量。当 K_1 较小时，我们的方法所表现出的劣势，相较于 K_1 较大时观察到的收益而言相对较小。

C 证明

C.1 主要结果的证明

C.1.1 定理 3.1 的证明

对于 $\iota \in \{1,2\}$，定义筛逼近误差矩阵为 $R_{G_\iota} = G_\iota(X) - \Phi(X)B_\iota$和 $R_{Q_\iota} = Q_\iota(Z) - \Psi(Z)A_\iota$，其中 A_ι 和 B_ι 分别是由假设 3.3中的a_ι和k 组成的筛系数矩阵。首先，我们注意到

$$\begin{aligned}\widehat{M}_1 &= P_X G_1 Q_1^\top P_Z + P_X G_2 V_1^\top P_Z + P_X W_1 Q_2^\top P_Z + P_X W_2 V_2^\top P_Z + P_X E P_Z \\ &= G_1 Q_1^\top + G_2 V_1^\top P_Z + P_X W_1 Q_2^\top + P_X W_2 V_2^\top P_Z + P_X E P_Z + \textit{smoothing error_1},\end{aligned} \tag{5}$$

其中

$$\begin{aligned}\text{平滑误差}\ \ 1 =& (P_X - I_N) R_{G_1} Q_1^\top + G_1 R_{Q_1}^\top (P_Z - I_T) + (P_X - I_N) R_{G_1} R_{Q_1}^\top (P_Z - I_T) \\ &+ (P_X - I_N) R_{G_2} V^\top P_Z + P_X W_1 R_{Q_2}^\top (P_Z - I_T).\end{aligned}$$

接着，由于根据假设 3.3，我们有 $\|R_{G_1}\|_F = O_p\left(\frac{\sqrt{NK_1}}{J^{\gamma_1^G}}\right)$，$\|R_{G_2}\|_F = O_p\left(\frac{\sqrt{NK_2}}{J^{\gamma_1^G}}\right)$，$\|R_{Q_1}\|_F = O_p\left(\frac{\sqrt{TK_1}}{J^{\gamma_1^Q}}\right)$，并且 $\|R_{Q_2}\|_F = O_p\left(\frac{\sqrt{TK_3}}{J^{\gamma_2^Q}}\right)$，我们知道

$$\|smoothing\ error_1\|_F = O_p\left(\frac{\sqrt{NT}K_1}{J^{\gamma_1^G}} + \frac{\sqrt{NT}K_1}{J^{\gamma_1^Q}} + \frac{\sqrt{NT}K_2}{J^{\gamma_2^G}} + \frac{\sqrt{NT}K_3}{J^{\gamma_2^Q}}\right). \tag{6}$$

接下来，对于 M_2，注意到

$$\begin{aligned} P_X Y(I_T - P_Z) &= P_X G_1 Q_1^\top (I_T - P_Z) + P_X G_2 V_1^\top (I_T - P_Z) + P_X W_1 Q_2^\top (I_T - P_Z) \\ &\quad + P_X W_2 V_2^\top (I_T - P_Z) + P_X E(I_T - P_Z) \\ &= G_2 V_1^\top (I_T - P_Z) + P_X W_2 V_2^\top (I_T - P_Z) + P_X E(I_T - P_Z) + smoothing\ error_2, \end{aligned}$$

其中

$$smoothing\ error_2 = P_X G_1 R_{Q_1}^\top (I_T - P_Z) + (P_X - I_N) R_{G_2} V_1^\top (I_T - P_Z) + P_X W_1 R_{Q_2}^\top (I_T - P_Z).$$

根据假设 3.3，我们有

$$\|smoothing\ error_2\| = O_p\left(\frac{\sqrt{NT}K_1}{J^{\gamma_1^Q}} + \frac{\sqrt{NT}K_2}{J^{\gamma_2^G}} + \frac{\sqrt{NT}K_3}{J^{\gamma_2^Q}}\right) = o_p\left(\sqrt{T}\right).$$

然后，由于 $\|P_X E(I_T - P_X)\| \lesssim \sqrt{T}$ 根据引理 C.4 以高概率成立，因此我们有 $\|P_X E(I_T - P_X) + 平滑误差\ 2\| \le \nu_2 = C_2\sqrt{T}$ 对于某个较大的 $C_2 > 0$ 以高概率成立。因此，通过设定 $S = P_X E(I_T - P_X) +$ 平滑误差 2 以及 $L = M_2(I_T - P_Z) + P_X M_4(I_T - P_Z)$，根据引理 C.1 可得

$$\left\|\widehat{M}_2 - M_2(I_T - P_Z) - P_X M_4(I_T - P_Z)\right\|_F = O_p\left(\sqrt{K_2 + K_4}\min\{\sqrt{T}, \|M_2\|_F + \|P_X M_4\|_F\}\right). \tag{7}$$

^对于 M_3，注意到

$$(I_N - P_X)YP_Z = (I_N - P_X)W_1Q_2^\top + (I_N - P_X)W_2V_2^\top P_Z + (I_N - P_X)EP_Z + \textit{smoothing error_3},$$

其中

$$\textit{smoothing error_3} = (I_N - P_X)R_{G_1}Q_1^\top P_Z + (I_N - P_X)R_{G_2}V_1^\top P_Z + (I_N - P_X)W_1R_{Q_2}^\top(P_Z - I_T).$$

根据假设3.3，我们有

$$\|\textit{smoothing error_3}\| = O_p\left(\frac{\sqrt{NT}K_1}{J^{\gamma_1^Q}} + \frac{\sqrt{NT}K_2}{J^{\gamma_2^G}} + \frac{\sqrt{NT}K_3}{J^{\gamma_2^Q}}\right) = o_p\left(\sqrt{N}\right).$$

接下来，由于 $\|(I_N - P_X)EP_X\| \lesssim \sqrt{N}$根据引理C.4以高概率成立，因此我们有$\|($
$I_N - P_X)EP_X +$ 平滑误差$3\| \leq \nu_3 = C_3\sqrt{N}$ 对于某个较大的 $C_3 > 0$ 以高概率成立。因此，令 $S = (I_N - P_X)EP_X +$ 平滑误差 3 以及 $L = (I_N - P_X)W_1Q_2^\top + (I_N - P_X)W_2V_2^\top P_Z$，由引理C.1可得

$$\left\|\widehat{M}_3 - (I_N - P_X)M_3 - (I_N - P_X)M_4P_Z\right\|_F = O_p\left(\sqrt{K_3 + K_4}\min\{\sqrt{N}, \|M_3\|_F + \|M_4P_Z\|_F\}\right). \tag{8}$$

^最后，对于 M_4，注意到

$$(I_N - P_X)Y(I_T - P_Z) = (I_N - P_X)W_2V_2^\top(I_T - P_Z) + (I_N - P_X)E(I_T - P_Z) + \textit{smoothing error_4},$$

其中

$$\begin{aligned}\textit{smoothing error_4} = {} & (I_N - P_X)R_{G_1}R_{Q_1}^\top(I_T - P_Z) + (I_N - P_X)R_{G_2}V_1^\top(I_T - P_Z) \\ & + (I_N - P_X)W_1R_{Q_2}^\top(I_T - P_Z).\end{aligned}$$

因此，根据假设 3.3，我们有

$$\|smoothing\ error_4\| = O_p\left(\frac{\sqrt{NT}K_1}{J^{(\gamma_1^G+\gamma_1^Q)}} + \frac{\sqrt{NT}K_2}{J^{\gamma_2^G}} + \frac{\sqrt{NT}K_3}{J^{\gamma_2^Q}}\right) = o_p\left(\sqrt{N}+\sqrt{T}\right).$$

然后，由引理 C.4 可知，$\|(I_N - P_X)E(I_T - P_Z)\| \lesssim \sqrt{N} + \sqrt{T}$以高概率成立，因此我们有 $\|(I_N - P_X)E(I_T - P_Z)$ +平滑误差 $4\| \le \nu_4 = C_4\sqrt{N+T}$对某个较大的 $C_4 > 0$ 以高概率成立。因此，通过设定 S =$(I_N - P_X)E(I_T - P_Z)$ +平滑误差 4 和 L =$(I_N - P_X)W_2V_2^\top(I_T - P_Z)$，由引理 C.1 可推得

$$\left\|\widehat{M}_4 - (I_N - P_X)M_4(I_T - P_Z)\right\|_F = O_p\left(\sqrt{K_4}\min\{\sqrt{N+T}, \|M_4\|_F\}\right). \tag{9}$$

