不完整矩阵值时间序列的网络自回归

Xuening Zhu[1], Feifei Wang[2*]、Zeng Li[3] 和 Yanyuan Ma[4]

[1]复旦大学，中国； [2]中国人民大学，中国；

[3]南方科技大学，中国；

[4]宾夕法尼亚州立大学，美国

arXiv:2302.02768v1 [stat.ME] 6 Feb 2023

摘要

我们通过提出一种矩阵网络自回归模型，研究具有已观测网络结构的矩阵值时间序列的动态特征。该模型包含研究对象的行网络和列网络，并纳入协变量信息和低秩截距矩阵。我们允许矩阵中存在不完整观测，且缺失机制可以依赖于协变量。为估计该模型，我们提出了一种两步估计程序。第一步旨在估计网络自回归系数，第二步旨在估计回归参数，这些参数本身也是矩阵。从理论上说，我们首先分别建立自回归系数的渐近性质，并推导回归参数的误差界。随后，我们提出一种偏差缩减程序，以降低渐近偏差，并研究去偏估计量的理论性质。最后，我们通过若干数值研究和对Yelp数据集的分析，展示所提方法的有用性。

关键词：偏差降低；不完整矩阵观测；矩阵值时间序列；网络自回归。

*Feifei Wang 是通讯作者。

1 引言

随着数据收集能力的提升，矩阵值时间序列数据在经济学、金融学、环境科学等诸多领域越来越普遍（Leng 和 Tang，2012年；Zhou 和 Li，2014年；Zhou，2014年；Wang 等人，2019年；Chen 等人，2021b）。例如：一段时间内一组国家之间的进出口量；一年间空气污染监测站记录的环境指标（如 $PM_{2.5}$、SO_2, CO、NO_2）；以及不同国家每年公布的经济指标（如国内生产总值、居民消费价格指数、基准利率、失业率）。鉴于矩阵值时间序列数据应用广泛，了解这类数据的内部动态模式对于更好地开展实际应用至关重要。

为了对矩阵值数据的动态特征建模，一种直接的方法是将其堆叠成长向量，并对向量形式数据使用标准时间序列分析工具（L¨utkepohl，2005）。然而，这种方法忽略了矩阵行与列之间的内在关系，且缺乏可解释性（Chen 等人，2021b）。此外，它会增加待估计的模型参数数量，因此导致模型表现较差。为了增强模型的可解释性并减少模型参数，Chen 等人（2021a）提出了一种采用双线性形式的矩阵自回归模型。Wang 等人（2021）将矩阵自回归扩展为张量自回归模型，并利用低秩Tucker分解进行降维。为了对矩阵值时间序列的截面时间依赖性建模，Wang 等人（2019）、Chang 等人（2021b）、Chen 和 Fan（2021）、Yu 等人（2021）以及 Kong 等人（2022）采用动态因子和低秩结构进行降维。尽管这些方法有用，矩阵值时间序列模型仍存在一些问题。首先，尽管与向量模型方法相比，Chen 等人（2021a）的矩阵自回归模型能够显著减少模型参数，但对于一个 $N \times N$ 矩阵值时间序列，仍需估计 $O(N^2)$ 个参数。这会导致估计结果并非最优，尤其是在 N 极大的情况下。其次，尽管因子模型能够刻画未观测依赖结构，却没有考虑观测依赖结构（例如网络结构）。第三，

当前针对矩阵数据的模型估计方法通常假设矩阵中的所有元素均被完整观测。这些方法无法直接用于不完整的矩阵值数据，而这类数据在实践中也经常遇到（Cai 等人，2016年；Bi 等人，2017年；Mao等人，2019年）。

受上述问题的启发，我们提出了一种针对不完整矩阵值时间序列数据的网络自回归模型。首先，我们在建模过程中利用研究对象之间观测到的网络结构。例如，假设我们收集了若干空间区域内到访商店的个体所产生的矩阵值数据，并希望通过用户画像分析刻画用户动态的到访行为。在这种情况下，我们可以收集个体之间的社交网络，以及区域之间的空间网络。这两类网络分别称为行网络和列网络。这些网络信息有助于我们在预测个体的到访行为时，考虑其相连的朋友或邻近区域。因此，我们将观测到的网络结构信息明确地嵌入建模过程中。同时，这使我们能够采用Zhu 等人（2017）的方法进一步减少参数数量。

其次，我们考虑不完整矩阵值时间序列数据的模型估计。近年来，文献已在多种缺失机制下研究矩阵补全问题（Koltchinskii 等人，2011；Rohde 和 Tsybakov，2011；Cai 等人，2016；Mao等人，2019年）。然而，这些研究通常聚焦于静态数据设定。本文针对矩阵值时间序列开发了一种网络自回归模型，以研究其动态模式。受矩阵补全文献的启发，我们假设矩阵自回归中的截距（矩阵）为低秩。此外，我们还利用不随时间变化的协变量信息，并以线性形式将其纳入模型，其维度可能随样本量增长。这使我们能够进行动态矩阵补全，正如我们在实证研究中所展示的那样。

我们的模型估计分为两个步骤。第一步，我们使用剖面目标函数估计自回归系数。具体而言，我们使用逻辑回归对缺失机制建模，并在剖面目标函数中采用逆概率。

第二步，我们通过单独的投影程序和奇异值分解（SVD），估计协变量的回归系数以及低秩截距矩阵。借助所设计的算法，我们的模型估计在计算上可行且高效。将两组参数的估计拆分为两个步骤并不常见。在理论方面，我们首先建立了网络自回归系数估计量的渐近性质。此外，我们提出了一种偏差缩减程序，以降低该估计量的估计偏差。最后，我们研究了去偏估计量的误差界。与静态数据设定不同，理论分析还进一步考虑了动态依赖和网络依赖信息。随后，我们通过大量数值研究以及一个来自 Yelp（$www.yelp.com$）的真实数据示例，展示所提方法。

本文的组织结构如下。第2 节介绍模型和符号。第3 节介绍模型估计程序。第 4 节阐述所提出模型的理论性质。第 5 节开展多种数值研究，以评估该方法在有限样本下的表现。第 6 节对一个 Yelp数据集进行分析。我们在第 7节作总结。所有技术证明均置于补充文件中。

2 模型与符号

令$\mathbf{Y}_t = (Y_{ijt}) \in \mathbb{R}^{N_1 \times N_2}$ 表示我们感兴趣的动态高维矩阵，其中包含两个方面的联合信息。例如， Y_{ijt} 表示第 t个时间点第 j个空间区域内第 i位用户的响应（例如访问次数）。在实践中，我们假设只有部分响应被观测到。为表示缺失机制，我们使用二元指标 $R_{ijt} \in \{0, 1\}$，其中 $R_{ijt} = 1$ 表示在第t个时间点观测到了第i， j个响应，否则 $R_{ijt} = 0$ 。我们假设 R_{ijt} 服从参数为 $P(R_{ijt} = 1) = p_i$的伯努利分布。这里假设 p_i 取决于第 i位用户的某些协变量信息，记为$\mathbf{X}_i \in \mathbb{R}^p$。具体而言，我们假设采用逻辑回归

模型来描述 R_{ijt}，即，

$$P(R_{ijt}=1)=p_i=\frac{\exp\left(\mathbf{X}_i^\top\boldsymbol{\alpha}\right)}{1+\exp\left(\mathbf{X}_i^\top\boldsymbol{\alpha}\right)}, \tag{2.1}$$

对于 $t=1,\ldots,T$，其中 $\boldsymbol{\alpha}$ 表示 $\mathbf{X}_i$ 对应的系数。为方便起见，令$\mathbf{X}\equiv(\mathbf{X}_1,\ldots,\mathbf{X}_{N_1})^\top\in\mathbb{R}^{N_1\times p}$。

除了用户的协变量信息外，我们还可以观测用户之间的网络关系。我们用 $\mathbf{A}_1=(a_{1ij})\in\mathbb{R}^{N_1\times N_1}$记录网络信息，其中 $a_{1ij}=1$ 表示第 i 个用户与第 j 个用户相连，否则为 $a_{1ij}=0$。类似地，区域的空间关系也可以通过 $\mathbf{A}_2=(a_{2ij})\in\mathbb{R}^{N_2\times N_2}$构建，其中 $a_{2ij}=1$ 表示第 i 个区域是第 j 个区域的空间邻居，否则为 $a_{2ij}=0$ 。按照惯例，对于 $1\le i\le N_1$ ，我们令 $a_{1ii}=0$ ；对于 $1\le j\le N_2$，我们令 $a_{2jj}=0$。一般而言，我们将 $\mathbf{A}_1$ 称为 N_1 行主体的行网络，并将 $\mathbf{A}_2$ 称为 N_2 列位置的列网络。当然，行网络和列网络的网络节点在现实生活中不一定是用户和位置。只要能够适当地构建行网络和列网络，它们可以是任意研究对象。定义 $\mathbf{W}_1\equiv(d_{1i}^{-1}a_{1ij})\in\mathbb{R}^{N_1\times N_1}$ 为 $\mathbf{A}_1$ 的行归一化邻接矩阵，其中 $d_{1i}=\sum_j a_{1ij}$。类似地，定义 $\mathbf{W}_2=(d_{2j}^{-1}a_{2ij})\in\mathbb{R}^{N_2\times N_2}$，其中 $d_{2j}=\sum_i a_{2ij}$。换言之，$\mathbf{W}_2$ 是 $\mathbf{A}_2$的列归一化邻接矩阵。在我们的分析中，我们允许 N_1、N_2 发散，并为技术上的便利假设 $N_1\asymp N_2$。定义 $m=N_1+N_2$。

本文研究如何针对不完整矩阵元素进行矩阵值网络自回归。我们考虑以下矩阵网络自回归（MNAR）模型，

$$\mathbf{Y}_t=\boldsymbol{\Lambda}\mathbf{W}_1\mathbf{Y}_{t-1}+\mathbf{Y}_{t-1}\mathbf{W}_2\boldsymbol{\Gamma}+\mathbf{X}\boldsymbol{\beta}+\mathbf{B}+\mathcal{E}_t, \tag{2.2}$$

其中，$\boldsymbol{\Lambda}=\mathrm{diag}\{\lambda_i:1\le i\le N_1\}\in\mathbb{R}^{N_1\times N_1}$, $\boldsymbol{\Gamma}=\mathrm{diag}\{\gamma_i:1\le i\le N_2\}\in\mathbb{R}^{N_2\times N_2}$,$\boldsymbol{\beta}\in\mathbb{R}^{p\times N_2}$, $\mathbf{B}=(b_{ij})\in\mathbb{R}^{N_1\times N_2}$ 是未知参数矩阵，$\mathcal{E}_t=(\varepsilon_{ijt})\in\mathbb{R}^{N_1\times N_2}$是噪声矩阵。我们假设 $\mathrm{var}(\varepsilon_{ijt})=\sigma^2$，所有 $\{\varepsilon_{ijt}\}$s彼此独立，且也独立于 $\mathbf{Y}_{t-1}$。遵循相关文献（Rohde 和 Tsybakov，2011；Mao

等人，2019年；Wang 等人，2021），我们对截距矩阵 $\mathbf{B}$施加低秩结构。具体而言，我们假设 $\mathbf{B} = \mathbf{U}\mathbf{V}^\top$，其中 $\mathbf{U} \in \mathbb{R}^{N_1 \times r_B}$，$\mathbf{V} \in \mathbb{R}^{N_2 \times r_B}$，且 r_B 是$\mathbf{B}$ 的秩，其中 $r_B \ll \min(N_1, N_2)$。为保证可辨识性，我们假设 $\mathbf{B}$与 $\mathbf{X}$ 的列空间彼此正交，即 $\mathbf{X}^\top\mathbf{B} = \mathbf{0}$。

注1。 右侧（2.2）中的前两项刻画了相连用户和位置所产生的局部网络效应。这是将网络/空间自回归模型（*Zhu* 等人，*2017*年）推广到高维矩阵数据的一种形式。为了更好地理解该模型，我们可以考虑 $\mathbf{\Gamma} = \mathbf{0}$这一特殊情形。于是得到

$$\mathbf{Y}_t = \sum_{k=0}^{\infty} \mathbf{\Lambda}^k \mathbf{W}_1^k \Big(\mathbf{X}\boldsymbol{\beta} + \mathbf{B} + \mathcal{E}_{t-k} \Big).$$

因此，所提出的模型不仅纳入了用户自身的历史信息，还纳入了其相连邻居的历史信息。如果 $\mathbf{\Lambda} = \mathbf{0}$ 且 $\mathbf{\Gamma} = \mathbf{0}$，那么$\mathbf{Y}_t$ 完全由协变量效应加低秩矩阵的结构所刻画。在这种情况下，问题可归结为带外生协变量的矩阵补全问题（*Mao*等人，*2019*年）。

注2。 最近，*Chen* 等人（*2021a*）针对矩阵值时间序列研究了双线性自回归模型。将其直接扩展到网络框架可得

$$\mathbf{Y}_t = \mathbf{\Lambda}\mathbf{W}_1\mathbf{Y}_{t-1}\mathbf{W}_2\mathbf{\Gamma} + \mathbf{X}\boldsymbol{\beta} + \mathbf{B} + \mathcal{E}_t. \tag{2.3}$$

这与我们的模型（2.2）不同，因为网络效应 $\mathbf{\Lambda}$ 和 $\mathbf{\Gamma}$ 在（2.3）中以乘法形式发挥作用。*Wu* 等人 （*2021*）也采用了类似的乘法效应，用于刻画向量数据中网络节点的向内和向外影响。相比之下，我们以加法形式考虑网络效应。这种建模差异导致了完全不同的估计程序和理论发展。

General Notations. 对于对称矩阵或厄米矩阵$\mathbf{A}$，我们用 $\lambda_{\max}(\mathbf{A})$和$\lambda_{\min}(\mathbf{A})$分别表示其最大和最小特征值。对于任意矩阵$\mathbf{M} = (M_{ij}) \in \mathbb{R}^{m \times n}$，定义 $\sigma_1(\mathbf{M})$为$\mathbf{M}$的最大奇异值，定义 $\rho(\mathbf{M})$为

$\mathbf{M}$ 的谱半径。定义 $\|\mathbf{M}\|_F = \text{tr}(\mathbf{M}^\top\mathbf{M})^{1/2}$ 为矩阵 $\mathbf{M}$ 的弗罗贝尼乌斯范数。令 $\mathbf{M}_{i\cdot} \in \mathbb{R}^{1\times n}$ 和 $\mathbf{M}_{\cdot j} \in \mathbb{R}^m$ 分别表示 $\mathbf{M}$ 的第 i个行向量或第 j个列向量。令 $\text{vec}(\mathbf{M}) \in \mathbb{R}^{mn}$ 表示将 $\mathbf{M}$ 的各列向量堆叠成一个长向量所得的向量化结果。令 $|\mathbf{M}|_e = (|M_{ij}|) \in \mathbb{R}^{m\times n}$，其中对 $\mathbf{M}$ 的每个元素取绝对值。此外，若$M_{1、ij} \le M_{2、ij}$ 对 $1 \le i \le m、1 \le j \le n$成立，则定义 $\mathbf{M}_1 \preccurlyeq \mathbf{M}_2$。令 $\{a_n\}$和 $\{b_n\}$ 为与 n相关的两个数列。对于某个常数 c，定义 $a_n \lesssim b_n$ 为 $a_n \le cb_n\ n \to \infty$。当 $a_n/b_n \to 0$ 时，定义 $a_n \ll b_n$ 为 $n \to \infty$。令 $\mathbf{e}_i \in \mathbb{R}^n$ 为一个向量，其第 i个元素为 1，其余元素为 0。定义 $\mathbf{1}_n \in \mathbb{R}^n$ 为一个 n维向量，其元素全为 1。令 $\mathbf{I}_n \in \mathbb{R}^{n\times n}$ 为 n维单位矩阵。此外，对任意整数 N，令$[N] = \{1, 2, \cdots, N\}$。

