arXiv:2302.02476v1 [stat.ME] 5 Feb 2023

高维时间序列的时变网络估计

Jia Chen*、DeguiLi†、Yuning Li‡、OliverLinton§

约克大学和剑桥大学

版本：2023年2月7日

摘要

我们在大型 VAR 模型框架下研究高维局部平稳时间序列的时变网络，其中转移矩阵和（误差）精度矩阵均随时间平滑变化。本文考察两类时变图：一类包含表示格兰杰因果关系连接的有向边，另一类包含表示偏相关连接的无向边。在稀疏结构假设下，我们提出一种惩罚局部线性方法，采用时变加权组 LASSO 联合估计转移矩阵并识别其显著元素；同时提出时变 CLIME 方法来估计精度矩阵。随后利用估计得到的转移矩阵和精度矩阵确定时变网络结构。在一些温和条件下，我们推导了所提出估计量的理论性质，包括一致性和 Oracle 性质。此外，我们将方法和理论扩展至高度相关的大规模时间序列；在这类序列中，稀疏性假设不再成立，因此我们在估计因子调整后的时变网络之前允许存在共同因子。我们通过大量模拟研究以及对美国大型宏观经济数据集的实证应用，展示了所提方法的有限样本表现。

关键词：CLIME、因子模型、格兰杰因果关系、LASSO、局部线性平滑、偏相关、时变网络、VAR。

*经济学与相关研究系，约克大学，英国。Jia Chen 的研究部分得到了经济与社会研究委员会（资助编号：ES/T01573X/1）的支持。

†数学系，约克大学，英国。

‡经济学与相关研究系，约克大学，英国。§经济学院，剑桥大学，剑桥，英国。

1 引言

近年来，网络分析已成为探索大量变量之间相互联系的有效工具，并应用于流行病学、经济学、金融学和社会网络等多个领域（例如，Newman，2002；Burt、Kilduff 和 Tasselli，2013；Diebold 和 Ylmaz，2014、2015；Hautsch、Schaumburg 和 Schienle，2014；Scott，2017；Barigozziand Brownlees，2019；Zhu 等，2019）。所谓图模型通常用于网络分析，以直观呈现大型面板中的连通性：顶点代表面板中的变量，边的存在则表示变量之间存在适当的（条件）依赖关系。过去几十年来，关于网络数据统计估计与推断的大多数现有文献都将关注点局限于静态网络，并假设网络不随时间变化（例如，Yuan 和 Lin，2007；Fan、Feng 和 Wu，2009；Loh 和 Wainwright，2013；Basu、Shojaie 和 Michailidis，2015；Zhao 等，2022）。然而，这一假设可能过于严格，在底层数据生成机制具有动态性的实际应用中往往不成立。近期文献中已有一些尝试放宽静态网络假设，允许连通结构呈现时变特征。例如，Kolar 等（2010）和 Zhou、Lafferty 和 Wasserman（2010）研究了具有平滑时变结构变化的动态网络模型；而 Wang、Yu 和 Rinaldo（2021）则研究了动态网络中的变点检测与估计。然而，上述大多数文献通常假设网络数据彼此独立，这一假设在实践中往往并不成立。我们旨在放宽这一严格假设，并在一般的时间依赖结构下对大规模网络数据进行建模。

向量自回归（VAR）是多元时间序列数据的一种基础建模工具（例如，L ̈utkepohl，2006）。近年来，人们越来越关注如何将有限维 VAR 扩展到高维情形。在对转移矩阵（或自回归系数矩阵）施加适当的稀疏性约束后，研究者提出了各种正则化方法，用于估计高维 VAR 模型并识别转移矩阵中的非零元素（例如，Basu 和 Michailidis，2015；Han、Lu 和 Liu，2015；Kock 和 Callot，2015；Davis、Zang 和 Zheng，2016）。Zhu 等。（2017）通过纳入邻接矩阵来捕捉网络效应，并采用普通最小二乘法估计模型，从而提出了网络 VAR 模型。近期，Chen、Fan 和 Zhu （2020）以及Miao、Phillips 和 Su（2022）进一步研究了含有潜在公共因子的高维 VAR 和网络 VAR，允许大型面板时间序列存在较强的截面依赖。这些论文所发展的方法和理论在很大程度上依赖于平稳性假设，即转移矩阵和波动率矩阵均不随时间变化。

稳定 VAR 模型无法捕捉底层的平滑结构变化和结构突变

数据生成过程中的这两种典型动态特征，常见于跨度较长的时间序列数据。为解决这一问题，Ding、Qiu 和 Chen (2017) 针对高维时间序列（允许变量数量以低于样本量次指数的速度增长）考虑了时变 VAR 模型，并结合核平滑与 ℓ_1-正则化来估计时变转移矩阵；而Safikhani 和 Shojaie (2022) 则通过采用全变差惩罚的三阶段程序，同时检测高维 VAR 中的结构突变并估计转移矩阵。Xu、Chen 和 Wu (2020) 检测高维时间序列（VAR 是其特例）中的结构突变，并估计协方差矩阵和精度矩阵的平滑变化（结构突变之间）。本文旨在局部平稳框架下，联合估计高维稀疏 VAR 中的时变转移矩阵和精度矩阵。受Barigozzi 和 Brownlees(2019) 对稳定网络时间序列的分析启发，我们使用估计得到的转移矩阵和精度矩阵，进一步构建两个时变网络：一个包含表示格兰杰因果关系的有向边，另一个包含表示偏相关连接的无向边。

通过 VAR 提出的时变网络与局部平稳模型自然相连，而后者在低维时间序列领域的文献中已得到系统研究。Dahlhaus (1997) 首先通过时变谱表示引入了局部平稳时间序列模型。Dahlhaus and Subba Rao(2006) 研究了时变 ARCH 模型，并提出了核加权准极大似然估计方法。Hafnerand Linton (2010) 进一步研究了 GARCH 模型的时变版本，并引入了一种半参数方法，以估计其中涉及的参数分量和非参数分量。Vogt (2012) 和 Zhangand Wu(2012) 研究了一般局部平稳时间序列中的非参数核估计与推断。Koo and Linton (2012) 将局部平稳模型框架扩展到了扩散过程。Yan, GaoandPeng (2020) 为时变向量移动平均模型发展了核估计方法及其理论。本文进一步探讨高维动态网络结构，从而对局部平稳时间序列文献形成补充。

我们研究大规模时间序列的时变 VAR 与网络模型，允许变量数量远大于时间序列长度。在转移矩阵和精度矩阵满足稀疏性假设且具有平滑结构变化的条件下，我们提出一种三阶段估计程序：（i）使用时变 LASSO 对转移矩阵及其导数进行初步局部线性估计；（ii）使用加权组 LASSO 对时变转移矩阵进行联合局部线性估计和特征选择；（iii）使用时变 CLIME 估计精度矩阵。为保证预言机性质，第二阶段估计中 LASSO 的权重通过对 SCAD 惩罚进行局部线性近似来构造（例如，Zou 和 Li，2008），并使用第一阶段获得的一致初步估计。我们的

时变转移矩阵的惩罚估计方法与为高维函数系数模型开发的各种非参数筛选和收缩方法相关（例如，Wang 和 Xia，2009；Lian，2012；Fan, Ma and Dai，2014；Liu, Li and Wu，2014；Li, Ke and Zhang，2015），而时变 CLIME 则是传统 CLIME 在静态精度矩阵估计中的自然扩展（例如，Cai、Liu 和 Luo，2011）。上述文献中所开发方法的理论性质（如预言机性质和极小极大最优收敛速率）依赖于独立数据假设。将方法和理论扩展至高维局部平稳时间序列并非易事，需要新的技术工具，例如针对时变 VAR 的集中不等式。在一些正则条件下，我们证明，所提出的加权组 LASSO 局部线性估计等于利用时变转移矩阵显著元素先验信息得到的不可行预言估计；而使用时变 CLIME 得到的精度矩阵估计，在各种矩阵范数下均一致，并具有合理的收敛速率。估计得到的转移矩阵用于一致地估计具有有向格兰杰因果关系的整体网络结构，而估计得到的精度矩阵则用于构建具有无向偏相关连接的网络结构。

我们进一步考虑高度相关的大规模时间序列。在这种情况下，稀疏模型假设不再成立，因此需要对方法和理论进行实质性修改。近似因子模型（例如，Chamberlain and Rothschild，1983年）或其时变版本（例如，Su and Wang，2017）被用于处理大量时间序列之间的强横截面依赖。具体而言，我们假设近似因子模型中的高维特质误差过程满足时变 VAR 结构，并对其转移矩阵和精度矩阵施加稀疏性约束。潜在的共同成分和特质成分需要得到一致估计。利用近似得到的特质误差向量，采用加权组 LASSO 惩罚局部线性估计方法和时变 CLIME 来估计时变转移矩阵和精度矩阵。随后，得到包含有向格兰杰因果关系和无向偏相关连接的因子调整的时变网络估计。因此，我们的论文大幅拓展了近期关于因子调整的稳定 VAR 模型估计的研究（例如，Fan,Masini and Medeiros，2021；Barigozzi, Cho andOwens，2022；Krampe and Margaritella，2022）。

我们的模拟研究表明，当时间序列变量的数量与样本量相当时，所提出的方法能够准确估计时变格兰杰因果关系网络和偏相关网络。特别是，在时变转移矩阵估计方面，采用加权组 LASSO 的惩罚局部线性方法优于传统局部线性方法（后者在高维时间序列情境下往往会失效），且其数值结果与预言机估计相近。对于时变误差精度