由关系式（5），可得

$$\begin{aligned}\widehat{M} - M &= \widehat{M}_1 + \widehat{M}_2 + \widehat{M}_3 + \widehat{M}_4 - (M_1 + M_2 + M_3 + M_4)\\ &= \widehat{M}_2 - M_2(I_T - P_Z) - P_X M_4(I_T - P_Z) + \widehat{M}_3 - (I_N - P_X)M_3 - (I_N - P_X)M_4P_Z\\ &\quad + \widehat{M}_4 - (I_N - P_X)M_4(I_T - P_Z) + P_X E P_Z + smoothing\ error_1.\end{aligned}$$

因此，根据界（6）、（7）、（8）和（9），结合引理（C.4），可得

$$\begin{aligned}\left\|\widehat{M} - M\right\|_F = O_p\Big(&J + \sqrt{K_2+K_4}\min\left\{\sqrt{T}, \|M_2\|_F + \|P_X M_4\|_F\right\}\\ &+ \sqrt{K_3+K_4}\min\left\{\sqrt{N}, \|M_3\|_F + \|M_4P_Z\|_F\right\} + \sqrt{K_4}\min\left\{\sqrt{N+T}, \|M_4\|_F\right\}\\ &+\sqrt{NT}\left[\frac{K_1}{J^{\gamma_1^G}} + \frac{K_1}{J^{\gamma_1^Q}} + \frac{K_2}{J^{\gamma_2^G}} + \frac{K_3}{J^{\gamma_2^Q}}\right]\Big). \ \square\end{aligned}$$

C.1.2 推论 3.2的证明

这很容易由 Davis-Kahan 定理推得（例如，参见Chen 等人（2021））。 □

C.1.3 定理的证明 4.1

首先，注意到

$$\begin{aligned}\widehat{M}_1 &= \frac{1}{p}P_X(\Omega\circ M)P_Z + \frac{1}{p}P_X(\Omega\circ E)P_Z \\ &= P_X M P_Z + \frac{1}{p}P_X((\Omega - p\mathbf{1}\mathbf{1}^\top)\circ M)P_Z + \frac{1}{p}P_X(\Omega\circ E)P_Z \\ &= G_1Q_1^\top + G_2V_1^\top P_Z + P_XW_1Q_2^\top + P_XW_2V_2^\top P_Z + \textit{smoothing error_1} \\ &\quad + \frac{1}{p}P_X((\Omega - p\mathbf{1}\mathbf{1}^\top)\circ M)P_Z + \frac{1}{p}P_X(\Omega\circ E)P_Z,\end{aligned} \tag{10}$$

其中平滑误差 *1*定义于(5) 中。注意

$$\left\|\frac{1}{p}P_X((\Omega - p\mathbf{1}\mathbf{1}^\top)\circ M)P_Z\right\|_F \le \frac{1}{p}\left\|\Phi(\Phi^\top\Phi)^{-1}\right\|\left\|\Phi^\top((\Omega - p\mathbf{1}\mathbf{1}^\top)\circ M)\Psi\right\|_F\left\|\Psi(\Psi^\top\Psi)^{-1}\right\|.$$

根据假设 3.2，我们知道 $\left\|\Phi(\Phi^\top\Phi)^{-1}\right\| \lesssim \frac{1}{\sqrt{N}}$ 且 $\left\|\Psi(\Psi^\top\Psi)^{-1}\right\| \lesssim \frac{1}{\sqrt{T}}$。此外，因为

$$\begin{aligned}\mathbb{E}\left[\left\|\Phi^\top((\Omega - p\mathbf{1}\mathbf{1}^\top)\circ M)\Psi\right\|_F^2 \middle| M, X, Z\right] &= \sum_{j_1,j_2}\mathbb{E}\left[\left(\sum_{it}(\omega_{it}-p)m_{it}\phi_{i,j_1}\psi_{t,j_2}\right)^2 \middle| M, X, Z\right] \\ &= \sum_{j_1,j_2}\sum_{it}\mathbb{E}[(\omega_{it}-p)^2]m_{it}^2\phi_{i,j_1}^2\psi_{t,j_2}^2 \\ &= p\sum_{j_1,j_2}\sum_{it}m_{it}^2\phi_{i,j_1}^2\psi_{t,j_2}^2 = p\sum_{it}m_{it}^2\left\|\phi_i\right\|^2\left\|\psi_t\right\|^2 \\ &= O_p(pNTJ^2)\end{aligned}$$

根据假设，$\mathbb{E}[m_{it}^4]$、$\mathbb{E}[\phi_{ij}^4]$ 和 $\mathbb{E}[\psi_{tj}^4]$ 有界，我们有 $\left\|\Phi^\top((\Omega - p\mathbf{1}\mathbf{1}^\top)\circ M)\Psi\right\|_F = O_p(\sqrt{pNTJ})$。因此，我们有

$$\left\|\frac{1}{p}P_X((\Omega - p\mathbf{1}\mathbf{1}^\top)\circ M)P_Z\right\|_F = O_p\left(\frac{J}{\sqrt{p}}\right).$$

同理，我们有 $\left\|\Phi^{\top}(\Omega \circ E)\Psi\right\|_F = O_p(\sqrt{pNTJ})$，并且

$$\left\|\frac{1}{p}P_X(\Omega \circ E)P_Z\right\|_F = O_p\left(\frac{J}{\sqrt{p}}\right).$$

此外，由（10），我们有

$$\begin{aligned} Y - \widehat{M}_1 = {} & G_2V_1^{\top}(I_T - P_Z) + (I_N - P_X)W_1V_2^{\top} + W_2V_2^{\top} - P_XW_2V_2^{\top}P_Z \\ & + E - \frac{1}{p}P_X((\Omega - p\mathbf{1}\mathbf{1}^{\top}) \circ M)P_Z - \frac{1}{p}P_X(\Omega \circ E)P_Z - \textit{smoothing error_1}. \end{aligned}$$

注意，在我们的假设下，$\left\|\frac{1}{p}P_X((\Omega - p\mathbf{1}\mathbf{1}^{\top}) \circ M)P_Z\right\|_F$，$\left\|\frac{1}{p}P_X(\Omega \circ E)P_Z\right\|_F$，且平滑误差 $\textit{1}\|_F$为 $o_p(p^{1/2}\sqrt{N+T})$。因此，我们有

$$\|\Omega \circ S'\| \le \|\Omega \circ S'\|_F \le \|S'\|_F = o_p\left(p^{1/2}\sqrt{N+T}\right)$$

其中 $S' = \frac{1}{p}P_X((\Omega - p\mathbf{1}\mathbf{1}^{\top}) \circ M)P_Z + \frac{1}{p}P_X(\Omega \circ E)P_Z +$平滑误差 1。此外，根据引理 C.3，由于 $\|\omega_{it}\epsilon_{it}\|_{\psi_2} \le p^{1/2}\sigma$，我们以高概率有 $\|\Omega \circ E\| \lesssim p^{1/2}\sqrt{N+T}$。因此，$\|\Omega \circ (E - S')\| \lesssim p^{1/2}\sqrt{N+T}$ 以高概率成立，且对某个较大的 $C > 0$，$\|\Omega \circ (E - S')\| \le \nu = Cp^{1/2}\sqrt{N+T}$ 以高概率成立。然后，通过设定 $S = E - S'$和 $L = M_2(I_T - P_Z) + (I_N - P_X)M_3 + M_4 - P_XM_4P_Z$，根据引理 C.2 可得

$$\begin{aligned} & \left\|\widehat{M}_{rest} - (M_2(I_T - P_Z) + (I_N - P_X)M_3 + M_4 - P_XM_4P_Z)\right\|_F \\ & = O_p\left(\sqrt{K^*}\min\left\{\frac{\sqrt{N+T}\,(1 + M_{\max})}{\sqrt{p}}, \|M_2\|_F + \|M_3\|_F + \|M_4\|_F\right\}\right). \end{aligned}$$

然后，由于

$$\begin{aligned} \left\|\widehat{M}_1 + \widehat{M}_{rest} - M\right\|_F \le {} & \left\|\widehat{M}_{rest} - (M_2(I_T - P_Z) + (I_N - P_X)M_3 + M_4 - P_XM_4P_Z)\right\|_F \\ & + \left\|\frac{1}{p}P_X((\Omega - p\mathbf{1}\mathbf{1}^{\top}) \circ M)P_Z + \frac{1}{p}P_X(\Omega \circ E)P_Z\right\|_F + \|\textit{smoothing error_1}\|_F, \end{aligned}$$

and

$$\left\|\frac{1}{p}P_X((\Omega-p\mathbf{1}\mathbf{1}^\top)\circ M)P_Z+\frac{1}{p}P_X(\Omega\circ E)P_Z\right\|_F=O_p\left(\frac{J}{\sqrt{p}}\right),$$
$$\|smoothing\ error_1\|_F=O_p\left(\frac{\sqrt{NT}K_1}{J^{\gamma_1^G}}+\frac{\sqrt{NT}K_1}{J^{\gamma_1^Q}}+\frac{\sqrt{NT}K_2}{J^{\gamma_2^G}}+\frac{\sqrt{NT}K_3}{J^{\gamma_2^Q}}\right),$$