NOTATION OF NORMS. 对于向量 $\mathbf{a}$，我们用 $\|\mathbf{a}\|$ 表示其 l_2 范数。对于任意矩阵$\mathbf{M}$，定义 $\|\mathbf{M}\|_{\max} = \text{最大}_{i,\ j}|M_{ij}|$ 以及 $\|\mathbf{M}\| = \sigma_1(\mathbf{M})$。进一步地，将 $\|\mathbf{M}\|_* = \sum_k \sigma_k(\mathbf{M})$ 表示为矩阵 $\mathbf{M}$ 的核范数。

3 模型估计

3.1 两步估计程序

为估计 MNAR 中的未知参数，我们首先假设响应的矩阵元素均被完整观测，并考虑一种最小二乘型目标函数。随后，我们针对不完整矩阵数据推导剖面目标函数，并提出两步估计程序。令 $\boldsymbol{\theta} = (\boldsymbol{\lambda}^\top,\ \boldsymbol{\gamma}^\top,\ \text{vec}(\boldsymbol{\beta})^\top,\ \text{vec}(\mathbf{B})^\top)^\top$ 表示所有待估参数，其中 $\boldsymbol{\lambda} = \text{diag}(\boldsymbol{\Lambda})$ 且 $\boldsymbol{\gamma} = \text{diag}(\boldsymbol{\Gamma})$。若我们能够观测到整个矩阵 $\mathbf{Y}_t$，则可以最小化以下最小二乘目标函数

$$Q^F(\boldsymbol{\theta}) = \sum_{t=1}^{T} \left\| \mathbf{Y}_t - \boldsymbol{\Lambda}\mathbf{W}_1\mathbf{Y}_{t-1} - \mathbf{Y}_{t-1}\mathbf{W}_2\boldsymbol{\Gamma} - \mathbf{X}\boldsymbol{\beta} - \mathbf{B} \right\|_F^2 \tag{3.1}$$

以得到参数 $\boldsymbol{\theta}$，其中 F 表示完整数据。定义

$$\Delta_{ijt}^{F}(\boldsymbol{\theta}) = Y_{ijt} - \lambda_i \sum_{k=1}^{N_1} W_{1ik} Y_{kj(t-1)} - \sum_{k=1}^{N_2} Y_{ik(t-1)} W_{2kj} \gamma_j - \xi_{ij},$$

其中 $\xi_{ij} = \mathbf{X}_i^{\top} \boldsymbol{\beta}_{\cdot j} + b_{ij}$。于是我们有 $Q^{F}(\boldsymbol{\theta}) = \sum_{i,\ j,\ t} \Delta_{ijt}^{F}(\boldsymbol{\theta})^2$。

尽管目标函数（3.1）很直观，但由于两个主要问题，无法直接应用。第一个问题是矩阵元素存在缺失值。事实上，$\mathbf{Y}_t$ 的并非所有元素都可观测。为处理$\mathbf{Y}_t$中的缺失机制，我们定义$\mathbf{Z}_t = (Z_{ijt}) \in \mathbb{R}^{N_1 \times N_2}$，其中 $Z_{ijt} = R_{ijt} Y_{ijt} / p_i$ 是逆概率加权响应。随后，我们针对存在不完整矩阵元素的情形，基于$\mathbf{Z}_t$ 矩阵设计估计方法。第二个问题是目标函数形式复杂，包含不同参数，因此需要采用不同的计算处理方法，并具有不同的理论性质。例如，为了考虑模型性质，我们需要对部分参数施加若干惩罚项，这使得难以同时对所有参数进行优化，并可能导致计算效率低下。此外，从理论角度看，各参数的统计收敛速率本质上不同，而分析同时优化过程较为困难，还可能人为降低某些参数的收敛速率。

为了解决这两个问题，我们设计了一个两步估计程序。第一步，通过最小化一个仅涉及参数 $\boldsymbol{\Lambda}$ 和 $\boldsymbol{\Gamma}$ 的目标函数，估计网络效应（$\boldsymbol{\Lambda}$ 和 $\boldsymbol{\Gamma}$）。第二步，将 $\boldsymbol{\Lambda}$ 和 $\boldsymbol{\Gamma}$ 的估计值代入另一个目标函数并将其最小化，从而估计回归矩阵参数（$\boldsymbol{\beta}$ 和 $\mathbf{B}$）。两步程序为模型估计带来了计算效率更高的算法。此外，在这样的两步程序中分开估计参数，有助于开展自然且方便的理论研究，并确保两组参数都达到各自适当的收敛速率。接下来，我们介绍

详细介绍两步估计程序。定义

$$\begin{aligned}\Delta_{ijt}(\boldsymbol{\theta}) &= Z_{ijt} - \lambda_i \sum_{k=1}^{N_1} W_{1ik} Z_{kj(t-1)} - \sum_{k=1}^{N_2} Z_{ik(t-1)} W_{2kj} \gamma_j - \xi_{ij} \\ &= Z_{ijt} - \lambda_i \mathbf{W}_{1i\cdot} \mathbf{Z}_{\cdot j(t-1)} - \mathbf{Z}_{i\cdot(t-1)} \mathbf{W}_{2\cdot j} \gamma_j - \xi_{ij}. \end{aligned} \tag{3.2}$$

我们首先关注网络效应的估计，即 $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_{N_1})^\top$ 和 $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_{N_2})^\top$。为此，我们通过构造以下形式剖面消去 ξ_{ij}

$$\begin{aligned}\widetilde{\Delta}_{ijt}(\boldsymbol{\theta}) &= \Delta_{ijt}(\boldsymbol{\theta}) - \overline{\Delta}_{ij}(\boldsymbol{\theta}) \\ &= Z_{ijt} - \overline{Z}_{ij} - \lambda_i \mathbf{W}_{1i\cdot} (\mathbf{Z}_{\cdot j(t-1)} - \overline{\mathbf{Z}}_{\cdot j,\text{lag}}) - (\mathbf{Z}_{i\cdot(t-1)} - \overline{\mathbf{Z}}_{i\cdot,\text{lag}}) \mathbf{W}_{2\cdot j} \gamma_j \\ &= \widetilde{Z}_{ijt} - \lambda_i \mathbf{W}_{1i\cdot} \widetilde{\mathbf{Z}}_{\cdot j(t-1),\text{lag}} - \widetilde{\mathbf{Z}}_{i\cdot(t-1),\text{lag}} \mathbf{W}_{2\cdot j} \gamma_j, \end{aligned}$$

其中 $\overline{\Delta}_{ij}(\boldsymbol{\theta}) = T^{-1} \sum_{t=1}^{T} \Delta_{ijt}(\boldsymbol{\theta})$，$\overline{\mathbf{Z}} = T^{-1} \sum_{t=1}^{T} \mathbf{Z}_t = (Z_{ij})$ 且 $\overline{\mathbf{Z}^{\text{lag}}} = (T-1)^{-1} \sum_{t=2}^{T} \mathbf{Z}_{t-1} = \widetilde{Z}_{ij,\text{lag}}$。令 $\widetilde{\mathbf{Z}}_t = \mathbf{Z}_t - \overline{\mathbf{Z}} = (\widetilde{Z}_{ijt})$ 且 $\widetilde{\mathbf{Z}}_{t-1,\text{lag}} = \mathbf{Z}_{t-1} - \overline{\mathbf{Z}^{\text{lag}}} = (\widetilde{Z}_{ij(t-1),\text{lag}})$。不难验证，$E\{\Delta_{ijt}(\boldsymbol{\theta}) | \mathcal{D}\} = \Delta^F_{ijt}(\boldsymbol{\theta})$，其中 $\mathcal{D} = \{(\mathbf{X}_t, \mathbf{Y}_t) : \ 1 \le t \le T\}$ 且 $\widetilde{\Delta}^F_{ijt}(\boldsymbol{\theta}) = \Delta^F_{ijt}(\boldsymbol{\theta}) - \Delta^F_{ij}(\boldsymbol{\theta})$。然而，我们指出，$E\{\widetilde{\Delta}_{ijt}(\boldsymbol{\theta})^2 | \mathcal{D}\} \neq \widetilde{\Delta}^F_{ijt}(\boldsymbol{\theta})^2$,，因为 $E(Z^2_{ijt} | \mathcal{D}) = p_i^{-1} Y^2_{ijt} \neq Y^2_{ijt}$。因此，我们不能直接使用 $\widetilde{\Delta}_{ijt}(\boldsymbol{\theta})^2$ 来替换目标函数中的 $\widetilde{\Delta}^F_{ijt}(\boldsymbol{\theta})^2$。注意到 $\sum_t \Delta_{ijt}(\boldsymbol{\theta})^2 = \sum_t \Delta_{ijt}(\boldsymbol{\theta})^2 - T\Delta_{ij}(\boldsymbol{\theta})^2$,，我们将 $\sum_t \Delta^2_{ijt}(\boldsymbol{\theta})$ 修改为

$$\begin{aligned}\widetilde{\Delta}^{c2}_{ij}(\boldsymbol{\theta}) = \sum_t \widetilde{\Delta}^2_{ijt}(\boldsymbol{\theta}) &+ \lambda_i^2 (1 - T^{-1}) \sum_t \sum_{k=1}^{N_1} W^2_{1ik} Z_{kj(t-1)} \Big(Y_{kj(t-1)} - Z_{kj(t-1)} \Big) \\ &+ \gamma_j^2 (1 - T^{-1}) \sum_t \sum_{k=1}^{N_2} W^2_{2kj} Z_{ik(t-1)} \Big(Y_{ik(t-1)} - Z_{ik(t-1)} \Big) + (1 - T^{-1}) \sum_t Z_{ijt} (Y_{ijt} - Z_{ijt}), \end{aligned}$$

其中，上标 c 表示校正。我们可以验证，$E\{\Delta^{c2}_{ij}(\boldsymbol{\theta})\} = \sum_t E\{\Delta_{ijt}(\boldsymbol{\theta})^2\}$。基于上述分析，令 $Q(\boldsymbol{\theta}) = \sum_{i,\ j} \Delta^{c2}_{ij}(\boldsymbol{\theta})$ 为剖面目标函数。我们还注意到，由于网络规模 N_1和 N_2不断增大，$\boldsymbol{\lambda}$ 和 $\boldsymbol{\gamma}$ 的维数较高。因此，自然需要考虑惩罚方法。具体而言，

我们考虑　　通过岭惩罚最小化以下剖面目标函数　　n

$$\widetilde{Q}_p(\boldsymbol{\theta}) = \widetilde{Q}(\boldsymbol{\theta}) + \nu_1 \|\boldsymbol{\Lambda}\|_F^2 + \nu_2 \|\boldsymbol{\Gamma}\|_F^2, \tag{3.3}$$

~ ~　　~其中，ν_1 和 ν_2 是两个调优参数。令（$\boldsymbol{\Lambda}, \boldsymbol{\Gamma}$）$= \arg\min_{\tilde{\boldsymbol{\Lambda}}, \tilde{\boldsymbol{\Gamma}}} Q_p(\boldsymbol{\theta})$ 为惩罚剖面估计量。

~ ~接下来，利用惩罚剖面估计量（$\boldsymbol{\Lambda}, \boldsymbol{\Gamma}$），我们在第二步中继续估计 $\boldsymbol{\beta}$ 和$\mathbf{B}$ 。定义

$$\widetilde{\mathbf{E}}_t = \mathbf{Z}_t - \widetilde{\boldsymbol{\Lambda}} \mathbf{W}_1 \mathbf{Z}_{t-1} - \mathbf{Z}_{t-1} \mathbf{W}_2 \widetilde{\boldsymbol{\Gamma}} \tag{3.4}$$

~为残差。接着，我们将$\mathbf{E}_t$ 视为响应，以估计 $\{\boldsymbol{\beta}$,$\mathbf{B}\}$。由于参数具有高维性，我们遵循Mao等人(2019年)的做法，考虑以下惩罚目标函数，

$$Q_p^{(2)}(\boldsymbol{\theta}) = \Big\| T^{-1} \sum_{t=1}^{T} \widetilde{\mathbf{E}}_t - \mathbf{X}\boldsymbol{\beta} - \mathbf{B} \Big\|_F^2 + \nu_3 \|\boldsymbol{\beta}\|_F^2 + \nu_4 \Big(\alpha \|\mathbf{B}\|_* + (1-\alpha) \|\mathbf{B}\|_F^2 \Big), \tag{3.5}$$

其中，$\|\mathbf{B}\|_*$ 表示$\mathbf{B}$的核范数，ν_3、ν_4 和 α 是调优参数。通过使用惩罚项 $\nu_3 \|\boldsymbol{\beta}\|_F^2$和 ν_4($1-\alpha$)$\|\mathbf{B}\|_F^2$，我们旨在对 $\boldsymbol{\beta}$ 和$\mathbf{B}$进行 L_2收缩，这种方法计算效率高，有助于处理高维问题。此外，使用核惩罚 $\nu_4 \alpha \|\mathbf{B}\|_*$ 对$\mathbf{B}$ 的奇异值进行惩罚，以促成低秩结构。调优参数 $\alpha \in [0,1]$ 用于平衡核惩罚与$\mathbf{B}$上的 L_2收缩。在我们的数值研究中，所有调优参数均通过五折交叉验证法选取。

3.2 一种迭代优化算法

本节讨论两步估计法的实现与优化算法。注意，惩罚剖面目标函数（3.3）具有最小二乘形式，因此可以通过解析方法求得估计量。

然而，可以注意到，其解析形式的推导涉及对一个维度为 $m \times m$ 的高维矩阵求逆，其中 $m = N_1 + N_2$。因此，这会带来巨大的计算挑战，尤其是对于大规模网络。

为解决这一问题，我们提出了一种用于模型估计的迭代算法。值得注意的是，一旦将其中一个参数固定为已知量，参数 $\boldsymbol{\lambda}$ 或 $\boldsymbol{\gamma}$ 就可以很容易地得到。以将$\sim\boldsymbol{\gamma}$ 固定为例。在这种情况下，我们可以通过最小化 $Q(\boldsymbol{\theta}) + \nu_1\|\boldsymbol{\Lambda}\|_F^2$来估计 $\boldsymbol{\lambda}$。定义 $\Delta_{1it}^{\gamma} = (\widetilde{\mathbf{Z}}_{i\cdot t} - \widetilde{\mathbf{Z}}_{i\cdot(t-1),\text{lag}}\mathbf{W}_2\boldsymbol{\Gamma})^\top \in \mathbb{R}^{N_2}$，进而得到

$$\widetilde{\lambda}_i = \Big(\sum_t \mathbf{W}_{1i\cdot}\widetilde{\mathbf{Z}}_{t-1,\text{lag}}\widetilde{\mathbf{Z}}_{t-1,\text{lag}}^\top\mathbf{W}_{1i\cdot}^\top + \kappa_i + \nu_1\Big)^{-1}\Big(\sum_t \mathbf{W}_{1i\cdot}\widetilde{\mathbf{Z}}_{t-1,\text{lag}}\Delta_{1it}^{\gamma}\Big), \tag{3.6}$$