矩阵估计，所提出的时变 CLIME 的数值表现与时变图形 LASSO 相当。我们进一步将所开发的方法应用于 FRED-MD宏观经济数据集，并通过所提出的时变 VAR模型估计格兰杰因果关系网络和偏相关网络。

本文其余部分安排如下。第 2 节介绍时变 VAR 和网络模型结构。第 3 节介绍时变转移矩阵和精度矩阵的估计方法，第4节给出所提出估计量的渐近性质。第 5 节讨论因子调整时变向量自回归模型和网络估计。第 6 节和第 7 节分别报告模拟研究和一项实证应用。第 8 节总结全文。补充文档包含主要定理的证明、若干带证明的技术引理、对一项关键假设的验证，以及关于调优参数选择的讨论。全文中，我们令 $|\cdot|_0$、$|\cdot|_1$、$\|\cdot\|$ 和 $|\cdot|_{\max}$ 分别表示向量的 L_0、L_1、L_2（欧几里得）范数和最大范数。令 I_d 和 $O_{d\times d}$ 分别表示 $d\times d$ 单位矩阵和零矩阵。对于 $d\times d$ 矩阵 $W=(w_{ij})_{d\times d}$，令 $\|W\|=\lambda_{\max}^{1/2}(W'W)$ 表示算子范数，$\|W\|_F=\left[\mathrm{Tr}\left(W'W\right)\right]^{1/2}$ 表示弗罗贝尼乌斯范数，$\|W\|_1=\max_{1\leqslant j\leqslant d}\sum_{i=1}^{d}|w_{ij}|$，$\|W\|_{\max}=\max_{1\leqslant i\leqslant d}\max_{1\leqslant j\leqslant d}|w_{ij}|$，以及 $|W|_1=\sum_{i=1}^{d}\sum_{j=1}^{d}|w_{ij}|$，其中 $\lambda_{\max}(\cdot)$ 是矩阵的最大特征值，$\mathrm{Tr}(\cdot)$ 表示迹。将方阵的行列式记为 $\det(\cdot)$。令 $a_n\sim b_n$、$a_n\propto b_n$ 和 $a_n\gg b_n$ 分别表示 $a_n/b_n\to 1$、$0<c\leqslant a_n/b_n\leqslant c<\infty$ 和$b_n/a_n\to 0$。

2 时变 VAR 与网络模型

本节首先介绍具有时变转移矩阵和精度矩阵的局部平稳 VAR 模型，然后分别定义两类时变网络结构，其连接关系分别基于格兰杰因果关系和偏相关。第5 节将进一步把它们推广到因子调整时变 VAR 与网络设定。

2.1 时变 VAR 模型

设 $(X_t:t=1,\ldots,n)$ 满足 $X_t=(x_{t,1},\ldots,x_{t,d})'$ 是由 d阶时变 VAR 模型生成的 d维随机向量序列：

$$X_t=\sum_{k=1}^{p}A_{t,k}X_{t-k}+e_t\ \text{ with }\ e_t=\Sigma_t^{1/2}\varepsilon_t,\ \ t=1,\ldots,n, \tag{2.1}$$

其中，$A_{t,k}=A_k(t/n), k=1$、。 。 。、 p是 d×d 时变转移矩阵，其每个元素都是缩放时间的光滑确定性函数； $\mathbf{\Sigma}_t=\mathbf{\Sigma}(t/n)$ 是 d × d 时变波动率矩阵； (ε_t) 是由均值为零且协方差矩阵为单位协方差矩阵的独立同分布（i.i.d.） d维随机向量组成的序列。定义 $\mathbf{\Omega}_t=\mathbf{\Omega}(t/n)$ 为 $\mathbf{\Sigma}_t$（时变精度矩阵）的逆矩阵。我们考虑超大规模时间序列情形，即允许维度 d 以样本量 n的指数速率增长。时变 VAR 模型（2.1）是有限维时变 VAR 模型向高维时间序列的自然扩展。如果将 $\mathbf{\Sigma}_t$ 替换为时不变协方差矩阵，则（2.1）就变为 Ding、Qiu 和 Chen（2017）所研究的同一模型。此外，当 $A_{t,k}$、$k=1$、。 。 。、 p以及 $\mathbf{\Sigma}_t$ 均为时不变常数矩阵时，（2.1）就变为高维稳定 VAR:

$$X_t=\sum_{k=1}^{p}\mathbf{A}_kX_{t-k}+\mathbf{\Sigma}^{1/2}\varepsilon_t, \tag{2.2}$$

这在近期文献中已得到广泛研究（例如，Basu 和 Michailidis，2015；Han、Lu 和 Liu，2015；Kock 和 Callot，2015；Barigozzi 和 Brownlees，2019；Liu 和 Zhang，2021）。本文始终假设以下条件成立。

假设 1。 *(i)* 对 $\tau\in[0,1]$一致地，有行列式 $\left(\mathbf{I}_d-\sum_{k=1}^{p}\mathbf{A}_k(\tau)z^k\right)\neq 0$ 对于模不大于 *1* 的任意 $z\in\mathbb{C}$均成立，其中 $\mathbb{C}$ 表示复数集。$\mathbf{A}_k(\cdot)$ 中的每个元素在 $[0,1]$上均为二阶连续可微。

(ii) 精度矩阵 $\mathbf{\Omega}(\tau)$ 在 $\tau\in[0,1]$上一致正定，且$\mathbf{\Sigma}(\tau)$ 的算子范数在 $\tau\in[0,1]$上一致有界。此外， $\mathbf{\Sigma}(\tau)$ 和 $\mathbf{\Omega}(\tau)$ 中的每个元素在 $[0,1]$上均为二阶连续可微。

(iii) 对任意满足 $\|u\|=1$ 的 d 维向量 u，$\mathsf{E}\left[\exp\left\{\iota_1(u\ \varepsilon_t)^2\right\}\right]\leqslant C_0<\infty$，其中 ι_1 和 C_0 为正常数。

假设 1(i) 中的第一个条件，是对常数转移矩阵所施加的稳定性假设的自然扩展（例如，L¨utkepohl，2006），表明时变 VAR 过程是局部平稳/稳定的，并由此得到以下 Wold 表示

$$X_t=\sum_{k=0}^{\infty}\mathbf{\Phi}_{t,k}e_{t-k}, \tag{2.3}$$

其中系数矩阵 $\mathbf{\Phi}_{t,\ k}$ 绝对可和（在适当的矩阵范数下）。例如，当 $p=1$ 时，我们有 $\mathbf{\Phi}_{t,\ 0}=\mathbf{I}_d$ 且 $\mathbf{\Phi}_{t,\ k}=\Pi_{j=1}^{k}\mathbf{A}_{t-j+1,\ 1}$ 对 $k\geqslant 1$ 成立。假设对于充分大的 k,

$$\max_{0\leqslant t\leqslant n}\|\mathbf{\Phi}_{t,k}\|\leqslant C_1\rho^k, \tag{2.4}$$

其中 C_1 是一个正 常数，且 $0<\rho<1$。类似的假设可见于Ding、Qiu

和 Chen (2017）。在某些特殊的模型设定中，(2.4) 可能不成立，感兴趣的读者可参阅Basu 和 Michailidis (2015) 以及 Liu 和 Zhang (2021) 中的讨论。事实上，通过施加一些高层次条件（例如，对 $x_{t,i}$ 施加次高斯条件，见引理 B.1 中的证明），可以去掉条件 (2.4)。假设 1(i)(ii) 中的光滑性条件是基于核的局部估计方法及其理论中的常见条件。假设1(iii) 中的次高斯矩条件在高维特征选择以及协方差/精度矩阵估计文献中并不罕见（例如，WAINWRIGHT，2019），且弱于高维 VAR 文献中经常采用的高斯假设（例如，Basu 和 Michailidis，2015；Kock 和 Callot，2015）。

2.2 时变网络结构

记 $A_{t,k} = \left(a_{k,ij|t}\right)_{d\times d}$, $\mathbf{\Omega}_t = \left(\omega_{ij|t}\right)_{d\times d}$、$A_k(\tau) = \left(a_{k,ij}(\tau)\right)_{d\times d}$ 和 $\mathbf{\Omega}(\tau) = \left(\omega_{ij}(\tau)\right)_{d\times d}$，其中 $1 \leqslant t \leqslant n$ 和 $0 \leqslant \tau \leqslant 1$。我们通过时变图 $\mathbb{G}_t = (\mathbb{V}, \mathbb{E}_t)$定义网络结构，其中 $\mathbb{V} = \{1, 2, 。..... d\}$表示顶点集，$\mathbb{E}_t = \left\{(i,j) \in \mathbb{V}\times\mathbb{V} : c_{ij|t} \neq 0,\ i \neq j\right\}$ 表示时变边集。$c_{ij|t}$ 的选择由连接关系的定义决定。$\mathbb{G}_t$ 的构造与 Kolar 等（2010）以及 Zhou、Lafferty 和 Wasserman（2010）针对独立网络数据所采用的方法类似。遵循 Barigozzi 和 Brownlees（2019）以及 Barigozzi、Cho 和 Owens（2022）的稳定网络分析，接下来我们考虑两类时变连接关系：有向格兰杰因果连接和无向偏相关连接。