如（6）中所述，我们得到了所需的结果。 □

C.1.4 定理4.2的证明

为简化记号，将下标'*tall*'和'*wide*'分别记为 $\mathcal{T}$ 和 $\mathcal{W}$。这里，$(U_\mathcal{T}, D_\mathcal{T}, V_\mathcal{T})$、$(U_\mathcal{W}, D_\mathcal{W}, V_\mathcal{W})$和$(U, D, V)$分别表示"高"、"宽"和"满"矩阵的奇异值分解。首先，分别对高矩阵和宽矩阵应用推论3.2，可得

$$\left\|\widehat{U}_\mathcal{W}O_\mathcal{W}-U_\mathcal{W}\right\|=O_p\left(\frac{\mathcal{R}_\mathcal{W}}{\lambda_{\min,\mathcal{W}}}\right),\quad \left\|\widehat{U}_\mathcal{T}O_\mathcal{T}-U_\mathcal{T}\right\|=O_p\left(\frac{\mathcal{R}_\mathcal{T}}{\lambda_{\min,\mathcal{T}}}\right).$$

此外，由引理C.6，可得$U_\mathcal{W}=U_{(N_0)}H_\mathcal{W}$其中$U_{(N_0)}=[u_1,\cdots,u_{N_0}]^\top$且$H_\mathcal{W}=(U^\top_{(N_0)}U_{(N_0)})^{-1/2}G_\mathcal{W}$对于某个 $K\times K$正交矩阵 $G_\mathcal{W}$。然后，由引理C.6 可得

$$\left\|\widehat{U}_\mathcal{W}O_\mathcal{W}-U_\mathcal{W}\right\|_F=\left\|\widehat{U}_\mathcal{W}O_\mathcal{W}-U_{(N_0)}H_\mathcal{W}\right\|_F=\left\|\widehat{U}_\mathcal{W}-U_{(N_0)}Q_\mathcal{W}^{-1}\right\|_F=O_p\left(\frac{\mathcal{R}_\mathcal{W}}{\lambda_{\min,\mathcal{W}}}\right),$$
$$\left\|\widehat{U}_\mathcal{W}Q_\mathcal{W}-U_{(N_0)}\right\|_F=\left\|\widehat{U}_\mathcal{W}-U_{(N_0)}Q_\mathcal{W}^{-1}\right\|_F\|Q_\mathcal{W}\|=O_p\left(\frac{\sqrt{N_0}}{\sqrt{N}}\frac{\mathcal{R}_\mathcal{W}}{\lambda_{\min,\mathcal{W}}}\right),$$

其中 $Q_\mathcal{W}^{-1}=H_\mathcal{W}O_\mathcal{W}^\top$。同样，由引理 C.6，我们有 $U_\mathcal{T}=UH_\mathcal{T}$其中 $H_\mathcal{T}=(U^\top U)^{-1/2}G_\mathcal{T}$对于某个 $K\times K$正交矩阵 $G_\mathcal{T}$。于是我们有

$$\left\|\widehat{U}_\mathcal{T}O_\mathcal{T}-U_\mathcal{T}\right\|_F=\left\|\widehat{U}_\mathcal{T}O_\mathcal{T}-UH_\mathcal{T}\right\|_F=\left\|\widehat{U}_\mathcal{T}-UQ_\mathcal{T}^{-1}\right\|_F=O_p\left(\frac{\mathcal{R}_\mathcal{T}}{\lambda_{\min,\mathcal{T}}}\right),$$
$$\left\|\widehat{U}_\mathcal{T}Q_\mathcal{T}-U\right\|_F=\left\|\widehat{U}_\mathcal{T}-UQ_\mathcal{T}^{-1}\right\|_F\|Q_\mathcal{T}\|=O_p\left(\frac{\mathcal{R}_\mathcal{T}}{\lambda_{\min,\mathcal{T}}}\right),$$

其中 $Q_{\mathcal{T}}^{-1} = H_{\mathcal{T}} O_{\mathcal{T}}^{\top}$。定义 $R_1 = \widehat{U}_{\mathcal{W}} Q_{\mathcal{W}} - U_{(N_0)}$ 以及 $R_2 = \widehat{U}_{\mathcal{T},(N_0)} Q_{\mathcal{T}} - U_{(N_0)}$ 其中$U_{\mathcal{T},(N_0)} = [u_{\mathcal{T},1}, \cdots, u_{\mathcal{T},N_0}]^{\top}$ 以及 $\widehat{U}_{\mathcal{T},(N_0)} = [\widehat{u}_{\mathcal{T},1}, \cdots, \widehat{u}_{\mathcal{T},N_0}]^{\top}$。于是，我们有

$$\widehat{U}_{\mathcal{W}} Q_{\mathcal{W}} - R_1 = \widehat{U}_{\mathcal{T},(N_0)} Q_{\mathcal{T}} - R_2 \quad \Longrightarrow \quad \widehat{U}_{\mathcal{W}} = \widehat{U}_{\mathcal{T},(N_0)} H_{adj} + R_1 Q_{\mathcal{W}}^{-1} - R_2 Q_{\mathcal{W}}^{-1},$$

其中 $H_{adj} = Q_{\mathcal{T}} Q_{\mathcal{W}}^{-1}$。因此，我们有

$$\begin{aligned}
\widehat{H}_{adj} - H_{adj} &= \left(\widehat{U}_{\mathcal{T},(N_0)}^{\top} \widehat{U}_{\mathcal{T},(N_0)}\right)^{-1} \widehat{U}_{\mathcal{T},(N_0)}^{\top} \widehat{U}_{\mathcal{W}} - H_{adj} \\
&= \left(\widehat{U}_{\mathcal{T},(N_0)}^{\top} \widehat{U}_{\mathcal{T},(N_0)}\right)^{-1} \widehat{U}_{\mathcal{T},(N_0)}^{\top} R_1 Q_{\mathcal{W}}^{-1} - \left(\widehat{U}_{\mathcal{T},(N_0)}^{\top} \widehat{U}_{\mathcal{T},(N_0)}\right)^{-1} \widehat{U}_{\mathcal{T},(N_0)}^{\top} R_2 Q_{\mathcal{W}}^{-1}.
\end{aligned}$$

那么，由于

$$\left\| \left(\widehat{U}_{\mathcal{T},(N_0)}^{\top} \widehat{U}_{\mathcal{T},(N_0)}\right)^{-1} \widehat{U}_{\mathcal{T},(N_0)}^{\top} \right\| = \left\| \left(\widehat{U}_{\mathcal{T},(N_0)}^{\top} \widehat{U}_{\mathcal{T},(N_0)}\right)^{-1} \right\|^{1/2} = O_p\left(\frac{\sqrt{N}}{\sqrt{N_0}}\right)$$

根据引理 C.6 以及 $\|R_1\| = O_p\left(\frac{\sqrt{N_0}}{\sqrt{N}} \frac{\mathcal{R}_{\mathcal{W}}}{\lambda_{\min,\mathcal{W}}}\right)$，$\|R_2\| = O_p\left(\frac{\mathcal{R}_{\mathcal{T}}}{\lambda_{\min,\mathcal{T}}}\right)$ 根据上述界，由引理 C.6 可得

$$\left\| \widehat{H}_{adj} - H_{adj} \right\| = O_p\left(\frac{\sqrt{N}}{\sqrt{N_0}} \frac{\mathcal{R}_{\mathcal{W}}}{\lambda_{\min,\mathcal{W}}} + \frac{N}{N_0} \frac{\mathcal{R}_{\mathcal{T}}}{\lambda_{\min,\mathcal{T}}} \right).$$

此外，由于 $\|H_{adj}\| = O_p\left(\frac{\sqrt{N}}{\sqrt{N_0}}\right)$ 根据引理 C.6，我们有

$$\left\| \widehat{H}_{adj} \right\| \leq \|H_{adj}\| + \left\| \widehat{H}_{adj} - H_{adj} \right\| = O_p\left(\frac{\sqrt{N}}{\sqrt{N_0}}\right).$$