其中 $\kappa_i = (1-T^{-1})\sum_t\sum_{j=1}^{N_2}\sum_{k=1}^{N_1} W_{1ik}^2 Z_{kj(t-1)}\big(Y_{kj(t-1)} - Z_{kj(t-1)}\big)$。特别地，我们注意到，当给定 $\boldsymbol{\gamma}$时，对于每个 λ_i，在分别给定 $1 \le i \le N_1$ 的情况下，都可以得到（3.6）。因此，该过程不涉及高维矩阵的逆矩阵运算，计算上较为友好。类似地，当给定 $\boldsymbol{\lambda}$ 时，我们可以推导出$\widetilde{\gamma}_j$ 在给定 $1 \le j \le N_2$条件下的解析形式。由此得到如下迭代算法。

步骤 1。获得初始估计量 $\boldsymbol{\Lambda}^{(0)}$ 和 $\boldsymbol{\Gamma}^{(0)}$。

步骤 2。令 M表示收敛所需的迭代次数。对于 $m = 1, \cdots M$，将第 m 个估计量记为 $\boldsymbol{\Lambda}^{(m)}$ 和 $\boldsymbol{\Gamma}^{(m)}$。然后重复以下步骤 2.1–2.2，直至收敛。

步骤 2.1。给定 $\boldsymbol{\Gamma}^{(m)}$，通过最小化 $\widetilde{Q}(\boldsymbol{\theta}) + \nu_1\|\boldsymbol{\Lambda}\|_F^2$来优化 $\boldsymbol{\Lambda}$。定义 $\Delta_{1it}^{(m)} = (\widetilde{\mathbf{Z}}_{i\cdot t} - \widetilde{\mathbf{Z}}_{i\cdot(t-1),\ \text{lag}}\mathbf{W}_2\boldsymbol{\Gamma}^{(m)})^\top \in \mathbb{R}^{N_2}$，然后得到

$$\lambda_i^{(m+1)} = \Big(\sum_t \mathbf{W}_{1i\cdot}\widetilde{\mathbf{Z}}_{t-1,\text{lag}}\widetilde{\mathbf{Z}}_{t-1,\text{lag}}^\top\mathbf{W}_{1i\cdot}^\top + \kappa_i + \nu_1\Big)^{-1}\Big(\sum_t \mathbf{W}_{1i\cdot}\widetilde{\mathbf{Z}}_{t-1,\text{lag}}\Delta_{1it}^{(m)}\Big),$$

$$\text{where } \kappa_i = (1-T^{-1})\sum_t\sum_{j=1}^{N_2}\sum_{k=1}^{N_1} W_{1ik}^2 Z_{kj(t-1)}\big(Y_{kj(t-1)} - Z_{kj(t-1)}\big).$$

步骤 2.2. 给定 $\boldsymbol{\Lambda}^{(m+1)}$，通过最小化 $\widetilde{Q}(\boldsymbol{\theta}) + \nu_2\|\boldsymbol{\Gamma}\|_F^2$ 来优化 $\boldsymbol{\Gamma}$。定义 $\Delta_{2jt}^{(m)} = \widetilde{\mathbf{Z}}_{\cdot jt} -$

$\mathbf{\Lambda}^{(m+1)}\mathbf{W}_1\widetilde{\mathbf{Z}}_{\cdot j(t-1),\ \mathrm{lag}} \in \mathbb{R}^{N_1}$，然后我们得到

$$\gamma_j^{(m+1)} = \Big(\sum_t \mathbf{W}_{2\cdot j}^\top \widetilde{\mathbf{Z}}_{t-1,\mathrm{lag}}^\top \widetilde{\mathbf{Z}}_{t-1,\mathrm{lag}} \mathbf{W}_{2\cdot j} + \xi_j + \nu_2\Big)^{-1}\Big(\sum_t \mathbf{W}_{2\cdot j}^\top \widetilde{\mathbf{Z}}_{t-1,\mathrm{lag}}^\top \Delta_{2jt}^{(m)}\Big),$$

where $\xi_j = (1-T^{-1})\sum_t \sum_{i=1}^{N_1}\sum_{k=1}^{N_2} W_{2kj}^2 Z_{ik(t-1)}\big(Y_{ik(t-1)} - Z_{ik(t-1)}\big)$.

设 $\widetilde{\mathbf{\Lambda}} = \mathbf{\Lambda}^{(M)}$、$\widetilde{\mathbf{\Gamma}} = \mathbf{\Gamma}^{(M)}$ 和 $\widetilde{\mathbf{E}}_t = \mathbf{Z}_t - \widetilde{\mathbf{\Lambda}}\mathbf{W}_1\mathbf{Z}_{t-1} - \mathbf{Z}_{t-1}\mathbf{W}_2\widetilde{\mathbf{\Gamma}}$。一旦得到 $\widetilde{\mathbf{\Lambda}}$ 和$\widetilde{\mathbf{\Gamma}}$，即可继续估计 $\boldsymbol{\beta}$ 和 $\mathbf{B}$。回顾一下，我们假设$\mathbf{X}$ 的列空间与$\mathbf{B}$的列空间正交。因此，我们可以通过以下方式估计 $\boldsymbol{\beta}$

$$\widehat{\boldsymbol{\beta}} = \Big(\mathbf{X}^\top\mathbf{X} + \nu_3\mathbf{I}_p\Big)^{-1}\Big(T^{-1}\sum_t \mathbf{X}^\top\widetilde{\mathbf{E}}_t\Big). \tag{3.7}$$

随后，我们通过 $\widehat{\mathbf{B}}^{(1)} = T^{-1}\sum_t \mathbf{P}_{\mathbf{X}}^{\perp}\widetilde{\mathbf{E}}_t$ 得到 $\widehat{\mathbf{B}}^{(1)}$，其中 $\mathbf{P}_{\mathbf{X}}^{\perp} = \mathbf{I} - \mathbf{P}_{\mathbf{X}}$，且 $\mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$。对于任意矩阵 $\mathbf{\Sigma}$，令 $\mathbf{UDV}^\top$ 为 $\mathbf{\Sigma}$的奇异值分解。将软阈值算子 τ_c 定义为 $\tau_c(\mathbf{\Sigma}) = \mathbf{U}\mathrm{diag}\{(\sigma_i - c)_+\}\mathbf{V}^\top$，其中 $\sigma_i = \mathbf{D}_{ii}$ 是 D 的第 i个对角元素。遵循 Mao 等人（2019）的方法，我们通过以下方式估计 B

$$\widehat{\mathbf{B}} = \frac{1}{1+(1-\alpha)\nu_4}\tau_{\alpha\nu_4/2}(\widehat{\mathbf{B}}^{(1)}). \tag{3.8}$$

上述分析表明，式（3.5）的解具有解析形式，便于高效实现。正如 Mao等人 (2019年) 所示，针对 3.5）关于 $\mathbf{B}$进行最小化，并满足相对于 $\mathbf{X}$ 的正交约束，等价于以下优化问题：

$$\arg\min_{\mathbf{B}\in\mathbb{R}^{N_1\times N_2}} \Big\|T^{-1}\sum_t \mathbf{P}_{\mathbf{X}}^{\perp}\widetilde{\mathbf{E}}_t - \mathbf{B}\Big\|_F^2 + \nu_4\Big(\alpha\|\mathbf{B}\|_* + (1-\alpha)\|\mathbf{B}\|_F^2\Big).$$

根据Mazumder 等人(2010的定理1，解析解由（3.8）给出。由于其具有简单的解析形式，与其他迭代型算法（Troyanskaya 等人，2001；Ma 等人，2011）相比，我们能够大幅降低计算负担。

4 理论性质

本节研究估计量的理论性质。首先，我们推导$\widetilde{\boldsymbol{\lambda}}$ 和 $\widetilde{\gamma}$的估计一致性与正态性。此外，我们还给出了 $\boldsymbol{\beta}$ 和$\mathbf{B}$ 的估计误差$\widehat{\quad}$ $\widehat{\quad}$界。随后，我们指出第一步估计量存在不可忽略的偏差。特别是在时间期数较少的大规模网络中，偏差可能很大。因此，我们进一步设计了偏差缩减程序，以降低估计偏差。

4.1 技术条件

为了分析估计量的理论性质，需要满足以下技术条件。

假设1. (Distribution) 假设 $\varepsilon_{ijt}(1 \leq i \leq N_1, 1 \leq j \leq N_2, 1 \leq t \leq T)$ 是相互独立且同分布的次高斯变量，均值为零，尺度因子为 $0 < \sigma < \infty$。也就是说，对于任意 t，有 $E\{\exp(t\varepsilon_{ijt})\} \leq \exp(\sigma^2 t^2/2)$ 。令$E(\varepsilon_{i_1 j_1 t_1}\varepsilon_{i_2 j_2 t_2}\varepsilon_{i_3 j_3 t_3}) = 0$，对于任意 $1 \leq i_k \leq N_1$, $1 \leq j_k \leq N_2, 1 \leq t_k \leq T$，满足 $k = 1,2$、3。

假设2. (Network Structure)

(a) (Connectivity) 将$\mathbf{W}_1$ 和$\mathbf{W}_2^\top$视为两个马尔可夫链的转移概率矩阵。对于$k = 1,2$, 状态空间分别定义为节点集合 $\{1, \cdots, N_k\}$。假设这些马尔可夫链不可约且非周期。进一步令$\boldsymbol{\pi}_k = (\pi_{k1}, \cdots, \pi_{kN_k})^\top \in \mathbb{R}^{N_k}$为第 k 个马尔可夫链的平稳分布，满足 *(i)* $\pi_{ki} \geq 0$ 和 $\sum_{i=1}^{N_k} \pi_{ki} = 1$, *(ii)* $\boldsymbol{\pi}_1 = \mathbf{W}_1^\top \boldsymbol{\pi}_1$ 和 $\boldsymbol{\pi}_2 = \mathbf{W}_2 \boldsymbol{\pi}_2$。假设存在有限整数 K 和常数 C，使得对于 $n \geq K$ 和 $k = 1,2$，有 $\mathbf{W}_k^n \preccurlyeq C\mathbf{1}_{N_k}\boldsymbol{\pi}_k^\top$。

(b) (Uniformity) 假设 $\sigma_1(\mathbf{W}_k) = O(\log N_k)$ 对 $k = 1,2$成立。此外，我们假设$\|\mathbf{W}_1^\top \mathbf{1}_{N_1}\|_\infty = O(\log N_1)$ 以及 $\|\mathbf{W}_2 \mathbf{1}_{N_2}\|_\infty = O(\log N_2)$。

假设 3. (Missing Rate) 假设 $\min_{i\in[N_1]} p_i \geq c_p$，其中 $c_p > 0$，并且我们允许$c_p \to 0$作为 $m \to \infty$。

假设 4。 (Stationarity) 令 $\kappa_1 = \|\mathbf{\Lambda}\|_{\max}$、$\kappa_2 = \|\mathbf{\Gamma}\|_{\max}$，并假设 $\kappa_1 + \kappa_2 < \kappa$且满足 $\kappa < 1$。

假设 5。 (Uniform Boundedness) 将 $\mathbf{X}$ 视为固定协变量，并假设 $|X_{ij}| < C < \infty$ 对所有 $i = 1$、...、N_1 以及 $j = 1$、...、p 均成立，其中 C是有限常数。令 $\|N_1^{-1}\mathbf{X}^\top\mathbf{X} - \mathbf{\Sigma}_X\| = o(1)$ 为 $N_1 \to \infty$，其中 $\mathbf{\Sigma}_X$ 是正定矩阵。此外，假设$\|\boldsymbol{\beta}_{\cdot j}\|_2 < C < \infty$ 对所有 $j = 1$、...、N_2 均成立，且 $\max\{\|\mathbf{X}\boldsymbol{\beta}\|_{\max}, \|\mathbf{B}\|_{\max}\} < C < \infty$。

假设 6. (Identification) 假设 $\mathbf{X}^\top\mathbf{B} = \mathbf{0}$。

假设 7. (Local Convexity) 将 $Q_p(\boldsymbol{\theta}, \boldsymbol{\alpha})$ 写成 $\boldsymbol{\theta}$ 和 $\boldsymbol{\alpha}$的函数。令 $H_{\theta\theta} = E\{\partial^2\widetilde{Q}_p(\boldsymbol{\theta}, \boldsymbol{\alpha})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top\}$和 $H_{\theta\alpha} = E\{\partial^2\widetilde{Q}_p(\boldsymbol{\theta}, \boldsymbol{\alpha})/\partial\boldsymbol{\theta}\partial\boldsymbol{\alpha}^\top\}$。假设 $\lambda_{\min}(H_{\theta\theta}/(mT)) \geq \tau_1$且 $\sigma_1(H_{\theta\alpha}/(mT)) \leq \tau_2$ 以$\min\{m, T\} \to \infty$ 为下界，其中 τ_1 和 τ_2 为正的常数。

假设 8. 令 $\mathbf{\Sigma}_{\boldsymbol{\alpha}}^{(m)} = (N_1c_p)^{-1}\sum_{i=1}^{N_1-1}\mathbf{X}_i\mathbf{X}_i^\top p_i(1-p_i)$。假设存在 $\mathbf{\Sigma}_\alpha$ 使得$\|\mathbf{\Sigma}_\alpha^{(m)} - \mathbf{\Sigma}_\alpha\| = o(1)$ 当 $m \to \infty$ 时成立，其中 $c_2 \leq \lambda_{\min}(\mathbf{\Sigma}_\alpha) \leq \lambda_{\max}(\mathbf{\Sigma}_\alpha) \leq c_1$，其中 c_1 和 c_2 为正的常数。

假设 1 意味着，对于向量 $\mathbf{a} = (a_1, \cdots, a_{N_1N_2})^\top$，有

$$P\big(\big|\mathbf{a}^\top\mathbb{E}_t\big| > x\big) \leq 2\exp\Big(-\frac{x^2}{2\sigma^2\|\mathbf{a}\|^2}\Big) \tag{4.1}$$

根据Wang 等人(2013)，其中 $\mathbb{E}_t = \text{vec}(\mathcal{E}_t)$。高维建模文献中广泛采用这一条件（Wang 等人，2013；Lugosi 和 Mendelson，2019年；Fan 等人，2021），它比正态性假设更宽松。接下来，假设2 是对网络结构施加的条件。假设2(a)要求所有网络节点彼此可达（即不可约性）。不可约性和非周期性同时成立的一个简单充分条件是：经过有限步后，网络始终保持完全连通。根据网络著名的小世界特性，各种现实世界网络都很容易满足这一条件（Watts 和 Strogatz，1998；

Newman 等人，2011，第4.2章）。接下来，我们假设$\mathbf{W}_k^n$ 的各元素均有界于 $C\mathbf{1}_{N_k}\boldsymbol{\pi}_k^\top$，其中 $n \geq K$；若马尔可夫链以较快速度一致收敛至其平稳分布，则可直接得出这一结论。该条件可以放宽，允许 K缓慢趋于无穷大，当 $m \to \infty$时。假设2(b)是一个均匀性条件，用于限制网络中的超级明星效应（Zhou 等人，2017；Zhu 等人，2021）。

随后，假设 3 关于非缺失率，该非缺失率允许随着 N_1 和 N_2 发散而趋于零。因此，我们允许存在稀疏观测条目的情形。Mao 等人 (2019年）也针对静态数据考虑了类似的设定。此外，对于 Chen 等人 (2021a）所考虑的矩阵自回归模型，缺失机制的设定更加灵活，该模型不允许矩阵元素存在缺失。