格兰杰因果关系的定义最早由 Granger（1969）提出，用于研究小型经济时间序列系统中的因果关系。在稳定 VAR（阶数为p）的背景下，如果存在 $k \in \{1, 2, ...、\ p\}$，使得$x_{t-k、\ j}$通过减少预测误差提高了对$x_{t、\ i}$的可预测性，我们就称$x_{t、\ j}$格兰杰导致$x_{t、\ i}$。使用（2.2）中的稳定转移矩阵 $A_k = \left(a_{k,ij}\right)_{d\times d}$ 来确定格兰杰因果结构，是一种自然的想法；也就是说，如果至少存在一个 k 使得$a_{k、\ ij} \neq 0$，那么$x_{t、\ j}$就格兰杰导致$x_{t、\ i}$。我们可以借助（2.1）将稳定的格兰杰因果结构扩展为更一般的时变版本。在给定时间点 t，如果至少存在一个 k 使得$a_{k、\ ij|t} \neq 0$，我们就称$x_{t、\ j}$的滞后项格兰杰导致$x_{t、\ i}$。因此，对于给定的 $\tau \in (0,1)$，我们将时变局部图 $\mathbb{G}_\tau^G = \left(\mathbb{V}、\ \mathbb{E}_\tau^G\right)$ 定义为

$$\mathbb{E}_\tau^G = \{(i,j) \in \mathbb{V}\times\mathbb{V} : \exists\, k \in \{1, 2, \ldots, p\},\ a_{k,ij}(\tau) \neq 0\}. \tag{2.5}$$

偏相关是网络时间序列中常用的条件依赖度量。接下来，我们使用 $\mathbf{\Omega}_t = \mathbf{\Omega}(t/n)$ 在（2.1）中将其扩展到时变设定。令$\rho_{ij|t} = \mathrm{cor}(e_{t,i}, e_{t,j} | e_{t,k}, k \neq i,j)$ 表示创新项$e_{t与i}$ 以及$e_{t之间的时变（同期）偏相关，\ 其中j,\ e_{t是i}}$ 的第 i个元素。根据Dempster （1972），可以证明 $\rho_{ij|t} \neq 0$ 等价于 $\omega_{ij|t} \neq 0$，当 $i \neq j$时。因此，我们可以通过以下方式构造

边：收集时变精度矩阵中非零元素的索引对。对于 $\tau \in (0,1)$，定义局部图 $\mathbb{G}_\tau^P = \left(\mathbb{V}, \mathbb{E}_\tau^P\right)$，其中

$$\mathbb{E}_\tau^P = \{(i,j) \in \mathbb{V} \times \mathbb{V} : \ \omega_{ij}(\tau) \neq 0, \ i \neq j\}. \tag{2.6}$$

实际上，人们通常关注的是整个时间区间内的完整网络结构。这需要构造 $\mathbb{G}_\tau^G$和 $\mathbb{G}_\tau^P$ 的统一版本。将统一图记为 $\mathbb{G}^G = \left(\mathbb{V}、\ \mathbb{E}^G\right)$ 和 $\mathbb{G}^P = \left(\mathbb{V}、\ \mathbb{E}^P\right)$，其中

$$\mathbb{E}^G = \{(i,j) \in \mathbb{V} \times \mathbb{V} : \ \exists\, k \in \{1,2,\ldots,p\} \text{ and } \tau \in (0,1), \ a_{k,ij}(\tau) \neq 0\} \tag{2.7}$$

and

$$\mathbb{E}^P = \{(i,j) \in \mathbb{V} \times \mathbb{V} : \ \exists\, \tau \in (0,1), \ \omega_{ij}(\tau) \neq 0, \ i \neq j\}. \tag{2.8}$$

容易验证， $\mathbb{E}_\tau^G \subset \mathbb{E}^G$ 和 $\mathbb{E}_\tau^P \subset \mathbb{E}^P$ 对任意 $\tau \in (0,1)$均成立。下文第 3.4 节定义上述统一网络的离散版本，并给出其估计量。

3 方法

令 $A_{k,\cdot i}^\intercal(\cdot)$ 和 $C_i^\intercal(\cdot)$ 分别为 i 的第$\mathbf{A}_k(\cdot)$ 和 $\mathbf{\Omega}^{-1/2}(\cdot)$ 行，

$$\boldsymbol{\alpha}_{i\bullet}(\cdot) = \left[A_{1,i}^\intercal(\cdot), \ldots, A_{p,i}^\intercal(\cdot)\right]^\intercal, \ \mathbf{X}_t = \left(X_t^\intercal, \ldots, X_{t-p+1}^\intercal\right)^\intercal, \tag{3.1}$$

and $\tau_t = t/n$。时变 VAR 模型（2.1）可等价地写为

$$x_{t,i} = \boldsymbol{\alpha}_{i\bullet}^\intercal(\tau_t)\mathbf{X}_{t-1} + e_{t,i} \ \text{ with } \ e_{t,i} = C_i^\intercal(\tau_t)\varepsilon_t, \ \ i = 1, \ldots, d, \tag{3.2}$$

这是一个高维时变系数自回归模型，具有标量响应和 pd 个候选预测变量，分别对应每个 i。由于预测变量的维度可能极大，我们需要对时变参数向量 $\boldsymbol{\alpha}_{i\bullet}(\cdot)$ 施加适当的稀疏性约束，以限制其显著元素的数量。文献已对高维变系数模型进行了系统研究，并提出了各种非参数筛选和收缩方法，以便在独立数据假设下选择显著协变量、估计系数函数并识别模型结构（例如，Wang、Li 和 Huang，2008；Wang and Xia，2009；Lian，2012；Cheng 等。，2014；Fan, Ma and Dai，2014；Liu, Li andWu，2014；Li, Ke and Zhang，2015）。本节在高维局部平稳时间序列框架下，提出一个三阶段程序来估计格兰杰因果关系和偏相关网络结构：（i）首先获得初步局部线性

估计值 $\boldsymbol{\alpha}_{i\bullet}(\cdot)$（及其导数），使用时变 LASSO 对$\mathbf{X}_{t-1}$中的预测变量进行第一阶段筛选；（ii）使用加权组 LASSO 进行局部线性估计和特征选择，其中权重通过对阶段（i）中 $\boldsymbol{\alpha}_{i\bullet}(\cdot)$ 的初步估计应用 SCAD 惩罚的局部线性近似构造；（iii）通过时变 CLIME 方法估计误差精度矩阵 $\boldsymbol{\Omega}(\cdot)$。最后，利用估计得到的转移矩阵和精度矩阵构造统一网络结构。

3.1 初步时变 LASSO 估计

对于 $\tau \in (0,1)$，根据假设1(i)中关于转移矩阵的光滑性条件，我们有如下对 $\boldsymbol{\alpha}_{i\bullet}(\tau_t)$的局部线性近似：

$$\boldsymbol{\alpha}_{i\bullet}(\tau_t) \approx \boldsymbol{\alpha}_{i\bullet}(\tau) + \boldsymbol{\alpha}'_{i\bullet}(\tau)(\tau_t - \tau), \quad i = 1, \ldots, d,$$

当 τ_t 处于 τ的一个小邻域内时，其中 $\boldsymbol{\alpha}'_{i\bullet}(\cdot)$ 是一个 (pd)维向量，其元素为 $\boldsymbol{\alpha}_{i\bullet}(\cdot)$中各元素的一阶导数。因此，对于每个 $i \in \{1, 2, \ldots\ldots, d\}$ 以及给定的$\tau \in (0,1)$，我们定义如下局部线性目标函数（例如，Fan 和 Gijbels，1996）：

$$\mathcal{L}_i(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau) = \frac{1}{n}\sum_{t=1}^{n}\left\{x_{t,i} - [\boldsymbol{\alpha} + \boldsymbol{\beta}(\tau_t - \tau)]^{\intercal}\mathbf{X}_{t-1}\right\}^2 K_h(\tau_t - \tau), \tag{3.3}$$

其中 $K_h(\cdot) = \frac{1}{h}K(\cdot/h)$， $K(\cdot)$ 是核函数，而 h 是带宽或平滑参数。通过最小化 $\mathcal{L}_i(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau)$，并分别对 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$进行优化，可得到 $\boldsymbol{\alpha}_{i\bullet}(\tau)$ 和 $\boldsymbol{\alpha}'_{i\bullet}(\tau)$ 的估计。然而，只有当预测变量的维度固定或显著小于样本量 n时，这种局部线性估计才可行（例如，Cai，2007；Li、Chen 和 Gao，2011）。在我们的高维情形中，由于预测变量的数量可能超过 n，直接最小化 $\mathcal{L}_i(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau)$以获得令人满意的估计颇具挑战。为解决这一问题，我们假设 $\boldsymbol{\alpha}_{i\bullet}(\tau)$ 中的显著分量数量远小于 n，并在局部线性目标函数中加入 LASSO 惩罚项（3.3）。

LASSO 估计最早由 Tibshirani （1996）在线性回归背景下提出，现已成为高维变量和特征选择中最常用的工具之一。接下来，我们采用 LASSO 估计的时变版本。定义如下：

$$\mathcal{L}_i^*(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau) = \mathcal{L}_i(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau) + \lambda_1\left(|\boldsymbol{\alpha}|_1 + h|\boldsymbol{\beta}|_1\right), \tag{3.4}$$

其中， λ_1 为调节参数。令 $\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau)$ 和 $\widetilde{\boldsymbol{\alpha}}'_{i\bullet}(\tau)$ 分别为关于 $\boldsymbol{\alpha}$和 $\boldsymbol{\beta}$最小化$\mathcal{L}_i^*(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau)$ 所得的解。我们称其为初步时变 LASSO 估计。这种 LASSO 估计可能无法准确识别真正的显著预测变量，但可以剔除

大量无关预测变量，因此可作为初步筛选步骤。此外，第一阶段的估计值将用于在第二阶段的加权组 LASSO 中构造权重，从而更精确地估计时变参数并准确选择显著预测变量。