此外，请注意

$$\begin{aligned}
\left\| \widehat{V}_{\mathcal{W}} \widehat{D}_{\mathcal{W}} - V_{\mathcal{W}} D_{\mathcal{W}} O_{\mathcal{W}}^{\top} \right\|_F &= \left\| \widehat{D}_{\mathcal{W}} \widehat{V}_{\mathcal{W}}^{\top} - O_{\mathcal{W}} D_{\mathcal{W}} V_{\mathcal{W}}^{\top} \right\|_F = \left\| \widehat{U}_{\mathcal{W}}^{\top} \widehat{M}_{\mathcal{W}} - O_{\mathcal{W}} U_{\mathcal{W}}^{\top} M_{\mathcal{W}} \right\|_F \\
&= \left\| \widehat{U}_{\mathcal{W}}^{\top} \left(\widehat{M}_{\mathcal{W}} - M_{\mathcal{W}}\right) + \left(\widehat{U}_{\mathcal{W}}^{\top} - O_{\mathcal{W}} U_{\mathcal{W}}^{\top}\right) M_{\mathcal{W}} \right\|_F \\
&\leq \left\| \widehat{M}_{\mathcal{W}} - M_{\mathcal{W}} \right\|_F + \left\| \widehat{U}_{\mathcal{W}}^{\top} - O_{\mathcal{W}} U_{\mathcal{W}}^{\top} \right\|_F \|M_{\mathcal{W}}\|
\end{aligned}$$

$$= O_p\left(\mathcal{R}_{\mathcal{W}} + \frac{\lambda_{\max,\mathcal{W}}}{\lambda_{\min,\mathcal{W}}}\mathcal{R}_{\mathcal{W}}\right).$$

然后，因为

$$D_{\mathcal{W}}V_{\mathcal{W}}^{\top} = U_{\mathcal{W}}^{\top}M_{\mathcal{W}} = U_{\mathcal{W}}^{\top}(U_{(N_0)}H_{\mathcal{W}})H_{\mathcal{W}}^{-1}DV^{\top} = U_{\mathcal{W}}^{\top}U_{\mathcal{W}}H_{\mathcal{W}}^{-1}DV^{\top} = H_{\mathcal{W}}^{-1}DV^{\top},$$

我们有

$$\left\|\widehat{V}_{\mathcal{W}}\widehat{D}_{\mathcal{W}} - VDQ_{\mathcal{W}}^{\top}\right\|_F = \left\|\widehat{V}_{\mathcal{W}}\widehat{D}_{\mathcal{W}} - V_{\mathcal{W}}D_{\mathcal{W}}O_{\mathcal{W}}^{\top}\right\|_F = O_p\left(\mathcal{R}_{\mathcal{W}} + \frac{\lambda_{\max,\mathcal{W}}}{\lambda_{\min,\mathcal{W}}}\mathcal{R}_{\mathcal{W}}\right).$$

最后，我们有如下分解：

$$\begin{aligned}\left\|\widehat{M} - M\right\|_F &= \left\|\widehat{U}_{\mathcal{T}}\widehat{H}_{adj}\widehat{D}_{\mathcal{W}}\widehat{V}_{\mathcal{W}}^{\top} - UDV^{\top}\right\|_F \\ &\lesssim \left\|UQ_{\mathcal{T}}^{-1}\widehat{H}_{adj}\left(\widehat{V}_{\mathcal{W}}\widehat{D}_{\mathcal{W}} - VDQ_{\mathcal{W}}^{\top}\right)^{\top}\right\|_F + \left\|\left(\widehat{U}_{\mathcal{T}} - UQ_{\mathcal{T}}^{-1}\right)\widehat{H}_{adj}Q_{\mathcal{W}}DV^{\top}\right\|_F \\ &\quad + \left\|U\left(Q_{\mathcal{T}}^{-1}\widehat{H}_{adj}Q_{\mathcal{W}} - I_K\right)DV^{\top}\right\|_F.\end{aligned}$$

第一项可以如下界定

$$\left\|UQ_{\mathcal{T}}^{-1}\widehat{H}_{adj}\left(\widehat{V}_{\mathcal{W}}\widehat{D}_{\mathcal{W}} - VDQ_{\mathcal{W}}^{\top}\right)^{\top}\right\|_F \le \left\|Q_{\mathcal{T}}^{-1}\right\|\left\|\widehat{H}_{adj}\right\|\left\|\widehat{V}_{\mathcal{W}}\widehat{D}_{\mathcal{W}} - VDQ_{\mathcal{W}}^{\top}\right\|_F = O_p\left(\frac{\sqrt{N}}{\sqrt{N_0}}\kappa_{\mathcal{W}}\mathcal{R}_{\mathcal{W}}\right),$$

其中 $\kappa_{\mathcal{W}} = \frac{\lambda_{\max,}\ \lambda_{\min}}{\mathcal{W}}$ 。第二项可以如下界定

$$\left\|\left(\widehat{U}_{\mathcal{T}} - UQ_{\mathcal{T}}^{-1}\right)\widehat{H}_{adj}Q_{\mathcal{W}}DV^{\top}\right\|_F \le \left\|\widehat{U}_{\mathcal{T}} - UQ_{\mathcal{T}}^{-1}\right\|\left\|\widehat{H}_{adj}\right\|\|Q_{\mathcal{W}}\|\|D\| = O_p\left(\frac{\lambda_{\min}}{\lambda_{\min,\mathcal{T}}}\kappa\mathcal{R}_{\mathcal{T}}\right).$$

此外，最后一项可以按如下方式进行界定

$$\begin{aligned}\left\|U\left(Q_{\mathcal{T}}^{-1}\widehat{H}_{adj}Q_{\mathcal{W}} - I_K\right)DV^{\top}\right\|_F &= \left\|UQ_{\mathcal{T}}^{-1}\left(\widehat{H}_{adj} - H_{adj}\right)Q_{\mathcal{W}}DV^{\top}\right\|_F \\ &\le \left\|Q_{\mathcal{T}}^{-1}\right\|\left\|\widehat{H}_{adj} - H_{adj}\right\|\|Q_{\mathcal{W}}\|\|D\|_F\end{aligned}$$

$$= O_p\left(\kappa\frac{\lambda_{\min}}{\lambda_{\min,\mathcal{W}}}\mathcal{R}_{\mathcal{W}} + \kappa\frac{\sqrt{N}}{\sqrt{N_0}}\frac{\lambda_{\min}}{\lambda_{\min,\mathcal{T}}}\mathcal{R}_{\mathcal{T}}\right).$$

此外，根据Choi 和 Yuan (2024) 的引理 F.1. $\kappa_{\mathcal{W}} \lesssim \kappa$, $\frac{\lambda_{\min}}{\lambda_{\min,\mathcal{T}}} \asymp \frac{\sqrt{T}}{\sqrt{T_0}}$, 以及$\frac{\lambda_{\min}}{\lambda_{\min,\mathcal{W}}} \asymp \frac{\sqrt{N}}{\sqrt{N_0}}$。综上，我们得到了所需结果。 □

C.2 辅助引理

考虑以下一般模型， $Z = L+S$，其中 $rank(L) = K_L$。将核范数惩罚估计量记为

$$\widehat{L} := \arg\min_{A} \|Z - A\|_F^2 + \lambda \|A\|_* .$$

那么，估计量满足以下界。

Lemma C.1. 设 $\lambda \geq C\|S\|$，其中某个较大的常数为 $C > 0$。那么，我们有 e

$$\left\|\widehat{L} - L\right\|_F \lesssim \min\left\{\sqrt{K_L}\lambda, \sqrt{K_L}\|L\|_F\right\}.$$

^此外，如果 $L = 0$，那么 $L = 0$。

^**证明**。 设 $\Delta = L - L$。那么，我们有

$$\left\|Z - \widehat{L}\right\|_F^2 = \|S - \Delta\|_F^2 = \|S\|_F^2 + \|\Delta\|_F^2 - 2tr(\Delta^\top S).$$

此外，对于某个常数 $0 < c < 1$，我们有

$$\left|2tr(\Delta^\top S)\right| \leq 2\|\Delta\|_* \|S\| \leq (1-c)\lambda\|\Delta\|_*$$

由于 $\lambda \geq \frac{2}{1-c}\|S\|_{\circ}$ 因此，我们有

$$\begin{aligned}
&\left\|Z-\widehat{L}\right\|_F^2+\lambda\left\|\widehat{L}\right\|_* \leq \|Z-L\|_F^2+\lambda\|L\|_*, \\
&\|\Delta\|_F^2-2tr(\Delta^\top S)+\lambda\left\|\widehat{L}\right\|_* \leq \lambda\|L\|_*, \\
&\|\Delta\|_F^2-(1-c)\lambda\|\Delta\|_*+\lambda\left\|\widehat{L}\right\|_* \leq \lambda\|L\|_*. \qquad (11)
\end{aligned}$$