假设4旨在确保矩阵值时间序列的平稳性。这里，矩阵值时间序列的平稳性是指相应向量化时间序列的平稳性，即 vec($\mathbf{Y}_t$)。Zhu 等人（2017）针对动态网络数据也施加了类似条件。为理论上的便利，假设5将 X 视为有界固定协变量。假设6是一个识别条件，使我们能够分别估计 $\boldsymbol{\beta}$ 和 B（Mao 等，2019）。接下来，假设7假定目标函数具有局部凸性。假设8假定我们可以通过使用最大似然估计，并以对数似然函数为 $\mathcal{L}(\boldsymbol{\alpha}) = \sum_{i,j,t}\{R_{ijt}\mathbf{X}_i^\top\boldsymbol{\alpha} - \log(1+\exp(\mathbf{X}_i^\top\boldsymbol{\alpha}))\}$，获得(2.1)中 $\boldsymbol{\alpha}$ 的可靠估计，其中 $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{(m)}$为 Hessian矩阵。该假设在此使用$c_p$ 来限制不平衡数据的影响（即，$P(R_{ijt}=1) \to 0$）。近期关于不平衡数据学习问题的文献（Wang，2020）中也采用了类似假设。

4.2 第一步估计性质

~ ~ ~定义 $\boldsymbol{\theta} = \text{argmin}_{\boldsymbol{\theta}} Q_p(\boldsymbol{\theta})$，其中 $Q_p(\boldsymbol{\theta})$ 如（3.3）所示。我们首先在以下定理中确立剖面估计量 $\boldsymbol{\theta}$ 的渐近性质。

定理1。 假设假设条件 *1–8* 成立。设 $\boldsymbol{\eta}$ 是 $\mathbb{R}^m$ 中的任意向量，满足~ ~ $c_1 \leq \|\boldsymbol{\eta}\| \leq \|\boldsymbol{\eta}\|_1 \leq c_2$，其中 c_1, c_2 是正常数。回顾 $\boldsymbol{\Sigma}_2 = (mT)^{-1}Q_p(\boldsymbol{\theta}_0)$

~ ~和 $\mathbf{\Sigma}_2 = E(\mathbf{\Sigma}_2)$。 $\mathbf{\Sigma}_2$ 的显式形式见补充材料中的 *A.2*节。此外，设 $(mT)^{-1}c_p^2\boldsymbol{\eta}^\top cov\{\widetilde{Q}_p(\boldsymbol{\theta}_0)\}\boldsymbol{\eta} \to \boldsymbol{\eta}^\top\mathbf{\Sigma}_1\boldsymbol{\eta}$，其中 $\mathbf{\Sigma}_1$ 是一个正定矩阵。 $\mathbf{\Sigma}_1$ 的具体形式也见补充材料中的 *A.2*节。进一步假设

$$\frac{1}{Tc_p^4}\Big[mr_\pi^2(logm)^{4K+1} + mr_\pi(logm)^{2K+3} + (logm)^{2K} + c_p^2m^2r_\pi^2(logm)^{4K} + c_p^3\big\{m^4r_\pi^4(logm)^4 + m^2r_\pi^2(logm)^{4K+4}\big\}\Big] \to 0, \tag{4.2}$$

其中 $r_\pi = max(r_{1\pi}, r_{2\pi})$, $r_{1\pi} = \boldsymbol{\pi}_1^\top\boldsymbol{\pi}_1$, $r_{2\pi} = \boldsymbol{\pi}_2^\top\boldsymbol{\pi}_2$ 和 K 在假设*2(a)*中给出。定义 $\boldsymbol{b}^{(1)} = (\boldsymbol{b}^{(1)\lambda\top}, \boldsymbol{b}^{(1)\gamma\top})^\top$，其中 $\boldsymbol{b}^{(1)\lambda} = (b_i^{(1)\lambda} : 1 \le i \le N_1)^\top$ 和 $\boldsymbol{b}^{(1)\gamma} = (b_j^{(1)\gamma} : 1 \le j \le N_2)^\top$，即，

$$b_i^{(1)\lambda} = \frac{2}{mT}\lambda_i\sum_j\sum_k W_{1ik}^2E\Big\{Z_{kjt}(Y_{kjt} - Z_{kjt})\Big\}, \tag{4.3}$$

$$b_j^{(1)\gamma} = \frac{2}{mT}\gamma_j\sum_i\sum_k W_{2kj}^2E\Big\{Z_{ikt}(Y_{ikt} - Z_{ikt})\Big\}. \tag{4.4}$$

那么我们有

$$\sqrt{mT}c_p\boldsymbol{\eta}^\top(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta} - \mathbf{b}_1) \to_d N(0, \boldsymbol{\eta}^\top\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1\mathbf{\Sigma}_2^{-1}\boldsymbol{\eta}), \tag{4.5}$$

其中 $\boldsymbol{b}_1 = \widetilde{\mathbf{\Sigma}}_2^{-1}(\boldsymbol{b}^{(1)} - \boldsymbol{r}/(mT))$ 且 $\boldsymbol{r} = 2(\nu_1\boldsymbol{\lambda}^\top, \nu_2\boldsymbol{\gamma}^\top)^\top$。进一步假设 $max\{\nu_1, \nu_2\} = o(c_p^{-1}m(logm)^{2K})$，则有 $\|\boldsymbol{b}^{(1)}\|_\infty = O(c_p^{-1}(logm)^{2K}/T)$ 和 $\|\boldsymbol{b}_1\|_\infty = O_p(c_p^{-1}(logm)^{2K}/T)$。

关于理论结果，我们有以下几点评论。首先，条件（4.2）是一个关键条件，它要求 T 足够大，以便对 $\boldsymbol{\theta}$进行可靠估计。如果 m （网络规模 $N_1 + N_2$）和 r_π（与$\mathbf{W}_1$ 和 $\mathbf{W}_2$的平稳分布相关的量）较小，而 c_p （观测率的下界）较高，则该条件更容易成立。回想一下， $\boldsymbol{\pi}_1$ 和 $\boldsymbol{\pi}_2$ 分别是对应于$\mathbf{W}_1$ 和$\mathbf{W}_2$, 的平稳分布向量。以 $\boldsymbol{\pi}_1$ 为例。如果平稳分布相对均匀，那么应有 $\pi_{1i} \approx N_1^{-1}$（$1 \le i \le N_1$），从而得到

$\boldsymbol{\pi}_1^\top \boldsymbol{\pi}_1 \approx N_1^{-1} \to 0$。在这种情况下， $r_{1\pi}$ 会快速收敛到零。在网络文献中， $\boldsymbol{\pi}_1$ 和 $\boldsymbol{\pi}_2$ 也称为网络节点的特征向量中心性（Jackson，2010）。特征向量中心性通常用于刻画网络节点的影响力。如果网络节点影响力的分布相对均匀，则有 $\pi_{1i} \approx N_1^{-1}$和 $\pi_{2i} \approx N_2^{-1}$。换言之，当网络超级明星效应较弱时， r_π 将快速收敛到零。同时，log(m)这一项也与均匀性假设2(b)密切相关。在假设2(b)中，我们基本上假定 σ_1(W$_k$) $= O$(log$m$)，且 $\|$W$_1^\top \mathbf{1}\|_\infty = O$(log$m$)。例如，考虑一个极端情况。令$a_{1,i1} = 1$ （对 $2 \le i \le N_1$ 而言），并令 A$_1$ 中的其他元素为零。在这种情况下，第一个节点是网络中的超级明星，可以验证 σ_1(W$_1$) $= O(m^{1/2})$，且 $\|\mathbf{W}_1^\top \mathbf{1}\|_\infty = O(m)$。这一极端情况破坏了节点的均匀性，并违反假设2(b)。因此，只要控制网络超级明星效应和缺失率，条件（4.2）就可以成立。

其次，(4.5)中的渐近收敛结果表明，渐近偏差为b$_1$ ，收敛速率为$\sqrt{mTc_p}$。渐近偏差主要由时间长度 T、与网络结构相关的量(log$m$)^{2K}以及观测率 c_p决定。令 $\boldsymbol{\eta} =$ e$_i$ （ $1 \le i \le m$），考察逐元素收敛。只要 $T^{-1}m$(log$m$)$^{4K} \to 0$，渐近偏差项就会消失，这要求时间期数 T相对较大。注意，当 m 或 T增大时，收敛速率会提高。 T 增大带来的提升直接源于时间期数增加。 m 增大带来的提升则可理解为矩阵乘积W$_1$Y$_{t-1}$ 和Y$_{t-1}$W$_2$, 所贡献的累积信息所致，这些乘积分别对应于结合 N_1 和 N_2个观测。接下来，观测率 c_p 越高，收敛速率就越快。这是因为更高的观测率 c_p 会带来更大的有效样本量，从而改善估计性能。该渐近结果拓展了包含固定效应的经典面板数据模型（Arellano等，2007）的结果，而我们的研究重点是具有不完整观测的矩阵值时间序列。最后，为处理参数 $\boldsymbol{\theta}$维度的发散，我们通过将(4.5)投影到满足$c_1 \le \|\boldsymbol{\eta}\| \le \|\boldsymbol{\eta}\|_1 \le c_2$的任意向量 $\boldsymbol{\eta}$ 上，来描述其渐近正态性。之所以可以考虑范数有界的任意 $\boldsymbol{\eta}$ ，是因为

且为正的 L_1 和 L_2 范数，是因为尽管我们的参数维度不断增大，但并非超高维。在这种情况下，估计参数在任意方向上的投影仍渐近正态。若考虑超高维参数，情况则不再如此。关于超高维下的统计推断，可参阅Chernozhukov等(2017)、Chang等(2021a)以及Koike (2022)，了解用于建立大样本性质的新型理论工具。

推论1. 假设与定理*1*中的条件相同。记 $\mathbf{b}_1 = (\mathbf{b}_1^{\lambda\top}, \mathbf{b}_1^{\gamma\top})^\top$，其中$\mathbf{b}_1^{\lambda} \in \mathbb{R}^{N_1}$ 和 $\mathbf{b}_1^{\gamma} \in \mathbb{R}^{N_2}$。则有

$$N_1^{-1/2}\|\widetilde{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0 - \mathbf{b}_1^{\lambda}\| = O_p((mT)^{-1/2}c_p^{-1}),$$
$$N_2^{-1/2}\|\widetilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0 - \mathbf{b}_1^{\gamma}\| = O_p\big((mT)^{-1/2}c_p^{-1}\big).$$

推论1是定理1的直接结论。它建立了 L_2-收敛结果，即 $\widetilde{\ }$ $\widetilde{\ }$参数 $\boldsymbol{\theta}$。根据 $\boldsymbol{\theta}$的收敛结果，我们能够在下述定理中分别建立$\widehat{\ }$ $\widehat{\ }$关于 $\boldsymbol{\beta}$ 和 $\mathbf{B}$ 的估计误差界。

定理2. 假设与定理*1* 中的条件相同，且 $\nu_3 = o(m)$，则有

$$\frac{1}{N_2 p}\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_F = O_p\Big\{\frac{1}{N_2 p}\Big(m^{-1/2}\nu_3 + (Tc_p)^{-1/2}p^{1/2}(logm)^K + p^{1/2}(\|\mathbf{b}_1^{\lambda}\| + \|\mathbf{b}_1^{\gamma}\|)\Big)\Big\}. \quad (4.6)$$

$\widehat{\ }$定理2为 $\boldsymbol{\beta}$建立了误差界。这里，$N_2 \times p$ 是 $\boldsymbol{\beta}$ 矩阵的维度。因此，$\|\boldsymbol{\beta} - \boldsymbol{\beta}\|_F/(N_2 p)$是 $\boldsymbol{\beta}$矩阵的平均误差界。如式（4.6）所示，$\widehat{\boldsymbol{\beta}}$ 的误差由三个主要项构成。第一项（$m^{-1/2}\nu_3$）/（$N_2 p$）源于使用惩罚因子为 ν_3的岭回归所导致的估计偏差。令 $\nu_3 = o$（m），则有（$m^{-1/2}\nu_3$）/（$N_2 p$）$= o$（$m^{-1/2}p^{-1}$）。其余两项源于第一步中对 $\boldsymbol{\theta}$的估计。具体而言，第二项 {（Tc_p）$^{-1/2}$（log$m$）^{K}}/（$N_2 p^{1/2}$）刻画了 $\boldsymbol{\theta}$估计变异性的影响，而第三项（‖b$_1^{\lambda}$‖ + ‖b$_1^{\gamma}$‖）/（$N_2 p^{1/2}$）刻画了 $\boldsymbol{\theta}$估计偏差的影响。值得注意的是，在条件（4.2）下，第二项占主导地位。然而，为了明确体现这两项，我们仍在式（4.6）中同时保留它们

的影响。显然，增大 T 有助于减小第二项和第三项，但不会减小第一项。与此前的结果类似，与网络结构相关的量（log$m$）^{K} 以及观测率 c_p 在最终误差界中发挥着重要作用。

定理 3。 假设与定理 *1* 相同的条件。此外，令 $\nu_4'\alpha\Delta_E^{-1}$ ，$_{N_1N_2T}\to\infty$ 以及 $p \ll (c_pN_1N_2T)^{1/2}$，其中 $\nu_4' = \nu_4/(N_1N_2)$，且

$$\Delta_{E,N_1N_2T} = \frac{1}{m^2}\Big\{\frac{p+m^{1/2}log(m)}{T^{1/2}c_p} + \frac{\sqrt{mlog(m)}\{1+log(T)r_\pi^{1/2}(logm)^{K+1}\}}{\sqrt{c_pT}} + m\|\mathbf{b}_1\|_\infty\Big\}.$$

则有

$$\frac{1}{N_1N_2}\|\widehat{\mathbf{B}}-\mathbf{B}\|_F^2 = O_p\Big(\max[\min\{\nu_4'\alpha\|\mathbf{B}\|_*, N_2N_2r_B(\nu_4'\alpha)^2\}, \nu_4'(1-\alpha)\|\mathbf{B}\|_F^2]\Big), \tag{4.7}$$

其中， r_B 是$\mathbf{B}$的秩。

定理 3 表明，B 的误差界与 B 的低秩结构（即 $\|\mathrm{B}\|_*$ 和 r_B）、惩罚因子（即 ν_4 和 α）以及 B 的弗罗贝尼乌斯范数（即$\|\mathrm{B}\|_F^2$）有关。一般性证明遵循 Mao 等人（2019）的思路，结论（4.7）也与 Mao 等人（2019）的定理1一致。主要区别在于 Δ_{E,N_1N_2T} 的构造，其实际上是 $(N_1N_2)^{-1}\|\mathrm{X}\boldsymbol{\beta}+\mathrm{B}-\mathrm{E}\|$ 的误差界，其中 $\mathrm{E} = T^{-1}\sum_t \mathrm{E}_t$。在我们的情形下，建立这一误差界更具挑战性，因为还需要进一步处理矩阵值时间序列的时间依赖性。为此，我们使用了针对矩阵值数据的鞅差数组工具（Tropp，2011），以及随机矩阵理论（Ajanki等人，2017；Alt 等人，2019；Erdős 和 Müuhlbacher，2019）来建立误差界。我们注意到，Δ_{E,N_1N_2T} 表达式中的第二项和第三项是由第一步中 $\boldsymbol{\theta}$ 的估计误差引起的。特别地，与 Mao 等人（2019）的静态矩阵补全结果相比，在我们的情形下，通过增加 T 可以进一步降低误差率。

最后，我们指出，估计偏差$\mathbf{b}_1$ 会影响 $\boldsymbol{\theta}$的渐近性质。它也与定理 $\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|$ 中的2 以及 $\|\widehat{\mathbf{B}}-\mathbf{B}\|_F^2$中的定理3有关。因此，研究如何降低估计偏差以获得更好的估计量颇有意义。