3.2 使用加权组 LASSO 的惩罚局部线性估计

为了估计统一的格兰杰因果网络，接下来我们引入一种全局惩罚方法，以同时估计 $\tau_t, t = 1, \ldots, n$ 处的时变参数，并识别非零索引集 $\mathcal{J}_i = \bigcup_{t=1}^{n} \mathcal{J}_i(\tau_t)$ 和 $\mathcal{J}_i' = \bigcup_{t=1}^{n} \mathcal{J}_i'(\tau_t)$，其中

$$\mathcal{J}_i(\tau) = \{1 \leqslant j \leqslant pd : \ \alpha_{i,j}(\tau) \neq 0\} \ \text{ and } \ \mathcal{J}_i'(\tau) = \left\{1 \leqslant j \leqslant pd : \ \alpha_{i,j}'(\tau) \neq 0\right\}$$

其中 $\alpha_{i,j}(\cdot)$ 和 $\alpha_{i,j}'(\cdot)$ 分别是 $\boldsymbol{\alpha}_{i\bullet}(\cdot)$ 和 $\boldsymbol{\alpha}_{i\bullet}'(\cdot)$ 的第 j 个元素。对于每个 i，请注意，在 $\boldsymbol{\alpha}_{i\bullet}'(\tau_t)$ 中识别零元素（在 t 上一致）等价于识别索引 $j, 1 \leqslant j \leqslant pd$，使得 $D_{i,j} = 0$，其中

$$D_{i,j}^2 = \sum_{t=1}^{n} \left[\alpha_{i,j}(\tau_t) - \frac{1}{n} \sum_{s=1}^{n} \alpha_{i,j}(\tau_s) \right]^2.$$

在实践中，$D_{i,j}^2$ 可以通过以下方式估计：

$$\widetilde{D}_{i,j}^2 = \sum_{t=1}^{n} \left[\widetilde{\alpha}_{i,j}(\tau_t) - \frac{1}{n} \sum_{s=1}^{n} \widetilde{\alpha}_{i,j}(\tau_s) \right]^2,$$

使用初步时变 LASSO 估计 $\widetilde{\alpha}_{i,j}(\tau_t), t = 1, \ldots, n$。令 $\mathbf{A} = (\boldsymbol{\alpha}_{\bullet 1}, \ldots, \boldsymbol{\alpha}_{\bullet n})$，其中 $\boldsymbol{\alpha}_{\bullet t} = (\alpha_{1|t}, \ldots, \alpha_{pd|t})^\intercal$；并令 $\mathbf{B} = (\boldsymbol{\beta}_{\bullet 1}, \ldots, \boldsymbol{\beta}_{\bullet n})$，其中 $\boldsymbol{\beta}_{\bullet t} = (\beta_{1|t}^\intercal, \ldots, \beta_{pd|t})$。我们定义带加权组 LASSO 的惩罚目标函数的全局版本：

$$Q_i(\mathbf{A}, \mathbf{B}) = \sum_{t=1}^{n} \mathcal{L}_i(\boldsymbol{\alpha}_{\bullet t}, \boldsymbol{\beta}_{\bullet t} \mid \tau_t) + \sum_{j=1}^{pd} p_{\lambda_2}'(\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|) \|\boldsymbol{\alpha}_j\| + \sum_{j=1}^{pd} p_{\lambda_2}'\left(\widetilde{D}_{i,j}\right) \|h\boldsymbol{\beta}_j\|, \tag{3.5}$$

其中

$$\widetilde{\boldsymbol{\alpha}}_{i,j} = [\widetilde{\alpha}_{i,j}(\tau_1), \ldots, \widetilde{\alpha}_{i,j}(\tau_n)]^\intercal, \ \boldsymbol{\alpha}_j = \left(\alpha_{j|1}, \ldots, \alpha_{j|n}\right)^\intercal, \ \boldsymbol{\beta}_j = \left(\beta_{j|1}, \ldots, \beta_{j|n}\right)^\intercal,$$

其中，λ_2 是一个调节参数，$p_\lambda'(\cdot)$ 是 SCAD 惩罚函数的导数：

$$p_\lambda'(z) = \lambda \left[I(z \leqslant \lambda) + \frac{(a_0\lambda - z)_+}{(a_0 - 1)\lambda} I(z > \lambda) \right],$$

取 $a_0 = 3.7$，正如 Fan 和 Li (2001) 所建议的那样，$I(\cdot)$ 为指示函数。(3.5) 中的惩罚项源于 SCAD 惩罚函数的局部线性近似（Zou 和 Li，2008）。式（3.5）中的项 $p'_{\lambda_2}(\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|)$ 和 $p'_{\lambda_2}\left(\widetilde{D}_{i,j}\right)$ 用作组 LASSO 的权重，其取值由第3.1节的初步估计确定，即，当 $\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|$ 或 $D_{i,j}$ 接近零时，~对应的权重较大；而当 $\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|$ 或 $D_{i,j}$ 较大时，权重较小或等于零。在第二个惩罚项中使用 $D_{i,j}$ 而非 L_2-范数的 $\widetilde{\boldsymbol{\alpha}}'_j = [\widetilde{\alpha}'_{i,j}(\tau_1), \ldots, \widetilde{\alpha}'_{i,j}(\tau_n)]$ 的一个优点是，涉及 $D_{i,j}$ 的时变参数估计~通常比其导数对应项更稳定。

^ ^令 $\mathbf{A}_i$ 和 $\mathbf{B}_i$ 为关于 $\mathbf{A}$ 和 $\mathbf{B}$ 的 $\mathbf{A},\mathbf{B}$) 的最小化解，其中

$$\widehat{\mathbf{A}}_i = (\widehat{\boldsymbol{\alpha}}_{i,1}, \ldots, \widehat{\boldsymbol{\alpha}}_{i,pd}) \quad \text{with} \quad \widehat{\boldsymbol{\alpha}}_{i,j} = [\widehat{\alpha}_{i,j}(\tau_1), \ldots, \widehat{\alpha}_{i,j}(\tau_n)]^\intercal,$$
$$\widehat{\mathbf{B}}_i = (\widehat{\boldsymbol{\alpha}}'_{i,1}, \ldots, \widehat{\boldsymbol{\alpha}}'_{i,pd}) \quad \text{with} \quad \widehat{\boldsymbol{\alpha}}'_{i,j} = [\widehat{\alpha}'_{i,j}(\tau_1), \ldots, \widehat{\alpha}'_{i,j}(\tau_n)]^\intercal.$$

指标集 $\mathcal{J}_i$ 通过 $\widehat{\mathcal{J}}_i = \{j : \widehat{\boldsymbol{\alpha}}_{i,j} \neq \mathbf{0}_n\}$估计，而 $\mathcal{J}'_i$ 通过 $\widehat{\mathcal{J}}'_i = \{j : \widehat{\boldsymbol{\alpha}}'_{i,j} \neq \mathbf{0}_n\}$估计，其中 $\mathbf{0}_k$ 是一个 k维零向量。Li, Ke and ZhangLi,Ke and Zhang （2015）以及Chen等人。（2021）采用了类似的收缩估计方法，以识别独立数据的高维半变系数模型结构。据我们所知，目前尚无关于此类惩罚技术及其相关理论在高维局部平稳时间序列数据中的研究。

3.3 时变精度矩阵的估计

在本节中，我们研究 $\boldsymbol{\Omega}(\cdot)$ 在模型（2.1）中的估计，这对于揭示偏相关的时变且统一的网络结构至关重要。在稀疏性假设下，大型静态精度矩阵的估计已得到广泛研究，文献中也提出了各种估计技术，例如惩罚似然、图形 Danzig 选择器和 CLIME（例如，Lam 和 Fan，2009；Yuan，2010；Cai、Liu 和 Luo，2011）。Xu、Chen 和 Wu(2020) 进一步为可观测的高维局部平稳时间序列提出了时变 CLIME 方法。请注意，在本文中，$\boldsymbol{\Omega}(\cdot)$ 是高维不可观测误差向量 e_t 的时变精度矩阵，因此其估计需要对时变 CLIME 方法及理论作出重大修改。

根据 $\widehat{\boldsymbol{\alpha}}_{i\bullet}(\cdot)$，$i = 1、\ldots\ldots、d$，利用第3.2节的结果，我们可以提取时变^转移矩阵的估计值，记为 $\mathbf{A}_k(\tau_t)$，$t = 1、\ldots\ldots、n$，$k = 1、\ldots\ldots、p$，并用 e_t 近似为

$$\widehat{e}_t = (\widehat{e}_{t,1}, \ldots, \widehat{e}_{t,d})^\intercal = X_t - \sum_{k=1}^{p} \widehat{\mathbf{A}}_k(\tau_t) X_{t-k}, \quad t = 1, \ldots, n. \tag{3.6}$$

近似 ac　　准确性取决于时变加权

groupLASSO 估计量的一致预测率。为了应用时变 CLIME，我们假设 $\boldsymbol{\Omega}(\cdot)$ 满足一致稀疏性假设，这是将经典稀疏性假设自然推广到局部平稳时间序列情形的一种方式。具体而言，我们假设 $\{\boldsymbol{\Omega}(\tau) : 0 \leqslant \tau \leqslant 1\} \in \mathcal{S}(q, \xi_d)$，其中

$$\mathcal{S}(q,\xi_d) = \left\{ \mathbf{W}(\tau) = \left[w_{ij}(\tau)\right]_{d\times d}, 0 \leqslant \tau \leqslant 1 : \mathbf{W}(\tau) \succ 0,\ \sup_{0\leqslant\tau\leqslant 1} \|\mathbf{W}(\tau)\|_1 \leqslant C_2,\ \sup_{0\leqslant\tau\leqslant 1} \max_{1\leqslant i\leqslant d} \sum_{j=1}^{d} |w_{ij}(\tau)|^q \leqslant \xi_d \right\}, \tag{3.7}$$