(1) 当 $L=0_{\circ}$

$\widehat{}$由于 $L=\Delta$ 且 $\|L\|_*=0$，由 (11) 可得

$$\|\Delta\|_F^2+2c\lambda\|\Delta\|_* \leq 0.$$

由于 $c>0$，我们有 $\|\Delta\|_F=0_{\circ}$

(2) 当 $L\neq 0_{\circ}$

注意到

$$\left\|\widehat{L}\right\|_* = \|\Delta+L\|_* \geq \|\Delta\|_*-\|L\|_*.$$

因此，由 (11) 可得

$$\begin{aligned}
&\|\Delta\|_F^2-(1-c)\lambda\|\Delta\|_*+\lambda\|\Delta\|_*-\lambda\|L\|_* \leq \|\Delta\|_F^2-(1-c)\lambda\|\Delta\|_*+\lambda\left\|\widehat{L}\right\|_* \leq \lambda\|L\|_*, \\
&\|\Delta\|_F^2+c\lambda\|\Delta\|_*-\lambda\|L\|_* \leq \lambda\|L\|_*, \\
&\|\Delta\|_F^2+c\lambda\|\Delta\|_* \leq 2\lambda\|L\|_*.
\end{aligned}$$

因此，我们有 $\|\Delta\|_* \leq \frac{2}{c}\|L\|_{*\circ}$ 接着，我们有 $\|\Delta\|_F \leq \|\Delta\|_* \leq \frac{2}{c}\|L\|_* \leq \frac{2}{c}\sqrt{K_L}\|L\|_{F\circ}$ 然后，我们推导 $\sqrt{K_L}\lambda$ 的界。将 L 的奇异值分解记为 $L=UDV^\top$，其中 $U=(U_o, U_c)$ 和 $V=(V_o, V_c)$。这里，(U_c, V_c) 是 U 中对应于零奇异值的列，(U_o, V_o) 则表示 U 中对应于非零奇异值的列，V

与非零奇异值相关联。此外，令

$$\mathcal{P}(A) = U_c U_c^\top A V_c V_c^\top, \qquad \mathcal{M}(A) = A - \mathcal{P}(A).$$

注意到

$$\begin{aligned} \left\|\widehat{L}\right\|_* &= \|L + \Delta\|_* = \|L + \mathcal{P}(\Delta) + \mathcal{M}(\Delta)\|_* \\ &\geq \|L + \mathcal{P}(\Delta)\|_* - \|\mathcal{M}(\Delta)\|_* = \|L\|_* + \|\mathcal{P}(\Delta)\|_* - \|\mathcal{M}(\Delta)\|_*. \end{aligned} \tag{12}$$

因此，利用这一关系与（11）以及 $(1-c)\ \lambda\|\Delta\|_* \leq (1-c)\lambda\|\mathcal{P}(\Delta)\|_* + (1-c)\lambda\|\mathcal{M}(\Delta)\|_*$这一事实，我们得到 $\|\Delta\|_F^2 + c\lambda\|\mathcal{P}(\Delta)\|_* \leq (2-c)\lambda\|\mathcal{M}(\Delta)\|_*$。因此，我们有

$$\|\Delta\|_F^2 \leq (2-c)\lambda\|\mathcal{M}(\Delta)\|_* \leq \lambda\|\mathcal{M}(\Delta)\|_F\sqrt{2K_L} \leq \lambda\|\Delta\|_F\sqrt{2K_L}. \ \square$$

另一方面，如果我们只能观测 $\Omega \circ Z$ 而非 Z，其中 $\Omega = (\omega_{it})_{i\leq N,\ t\leq T}$且观测到 $\omega_{it} = 1\{z_{it}\}$，那么核范数惩罚估计量变为

$$\widehat{L} := \underset{A\in\mathcal{A}}{\arg\min}\ \|\Omega \circ (Z - A)\|_F^2 + \lambda\|A\|_*,$$

其中 $\mathcal{A} = \{A : \|A\|_\infty \leq L_{\max}\}$。那么，估计量满足以下界。

引理 C.2. 设 $\lambda \geq C\|\Omega \circ S\|$，其中某个较大的常数为 $C > 0$。那么，如果 $\|L\|_\infty \leq L_{\max}$，以趋近于1的概率，我们有

$$\left\|\widehat{L} - L\right\|_F \lesssim \min\left\{\frac{\sqrt{K_L}\lambda}{p} + \frac{\sqrt{K_L(N+T)}L_{\max}}{\sqrt{p}}, \sqrt{K_L}\|L\|_F\right\},$$

其中 $p = \mathbb{E}[\omega_{it}]$。

^证明。(i) 令 $\Delta = L - L_{\circ}$ 于是，我们有

$$\left\|\Omega \circ (Z - \widehat{L})\right\|_F^2 = \|\Omega \circ (S - \Delta)\|_F^2 = \|\Omega \circ S\|_F^2 + \|\Omega \circ \Delta\|_F^2 - 2tr((\Omega \circ \Delta)^\top (\Omega \circ S)).$$

此外，对于某个常数 $0 < c < 1$，我们有

$$\left|2tr((\Omega \circ \Delta)^\top (\Omega \circ S))\right| = \left|2tr(\Delta^\top (\Omega \circ S))\right| \leq 2 \|\Delta\|_* \|\Omega \circ S\| \leq (1-c)\lambda \|\Delta\|_*$$

由于 $\lambda \geq \frac{2}{1-c}\|\Omega \circ S\|_{\circ}$ 因此，我们有

$$\begin{aligned}
&\left\|\Omega \circ (Z - \widehat{L})\right\|_F^2 + \lambda \left\|\widehat{L}\right\|_* \leq \|\Omega \circ (Z - L)\|_F^2 + \lambda \|L\|_*, \\
&\|\Omega \circ \Delta\|_F^2 - 2tr((\Omega \circ \Delta)^\top (\Omega \circ S)) + \lambda \left\|\widehat{L}\right\|_* \leq \lambda \|L\|_*, \\
&\|\Omega \circ \Delta\|_F^2 - (1-c)\lambda \|\Delta\|_* + \lambda \left\|\widehat{L}\right\|_* \leq \lambda \|L\|_*. \qquad (13)
\end{aligned}$$

注意

$$\left\|\widehat{L}\right\|_* = \|\Delta + L\|_* \geq \|\Delta\|_* - \|L\|_*.$$

因此，由（13）可得

$$\|\Omega \circ \Delta\|_F^2 - (1-c)\lambda \|\Delta\|_* + \lambda \|\Delta\|_* - \lambda \|L\|_* \leq \|\Omega \circ \Delta\|_F^2 - (1-c)\lambda \|\Delta\|_* + \lambda \left\|\widehat{L}\right\|_* \leq \lambda \|L\|_*,$$

$$\|\Omega \circ \Delta\|_F^2 + c\lambda \|\Delta\|_* \leq 2\lambda \|L\|_*.$$

因此，我们有 $\|\Delta\|_* \leq \frac{2}{c}\|L\|_{*\circ}$ 接着，我们有 $\|\Delta\|_F \leq \|\Delta\|_* \leq \frac{2}{c}\|L\|_* \leq \frac{2}{c}\sqrt{K_L}\|L\|_{F\circ}$
（2）利用关系（12）和（13），并结合（1−c）$\lambda\|\Delta\|_* \leq (1-c)\lambda\|\mathcal{P}(\Delta)\|_* + (1-c)$
$\lambda\|\mathcal{M}(\Delta)\|_*$，我们有

$$c\lambda \|\mathcal{P}(\Delta)\|_* \leq \|\Omega \circ \Delta\|_F^2 + c\lambda \|\mathcal{P}(\Delta)\|_* \leq (2-c)\lambda \|\mathcal{M}(\Delta)\|_*.$$

因此，我们有 $\|\mathcal{P}(\Delta)\|_* \leq \frac{2-c}{c}\|\mathcal{M}(\Delta)\|_{*\circ}$ 此外，我们知道 $\|\Delta\|_\infty \leq 2L_{\max\circ}$ 令

$\bar{L} = 2L_{\max}$ 和 $C_1 = \frac{2-c}{c}$。如果对于某个足够大的 $B > 0$，有 $\|\Delta\|_F^2 > 2B\frac{\bar{L}^2}{p}\sqrt{NT}$，那么 $\Delta \in \mathcal{C}\left(C_1, 2B\frac{\bar{L}^2}{p}\right)$，并且我们以高概率得到

$$p\|\Delta\|_F^2 < 2\|\Omega \circ \Delta\|_F^2 + 8BK_L(N+T)L_{\max}^2$$

根据引理 C.5。因此，由于 $\|\Omega \circ \Delta\|_F^2 \leq (2-c)\lambda\|\mathcal{M}(\Delta)\|_*$，我们有

$$\begin{aligned} p\|\Delta\|_F^2 &< 2(2-c)\lambda\|\mathcal{M}(\Delta)\|_* + 8BK_L(N+T)L_{\max}^2 \\ &< 2(2-c)\lambda\sqrt{2K_L}\|\Delta\|_F + 8BK_L(N+T)L_{\max}^2. \end{aligned}$$