4.3 偏差降低与校正

本节研究如何降低第一步估计中的估计偏差，这^ ^使我们能够进一步提高 $\boldsymbol{\beta}$ 和$\mathbf{B}$的估计效率。具体而言，我们可以利用矩阵数据估计式（4.3）和式（4.4）中的偏差项

$$\widehat{b}_i^{(1)\lambda} = \frac{2}{mT}\widetilde{\lambda}_i \sum_j \sum_k W_{1ik}^2 \frac{1}{T}\sum_t \left\{\widehat{Z}_{kjt}^2(\widehat{p}_k - 1)\right\}, \tag{4.8}$$

$$\widehat{b}_j^{(1)\gamma} = \frac{2}{mT}\widetilde{\gamma}_j \sum_i \sum_k W_{2kj}^2 \frac{1}{T}\sum_t \left\{\widehat{Z}_{ikt}^2(\widehat{p}_i - 1)\right\}, \tag{4.9}$$

其中 $\widehat{Z}_{kjt} = Y_{kjt}R_{kjt}/\widehat{p}_k$ 和 $\widehat{\mathbf{b}}^{(1)} = (\widehat{\mathbf{b}}^{(1)\lambda\top}, \widehat{\mathbf{b}}^{(1)\gamma\top})^\top$。将去偏估计量定义为 $\widehat{\boldsymbol{\theta}} = \widetilde{\boldsymbol{\theta}} - \widehat{\mathbf{b}}_{1,}$，其中 $\widehat{\mathbf{b}}_1 = \widetilde{\boldsymbol{\Sigma}}_2^{-1}(\widehat{\mathbf{b}}^{(1)} - (mT)^{-1}\widehat{\mathbf{r}})$ 且 $\widehat{\mathbf{r}} = 2(\nu_1\widetilde{\boldsymbol{\lambda}}^\top, \nu_2\widetilde{\boldsymbol{\gamma}}^\top)^\top$。这里，$\widehat{\mathbf{b}}_1$是式 (4.5) 中定义的 $\mathbf{b}_1$ 的估计量。以下定理给出了去偏估计量 $\boldsymbol{\theta}$ 的渐近性质^。

定理 4. (Bias Reduction) 假设满足定理 1中的相同条件。进一步假设 $T \gg \max\{c_p^{-1/2}p(logm)^{2K}, (\nu_1+\nu_2)^{2/3}(logm)^{4K/3}\}$。定义

$$b_i^{(2)\lambda} = \frac{1}{mT}\sum_k W_{1ik}^2 \sum_j E\{Z_{kjt}(Y_{kjt} - Z_{kjt})\}b_{1i}^\lambda,$$

$$b_j^{(2)\gamma} = \frac{1}{mT}\sum_k W_{2kj}^2 \sum_i E\{Z_{ikt}(Y_{ikt} - Z_{ikt})\}b_{1j}^\gamma,$$

以及 $\mathbf{b}^{(2)} = (\mathbf{b}^{(2)\lambda\top}, \mathbf{b}^{(2)\gamma\top})^\top$，其中 $\mathbf{b}^{(2)\lambda} = (b_i^{(2)\lambda} : 1 \le i \le N_1)^\top$，且 $\mathbf{b}^{(2)\gamma} = (b_i^{(2)\gamma} : 1 \le i \le N_2)^\top$。进一步定义 $\mathbf{b}_2 = -\widetilde{\boldsymbol{\Sigma}}_2^{-1}\mathbf{b}^{(2)}$。于是我们有

$$\sqrt{mT}c_p\boldsymbol{\eta}^\top\widetilde{\boldsymbol{\Sigma}}_2\big(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} - \mathbf{b}_2\big) \to N(0, \boldsymbol{\eta}^\top\boldsymbol{\Sigma}_1\boldsymbol{\eta}). \tag{4.10}$$

此外，我们有 $\|\mathbf{b}_2\|_\infty = O_p(T^{-2}(logm)^{4K}c_p^{-2})$。

^定理 4 表明，去偏估计量的渐近偏差 $\boldsymbol{\theta}$ 为$\mathbf{b}_2$，且收敛速率仍为 $\sqrt{mT}c_p$，与定理 1 中的相同。特别地，我们注意到$\|\mathbf{b}_2\|_\infty = \|\mathbf{b}_1\|_\infty O_p(c_p^{-1}T^{-1}(\log m)^{2K}) = O_p(\|\mathbf{b}_1\|_\infty^2)$。条件（4.2）表明

$c_p^{-1}T^{-1}(\log m)^{2K} \to 0$。因此，与$\mathbf{b}_1$相比，偏差$\mathbf{b}_2$进一步降低。此外，受上述结果启发，我们可以多次应用偏差缩减程序$\widehat{}$，以进一步降低偏差。假设在第 r步（$r \geq 2$）中，偏差估计量为$\mathbf{b}_r = (\widehat{}\mathbf{b}_r^{\lambda\top}, \widehat{}\mathbf{b}_r^{\gamma\top})^\top$。我们有 $\widehat{}\mathbf{b}_r = (-1)^{r+1}\widetilde{\mathbf{\Sigma}}_2^{-1}\widehat{}\mathbf{b}^{(r)}$，其中 $\widehat{}\mathbf{b}^{(r)} = (\widehat{}\mathbf{b}^{(r)\lambda\top}, \widehat{}\mathbf{b}^{(r)\gamma\top})^\top$，其中

$$\widehat{b}_i^{(r)\lambda} = \frac{1}{mT^2}\sum_k W_{1ik}^2 \sum_j \sum_t \{\widehat{Z}_{kjt}^2(\widehat{p}_k - 1)\}\widehat{b}_{r-1,i}^{\lambda},$$
$$\widehat{b}_j^{(r)\gamma} = \frac{1}{mT^2}\sum_k W_{2kj}^2 \sum_i \sum_t \{\widehat{Z}_{ikt}^2(\widehat{p}_i - 1)\}\widehat{b}_{r-1,j}^{\gamma}. \tag{4.11}$$

$\widehat{}$ $\widehat{}$ $\widehat{}$ $\widehat{}$将第 r步的去偏估计量记为 $\boldsymbol{\theta}_r$ $(r \geq 1)$，我们有 $\boldsymbol{\theta}_{r+1} = \boldsymbol{\theta}_r - \mathbf{b}_{r+1}$,$\widehat{}$ $\widehat{}$其中 $\boldsymbol{\theta}_1$ 设为 $\boldsymbol{\theta}$。当去偏迭代次数 r 足够大时，便可以得到偏差校正估计量。我们在下面的定理中陈述这一结果。

定理*5.* (Bias Correction)假设与定理*1*相同的条件。对于整数 $r = 1$、*2*、...，定义 $\boldsymbol{b}_r = (\boldsymbol{b}_r^{\lambda\top}, \boldsymbol{b}_r^{\gamma\top})^\top$ 和$\boldsymbol{b}_r = (-1)^{r+1}\widetilde{\mathbf{\Sigma}}_2^{-1}\boldsymbol{b}^{(r)}$，其中$\boldsymbol{b}^{(r)} = (\boldsymbol{b}^{(r)\lambda\top}, \boldsymbol{b}^{(r)\gamma\top})^\top$，其中

$$b_i^{(r)\lambda} = \frac{1}{mT}\sum_k W_{1ik}^2 \sum_j E\{Z_{kjt}(Y_{kjt} - Z_{kjt})\}b_{r-1,i}^{\lambda}, \tag{4.12}$$
$$b_j^{(r)\gamma} = \frac{1}{mT}\sum_k W_{2kj}^2 \sum_i E\{Z_{ikt}(Y_{ikt} - Z_{ikt})\}b_{r-1,i}^{\gamma}. \tag{4.13}$$

然后 $\|\mathbf{b}_{r+1}\|_\infty = \|\mathbf{b}_r\|_\infty O(c_p^{-1}(\text{对数}m)^{2K}/T)$。令 $\widehat{\boldsymbol{\theta}}_r = \widehat{\boldsymbol{\theta}}_{r-1} - \widehat{}\mathbf{b}_r$，其中 $\widehat{}\mathbf{b}_r$ 由（*4.11*）给出。假设 $\Delta_{N_1N_2} = o(T)$。那么我们有

$$\sqrt{mT}c_p\boldsymbol{\eta}^\top\widetilde{\mathbf{\Sigma}}_2\left\{\widehat{\boldsymbol{\theta}}_r - \boldsymbol{\theta} - \mathbf{b}_{r+1}\right\} \to_d N(0, \boldsymbol{\eta}^\top\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1\mathbf{\Sigma}_2^{-1}\boldsymbol{\eta}). \tag{4.14}$$

此外，我们有

$$\sqrt{mT}c_p\boldsymbol{\eta}^\top(\widehat{\boldsymbol{\theta}}_r - \boldsymbol{\theta}) \to_d N(0, \boldsymbol{\eta}^\top\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_1\mathbf{\Sigma}_2^{-1}\boldsymbol{\eta}) \tag{4.15}$$

对于 $r >$对数$\{m(\log m)^{2K}/T^{1/2}\}/\log[Tc_p/\{C(\log m)^{2K}\}]$，其中 $\|\mathbf{b}^{(1)}\|_\infty \lesssim C(\log m)^{2K}/(Tc_p)$.

定理 5 指出，当进行足够多轮的去偏处理时，渐近偏差可以忽略不计。根据该结果，令 $T = O(m^{\delta})$，其中 $\delta > 0$ 且 $c_p \geq c_0$为正常数 c_0，则只需有限轮次即可校正偏差。在我们的数值研究中发现，通常进行 2–3 轮偏差降低即可获得可靠估计量。

5 数值研究

5.1 模拟设计

为了展示 MNAR 模型在有限样本下的表现，本节将介绍多种模拟实验。我们首先讨论如何生成网络 $\mathbf{A}_1 \in \mathbb{R}^{N_1 \times N_1}$和$\mathbf{A}_2 \in \mathbb{R}^{N_2 \times N_2}$。下文分别将这两个网络称为行网络和列网络。在整个模拟研究中，我们假设行网络和列网络均遵循幂律分布模型。也就是说，我们考虑网络节点入度的幂律分布，这反映了实践中常见的一种网络现象。这意味着，网络中只有少数节点拥有大量关注者（即入度较大），而大多数节点的关注者很少。我们按照 Clauset 等人 （2009）的方法生成具有幂律分布模式的行邻接矩阵和列邻接矩阵$\mathbf{A}_1 \in \mathbb{R}^{N_1 \times N_1}$ 和$\mathbf{A}_2 \in \mathbb{R}^{N_1 \times N_1}$ 。具体而言，对于每个节点 i （$1 \leq i \leq N_1$），我们首先根据离散幂律分布生成其入度 $d_{1i} = \sum_j a_{1,\ ji}$ ，即 $P\left(d_{1i} = h\right) = ch^{-\upsilon}$，其中 c 是归一化常数， υ 是指数参数。然后，我们随机选取 d_{1i} 个节点作为节点 i的关注者。列网络 $\mathbf{A}_2$ 的生成方式类似。在两个网络中，我们设定 $\upsilon = 2.5$。生成$\mathbf{A}_1$ 和$\mathbf{A}_2$, 后，相应地计算行归一化邻接矩阵$\mathbf{W}_1$ 和列归一化邻接矩阵$\mathbf{W}_2$ 。

给定网络结构后，我们开始生成矩阵值时间序列 $\{\mathbf{Y}_t : 1 \leq t \leq T\}$。首先，令 $p = 6$ 并将协变量矩阵记为 $\mathbf{X} = (\mathbf{1}_{N_1},\ \mathbf{X}_{\text{sub}}) \in \mathbb{R}^{N_1 \times p}$,

其中，$\mathbf{X}_{\text{sub}}$ 服从标准正态分布。随后，我们按如下方式生成相应的系数矩阵 $\boldsymbol{\beta}$。首先，生成一个 $p \times N_2$ 维矩阵，其中每个元素独立服从 $U(-0.01, 0.01)$。然后，随机将其中 95% 的元素设为零，从而得到稀疏系数矩阵 $\boldsymbol{\beta}$。为生成低秩矩阵 B，我们先令 $r = 10$，并从均值为零、标准差为 0.5 的正态分布中生成 $\mathbf{U} \in \mathbb{R}^{N_1 \times r}$ 和 $\mathbf{V} \in \mathbb{R}^{N_2 \times r}$。接着，计算 $\mathbf{B} = \mathbf{P}_{\mathbf{X}}^{\perp}\mathbf{U}\mathbf{V}^{\top}$，以确保 B 和 X 的列空间彼此正交。接下来，对于对角矩阵 $\boldsymbol{\Lambda}$ 和 $\boldsymbol{\Gamma}$，我们令 $\lambda_i = 0.45$ 用于 $1 \le i \le N_1$，并令 $\gamma_i = 0.45$ 用于 $1 \le i \le N_2$。最后，噪声矩阵 $\mathcal{E}_t \in \mathbb{R}^{N_1 \times N_2}$ 中满足 $1 \le t \le T$ 的元素独立生成自标准正态分布。生成 $\boldsymbol{\Lambda}$、$\boldsymbol{\Gamma}$、X、$\boldsymbol{\beta}$、B 和 $\mathcal{E}_t$后，我们根据 MNAR 模型 (2.2) 生成矩阵 $\mathbf{Y}_t$。

在获得 $\mathbf{Y}_t$后，我们遵循 Mao 等人（2019）的做法，考虑用于生成观测矩阵 $\mathbf{Z}_t$的两种缺失机制。这里考虑的第一种缺失机制是随机缺失（MAR），其中我们采用逻辑回归模型生成 $R_{ijt} = 1$ for $1 \le i \le N_1$、$1 \le j \le N_2$ 和 $1 \le t \le T$ 的观测指示变量。具体而言，假设系数向量 $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_p)^{\top} \in \mathbb{R}^{p+1}$，并令 $\alpha_0 = -1.3$、$\alpha_i = 0.1$for $1 \le i \le p$。然后，观测概率 p_i 计算为 $p_i = \exp(\mathbf{X}_i^{\top}\boldsymbol{\alpha})/\{1 + \exp(\mathbf{X}_i^{\top}\boldsymbol{\alpha})\}$。随后，根据参数为 p_i 的伯努利分布生成观测指示变量 R_{ijt}。第二种缺失机制是均匀缺失（UNI），其中我们假设所有观测具有相同的缺失概率。为此，我们将观测概率设为 $p_i = 0.2$ for $1 \le i \le N_1$。选择观测概率 0.2，是为了近似 MAR 缺失机制下的平均观测概率。

我们设定 $N_1 = N_2 = N$，并考虑 $N = (100, 200, 400, 600)$。对于时间跨度，我们考虑 $T = (30, 60, 100)$。因此，在每种缺失机制下，共有 $4 \times 3 = 12$ 种实验设置。在每种实验设置下，我们重复实验 $R = 200$ 次，以获得可靠的评估结果。

5.2 性能测量

为评估 MNAR模型的估计性能，我们将其与 Mao等人 (2019年)提出的奇异值软阈值（SVT）方法进行比较，因为该方法在若干基准矩阵补全方法中表现出优势。由于 SVT 方法是一种静态方法，无法直接处理动态矩阵补全问题，因此我们对原始 SVT 方法进行了修改，以用于比较。具体而言，我们考虑了 SVT 方法的三个修改版本。首先，在每个单独的时间点上对数据集应用 SVT 方法，并将不同时间点得到的估计量作为 MNAR 的竞争方法。为方便起见，我们将此方法称为 SVT-SEP。其次，我们对不同时间点的 SVT 估计量取平均值，以获得更稳定的估计量，并称之为 SVT-AVG 估计量。最后，我们忽略时间依赖性，通过最小化以下目标函数来估计参数

$$Q_p^{(3)}(\boldsymbol{\beta}, \mathbf{B}) \stackrel{\text{def}}{=} \frac{1}{T}\sum_{t=1}^{T} \left\|\mathbf{Y}_t - \mathbf{X}\boldsymbol{\beta} - \mathbf{B}\right\|_F^2 + \nu_3\|\boldsymbol{\beta}\|_F^2 + \nu_4\Big(\alpha\|\mathbf{B}\|_* + (1-\alpha)\|\mathbf{B}\|_F^2\Big)$$