其中 $0 \leqslant q < 1$，“$\mathbf{W} \succ 0$”表示 $\mathbf{W}$ 是正定的，且 C_2 是一个有界正常数。定义

$$\widehat{\boldsymbol{\Sigma}}(\tau) = [\widehat{\sigma}_{ij}(\tau)]_{d\times d} \quad \text{with } \widehat{\sigma}_{ij}(\tau) = \sum_{t=1}^{n} \omega_{n,t}(\tau)\widehat{e}_{t,i}\widehat{e}_{t,j} / \sum_{t=1}^{n} \omega_{n,t}(\tau), \tag{3.8}$$

whe其中，权重函数$\omega_{n,\ t}(\cdot)$ 通过局部线性平滑构造:

$$\omega_{n,t}(\tau) = K\left(\frac{\tau_t - \tau}{b}\right) s_{n,2}(\tau) - K_1\left(\frac{\tau_t - \tau}{b}\right) s_{n,1}(\tau),$$

其中$s_{n,\ j}(\tau) = \sum_{t=1}^{n} K_j\left(\frac{\tau_t-\tau}{b}\right)$，$K_j(x) = x^j K(x)$，且 b 是一个带宽。在一致稀疏性假设（3.7）下，我们通过时变 CLIME 方法估计 $\boldsymbol{\Omega}(\tau)$：

$$\widetilde{\boldsymbol{\Omega}}(\tau) = [\widetilde{\omega}_{ij}(\tau)]_{d\times d} = \arg\min_{\boldsymbol{\Omega}} |\boldsymbol{\Omega}|_1 \quad \text{subject to } \left\|\widehat{\boldsymbol{\Sigma}}(\tau)\boldsymbol{\Omega} - \mathbf{I}_d\right\|_{\max} \leqslant \lambda_3, \tag{3.9}$$

其中 λ_3 是一个调节参数。由于潜在的时变精度矩阵是对称的，因此需要对由（3.9）得到的矩阵估计进行对称化，以获得最终估计，记为$\widehat{\ }\boldsymbol{\Omega}(\tau) = [\widehat{\omega}_{ij}(\tau)]_{d\times d}$，其中

$$\widehat{\omega}_{ij}(\tau) = \widehat{\omega}_{ji}(\tau) = \widetilde{\omega}_{ij}(\tau) I\left(|\widetilde{\omega}_{ij}(\tau)| \leqslant |\widetilde{\omega}_{ji}(\tau)|\right) + \widetilde{\omega}_{ji}(\tau) I\left(|\widetilde{\omega}_{ij}(\tau)| > |\widetilde{\omega}_{ji}(\tau)|\right). \tag{3.10}$$

3.4 均匀时变网络的估计

在实践中，当样本量 n 足够大时，通常可以用以下离散版本近似一致边集，$\mathbb{E}^G$ 和 $\mathbb{E}^P$:

$$\mathbb{E}_n^G = \{(i,j) \in \mathbb{V}\times\mathbb{V} : \exists\, k \in \{1,2,\ldots,p\} \text{ and } t \in \{1,\ldots,n\},\ a_{k,ij}(\tau_t) \neq 0\} \tag{3.11}$$

and

$$\mathbb{E}_n^P = \{(i,j) \in \mathbb{V}\times\mathbb{V} : \exists\, t \in \{1,\ldots,n\},\ \omega_{ij}(\tau_t) \neq 0,\ i \neq j\}. \tag{3.12}$$

因此，接下来我们估计 $\mathbb{E}_n^G$ 和 $\mathbb{E}_n^P$ 而非 $\mathbb{E}^G$ 和 $\mathbb{E}^P$。借助第3.2 节和3.3节中的时变转移矩阵和精度矩阵估计，我们可以通过以下方式估计 $\mathbb{E}_n^G$：

$$\widehat{\mathbb{E}}_n^G = \left\{ (i,j) \in \mathbb{V} \times \mathbb{V} : \exists\, k \in \{1,2,\ldots,p\},\ \sum_{t=1}^{n} \widehat{a}_{k,ij}^2(\tau_t) > 0 \right\}, \tag{3.13}$$

其中，$\widehat{a}_{k,\ ij}(\tau_t)$ 是 $\widehat{\ }\mathbf{A}_k(\tau_t)$的第 (i,j)个元素，估计量 $\mathbb{E}_n^P$ 为

$$\widehat{\mathbb{E}}_n^P = \{(i,j) \in \mathbb{V} \times \mathbb{V} : \exists\, t \in \{1,\ldots,n\},\ |\widehat{\omega}_{ij}(\tau_t)| \geqslant \lambda_3,\ i \neq j\}, \tag{3.14}$$

其中 λ_3 是时变 CLIME 中使用的调节参数。

4 主要理论结果

为简化记号，本节始终聚焦于时变 VAR(1) 模型：

$$X_t = \mathbf{A}(\tau_t) X_{t-1} + \mathbf{\Sigma}_t^{1/2} \varepsilon_t, \tag{4.1}$$

其中，$\mathbf{A}(\tau) = [\alpha_{ij}(\tau)]_{d\times d}$。对于一般的时变 VAR(p) 模型（2.1），它可以等价地改写为一个 (pd) 维 VAR(1) 模型，如下所示：

$$\mathbf{X}_t = \mathbf{A}_t^* \mathbf{X}_{t-1} + \mathbf{e}_t,$$

其中，$\mathbf{X}_t$ 定义见（3.1），$\mathbf{e}_t = \left(e_t,\ {}_\tau 0_d,\ \ldots\ldots,\ {}_\tau \theta_d\right)$，且$\mathbf{A}_t^*$是一个 $(pd)\times(pd)$ 时变转移矩阵：

$$\mathbf{A}_t^* = \begin{pmatrix} \mathbf{A}_{t,1} & \mathbf{A}_{t,2} & \cdots & \mathbf{A}_{t,p-1} & \mathbf{A}_{t,p} \\ \mathbf{I}_d & \mathbf{O}_{d\times d} & \cdots & \mathbf{O}_{d\times d} & \mathbf{O}_{d\times d} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{O}_{d\times d} & \mathbf{O}_{d\times d} & \cdots & \mathbf{I}_d & \mathbf{O}_{d\times d} \end{pmatrix}.$$

4.1 时变 LASSO 估计的一致性

定义

$$\mathbf{\Psi}(\tau) = \begin{bmatrix} \mathbf{\Psi}_0(\tau) & \mathbf{\Psi}_1(\tau) \\ \mathbf{\Psi}_1(\tau) & \mathbf{\Psi}_2(\tau) \end{bmatrix} \text{ with } \mathbf{\Psi}_k(\tau) = \frac{1}{n}\sum_{t=1}^{n} \left(\frac{\tau_t - \tau}{h}\right)^k X_{t-1} X_{t-1}^\tau K_h(\tau_t - \tau),\ k = 0,1,2, \tag{4.2}$$

and

$$\mathcal{B}_i(\tau) = \left\{ \left(u_1^\intercal, u_2^\intercal\right)^\intercal : \|u_1\|^2 + \|u_2\|^2 = 1,\ \sum_{j=1}^{d} \left(|u_{1,j}| + |u_{2,j}|\right) \leqslant 3 \left(\sum_{j \in \mathcal{J}_i(\tau)} |u_{1,j}| + \sum_{j \in \mathcal{J}_i'(\tau)} |u_{2,j}| \right) \right\},$$

其中 $\mathcal{J}_i(\tau)$ 和 $\mathcal{J}_i'(\tau)$ 的定义与第 3.2 节 相同，但 $p = 1$有所不同。为推导第 3.1 节中定义的初步时变 LASSO 估计的一致一致性性质，我们需要以下假设，其中一些假设可以弱化，但代价是需要更冗长的证明。

假设2。 *(i)* 核 $K(\cdot)$ 是一个有界、连续且对称的概率密度函数，其具有紧支撑 $[-1, 1]$。

(ii) 带宽 h 满足

$$nh/\log^2(n \vee d) \to \infty \ \textit{and}\ \ sh^2 \log(n \vee d) \to 0,$$

其中 $s = \text{最大值}_{1 \leqslant i \leqslant d}\, s_i$， s_i 表示指标集 $\mathcal{J}_i$的基数。

假设3。 *(i)* 调节参数 λ_1 满足

$$\zeta_{n,d} := \log(n \vee d)\left[(nh)^{-1/2} + sh^2\right] = o(\lambda_1) \ \text{ and } \ \sqrt{s}\lambda_1/h \to 0.$$

(ii) 存在一个正常数 κ_0，使得以概率趋于*1*（*w.p.a.1*），

$$\min_{1 \leqslant i \leqslant d} \min_{1 \leqslant t \leqslant n} \inf_{u \in \mathcal{B}_i(\tau_t)} u^\intercal \mathbf{\Psi}(\tau_t) u \geqslant \kappa_0. \tag{4.3}$$