如果第一项占主导地位，我们有

$$p\|\Delta\|_F^2 < 4(2-c)\lambda\sqrt{2K_L}\|\Delta\|_F \quad \Longrightarrow \quad \|\Delta\|_F < 4(2-c)\lambda\sqrt{2K_L}/p.$$

如果第二项占主导地位，我们有

$$p\|\Delta\|_F^2 < 16BK_L(N+T)L_{\max}^2 \quad \Longrightarrow \quad \|\Delta\|_F < 4B^{1/2}\sqrt{K_L(N+T)}L_{\max}/\sqrt{p}.$$

此外，如果 $\|\Delta\|_F^2 \leq 2B\frac{\bar{L}^2}{p}\sqrt{NT}$，我们有

$$\|\Delta\|_F \leq 2\sqrt{2}B^{1/2}\frac{L_{\max}\sqrt{N+T}}{\sqrt{p}}.$$

因此，以趋于 1 的概率，我们有

$$\left\|\widehat{L} - L\right\|_F \lesssim \frac{\sqrt{K_L}\lambda}{p} + \frac{\sqrt{K_L(N+T)}L_{\max}}{\sqrt{p}}. \quad \square$$

引理 C.3。设 E 为一个由独立次高斯条目构成的 $N \times T$ 矩阵，满足 $\|\epsilon_{it}\|_{\psi_2} \leq \sigma$。此外，设 L 和 R 分别为 $N \times J_1$ 和 $T \times J_2$ 正交矩阵，

分别。因此，我们有

$$\left\|L^{\top}ER\right\| \lesssim \sigma\sqrt{J_1 + J_2 + \log(\max\{N,T\})},$$

概率至少为 $1 - O(\text{最大}\{N,T\}^{-5})$。

证明。首先，根据Vershynin(2018年的推论 4.2.13，我们可以找到一个 $1/4$-网 $\mathcal{N}$ 覆盖单位球面 S^{J_1-1}，以及一个 $1/4$-网 $\mathcal{M}$ 覆盖单位球面 S^{J_2-1}，其基数分别为

$$|\mathcal{N}| \leq 9^{J_1}, \quad |\mathcal{M}| \leq 9^{J_2}.$$

注意

$$\left\|L^{\top}ER\right\| \leq 2 \max_{a\in\mathcal{N},b\in\mathcal{M}} \sum_{i=1}^{N}\sum_{t=1}^{T} \epsilon_{it}(a^{\top}L_i)(b^{\top}R_t)$$

其中 $L_i^{\top}$是 i 的第 L 行，而 $R_t^{\top}$是 t 的第 R 行（参见Vershynin(2018年）的第 4.4.1 节）。固定 $a_o \in \mathcal{N}$ 和 $b_o \in \mathcal{M}$。然后，根据独立次高斯假设，利用 Hoeffding 不等式，我们以至少 $1-u$ 的概率得到：

$$\sum_{i=1}^{N}\sum_{t=1}^{T} \epsilon_{it}(a_o^{\top}L_i)(b_o^{\top}R_t) \lesssim \sigma \left\|La_o\right\|_2 \left\|Rb_o\right\|_2 \sqrt{\log(u^{-1})} = \sigma\sqrt{\log(u^{-1})},$$

并令 $u = \text{最大}\{N,T\}^{-5}9^{-J_1-J_2}$，我们以至少 $1-\text{最大}\{N,T\}^{-5}9^{-J_1-J_2}$ 的概率得到，

$$\sum_{i=1}^{N}\sum_{t=1}^{T} \epsilon_{it}(a_o^{\top}L_i)(b_o^{\top}R_t) \lesssim \sigma\sqrt{J_1 + J_2 + \log(\max\{N,T\})},$$

于是，因为

$$\begin{aligned} P\left(\max_{a\in\mathcal{N},b\in\mathcal{M}} \sum_{i=1}^{N}\sum_{t=1}^{T} \epsilon_{it}(a^{\top}L_i)(b^{\top}R_t) > c\right) &\leq \sum_{a\in\mathcal{N},b\in\mathcal{M}} P\left(\sum_{i=1}^{N}\sum_{t=1}^{T} \epsilon_{it}(a^{\top}L_i)(b^{\top}R_t) > c\right) \\ &\leq 9^{J_1+J_2} P\left(\sum_{i=1}^{N}\sum_{t=1}^{T} \epsilon_{it}(a^{\top}L_i)(b^{\top}R_t) > c\right), \end{aligned}$$

通过设置 $c = \sigma\sqrt{J_1 + J_2 + \log(\max\{N,T\})}$，我们有

$$P\left(\max_{a\in\mathcal{N},b\in\mathcal{M}}\sum_{i=1}^{N}\sum_{t=1}^{T}\epsilon_{it}(a^\top L_i)(b^\top R_t) > \sigma\sqrt{J_1 + J_2 + \log(\max\{N,T\})}\right) \lesssim \max\{N,T\}^{-5}. \quad \square$$

引理 C.4. 以趋于1的概率，有

(i) $\|P_X E P_Z\| \lesssim \sigma\sqrt{J}$; *(ii)* $\|P_X E\| \lesssim \sigma\sqrt{T}$; *(iii)* $\|E P_Z\| \lesssim \sigma\sqrt{N}$; *(iv)* $\|E\| \lesssim \sigma\sqrt{N+T}$.

证明。 (i) 根据假设 3.2，以趋于 1 的概率，$\left\|\Phi(\Phi^\top\Phi)^{-1/2}\right\|$ 和 $\left\|\Psi(\Psi^\top\Psi)^{-1/2}\right\|$ 有界。因此，根据引理 C.3，以趋于 1 的概率，有

$$\|P_X E P_Z\| \le \left\|\Phi(\Phi^\top\Phi)^{-1/2}\right\| \left\|((\Phi^\top\Phi)^{-1/2}\Phi^\top)E(\Psi(\Psi^\top\Psi)^{-1/2})\right\| \left\|(\Psi^\top\Psi)^{-1/2}\Psi^\top\right\| \lesssim \sigma\sqrt{J}$$

因为 $\Phi(\Phi^\top\Phi)^{-1/2}$ 和 $\Psi(\Psi^\top\Psi)^{-1/2}$ 分别是 $N\times J$ 和 $T\times J$ 正交矩阵，且 $\log(\max\{N,T\}) \lesssim J$。

(ii) 根据引理C.3，以趋于 1 的概率，有

$$\|P_X E\| \le \left\|\Phi(\Phi^\top\Phi)^{-1/2}\right\| \left\|((\Phi^\top\Phi)^{-1/2}\Phi^\top)EI_T\right\| \lesssim \sigma\sqrt{T}$$

因为 $\Phi(\Phi^\top\Phi)^{-1/2}$ 和 I_T 分别是 $N \times J$ 和 $T \times T$ 正交矩阵，且$J \lesssim T$。

(iii) 证明与 (ii) 的证明对称。

(iv) 由引理 C.3 可得 $L = I_N$ 和 $R = I_T$。 $\square$

引理 C.5. 将受限方向集定义为

$$\mathcal{C}(c_1,c_2) = \left\{A \in \mathcal{A}^* : \|\mathcal{P}(A)\|_* \le c_1\|\mathcal{M}(A)\|_*, \|A\|_F^2 > c_2\sqrt{NT}\right\},$$

其中 $\mathcal{A}^* = \{A \in \mathbb{R}^{N\times T}: \ \|A\|_{\max} \le \bar{L}\}$。那么，对于任意 $C_1 > 0$，当 $B > 0$ 足够大时，以概率趋近于1，对 $A \in \mathcal{C}\left(C_1, 2B\frac{\bar{L}^2}{p}\right)$ 均有，

$$\|\Omega \circ A\|_F^2 > 0.5p\|A\|_F^2 - BK_L(N+T)\bar{L}^2.$$

证明。 这是 Chernozhukov 等人 (2023) 中引理 A.2 的推广。首先，令 $\Omega(A) = \|\Omega \circ A\|_F^2 = \sum_{it}\omega_{it}^2 A_{it}^2$。那么，我们有 $\mathbb{E}\Omega(A) = p\sum_{it}A_{it}^2 = p\|A\|_F^2$。此外，定义

$$\mathcal{E}(A) = \left\{|\Omega(A) - \mathbb{E}\Omega(A)| > 0.5\cdot\mathbb{E}\Omega(A) + BK_L(N+T)\bar{L}^2\right\}.$$

接下来，我们要证明 $P\left(\exists A \in \mathcal{C}\left(C_1, 2B\frac{\bar{L}^2}{p}\right): \ \mathcal{E}(A) \text{ 成立}\right) \to 0$。为使用标准的剥离论证，定义

$$\Gamma_l = \left\{A \in \mathcal{C}\left(C_1, 2B\frac{\bar{L}^2}{p}\right) : 2^l v_n \le \mathbb{E}\Omega(A) \le 2^{l+1}v_n\right\}$$