相应的估计量称为 SVT-SUM。

为估计 MNAR模型和三种 SVT类方法，必须先估计观测概率p_is。在 UNI 缺失机制下，使用$\mathbf{R}_t$的平均值作为观测概率的估计值。在 MAR 下，首先进行逻辑回归模型以估计 $\boldsymbol{\alpha}$。根据所得估计量 $\widehat{\boldsymbol{\alpha}}$，观测概率计算为 $\widehat{p_i} = \exp(\mathbf{X}_i^\top\widehat{\boldsymbol{\alpha}})/\{1+\exp(\mathbf{X}_i^\top\widehat{\boldsymbol{\alpha}})\}$。如$\widetilde{\quad}$$\widetilde{\quad}$前文所述，为实施非随机缺失方法，原始估计量 $\boldsymbol{\Lambda}$ 和 $\boldsymbol{\Gamma}$ 存在偏差。因此，我们进行偏差校正以获得更精确的估计量，在后续分析中将其记为 MNAR-ADJ。具体而言，对于短期数据集（即 $T = 30$），我们进行两次偏差缩减程序；而对于长期数据集（即 $T = 60$,100），只进行一次偏差降低。为便于比较，我们还计算了未采用偏差降低的原始估计量，并将其称为 MNAR-ORG。

对于每种方法（即 MNAR-ORG、MNAR-ADJ、SVT-SEP、SVT-AVG 和 SVT-

SUM），我们将 $\widehat{\boldsymbol{\beta}}^{(r)} = (\widehat{\beta}_{ij}^{(r)}) \in \mathbb{R}^{p\times N_2}$ 定义为第 r 次重复（$1 \le r \le R$）中 $\boldsymbol{\beta}$ 的估计量。随后，为评估各方法的估计效率，我们定义 $\boldsymbol{\beta}$ 的均方根误差（RMSE），即 $\text{RMSE}_{\boldsymbol{\beta}} = \{\sum_{r=1}^{R}\sum_{i,j}(\widehat{\beta}_{ij}^{(r)} - \beta_{ij})^2/(RpN_2)\}^{-1/2}$。其他参数（即 $\boldsymbol{\Lambda}$、 $\boldsymbol{\Gamma}$、$\mathbf{B}$）的估计性能也以类似方式计算。

接下来，我们评估矩阵补全任务的性能。定义 $\mathbf{Y}_t = \mathbf{A}_t + \mathcal{E}_t$，其中 $\mathbf{A}_t = E(\mathbf{Y}_t|\mathcal{F}_{t-1})$，且 $\mathcal{F}_{t-1} = \{\mathbf{Y}_s, \mathbf{R}_s : s \le t-1\}$。随后，我们评估 $\mathbf{A}_t$ 的估计性能，该矩阵是矩阵补全问题中的目标矩阵。对于 SVT类方法，$\mathbf{A}_t$ 的估计量可以直接计算。而在 MNAR模型中，我们有 $\mathbf{A}_t = \boldsymbol{\Lambda}\mathbf{W}_1\mathbf{Y}_{t-1} + \mathbf{Y}_{t-1}\mathbf{W}_2\boldsymbol{\Gamma} + \mathbf{X}\boldsymbol{\beta} + \mathbf{B}$。$\mathbf{A}_t$ 的估计需要首先恢复整个矩阵 $\mathbf{Y}_{t-1}$。为此，我们针对 $\mathbf{Y}_t$提出了一种滚动恢复$\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }$策略。具体而言，令 $\mathbf{Z}_0 = \mathbf{Z}_1$，并首先计算 $\mathbf{A}_1 = \boldsymbol{\Lambda}\mathbf{W}_1\mathbf{Z}_0 + \mathbf{Z}_0\mathbf{W}_2\boldsymbol{\Gamma} + \mathbf{X}\boldsymbol{\beta} + \mathbf{B}$，其中将 $\{\boldsymbol{\Lambda}, \boldsymbol{\Gamma}, \boldsymbol{\beta}, \mathbf{B}\}$ 相应地替换为其估计值。对于 $1 < t \le T$，我们对 $\mathbf{A}_t$ 进行滚动预测 $\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }$，即 $\mathbf{A}_t = \boldsymbol{\Lambda}\mathbf{W}_1\mathbf{A}_{t-1} + \mathbf{A}_{t-1}\mathbf{W}_2\boldsymbol{\Gamma} + \mathbf{X}\boldsymbol{\beta} + \mathbf{B}$。矩阵 $\mathbf{A}_t$ 的估计均方根误差（ RMSE ）计算为 $\text{RMSE}_{\mathbf{A}} = \{(RTN_1N_2)^{-1}\sum_{r=1}^{R}\sum_{t=1}^{T}\|\widehat{\ }\mathbf{A}_t^{(r)} - \mathbf{A}_t\|_F^2\}^{1/2}$，其中 $\widehat{\ }\mathbf{A}_t^{(r)}$ 是第 r 个模拟轮次中 $\mathbf{A}_t$ 的估计值。最后，为了纯粹量化矩阵补全的性能，我们采用测试误差这一度量（Mao 等，2019），其关注矩阵中的缺失值。测试误差计算为 $\sum_{i,j,t}\{R_{ijt}^{\dagger}(\widehat{A}_{ijt} - Y_{ijt})\}^2 / \sum_{i,j,t}(R_{ijt}^{\dagger}Y_{ijt})^2$, ，其中 $R_{ijt}^{\dagger} = 1 - R_{ijt}$。

5.3 模拟结果

表 3 和4 分别展示了 MAR 和 UNI 缺失机制下的模拟结果。总体而言，不同缺失机制下的模拟结果相似。具体来说，我们可以得出以下结论。首先，与 SVT 类方法相比，所提出的 MNAR-ORG 和 MNAR-ADJ 估计量在所有实验设置下都取得了更低的均方根误差和更低的测试误差。这些结果表明，在非随机缺失方法中考虑网络信息，可以获得更好的估计性能和矩阵补全性能。在 SVT 类估计量中，总体而言，SVT-AVG 表现最佳，其次是 SVT-SUM 和 SVT-SEP。其次，通过进行偏差缩减操作，

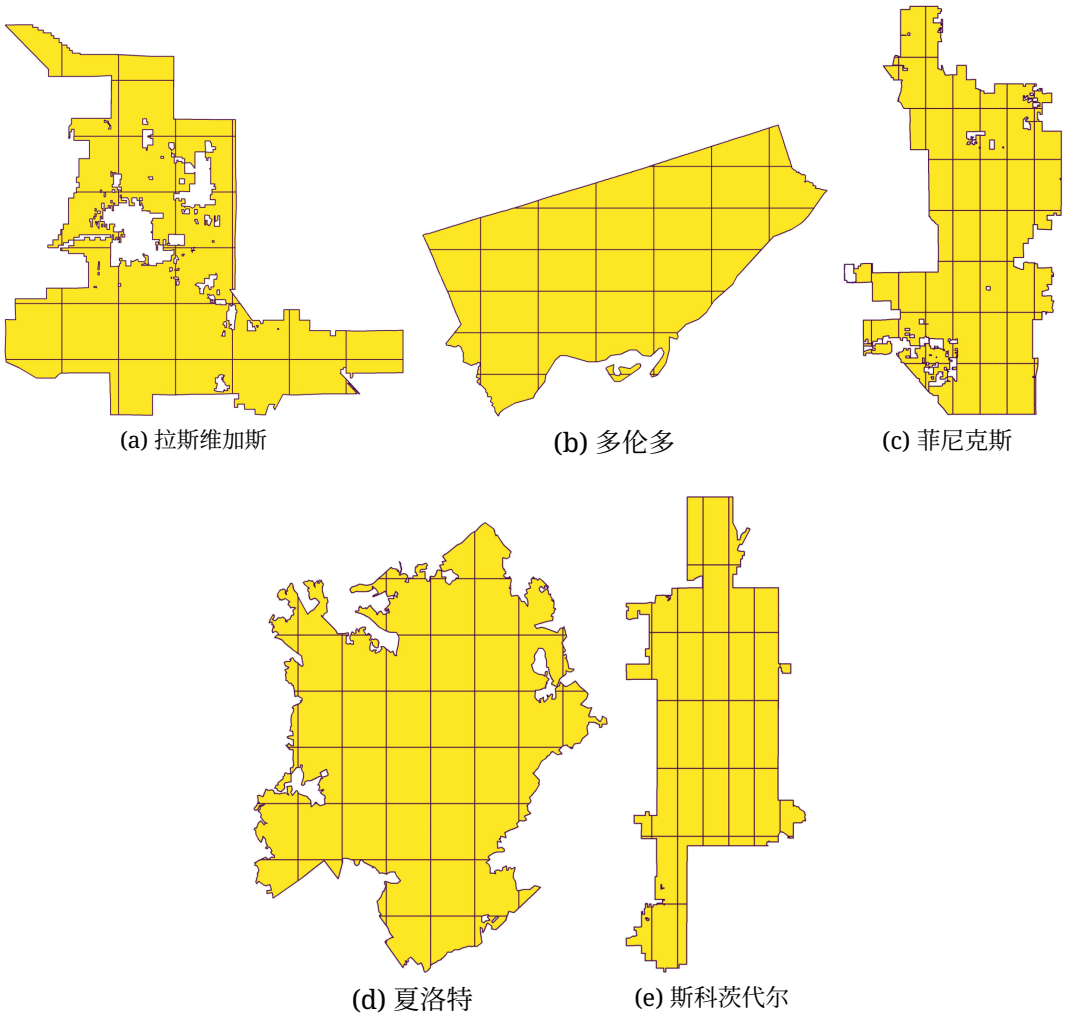
MNAR-ADJ 估计量的估计性能优于 MNAR-ORG 估计量。这一发现证明了偏差缩减操作的有用性。特别是，MNAR-ADJ 方法的均方根误差$_{\mathbf{A}}$ 和测试误差低于其他竞争方法，这说明了所提出的非随机缺失方法在处理矩阵补全任务方面的潜力。最后，随着样本量（N_1, N_2）或时间跨度 T增加，MNAR-ORG 和 MNAR-ADJ 中所有估计量的均方根误差都降低，这表明相应估计量具有一致性。

6 真实数据分析

为展示 MNAR 的实际表现，我们使用 Yelp 上的一个大型公开数据集开展实证研究。Yelp 是美国最大的网站评论平台。原始数据集可从 https://www.yelp.com/dataset 获取。该数据集包含 2010 年至 2018 年间 Yelp 商家、评论和用户的详细信息。整个数据集分为五个部分，分别是：(1) 商店信息（如餐馆、家庭服务等），(2) 用户信息，(3) 用户对商店的评论，(4) 用户为商店撰写的简短提示，以及 (5) 商家的汇总签到数据。

我们关注用户的评论行为，并试图探究有助于预测用户评论行为的影响因素。分析在城市层面进行。具体而言，我们选取商家店铺数量最多的五个城市（即拉斯维加斯、多伦多、菲尼克斯、夏洛特和斯科茨代尔）作为例子，这些店铺均来自该数据集。随后，我们将每个城市的所有店铺划分到不同区。为此，我们首先按升序排列所有店铺的经度，然后将它们均匀划分为 G 部分。对所有店铺的纬度也进行相同的操作。由此总共得到 $N_2 \leq G^2$ 个区。图 1 展示了每个城市划分后的地区。接着，为某个城市中的 N_2 个区构建空间邻接矩阵 $\mathbf{A}_2$。具体而言，如果区 $a_{2ij} = 1$ 与区i 相邻，则定义 j；否则，令 $a_{2ij} = 0$。随后，我们进行基本的数据清理，仅保留评论总数大于

20。然后，我们在活跃用户之间构建友谊网络 $\mathbf{A}_1$ 。具体而言，如果用户 i 是用户 j的朋友，则设定$a_{1ij} = 1$；否则设定 $a_{1ij} = 0$。



(a) 拉斯维加斯 (b) 多伦多 (c) 菲尼克斯

(d) 夏洛特 (e) 斯科茨代尔

图 1：各城市划分地区的地理地图。我们为拉斯维加斯、多伦多和斯科茨代尔设置了 $G = 7$, 为菲尼克斯和夏洛特设置了 $G = 8$ 。

各城市最终数据集的详细描述汇总于表1的A部分。为刻画各城市的网络结构，我们计算了 $\mathbf{A}_1$ 和 $\mathbf{A}_2$, 的网络密度，即 $\sum_i \sum_j a_{1ij}/N_1^2$和 $\sum_i \sum_j a_{2ij}/N_2^2$。如表1所示，所有城市的用户网络 $\mathbf{A}_1$ 都相当稀疏。在每个城市中，我们将响应变量 Y_{ijt} 定义为用户 i 在年份 t 评论区 j 内商店的平均得分。按照推荐系统中的常见做法（Berg 等，2017；Mao 等，2019；Zhang 和 Chen，2020），我们将用户的平均得分视为缺失值。具体而言，令 R_{ijt} 表示

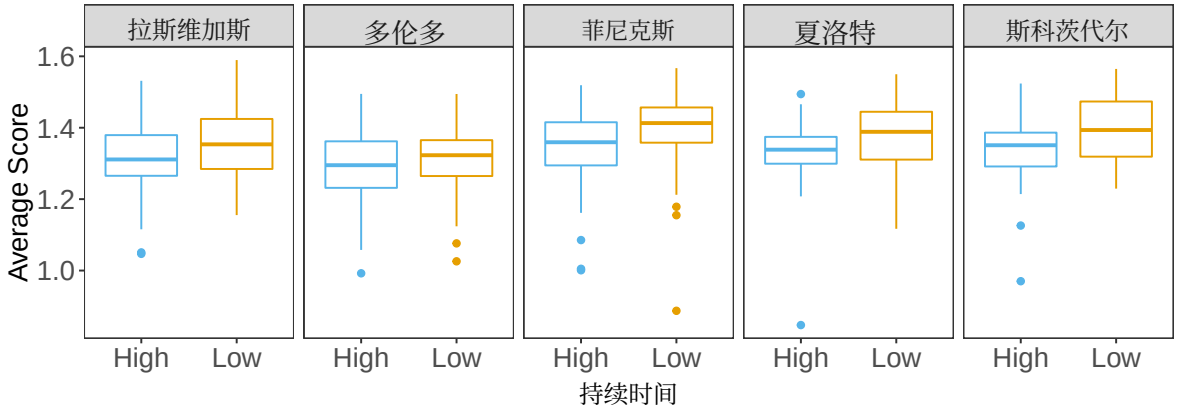
是否观测到用户 i 在区 j 于时间 t评论商店。于是，只有当 $R_{ijt} = 1$时，我们才能观测到 Y_{ijt}。随后，我们计算各城市的观测率，并将其列于表 1。由此可见，所有城市的观测率都很低，说明每个城市的整个矩阵都很稀疏。