假设 2(i) 是一个温和的限制条件，常用的核函数（如均匀核和 Epanechnikov 核）均可满足该条件。对核函数施加紧支撑假设并非必要，也可以用适当的尾部条件替代。假设 2(ii) 中的带宽条件对于推导基于核的量的一致收敛性质至关重要。当 s 有界且 d 以 n的多项式速率发散时，这些条件可简化为 $nh/\log^2 n \to \infty$和 $h^2 \log n \to 0$。假设3(ii) 可视为所谓受限特征值条件的一致版本，该条件广泛用于高维线性回归模型（例如，Bickel、Ritov 和 Tsybakov，2009；Basu 和 Michailidis，2015）。补充材料中的附录 D 给出了使高维局部平稳高斯时间序列满足假设 3(ii) 的充分条件。此外，借助针对时变（非高斯）VAR 过程的 Hanson-Wright 不等式（例如，Zhang 和 Wu，2021 年的命题6.2），我们可以证明 $\max_{1 \leqslant t \leqslant n} \|\mathbf{\Psi}(\tau_t) - E[\mathbf{\Psi}(\tau_t)]\|_{\max} = O_P\left(\sqrt{\log(n \vee d)/(nh)}\right)$。然后，利用附录 D 中的引理 D.1，并假设 $s\sqrt{\log(n \vee d)/(nh)} = o(1)$，一个充分条件

对于（4.3）是

$$\min_{1\leqslant i\leqslant d}\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}_i(\tau_t)} u^{\intercal}\mathsf{E}\left[\Psi(\tau_t)\right]u \geqslant \kappa_0.$$

定理 4.1。 假设 *1*–*3*均成立。则有

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant t\leqslant n}\|\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau_t)-\boldsymbol{\alpha}_{i\bullet}(\tau_t)\| = O_P\left(\sqrt{s}\lambda_1\right). \tag{4.4}$$

定理 4.1表明，转移矩阵的初步时变 LASSO 估计是一致的，其收敛速率取决于 s 和 λ_1。尽管变量的维度 d 允许以 n的指数速率增长，但为了保证一致性， $\boldsymbol{\alpha}_{i\bullet}(\cdot)$ 中的显著元素数量不能增长过快。此外，一致收敛结果（4.4）可以加强为

$$\max_{1\leqslant i\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\|\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau)-\boldsymbol{\alpha}_{i\bullet}(\tau)\| = O_P\left(\sqrt{s}\lambda_1\right). \tag{4.5}$$

一阶导数函数估计也具有类似的一致收敛性质，见定理 4.1证明中的 **(A.1)**。

4.2 加权组 LASSO 估计的预言机性质

分别将 $\mathcal{J}_i$ 和 $\mathcal{J}_i'$ 的补集记为 $\mathcal{J}_i$ 和 $\mathcal{J}_i'$，即⊢ $\mathcal{J}_i = \bigcap_{t=1}^{n}\{j:\ \alpha_{i,j}(\tau_t)=0\}$ 和 $\mathcal{J}_i' = \bigcap_{t=1}^{n}\left\{j:\ \alpha_{i,j}'(\tau_t)=0\right\}$ $^{o} = (\boldsymbol{\alpha}_{\bullet 1}^{o},\ \boldsymbol{\alpha}_{\bullet n}^{o})$ $^{\intercal}$ $^{o} = (\boldsymbol{\beta}_{\bullet 1}^{o},\ \boldsymbol{\beta}_{\bullet n}^{o})$ $^{\intercal}$ $\boldsymbol{\alpha}_{\bullet t}^{o} =$。令 $(\alpha_{1|t}^{o},\ \alpha_{d|t}^{o})$ $^{\intercal}$ $\alpha_{j|t}^{o}=0$ $j\in\mathcal{J}_i-$ $\boldsymbol{\beta}_{\bullet t}^{o} = (\beta_{1|t}^{o},\ \beta_{d|t}^{o})$ $^{\intercal}$ $\beta_{j|t}^{o}=0$ $j\in\mathcal{J}_i'$ − **A**.....，且 **B**.....，，其中.....，对于.....以及.....，对于.....。定义（不可行的）预言估计：

$$\widehat{\mathbf{A}}_i^{o} = \left(\widehat{\boldsymbol{\alpha}}_{i,1}^{o},\ldots,\widehat{\boldsymbol{\alpha}}_{i,d}^{o}\right)\ \text{ with }\ \widehat{\boldsymbol{\alpha}}_{i,j}^{o} = \left[\widehat{\alpha}_{i,j}^{o}(\tau_1),\ldots,\widehat{\alpha}_{i,j}^{o}(\tau_n)\right]^{\intercal}, \tag{4.6}$$

$$\widehat{\mathbf{B}}_i^{o} = \left(\widehat{\boldsymbol{\alpha}}_{i,1}^{\prime o},\ldots,\widehat{\boldsymbol{\alpha}}_{i,d}^{\prime o}\right)\ \text{ with }\ \widehat{\boldsymbol{\alpha}}_{i,j}^{\prime o} = \left[\widehat{\alpha}_{i,j}^{\prime o}(\tau_1),\ldots,\widehat{\alpha}_{i,j}^{\prime o}(\tau_n)\right]^{\intercal}, \tag{4.7}$$

作为使 $Q_i(\mathbf{A}^{o}、\mathbf{B}^{o})$ 最小化的 $\mathbf{A}^{o}$ 和 $\mathbf{B}^{o}$ 的取值。我们需要对调节参数 λ_2 以及转移矩阵中显著时变系数的下界施加以下条件。

假设 4。 *(i)* 调节参数 λ_2 满足

$$\sqrt{n}s\log(n\vee d)\zeta_{n,d}+\sqrt{ns}\lambda_1 = o(\lambda_2),$$

其中$\zeta_{n,\ d}$ 在假设*3* （*i*） 中定义。

（*ii*）满足

$$\min_{1\leqslant i\leqslant d}\min_{j\in\mathcal{J}_i}\left(\sum_{t=1}^{n}\alpha_{i,j}^2(\tau_t)\right)^{\frac{1}{2}}\geqslant(a_0+1)\lambda_2\ \ and\ \ \min_{1\leqslant i\leqslant d}\min_{j\in\mathcal{J}_i'}D_{i,j}\geqslant(a_0+1)\lambda_2,$$

其中 $a_0=3.7$ 在 *SCAD* 惩罚中定义。

当 s 是固定正整数，$h\propto n^{-1/5}$，$\lambda_1\propto n^{-2/5+\eta_0}$ 且 $0<\eta_0<1/5$，并且 $d\sim\exp\{n^{\eta_1}\}$ 且 $0<\eta_1<\eta_0$时，通过令 $\lambda_2\propto n^{1/2-\eta_2}$ 且$0<\eta_2<2/5-[\eta_0\vee(2\eta_1)]$，容易验证假设 4(i)。假设 4(ii) 对时变系数函数及其相对于均值的偏差的下界施加了限制。这些限制弱于 Li, Ke and Zhang(2015) 中的假设 6(ii) 以及 Chen 等人 (2021) 中的假设8，并确保能够以概率趋近于 1 检测到显著的系数函数及其导数*w.p.a.1*。

定理 4.2。 假设条件*1–4* 均成立。加权组 *LASSO* 目标函数的极小化解 $Q_i(\mathbf{A},\mathbf{B})$存在，且几乎必然等于式（*4.6*）和（*4.7*）中定义的预言估计量。此外，我们有如下均方收敛结果：

$$\max_{1\leqslant i\leqslant d}\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{d}[\widehat{\alpha}_{ij}(\tau_t)-\alpha_{ij}(\tau_t)]^2=O_P\left(s\zeta_{n,d}^2\right),\tag{4.8}$$

其中 s 在假设条件 *2*（*ii*）中定义，且$\zeta_{n,\,d}$ 在假设条件 *3*（*i*）中定义。

由于惩罚局部线性估计几乎必然与式（4.6）和（4.7）中定义的不可行预言估计量相同，因此第 *3.2*节提出的全局模型选择程序具有稀疏性，即时变转移矩阵中的零元素可以被准确估计为零。按照定理4.2的证明，我们可以验证 Fan、Xue 和 Zou （2014） 几乎必然满足折叠凹惩罚函数的性质（*i*）–（*iv*）。因此，定理 4.2可视为将Fan、Xue 和 Zou （2014）的定理 1 以及Li, Ke and Zhang （2015）的定理 3.1 推广至高维局部平稳时间序列。

根据定理 4.2 中的预言机性质，可以直接推导出以下关于格兰杰因果关系连接有向边的网络估计一致性性质。

推论 4.1。 在定理 *4.2*的假设下，我们有

$$P\left(\widehat{\mathbb{E}}_n^G=\mathbb{E}_n^G\right)\to1.\tag{4.9}$$

4.3 时变 CLIME 估计的一致性 s

为了推导时变 CLIME 估计的一致一致性性质，我们需要调节参数 b 和 λ_3满足以下条件。

假设 5. *(i)* 带宽 b 满足

$$b \to 0 \quad \textit{and} \quad nb/[\log(n \vee d)]^3 \to \infty.$$

此外，$s\zeta_{n,\,d}\sqrt{对数(n \vee d)} \to 0$，其中 $\zeta_{n,\,d}$ 定义见假设 *3(i)*。

(ii) 存在一个充分大的常数 C_3，使得 $\lambda_3 = C_3\left(\nu^{\diamond}_{n,\,d} + \nu^{*}_{n,\,d}\right)$，其中 *re*

$$\nu^{\diamond}_{n,d} = \left[\frac{\log(n \vee d)}{nb}\right]^{1/2} + b^2 \quad \text{and} \quad \nu^{*}_{n,d} = s\zeta_{n,d}\sqrt{\log(n \vee d)}.$$