其中 $v_n = B\bar{L}^2\sqrt{NT}$ 且 $l \in \mathbb{N}$。

第 1 部分。 我们要证明 $\mathcal{C}\left(C_1, 2B\frac{\bar{L}^2}{p}\right) \subset \cup_{l=1}^{\infty}\Gamma_l$。若 $A \in \mathcal{C}\left(C_1, 2B\frac{\bar{L}^2}{p}\right)$，则有

$$\mathbb{E}\Omega(A) = p\|A\|_F^2 \ge 2B\bar{L}^2\sqrt{NT} = 2v_n.$$

因此，存在 $l \in \mathbb{N}$，使得对于任意 $A \in \mathcal{C}\left(C_1, 2B\frac{\bar{L}^2}{p}\right)$，都有 $A \in \Gamma_l$。

第 2 部分。设

$$\mathcal{D}(x) = \left\{A \in \mathcal{C}\left(C_1, 2B\frac{\bar{L}^2}{p}\right) : \|A\|_F^2 \le x\right\},$$
$$\mathcal{F}(A) = \left\{|\Omega(A) - \mathbb{E}\Omega(A)| - BK_L(N+T)\bar{L}^2 > 0.25\cdot 2^{l+1}v_n\right\}.$$

我们要证明，如果 $A \in \Gamma_l$ 且 $\mathcal{E}(A)$ 成立，那么 $A \in \mathcal{D}(x_l)$ 成立，其中 $x_l = p^{-1}2^{l+1}v_n$ 且 $\mathcal{F}(A)$ 成立。这是因为

$$|\Omega(A) - \mathbb{E}\Omega(A)| - BK_L(N+T)\bar{L}^2 > 0.5 \cdot \mathbb{E}\Omega(A) \geq 0.25 \cdot 2^{l+1}v_n,$$

以及 $\|A\|_F^2 = p^{-1}\mathbb{E}\Omega(A) \leq p^{-1}2^{l+1}v_n$。

第3部分。令

$$Q(x) = \sup_{A \in \mathcal{D}(x)} \left| \frac{1}{NT} \sum_{it} \omega_{it}^2 A_{it}^2 - p A_{it}^2 \right|.$$

我们对 $\mathbb{E}Q(x)$ 进行界定。首先注意，对于任意 $A \in \mathcal{D}(x) \subset \mathcal{C}\left(C_1, 2B\frac{\bar{L}^2}{p}\right)$，有

$$\begin{aligned} \|A\|_* &= \|\mathcal{P}(A) + \mathcal{M}(A)\|_* \leq (1+C_1)\|\mathcal{M}(A)\|_* \leq (1+C_1)\sqrt{K_L}\|\mathcal{M}(A)\|_F \\ &\leq (1+C_1)\sqrt{K_L}\|A\|_F \leq (1+C_1)\sqrt{K_L x}. \end{aligned}$$

令 u_{it} 为独立同分布的拉德马赫随机变量。那么，$\mathbb{E}\|\Omega_u\| \lesssim p^{1/2}\sqrt{N+T}$ 其中 $\Omega_u = (\omega_{it}u_{it})_{N\times T}$。因此，利用集中不等式进行对称化论证（例如 Koltchinskii (2011) 中的 (2.3)），可得

$$\begin{aligned} \mathbb{E}Q(x) &\leq 2\mathbb{E}\sup_{A\in\mathcal{D}(x)}\left|\frac{1}{NT}\sum_{it}\omega_{it}^2A_{it}^2u_{it}\right| \leq c_3\bar{L}\mathbb{E}\sup_{A\in\mathcal{D}(x)}\left|\frac{1}{NT}\sum_{it}\omega_{it}A_{it}u_{it}\right| \\ &= c_3\bar{L}\mathbb{E}\sup_{A\in\mathcal{D}(x)}\left|\frac{1}{NT}tr(\Omega_u A^\top)\right| \leq c_3\bar{L}\mathbb{E}\sup_{A\in\mathcal{D}(x)}\frac{1}{NT}\|\Omega_u\|\|A\|_* \\ &\leq c_4 p^{1/2}\frac{\sqrt{N+T}}{NT}\bar{L}\sup_{A\in\mathcal{D}(x)}\|A\|_* \leq c_5 p^{1/2}\frac{\sqrt{N+T}}{NT}\bar{L}\sqrt{K_L x} \\ &= 2c_5\sqrt{8}\bar{L}\frac{\sqrt{K_L(N+T)}}{\sqrt{NT}} \times \sqrt{\frac{p}{32NT}x} \leq \frac{p}{32NT}x + 32c_5^2\frac{\bar{L}^2K_L(N+T)}{NT} \\ &\leq \frac{p}{32NT}x + B\frac{\bar{L}^2K_L(N+T)}{NT}, \end{aligned}$$

当 $B > 0$ 足够大时。这里，我们利用 ω_{it} 有界于 1，且 A_{it} 被 L 有界的事实，其中 $\bar{}$。

第 4 部分。接下来，我们对以下量的尾概率进行界定： $Q(x) - \mathbb{E}Q(x)$。由于 $A_{it}^2/\bar{L}^2$ 有界，我们可以使用 Massart 不等式（例如，B¨uhlmann 和 Van De Geer(2011）得到

$$P\left(\frac{1}{\bar{L}^2}Q(x) > \frac{1}{\bar{L}^2}\mathbb{E}Q(x) + t\right) \leq \exp(-c_6 NTt^2).$$

令 $t = \frac{7xp}{32\bar{L}^2NT}$。那么，由于

$$\frac{\mathbb{E}Q(x)}{\bar{L}^2} \leq \frac{p}{32\bar{L}^2NT}x + B\frac{K_L(N+T)}{NT},$$

我们有

$$\begin{aligned} P\left(Q(x) > B\frac{K_L(N+T)}{NT}\bar{L}^2 + 0.25\cdot\frac{xp}{NT}\right) &= P\left(\frac{1}{\bar{L}^2}Q(x) > B\frac{K_L(N+T)}{NT} + 0.25\cdot\frac{xp}{\bar{L}^2NT}\right) \\ &\leq \exp\left(-c_7\frac{p^2x^2}{\bar{L}^4NT}\right). \end{aligned}$$

第 5 部分。最后，我们使用剥离论证。注意，

$$\begin{aligned} &P\left(\exists A\in\mathcal{C}\left(C_1, 2B\frac{\bar{L}^2}{p}\right) : \mathcal{E}(A) \text{ holds}\right) \leq \sum_{l=1}^{\infty} P\left(\exists A\in\Gamma_l : \mathcal{E}(A) \text{ holds}\right) \\ &\leq \sum_{l=1}^{\infty} P\left(\exists A\in\mathcal{D}(x_l) : \mathcal{F}(A) \text{ holds}\right) \leq \sum_{l=1}^{\infty} P\left(\sup_{A\in\mathcal{D}(x_l)} |\Omega(A) - \mathbb{E}\Omega(A)| > BK_L(N+T)\bar{L}^2 + 0.25px_l\right) \\ &= \sum_{l=1}^{\infty} P\left(Q(x_l) > B\frac{K_L(N+T)}{NT}\bar{L}^2 + 0.25\frac{px_l}{NT}\right) \leq \sum_{l=1}^{\infty}\exp\left(-c_7\frac{p^2x_l^2}{\bar{L}^4NT}\right) \\ &= \sum_{l=1}^{\infty}\exp\left(-c_7 4^{l+1}B^2\right) \leq \frac{\exp(-16c_7B^2)}{1-\exp(-16c_7B^2)} < \varepsilon \end{aligned}$$

对于任意 $\varepsilon > 0$ ，当 B足够大时。这里，我们使用关系$x_l^2 = p^{-2}4^{l+1}v_n^2$ 和 $v_n^2 = B^2\bar{L}^4NT$。因此，以概率趋于 1 的方式，对于所有 $A\in\mathcal{C}\left(C_1, 2B\frac{\bar{L}^2}{p}\right)$，均有：

—

$$|\Omega(A) - \mathbb{E}\Omega(A)| \leq 0.5\cdot\mathbb{E}\Omega(A) + BK_L(N+T)\bar{L}^2,$$