表1：五个城市的数据和变量描述。

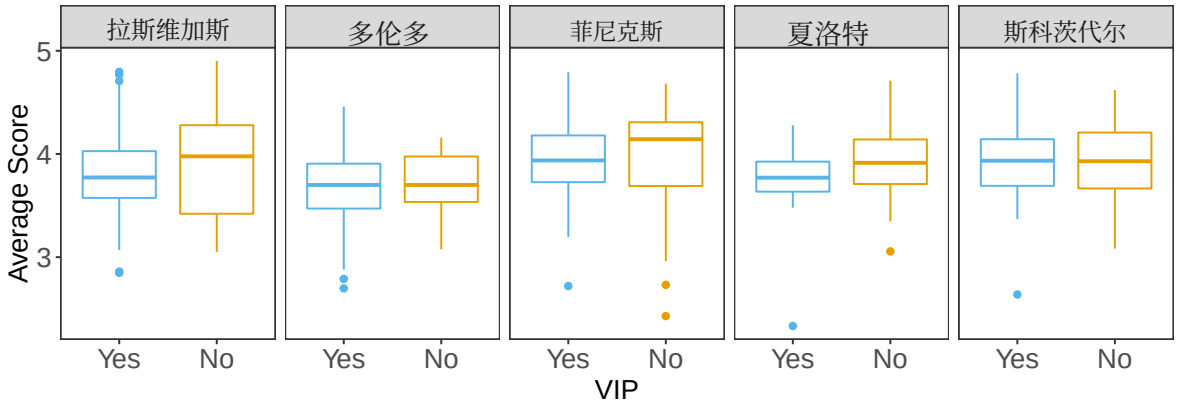
City	拉斯维加斯	多伦多	菲尼克斯	夏洛特	斯科茨代尔
*A*部分：数据基本统计					
N_1	248	269	168	101	112
N_2	49	49	64	64	49
Time	2010–2018年	2010–2018年	2010–2018年	2010–2018年	2010–2018年
A的密度$_1$	0.0022	0.0055	0.0074	0.0131	0.0061
A的密度$_2$	0.1531	0.1531	0.1181	0.1181	0.1531
观测率	0.2388	0.2932	0.1833	0.2813	0.1980
*B*面板：变量均值					
持续时间	0.3952	0.6171	0.6131	0.4059	0.6607
VIP%	0.4677	0.1599	0.3214	0.1881	0.3393
有用	1.8287	0.9139	2.0530	0.8669	1.9100
有趣	1.2570	0.3592	1.3300	0.3755	1.1880
Cool	1.4413	0.5176	1.3960	0.4638	1.3110
平均得分	3.8477	3.6765	3.9724	3.8973	3.9339

我们在分析中考虑了五个用户特定协变量，这些协变量是根据截至 2010 年的用户信息计算得出的。具体而言，第一个协变量是用户首次注册至 2010 年的年数，我们将其称为“持续时间”。第二个协变量表示用户是否为 VIP，分别用 1 和 0 编码。最后三个协变量分别是用户发布的“有用”“酷”和“有趣”评论的累计数量。对于观测响应和全部五个协变量，我们计算了它们在每个城市中的均值，汇总结果见 表 *1* 的 *B*面板。我们还探讨了响应与各协变量之间的关系。为此，我们首先计算了整个时间跨度内每位用户在所有地区发布的评论的平均得分，然后

研究不同协变量条件下每位用户平均得分的分布。作为说明，图 2 展示了五个城市中不同持续时间组（按中位数划分）和 VIP 组的每位用户平均得分箱线图。根据图 2，我们发现，在所有五个城市中，注册时间较短的用户往往给出较高的评分。此外，在拉斯维加斯、菲尼克斯和夏洛特，VIP 用户给出的评分往往低于非 VIP 用户。



(a) 每位用户的平均评分与两个时长等级的关系



(b) VIP 用户与非 VIP 用户的每位用户的平均评分

图2: 不同持续时间或 VIP 分组中每位用户平均评分的箱线图 .

随后，我们使用 MNAR 方法研究各城市用户的评分行为。为便于比较，我们将 SVT 类方法（即 SVT-AVG 和 SVT-SUM）（Mao 等人，2019年）作为竞争方法。鉴于 SVT-SEP 方法在模拟研究中的表现较差，因此不予考虑。对于 MNAR 方法，考虑到时间跨度相对较短，我们进行了两轮偏差缩减操作。MNAR-ORG

方法也用于比较。我们分别考虑两种缺失机制：MAR 和 UNI。在 MAR 缺失机制下，观测概率 p_i 通过上述五个用户特定协变量拟合逻辑回归模型进行估计。在 UNI 缺失机制下，直接使用表1 中所示的经验观测概率。为选择调优参数，我们将每个城市的整个数据集划分为三部分：（1）时间为 $t = 1$ 至 $T - 2$的训练数据集；（2）时间为 $T - 1$的验证数据集；（3）时间为 T的测试数据集。首先，我们使用不同的调优参数值，在训练数据集上估计各个模型。然后，我们在验证数据集上评估预测性能，以选择最佳调优参数。为评估预测性能，我们根据验证数据集中的观测响应及其对应的预测值计算均方根误差（RMSE）。选定调优参数后，将训练数据集和验证数据集合并以训练最终模型，并使用最后一年的测试数据进行方法比较。

两种缺失机制下的均方根误差详细结果汇总于表2中。在每种缺失机制下，所提出的 MNAR 方法在五个城市中的均方根误差均低于 SVT 类方法，表现更优。此外，在所有城市中，MNAR-ADJ 方法的预测性能均优于 MNAR-ORG 方法。这再次证明了偏差缩减操作的有效性。比较不同缺失机制下的结果，我们发现，MAR 缺失机制下的均方根误差均小于 UNI 缺失机制下的结果。这表明，对于该数据集的分析任务，考虑协变量信息的 MAR 缺失机制是必要的。

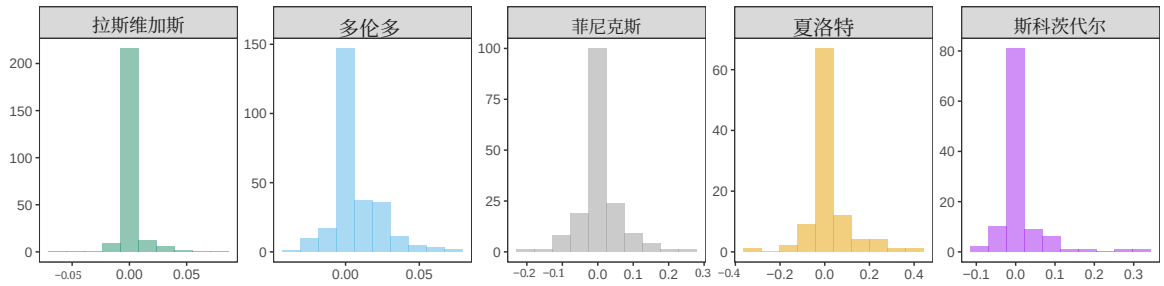
接下来，我们重点关注估计得到的网络效应。以 MAR 缺失机制下 MNAR-ADJ 估计的用户特定效应（即 $\widehat{\ }$）和区域特定效应（即 λ_i）为例。为展示网络效应，图 3 展示了五个城市中估计得到的用户特定效应和区域特定效应。如图所示，估计值$\widehat{\ }$ λ_i 和 $\widehat{\gamma}_j$ 在不同城市中有所不同。总体而言，大多数估计得到的用户特定效应$\backsim\lambda_i$ 介于 -0.4 和 0.4 之间。这一发现表明，朋友对

表2：不同方法在五个城市中得到的样本外均方根误差值

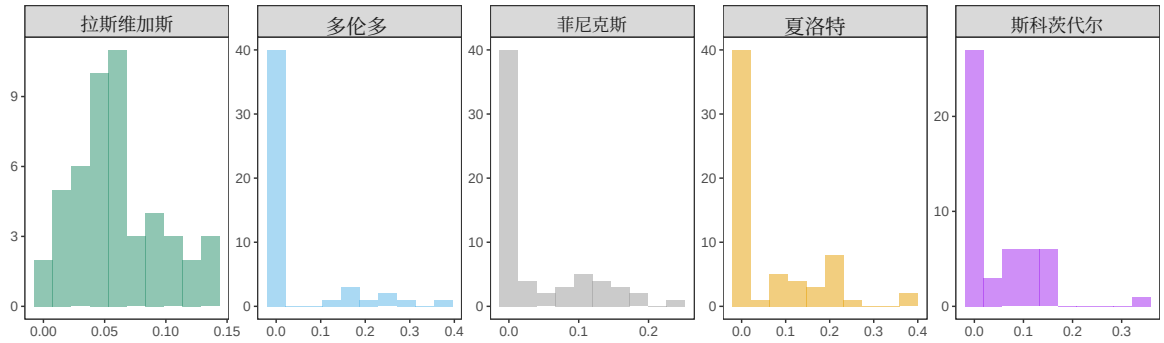
	City	拉斯维加斯	多伦多	菲尼克斯	夏洛特	斯科茨代尔
MAR	MNAR-ORG	0.5703	0.4319	0.4445	0.4821	0.5300
	MNAR-ADJ	0.5011	0.3644	0.3846	0.4149	0.4537
	SVT-AVG	0.6191	0.4392	0.4815	0.5239	0.5725
	SVT-SUM	0.6239	0.4646	0.4907	0.5386	0.5829
UNI	MNAR-ORG	0.9258	0.6617	1.1018	0.7363	1.0626
	MNAR-ADJ	0.7362	0.6125	0.9936	0.6153	0.9442
	SVT-AVG	1.1161	0.7077	1.2980	0.8326	1.3058
	SVT-SUM	1.1370	0.7779	1.3291	0.8853	1.3299

用户评论行为的影响可能是负面的，也可能是正面的。相比之下，估计得到的区域特定效应 $\widehat{\gamma}_j$ 均为非负值。这表明邻近商家总体上具有正向影响。此外，$\widehat{\gamma}_j$ 的偏斜分布反映了不同地区的非对称影响。

最后，我们检验估计得到的网络效应的显著性。如图3所示，许多用户或地区的网络效应非常小。通过设定阈值 $\delta = 0.05$，如果用户对应的用户特定效应大于该阈值，则将其视为活跃用户。具体而言，定义 $\Delta_u = \{i : \lambda_i > \delta\}$ 为活跃用户集合。类似地，定义 $\Delta_d = \{j : \widehat{\gamma}_j > \delta\}$ 为活跃地区集合。随后，我们将所有活跃用户的平均用户特定效应计算为 $\bar{\lambda} = \sum_{i\in\Delta_u} \lambda_i/|\Delta_u|$，其中 $|\Delta_u|$ 表示活跃用户总数。类似地，所有活跃地区的平均区域特异性效应可计算为 $\bar{\gamma} = \sum_{j\in\Delta_d} \widehat{\gamma}_j/|\Delta_d|$。然后，根据定理5检验 $\bar{\lambda}$ 和 $\bar{\gamma}$ 的显著性。具体而言，为计算 $\bar{\lambda}$的渐近方差，定理5中使用的向量 $\boldsymbol{\eta}$ 定义如下。设 $\boldsymbol{\eta}_1 = \{\eta_{11}, ..., \eta_{1N_1}\}^\top \in \mathbb{R}^{N_1}$。然后，当 $i \in \Delta_u$时定义 $\eta_{1i} = 1/|\Delta_u|$，否则定义为 $\eta_{1i} = 0$。然后，$\boldsymbol{\eta} = (\boldsymbol{\eta}_1^\top, \mathbf{0}_{N_2})^\top$。类似地，设 $\boldsymbol{\eta}_2 = \{\eta_{21}, ..., \eta_{2N_2}\}^\top \in \mathbb{R}^{N_2}$，其中当 $j \in \Delta_d$时 $\eta_{2j} = 1/|\Delta_d|$，否则为 $\eta_{2j} = 0$。随后，为计算 $\bar{\gamma}$的渐近方差，定理5中使用的向量 $\boldsymbol{\eta}$ 定义为 $\boldsymbol{\eta} = (\mathbf{0}_{N_1}, \boldsymbol{\eta}_2^\top)^\top$。



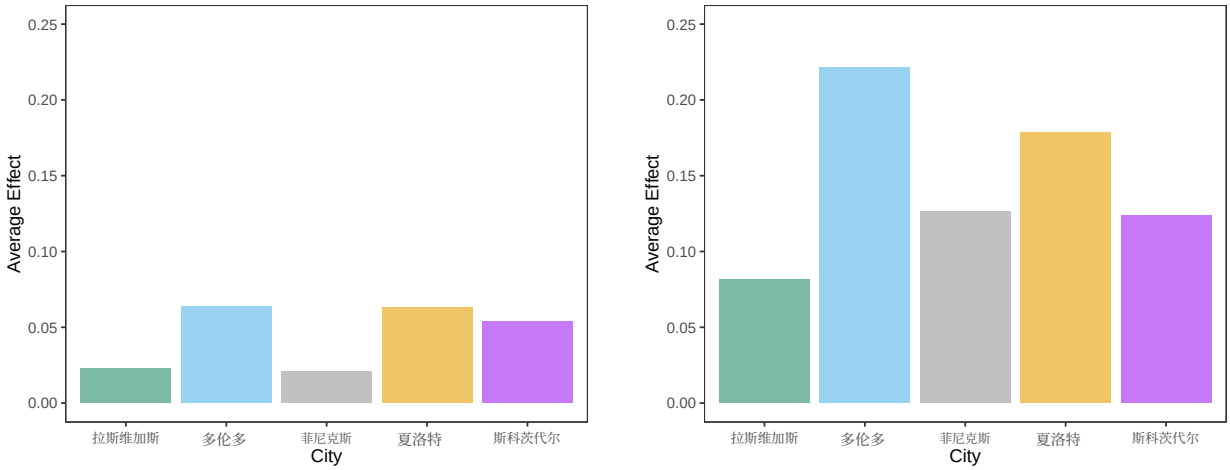
(a) $\widehat{\lambda}_i$ 与 $1 \leq i \leq N_1$



(b) $\widehat{\gamma}_j$ 与 $1 \leq j \leq N_2$

图3：五个城市中用户特定效应 λ_i 和区域特定效应 $\widehat{\gamma}_j$估计值的直方图。

图4 展示了 λ 每个城市中的平均用户特定效应和平均区域特异性效应 $\bar{\gamma}$。在显著性水平为5%的情况下，我们发现所有 λ和 $\bar{\gamma}$均具有显著性。具体而言，五个城市的平均用户特定效应相对较小，所有数值均小于0.1。与平均用户特定效应相比，平均区域特异性效应更大。这一发现表明，Yelp平台上用户的行为更有可能受到访问地点而非朋友的影响。此外，在五个城市中，多伦多和夏洛特的用户特定效应及区域特定效应比其他三个城市更为明显。这些结果再次证明了不同城市的异质性特征。



(a) 平均用户特定效应　　(b) 平均区域特异性效应

图4：五个城市的平均用户特定效应（λ）和区域特定效应（$\bar{\gamma}$）。

7 总结

在本文中，我们提出了矩阵网络自回归模型，该模型能够处理不完整的矩阵观测。现在，我们讨论一些对本文工作的潜在扩展，这些扩展可能会成为未来研究中有趣的课题。首先，为了更好地刻画矩阵值时间序列数据的动态特征，可以进一步考虑并研究矩阵因子结构（Wang 等，2019年；Chen 和 Fan，2021）。其次，可以引入非线性和非参数建模框架，以探索矩阵值时间序列数据更为灵活的动态特征。第三，可以采用更为精细的处理方法（Tsiatis，2009），以处理缺失响应；这可能会提升模型的性能，尤其是在 c_p 较小时。另一方面，考虑缺失机制进一步依赖于响应本身的情形也可能颇具意义。这种情形更具挑战性，但在某些应用中可能十分重要。

参考文献

Ajanki、O. H.、Erd¨os、L. 和 Kr¨uger、T.（2017年），“一般 Wigner 型矩阵的普适性”，概率论及相关领域，169，667–727。