下面的定理给出了不同矩阵范数下时变精确率^矩阵估计 $\boldsymbol{\Omega}(\tau)$ 的一致收敛速率。

定理 4.3. 假设 *1–5* 均成立，且 $\{\boldsymbol{\Omega}(\tau) : 0 \leqslant \tau \leqslant 1\} \in \mathcal{S}(q, \xi_d)$。则有

$$\sup_{0\leqslant\tau\leqslant1}\left\|\widehat{\boldsymbol{\Omega}}(\tau) - \boldsymbol{\Omega}(\tau)\right\|_{\max} = O_P\left(\nu^{\diamond}_{n,d} + \nu^{*}_{n,d}\right), \tag{4.10}$$

$$\sup_{0\leqslant\tau\leqslant1}\left\|\widehat{\boldsymbol{\Omega}}(\tau) - \boldsymbol{\Omega}(\tau)\right\| = O_P\left(\xi_d(\nu^{\diamond}_{n,d} + \nu^{*}_{n,d})^{1-q}\right), \tag{4.11}$$

$$\sup_{0\leqslant\tau\leqslant1}\frac{1}{d}\left\|\widehat{\boldsymbol{\Omega}}(\tau) - \boldsymbol{\Omega}(\tau)\right\|_F^2 = O_P\left(\xi_d(\nu^{\diamond}_{n,d} + \nu^{*}_{n,d})^{2-q}\right), \tag{4.12}$$

其中 ξ_d 定义见（3.7），$\nu^{\diamond}_{n,\,d}$ 和 $\nu^{*}_{n,\,d}$ 定义见假设 *5*（*ii*）。

定理 4.3中的一致收敛速率依赖于 $\nu^{\diamond}_{n,\,d}$ 和 $\nu^{*}_{n,\,d}$。第一种速率 $\nu^{\diamond}_{n,\,d}$ 是非参数基于核的量的常规一致收敛速率，而第二种速率 $\nu^{*}_{n,\,d}$ 源于 $\widehat{e}_t$ 对潜在 VAR 误差的近似误差 e_t。注意，维度 d 通过 ξ_d 和 $\log(n \vee d)$影响一致收敛速率；当 d 以 n的指数速率发散时，超高维情形下的一致性性质仍成立。定理 4.3 可视为将 Cai、Liu 和 Luo （2011）中的定理 1 推广到高维局部平稳时间序列情形。

根据定理4.3，我们很容易得到偏相关连接中无向边网络估计量的以下一致性。

推论 4.2. 无向　　在定理 4.3 的假设下，如果 $\min_{(i,j)\in\mathbb{E}^P}\min_{1\leqslant t\leqslant n}|\omega_{ij}(\tau_t)|\gg\lambda_3$，则有　　*ve*

$$\mathsf{P}\left(\widehat{\mathbb{E}}_n^P=\mathbb{E}_n^P\right)\to 1. \tag{4.13}$$

5 因子调整时变向量自回归与网络

本节中，令 $(Z_t:t=1,\ldots,n)$ 满足 $Z_t=(z_{t,1},\ldots,z_{t,\,d})$ 为一列观测到的 d 维随机向量。为处理实际收集的大规模时间序列中常见的强截面依赖，我们假设 Z_t 由近似因子模型生成：

$$Z_t=\boldsymbol{\Lambda}F_t+X_t,\quad t=1,\ldots,n, \tag{5.1}$$

其中 $\boldsymbol{\Lambda}=(\Lambda_1,\ldots,\Lambda_d)$ 是一个 $d\times k$ 阶因子载荷矩阵，F_t 是一个 k 维潜在因子向量，并且 (X_t) 假设其满足时变 VAR 模型（2.1）。更一般地，我们可以假设以下时变因子模型结构：

$$Z_t=\boldsymbol{\Lambda}_tF_t+X_t,\quad t=1,\ldots,n, \tag{5.2}$$

其中，$\boldsymbol{\Lambda}_t=\boldsymbol{\Lambda}(t/n)$ 是一个时变因子载荷矩阵，其每个元素都是缩放时间的平滑函数。近似因子模型及其时变推广已在文献中得到广泛研究（例如，Chamberlain and Rothschild，1983年；Bai 和 Ng，2002；Stock 和 Watson，2002；Motta、Hafner 和 von Sachs，2011；Su and Wang，2017）。本节主要关注估计特质误差向量 X_t 的时变网络。尽管 Z_t 的各个分量可能高度相关，但 X_t 的各个分量通常仅弱相关。因此，对特质误差过程的时变转移矩阵和精度矩阵施加稀疏性假设是合理的，这使得我们能够应用第 3 节提出的估计方法。然而，这并非易事，因为共同成分（$\boldsymbol{\Lambda}F_t$ 或 $\boldsymbol{\Lambda}_tF_t$）和特质误差成分都无法观测。受近期关于高维数据中因子模型与稀疏模型相结合的研究启发（例如，Fan, Masini and Medeiros，2021；Krampe and Margaritella，2022），接下来我们使用主成分分析（PCA）或其局部化版本，从观测到的时间序列数据中剔除由潜在因子驱动的共同成分。

设 $\mathbf{Z}=(Z_1,\ldots,Z_n)^\intercal$，$\mathbf{F}=(F_1,\ldots,F_n)^\intercal$ 以及 $\mathbf{X}=(X_1,\ldots,X_n)^\intercal$。对于传统因子模型（5.1），我们对 $n\times n$ 矩阵 $\mathbf{ZZ}^\intercal$ 进行特征值分析。$\mathbf{F}$的估计记作 $\widehat{\mathbf{F}}=\left(\widehat{F}_1,\ldots,\widehat{F}_n\right)^\intercal$，它由与 k 个最大特征值对应的 $\mathbf{ZZ}^\intercal$ 的特征向量组成的 $n\times k$ 矩阵（乘以 $\sqrt{n}$）得到。因子载荷矩阵的估计为

$\widehat{\mathbf{\Lambda}} = \left(\widehat{\Lambda}_1, \ldots, \widehat{\Lambda}_d\right)_\tau = \mathbf{Z}\ \widehat{}\mathbf{F}/n$。因此，$\mathbf{\Lambda}F_t$共同成分的估计为 $\widehat{\mathbf{\Lambda}}\widehat{F}_t$，而 X_t 特质误差成分的估计为

$$\widehat{X}_t = Z_t - \widehat{\mathbf{\Lambda}}\widehat{F}_t,\ \ t = 1, \ldots, n. \tag{5.3}$$

对于时变因子模型（5.2），上述主成分分析估计程序需要做一些修改。具体而言，令

$$K_{t,h_*}(\tau) = \frac{K_{h_*}(\tau_t - \tau)}{\sum_{s=1}^{n} K_{h_*}(\tau_s - \tau)},\ \ 0 < \tau < 1,$$

其中，h_* 是带宽，且 $K_{h_*}(\cdot)$ 定义同第3.1节，并定义局部化数据矩阵：

$$\mathbf{Z}(\tau) = [Z_1(\tau), \ldots, Z_n(\tau)]^\intercal \ \text{ with } \ Z_t(\tau) = Z_t K_{t,h_*}^{1/2}(\tau).$$

通过对矩阵 $\mathbf{Z}(\tau)\mathbf{Z}$ (τ)进行特征值分析，可以得到因子的局部主成分分析估计以及因子载荷矩阵，分别记为 $\widehat{}\mathbf{F}(\tau) = \left[\widehat{F}_1(\tau), \ldots\ldots \widehat{F}_n(\tau)\right]$ 和 $\widehat{\mathbf{\Lambda}}(\tau)$。随后，特质误差向量 X_t 可近似为

$$\widehat{X}_t = Z_t - \widehat{\mathbf{\Lambda}}(\tau_t)\widehat{F}(\tau_t),\ \ t = 1, \ldots, n, \tag{5.4}$$

w $\widehat{}$这里我们保留了与 X_t 式（5.3）相同的记号，以避免符号过于繁杂。

与第4节一样，我们仅考虑特质误差向量的时变 VAR(1) 模型。利用近似关系 X_t，我们可以应用第 3 节提出的三阶段估计程序。将初步时变LASSO估计记为 $\widetilde{\alpha}_{ij}^{\dagger}(\cdot)$，将第二阶段加权组LASSO估计记为 $\widehat{\alpha}_{ij}^{\dagger}(\cdot)$，将因子调整的时变精度矩阵估计记为 $\widehat{\mathbf{\Omega}}^{\dagger}(\cdot) = \left[\widehat{\omega}_{ij}^{\dagger}(\cdot)\right]_{d\times d}$。随后，我们可以构造一致网络估计$\widehat{\mathbb{E}}^{G}$、$_n{}^{\dagger}$和 $\widehat{\mathbb{E}}^{P}$、$_n{}^{\dagger}$，其定义与式（3.13）和（3.14）中的 $\widehat{\mathbb{E}}_n^{G}$ 和 $\widehat{\mathbb{E}}_n^{P}$ 类似，但分别以 $\widehat{\alpha}_{ij}^{\dagger}(\cdot)$ 和 $\widehat{\omega}_{ij}^{\dagger}(\cdot)$替换 $\widehat{\alpha}_{ij}(\cdot)$ 和 $\widehat{\omega}_{ij}(\cdot)$。为推导这些因子调整估计的收敛性质，我们需要以下假设；该假设对假设3–5进行了修改，以纳入特质误差成分的近似误差。

假设6。 *(i)* 记为 $\delta_X = \max_{1\leqslant t\leqslant n}\left|\widehat{X}_t - X_t\right|_{\max}$。有 $[\log(n\vee d)]^{1/2}s\delta_X = o_P(1)$。

(ii) 将 $\zeta_{n,d}$ 替换为 $\zeta_{n,d}^{\dagger} = \zeta_{n,d} + [\log(n\vee d)]^{1/2}s\delta_X$时，假设 *3(i)* 成立。