这意味着

$$\Omega(A) \geq 0.5 \cdot \mathbb{E}\Omega(A) - BK_L(N+T)\bar{L}^2.$$

T因此，根据 $\Omega(A)$ 和 $\mathbb{E}\Omega(A)$ 的定义，所需结果即得。 □

引理 C.6。 (i) 我们有 $U_{\mathcal{W}} = U_{(N_0)}H_{\mathcal{W}}$，其中 $U_{(N_0)} = [u_1, \cdots, u_{N_0}]^\top$，且 $H_{\mathcal{W}} = (U_{(N_0)}^\top U_{(N_0)})^{-1/2}G_{\mathcal{W}}$，其中某个 $K \times K$ 正交矩阵 $G_{\mathcal{W}}$ 满足条件； (ii) 我们有 $U_{\mathcal{T}} = UH_{\mathcal{T}}$，其中 $H_{\mathcal{T}} = (U^\top U)^{-1/2}G_{\mathcal{T}}$，其中某个 $K \times K$ 正交矩阵 $G_{\mathcal{T}}$ 满足条件； (iii) $\|H_{\mathcal{W}}\| \lesssim \sqrt{N}/\sqrt{N_0}$、$\left\|H_{\mathcal{W}}^{-1}\right\| \lesssim \sqrt{N_0}/\sqrt{N}$、$\|Q_{\mathcal{W}}\| \lesssim \sqrt{N_0}/\sqrt{N}$、$\left\|Q_{\mathcal{W}}^{-1}\right\| \lesssim \sqrt{N}/\sqrt{N_0}$ 以趋于 1 的概率成立。此外，$\|H_{\mathcal{T}}\|$、$\left\|H_{\mathcal{T}}^{-1}\right\|$、$\|Q_{\mathcal{T}}\|$、$\left\|Q_{\mathcal{T}}^{-1}\right\|$ 有界； (iv) $\left\| \left(\widehat{U}_{\mathcal{T}}^\top U_{(N_0)}\widehat{U}_{\mathcal{T}}^\top U_{(N_0)}\right)^{-1} \right\| = O_p\left(\frac{N}{N_0}\right)$。

证明。 (i) 设 $\Omega_{\mathcal{W}} = (U_{(N_0)}^\top U_{(N_0)})^{1/2}D^2(U_{(N_0)}^\top U_{(N_0)})^{1/2}$，且 $G_{\mathcal{W}}$ 是一个矩阵，其列为 $\Omega_{\mathcal{W}}$ 的特征向量，使得 $\Lambda_{\mathcal{W}} = G_{\mathcal{W}}^\top \Omega_{\mathcal{W}} G_{\mathcal{W}}$ 是 $\Omega_{\mathcal{W}}$ 的特征值按降序排列的对角矩阵。定义 $H_{\mathcal{W}} = (U_{(N_0)}^\top U_{(N_0)})^{-1/2}G_{\mathcal{W}}$。于是，我们有

$$\begin{aligned}
&(U_{(N_0)}D^2U_{(N_0)}^\top)U_{(N_0)}H_{\mathcal{W}} \\
&= U_{(N_0)}\left(U_{(N_0)}^\top U_{(N_0)}\right)^{-1/2}\left(U_{(N_0)}^\top U_{(N_0)}\right)^{1/2}D^2\left(U_{(N_0)}^\top U_{(N_0)}\right)^{1/2}\left(U_{(N_0)}^\top U_{(N_0)}\right)^{1/2}H_{\mathcal{W}} \\
&= U_{(N_0)}\left(U_{(N_0)}^\top U_{(N_0)}\right)^{-1/2}\left[\left(U_{(N_0)}^\top U_{(N_0)}\right)^{1/2}D^2\left(U_{(N_0)}^\top U_{(N_0)}\right)^{1/2}G_{\mathcal{W}}\right] \\
&= U_{(N_0)}\left(U_{(N_0)}^\top U_{(N_0)}\right)^{-1/2}\Omega_{\mathcal{W}}G_{\mathcal{W}} = U_{(N_0)}\left(U_{(N_0)}^\top U_{(N_0)}\right)^{-1/2}G_{\mathcal{W}}\Lambda_{\mathcal{W}} \\
&= U_{(N_0)}H_{\mathcal{W}}\Lambda_{\mathcal{W}}.
\end{aligned}$$

此外，注意到 $\left(U_{(N_0)}H_{\mathcal{W}}\right)^\top\left(U_{(N_0)}H_{\mathcal{W}}\right) = H_{\mathcal{W}}^\top U^\top{}_{(N_0)}U_{(N_0)}H_{\mathcal{W}} = G_{\mathcal{W}}^\top G_{\mathcal{W}} = I_K$。因此，$U_{(N_0)}H_{\mathcal{W}}$ 的列是 $U_{(N_0)}D^2U_{(N_0)}^\top$ 的特征向量，也是 $U_{(N_0)}DV^\top$ 的左奇异向量。

(ii) 证明与上述证明相同。
(iii) 由于 $\|H_{\mathcal{W}}\| \leq \left\|\left(U^\top{}_{(N_0)}U_{(N_0)}\right)^{-1/2}\right\| \|G_{\mathcal{W}}\|$ 且 $G_{\mathcal{W}}$是一个特征向量矩阵，我们有 $\|H_{\mathcal{W}}\| \lesssim \sqrt{N}/\sqrt{N_0}$ 且 $\left\|H_{\mathcal{W}}^{-1}\right\| \lesssim \sqrt{N_0}/\sqrt{N}$根据假设 4.3，以高概率成立。类似地，$\|H_{\mathcal{T}}\|$且 $\left\|H_{\mathcal{T}}^{-1}\right\|$ 有界，因为 $U^\top U = I_K$。此外，由于 $Q_{\mathcal{W}}^{-1} = H_{\mathcal{W}}O_{\mathcal{W}}^\top$，我们有

$\|Q_{\mathcal{W}}\| \lesssim \frac{\sqrt{N_0}}{\sqrt{N}}$且 $\left\|Q_{\mathcal{W}}^{-1}\right\| \lesssim \frac{\sqrt{N}}{\sqrt{N_0}}$以高概率成立。由于 $Q_{\mathcal{T}}^{-1} = H_{\mathcal{T}} O_{\mathcal{T}}^{\top}$，我们有$\|Q_{\mathcal{T}}\|$ 且 $\left\|Q_{\mathcal{T}}^{-1}\right\|$ 有界。

(iv) 首先，注意到

$$\begin{aligned}\left\|\widehat{U}_{\mathcal{T},(N_0)}^{\top}\widehat{U}_{\mathcal{T},(N_0)} - Q_{\mathcal{T}}^{-\top}U_{N_0}^{\top}U_{N_0}Q_{\mathcal{T}}^{-1}\right\| &\lesssim \left\|U_{N_0}Q_{\mathcal{T}}^{-1}\right\| \left\|\widehat{U}_{(N_0),\mathcal{T}} - U_{(N_0)}Q_{\mathcal{T}}^{-1}\right\|_F = O_p\left(\frac{\sqrt{N_0}}{\sqrt{N}}\frac{\mathcal{R}_{\mathcal{T}}}{\lambda_{\min,\mathcal{T}}}\right) \\ &= o_p\left(\frac{N_0}{N}\right)\end{aligned}$$

因为 $\|U_{N_0}\| = O_p\left(\frac{\sqrt{N_0}}{\sqrt{N}}\right)$ 和 $\left\|\widehat{U}_{\mathcal{T}} - UQ_{\mathcal{T}}^{-1}\right\|_F = O_p\left(\frac{\mathcal{R}_{\mathcal{T}}}{\lambda_{\min,\mathcal{T}}}\right)$。然后，由于

$$\lambda_{\min}(Q_{\mathcal{T}}^{-\top}U_{N_0}^{\top}U_{N_0}Q_{\mathcal{T}}^{-1}) \geq \lambda_{\min}^2(Q_{\mathcal{T}}^{-1})\lambda_{\min}(U_{N_0}^{\top}U_{N_0}) \geq c\frac{N_0}{N},$$

对于某个常数 $c > 0$，我们有：以趋近于 1 的概率，

$$\begin{aligned}\lambda_{\min}\left(\widehat{U}_{\mathcal{T},(N_0)}^{\top}\widehat{U}_{\mathcal{T},(N_0)}\right) &\geq \lambda_{\min}(Q_{\mathcal{T}}^{-\top}U_{N_0}^{\top}U_{N_0}Q_{\mathcal{T}}^{-1}) - \left\|\widehat{U}_{\mathcal{T},(N_0)}^{\top}\widehat{U}_{\mathcal{T},(N_0)} - Q_{\mathcal{T}}^{-\top}U_{N_0}^{\top}U_{N_0}Q_{\mathcal{T}}^{-1}\right\| \\ &\geq \frac{c}{2}\frac{N_0}{N}.\end{aligned}$$

因此，$\left\| \left(\widehat{U}_{\mathcal{T},(N_0)}^{\top}\widehat{U}_{\mathcal{T},\ (N_0)}\right)^{-1} \right\| = O_p\left(\frac{N}{N_0}\right)$。 —