Alt, J., Erd¨os, L., Kr¨uger, T., and Nemish, Y.（2019年），“Kronecker 随机矩阵的谱位置”，《亨利·庞加莱研究所年刊》´概率´与统计，55，661 – 696。

Arellano, M., Hahn, J., 等（2007），“理解非线性面板模型中的偏差：近期的一些进展”，计量经济学会专著，43，381。

Berg, R., Kipf, T. N., andWelling, M.（2017），“图卷积矩阵补全”，*arXiv* 预印本 *arXiv:1706.02263v2*。

Bi, X., Qu, A., Wang, J., and Shen, X.（2017），“群体特定推荐系统”，《美国统计学会杂志》，112，1344–1353。

Cai, T.、Cai, T. T. 和 Zhang, A.（2016），“应用于基因组数据整合的结构化矩阵补全”，《美国统计学会杂志》，111，621–633。

Chang, J.、Chen, X. 和 Wu, M.（2021a），“高维相关数据的中心极限定理”，*arXiv* 预印本 *arXiv:2104.12929*。

Chang, J.、He, J.、Yang, L. 和 Yao, Q.（2021b），“通过张量 CP 分解对矩阵时间序列进行建模”，*arXiv* 预印本 *arXiv:2112.15423*。

Chen, E. Y. 和 Fan, J.（2021），“高维矩阵值因子模型的统计推断”，《美国统计学会杂志》，1–18。

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

Chen, R.、Yang, D. 和 Zhang, C.-H.（2021b），《高维张量时间序列的因子模型》，《美国统计学会杂志》，1–23。

Chernozhukov, V.、Chetverikov, D. 和 Kato, K.（2017），《高维中心极限定理与自助法》，《概率年刊》，45，2309–2352。Clauset, A.、Shalizi, C. R. 和 Newman, M. E.（2009），《经验数据中的幂律分布》，《*SIAM*评论》，51，661–703。Erd¨os, L. 和 M¨uhlbacher, P.（2019），《Wigner 型随机矩阵范数的界》，《随机矩阵：理论与应用》，8，1950009。Fan, J.、Ke, Y. 和 Liao, Y.（2021），《增广因子模型：在验证市场风险因子和预测债券风险溢价中的应用》，《计量经济学杂志》，222，269–294。Jackson, M. O.（2010），社会与经济网络，普林斯顿大学出版社。

Koike, Y.(2022),“齐次和的高维中心极限定理”，《理论概率杂志》，1–45。

Koltchinskii, V., Lounici, K., 和Tsybakov, A. B. (2011),“含噪低秩矩阵补全的核范数惩罚与最优速率”，《统计学年刊》，39，2302– 2329。

Kong, X.-B., Liu, Y.-X., Yu, L., 和 Zhao, P. (2022),“矩阵分位数因子模型”，*arXiv*预印本 *arXiv:2208.08693*。

Leng, C. 和 Tang, C. Y. (2012),“稀疏矩阵图模型”，《美国统计学会杂志》，107，1187–1200。

Lugosi, G. 和 Mendelson, S.（2019年），《随机向量均值的次高斯估计量》，《统计学年刊》，47，783–794。L¨utkepohl, H.（2005），多元时间序列分析新导论，施普林格科学与商业媒体。Ma, S.、Goldfarb, D. 和 Chen, L.（2011），《矩阵秩最小化的定点迭代法和 Bregman 迭代法》，《数学规划》，128，321–353。

Mao, X.、Chen, S. X. 和 Wong, R. K.（2019年），《利用协变量信息进行矩阵补全》，《美国统计学会杂志》，114，198–210。

Mazumder, R., Hastie, T., 和 Tibshirani, R.（2010），《用于学习大型不完整矩阵的谱正则化算法》，机器学习研究期刊，11，2287–2322。

Newman, M., Barabasi, A.-L., 和 Watts, D. J.（2011），网络的结构与动力学，普林斯顿大学出版社。

Rohde, A. 和 Tsybakov, A. B.（2011），《高维低秩矩阵的估计》，《统计学年刊》，39，887–930。

Tropp, J.（2011），《矩阵鞅的 Freedman不等式》，《电子概率通讯》，16，262–270。

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., 和 Altman, R. B.（2001），《DNA微阵列缺失值估计方法》，生物信息学，17，520–525。

Tsiatis, A. A.（2009），半参数理论与缺失数据，纽约：施普林格。

Wang, D.、Liu, X. 和 Chen, R.（2019年），《矩阵值高维时间序列的因子模型》，《计量经济学杂志》，208，231–248。

Wang, D.、Zheng, Y. 和 Li, G.（2021），《高维低秩张量自回归时间序列建模》，*arXiv* 预印本 *arXiv:2101.04276*。

Wang, H.（2020），“针对稀有事件海量数据的逻辑回归”，载于国际机器学习大会，PMLR，第9829–9836页。Wang, L.、Kim, Y.和Li, R.（2013），“超高维中的非凸惩罚回归校准”，《统计年鉴》，41，2505–2536。Watts, D. J.和Strogatz, S. H.（1998），“小世界网络的集体动力学”，自然，393，440–442。

Wu, Y.、Lan, W.、Zou, T.和Tsai, C.-L.（2021），“内向和外向网络影响分析”，《商业与经济统计》，1–12。

Yu, L., He, Y., Kong, X., and Zhang, X. (2021), “大维矩阵因子模型的投影估计”，《计量经济学杂志》。

Zhang, M. and Chen, Y. (2020), “基于图神经网络的归纳式矩阵补全”，第*37*届国际机器学习会议。

Zhou, H. and Li, L. (2014), “正则化矩阵回归”，《皇家统计学会杂志：*B*辑》（统计方法学），76，463–483。

Zhou, J., Tu, Y., Chen, Y., and Wang, H. (2017), “利用抽样网络数据估计空间自相关”，《商业与经济统计》，35，130–138。

Zhou, S. (2014), “Gemini：利用矩阵变量正态实例进行图估计”，《统计学年刊》，42，532–562。

Zhu, X., Cai, Z., and Ma, Y. (2021), “网络函数变系数模型”，《美国统计学会杂志》，1–12。

Zhu, X., Pan, R., Li, G., Liu, Y., and Wang, H.（2017），“网络向量自回归”，统计年鉴，45，1096–1123。

Table 3: Simulation results for estimation performance under the missing mechanism of MAR. The empirical root mean square errors (RMSE) for $(\mathbf{\Lambda}, \mathbf{\Gamma}, \boldsymbol{\beta}, \mathbf{B}, \mathbf{A})$, as well as the test errors under different methods are reported, respectively. For illustration purpose, the RMSEs for $\boldsymbol{\beta}$ are multiplied by 100.

N_1	N_2	$\boldsymbol{\theta}$	SVT			MNAR		SVT			MNAR		SVT			MNAR	
			SEP	AVG	SUM	ORG	ADJ	SEP	AVG	SUM	ORG	ADJ	SEP	AVG	SUM	ORG	ADJ
			$T = 30$					$T = 60$					$T = 100$				
100	100	$\mathbf{\Lambda}$	–	–	–	0.383	0.337	–	–	–	0.331	0.282	–	–	–	0.280	0.230
		$\mathbf{\Gamma}$	–	–	–	0.385	0.340	–	–	–	0.331	0.281	–	–	–	0.284	0.234
		$\boldsymbol{\beta}$	4.577	1.309	1.282	1.153	0.996	4.581	1.166	1.153	0.906	0.765	4.581	1.104	1.096	0.733	0.577
		$\mathbf{B}$	1.619	0.260	0.263	0.255	0.254	1.619	0.202	0.203	0.184	0.180	1.619	0.175	0.175	0.146	0.141
		$\mathbf{A}$	2.650	0.310	0.313	0.302	0.295	2.651	0.264	0.265	0.244	0.235	2.650	0.245	0.245	0.213	0.204
		Error	1.491	0.158	0.162	0.151	0.143	1.492	0.116	0.117	0.098	0.091	1.492	0.099	0.099	0.075	0.069
200	200	$\mathbf{\Lambda}$	–	–	–	0.335	0.295	–	–	–	0.268	0.224	–	–	–	0.217	0.183
		$\mathbf{\Gamma}$	–	–	–	0.348	0.303	–	–	–	0.286	0.240	–	–	–	0.232	0.194
		$\boldsymbol{\beta}$	1.741	0.371	0.366	0.347	0.333	1.742	0.290	0.288	0.260	0.252	1.742	0.249	0.248	0.215	0.209
		$\mathbf{B}$	1.211	0.221	0.223	0.199	0.194	1.211	0.187	0.188	0.145	0.138	1.211	0.172	0.173	0.114	0.108
		$\mathbf{A}$	2.713	0.239	0.241	0.232	0.231	2.713	0.209	0.209	0.195	0.193	2.713	0.196	0.196	0.178	0.176
		Error	1.246	0.089	0.090	0.084	0.083	1.246	0.068	0.068	0.059	0.058	1.246	0.060	0.060	0.049	0.048
400	400	$\mathbf{\Lambda}$	–	–	–	0.293	0.247	–	–	–	0.222	0.189	–	–	–	0.174	0.152
		$\mathbf{\Gamma}$	–	–	–	0.283	0.239	–	–	–	0.212	0.182	–	–	–	0.165	0.145
		$\boldsymbol{\beta}$	0.721	0.281	0.275	0.212	0.185	0.721	0.267	0.264	0.173	0.153	0.721	0.260	0.258	0.152	0.140
		$\mathbf{B}$	0.791	0.176	0.178	0.145	0.140	0.791	0.156	0.156	0.104	0.101	0.791	0.147	0.147	0.081	0.079
		$\mathbf{A}$	2.733	0.228	0.230	0.204	0.197	2.733	0.214	0.214	0.176	0.172	2.733	0.207	0.207	0.163	0.161
		Error	1.193	0.081	0.082	0.065	0.060	1.194	0.071	0.071	0.049	0.046	1.193	0.067	0.067	0.042	0.040
600	600	$\mathbf{\Lambda}$	–	–	–	0.250	0.211	–	–	–	0.180	0.158	–	–	–	0.138	0.125
		$\mathbf{\Gamma}$	–	–	–	0.248	0.209	–	–	–	0.180	0.158	–	–	–	0.138	0.125
		$\boldsymbol{\beta}$	0.396	0.159	0.158	0.147	0.144	0.396	0.151	0.150	0.137	0.136	0.396	0.148	0.147	0.133	0.132
		$\mathbf{B}$	0.611	0.160	0.161	0.119	0.115	0.611	0.145	0.145	0.084	0.082	0.611	0.138	0.138	0.065	0.064
		$\mathbf{A}$	2.735	0.194	0.195	0.179	0.177	2.735	0.182	0.183	0.161	0.160	2.735	0.177	0.178	0.153	0.152
		Error	1.108	0.059	0.060	0.050	0.049	1.108	0.052	0.052	0.041	0.040	1.108	0.049	0.049	0.037	0.036

Table 4: Simulation results for estimation performance under the missing mechanism of UNI. The empirical root mean square errors (RMSE) for $(\mathbf{\Lambda}, \mathbf{\Gamma}, \boldsymbol{\beta}, \mathbf{B}, \mathbf{A})$, as well as the test errors under different methods are reported, respectively. For illustration purpose, the RMSEs for $\boldsymbol{\beta}$ are multiplied by 100.

N_1	N_2	$\boldsymbol{\theta}$	SVT			MNAR		SVT			MNAR		SVT			MNAR	
			SEP	AVG	SUM	ORG	ADJ	SEP	AVG	SUM	ORG	ADJ	SEP	AVG	SUM	ORG	ADJ
			$T = 30$					$T = 60$					$T = 100$				
100	100	$\mathbf{\Lambda}$	–	–	–	0.383	0.337	–	–	–	0.331	0.282	–	–	–	0.280	0.230
		$\mathbf{\Gamma}$	–	–	–	0.385	0.340	–	–	–	0.331	0.281	–	–	–	0.284	0.234
		$\boldsymbol{\beta}$	4.573	1.308	1.280	1.152	0.996	4.577	1.165	1.152	0.905	0.765	4.576	1.103	1.095	0.733	0.577
		$\mathbf{B}$	1.618	0.260	0.263	0.255	0.254	1.618	0.202	0.203	0.184	0.180	1.618	0.175	0.175	0.146	0.141
		$\mathbf{A}$	2.647	0.310	0.313	0.302	0.295	2.648	0.264	0.265	0.244	0.235	2.648	0.245	0.245	0.213	0.204
		Error	1.490	0.158	0.162	0.151	0.143	1.491	0.116	0.116	0.098	0.091	1.491	0.099	0.099	0.075	0.069
200	200	$\mathbf{\Lambda}$	–	–	–	0.335	0.295	–	–	–	0.268	0.224	–	–	–	0.217	0.183
		$\mathbf{\Gamma}$	–	–	–	0.348	0.303	–	–	–	0.286	0.240	–	–	–	0.232	0.194
		$\boldsymbol{\beta}$	1.741	0.371	0.366	0.347	0.333	1.742	0.290	0.288	0.260	0.252	1.742	0.249	0.248	0.215	0.209
		$\mathbf{B}$	1.211	0.221	0.223	0.199	0.194	1.211	0.187	0.188	0.145	0.138	1.211	0.172	0.173	0.114	0.108
		$\mathbf{A}$	2.712	0.239	0.241	0.232	0.231	2.712	0.209	0.209	0.195	0.193	2.712	0.196	0.196	0.178	0.176
		Error	1.246	0.088	0.090	0.084	0.083	1.246	0.068	0.068	0.059	0.058	1.246	0.060	0.060	0.049	0.048
400	400	$\mathbf{\Lambda}$	–	–	–	0.293	0.247	–	–	–	0.222	0.189	–	–	–	0.174	0.152
		$\mathbf{\Gamma}$	–	–	–	0.283	0.239	–	–	–	0.212	0.182	–	–	–	0.165	0.145
		$\boldsymbol{\beta}$	0.721	0.281	0.275	0.212	0.185	0.721	0.267	0.264	0.173	0.153	0.721	0.260	0.258	0.152	0.140
		$\mathbf{B}$	0.791	0.176	0.178	0.145	0.140	0.791	0.156	0.156	0.104	0.101	0.791	0.147	0.147	0.081	0.079
		$\mathbf{A}$	2.733	0.228	0.230	0.204	0.197	2.733	0.214	0.214	0.176	0.172	2.733	0.207	0.207	0.163	0.161
		Error	1.193	0.081	0.082	0.065	0.060	1.194	0.071	0.071	0.049	0.046	1.193	0.067	0.067	0.042	0.040
600	600	$\mathbf{\Lambda}$	–	–	–	0.250	0.211	–	–	–	0.180	0.158	–	–	–	0.138	0.125
		$\mathbf{\Gamma}$	–	–	–	0.248	0.209	–	–	–	0.180	0.158	–	–	–	0.138	0.125
		$\boldsymbol{\beta}$	0.396	0.159	0.158	0.147	0.144	0.396	0.151	0.150	0.137	0.136	0.396	0.148	0.147	0.133	0.132
		$\mathbf{B}$	0.611	0.160	0.161	0.119	0.115	0.611	0.145	0.145	0.084	0.082	0.611	0.138	0.138	0.065	0.064
		$\mathbf{A}$	2.735	0.194	0.195	0.179	0.177	2.735	0.182	0.183	0.161	0.160	2.735	0.177	0.178	0.153	0.152
		Error	1.108	0.059	0.060	0.050	0.049	1.108	0.052	0.052	0.041	0.040	1.108	0.049	0.049	0.037	0.036