(iii) 将$\zeta_{n,d}$ 替换为 $\zeta_{n,d}^{\dagger}$时，假设 *4(i)* 成立。

(iv) 将 $\zeta_{n,d}$ 和 $\nu_{n,d}^{*}$ 分别替换为 $\zeta_{n,d}^{\dagger}$ 和 $\nu_{n,d}^{\dagger} = s\zeta_{n,d}^{\dagger}\sqrt{\log(n\vee d)}$时，假设 *5* 成立。

假设6(i) 对因子模型中的潜在 X_t 的近似提出了一个高层次条件，即近似误差 δ_X 一致收敛至零，且收敛速度快于

$s^{-1}[\log(n \vee d)]^{-1/2}$。根据Fan、Liao 和 Mincheva(2013)中的推论1，对传统因子模型（5.1）进行主成分分析估计时，近似误差的典型收敛速度为

$$\delta_X = O_P\left((\log n)^{1/2}\left[(\log d)^{1/2} n^{-1/2} + n^{1/\upsilon} d^{-1/2}\right]\right), \tag{5.5}$$

其中，$\upsilon > 2$ 是一个与矩限制有关的正数。根据 Su和Wang（2017）中的定理3.5，当采用局部主成分分析估计时，我们可以得到时变因子模型（5.2）下 δ_X 的典型一致收敛速率。在假设 6(ii)–(iv)中，我们修改了假设3(i)、4(i)和5(ii)，以纳入近似误差 δ_X。然而，如果进一步假设$h \propto n^{-1/5}$ 和 d 以 n 的多项式速率发散，且满足 $d \gg n^{1+2/\upsilon}$，那么（5.5）中的速率可简化为 $\delta_X = O_P\left((\log d) n^{-1/2}\right) = o_P(h^2)$，从而$\zeta_{n、d} \propto \zeta^{\dagger}_{n、d}$。因此，我们可以去掉假设 6(ii)–(iv)，并且在模型（5.1）下，δ_X 不会影响估计收敛速率。

以下两个命题将第 4 节中的理论结果推广到因子调整时变向量自回归和网络。

命题 5.1. 假　　设因子模型（*5.1*）或（*5.2*）以及假设 *1*、*2* 和 *3*（*ii*）均成立。

（*i*）在假设 *6*（*i*）（*ii*）下，我们有

$$\max_{1 \leqslant i \leqslant d} \max_{1 \leqslant t \leqslant n} \sum_{j=1}^{d} \left[\widetilde{\alpha}^{\dagger}_{ij}(\tau_t) - \alpha_{ij}(\tau_t)\right]^2 = O_P\left(s\lambda_1^2\right). \tag{5.6}$$

（*ii*）在假设 *6*（*i*）–（*iii*）下，第二阶段加权组*LASSO*估计具有预言机性质，此外，

$$\max_{1 \leqslant i \leqslant d} \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{d} \left[\widehat{\alpha}^{\dagger}_{ij}(\tau_t) - \alpha_{ij}(\tau_t)\right]^2 = O_P\left(s\left(\zeta^{\dagger}_{n,d}\right)^2\right). \tag{5.7}$$

（*iii*）在假　假设 *6* 以及稀疏性条件 $\{\mathbf{\Omega}(\tau) : 0 \leqslant \tau \leqslant 1\} \in \mathcal{S}(q, \xi_d)$，我们有

$$\sup_{0 \leqslant \tau \leqslant 1} \left\|\widehat{\mathbf{\Omega}}^{\dagger}(\tau) - \mathbf{\Omega}(\tau)\right\|_{\max} = O_P\left(\nu^{\diamond}_{n,d} + \nu^{\dagger}_{n,d}\right), \tag{5.8}$$

$$\sup_{0 \leqslant \tau \leqslant 1} \left\|\widehat{\mathbf{\Omega}}^{\dagger}(\tau) - \mathbf{\Omega}(\tau)\right\| = O_P\left(\xi_d(\nu^{\diamond}_{n,d} + \nu^{\dagger}_{n,d})^{1-q}\right), \tag{5.9}$$

$$\sup_{0 \leqslant \tau \leqslant 1} \frac{1}{d} \left\|\widehat{\mathbf{\Omega}}^{\dagger}(\tau) - \mathbf{\Omega}(\tau)\right\|_F^2 = O_P\left(\xi_d(\nu^{\diamond}_{n,d} + \nu^{\dagger}_{n,d})^{2-q}\right). \tag{5.10}$$

命题 5.2。 *(i)* 在命题 *5.1(ii)* 的假设下，我们有

$$\mathsf{P}\left(\widehat{\mathbb{E}}_n^{G,\dagger} = \mathbb{E}_n^G\right) \to 1. \tag{5.11}$$

(ii) 在命题 *5.1(iii)* 的假设以及 $\min_{(i,j)\in\mathbb{E}^P} \min_{1\leqslant t\leqslant n} |\omega_{ij}(\tau_t)| \gg \lambda_3$成立时，我们有

$$\mathsf{P}\left(\widehat{\mathbb{E}}_n^{P,\dagger} = \mathbb{E}_n^P\right) \to 1. \tag{5.12}$$

6 蒙特卡洛模拟

本节提供四个模拟示例，以考察所提出的高维时变向量自回归模型和网络估计在有限样本下的数值表现。全文中，我们将所提出的时变加权组 LASSO 方法记为 tv-wgLASSO，将时变 CLIME 方法记为 tv-CLIME。我们将 tv-wgLASSO 的表现与（不可行的）时变预言机估计进行比较，后者记为 tv-Oracle，仅估计真实的显著系数函数（假设这些函数已知）；此外，还与无惩罚的完整时变估计进行比较，后者记为 tv-Full，在不施加惩罚的情况下估计所有系数函数。我们将 tv-CLIME 的表现与时变图形 LASSO 估计进行比较，后者记为 tv-GLASSO，并通过 R软件包“ glassoFast”基于向量自回归残差实现。此外，为考察 VAR 模型误差近似造成的估计精度损失，我们还报告了不可行时变 CLIME 的结果；该方法在估计精度矩阵时直接使用 VAR误差（而非残差）。

在模拟中，我们采用 Epanechnikov 核 $K(t) = 0$，带宽为$h = b = 0$。$75[\log(d)/n]^{1/5}$ 与 Li、Ke 和 Zhang（2015）中的设置相同。局部主成分分析的带宽设为 $h_* = (2, 35/\sqrt{12})[\sqrt{d/n}]^{1/5}$ 与 Su和Wang（2017）中的设置相同。我们将样本量 n 设为 200 和 400，将维度 d 设为 50 和 100。尽管这些维度小于样本量，但当 $n = 200$ 且 $d = 100$时，（3.3）中每次局部线性估计所用的“有效样本量”约为 $2nh \approx 140$，小于未知系数函数及其导数的总数 $2d = 200$。因此，在这种情况下，我们无法实施朴素的 tv-Full 估计。所提出的估计过程包含三个调节参数：λ_1 用于第一阶段的初步时变 LASSO 估计，λ_2 用于第二阶段的时变加权组 LASSO，λ_3 用于第三阶段的时变 CLIME。它们分别通过贝叶斯信息准则（BIC）、广义信息准则（GIC）和扩展贝叶斯信息准则（EBIC）选取。补充材料中的附录 E 给出了这些信息准则的定义。

为了评估 wh eth在时变模型结构得到准确估计后，我们报告

假阳性（FP）、假阴性（FN）、真正率（TPR）、真负率（TNR）、阳性预测值（PPV）、阴性预测值（NPV）、F1 分数（F1）和马修斯相关系数（MCC）。这些度量的定义见补充材料中的附录 E。为评估系数估计量的性能，我们报告所有维度上的平均决定系数（average R^2）、系数函数估计误差的平均缩放弗罗贝尼乌斯范数（EE_A），以及误差的均方根误差（$RMSE_e$）。以我们提出的用于时变 VAR(1) 的 tv-wgLASSO 估计量为例，

$$EE_A = \frac{1}{n\sqrt{d}} \sum_{t=1}^{n} \left\| \widehat{\mathbf{A}}_1(\tau_t) - \mathbf{A}_1(\tau_t) \right\|_F \quad \text{and} \quad RMSE_e = \sqrt{\frac{1}{nd} \sum_{i=1}^{d} \sum_{t=1}^{n} (\widehat{e}_{t,i} - e_{t,i})^2}.$$

为评估精度矩阵估计量的表现，我们报告估计误差（EE_Ω）的平均缩放弗罗贝尼乌斯范数，其定义为

$$EE_\Omega = \frac{1}{n\sqrt{d}} \sum_{t=1}^{n} \left\| \widehat{\boldsymbol{\Omega}}(\tau_t) - \boldsymbol{\Omega}(\tau_t) \right\|_F .$$

上述所有度量均针对每次蒙特卡洛重复分别计算，然后对 100 次重复的结果取平均。

例1。 数据由时变 VAR(1) 模型生成，其中 $\mathbf{A}_1(\tau)$ 对所有 $\tau \in [0,1]$ 均为对角矩阵。$\mathbf{A}_1(\tau)$ 的每个对角元素独立地以相同的 0.5 概率取值 $0.64\Phi(5(\tau - 1/2))$ 或 $0.64 - 0.64\Phi(5(\tau - 1/2))$，其中 $\Phi(\cdot)$ 为标准正态分布函数。我们将 $\boldsymbol{\Omega}(\tau)$ 设为分块对角矩阵：$\boldsymbol{\Omega}(\tau) = \mathbf{I}_{d/2} \otimes \boldsymbol{\Omega}_*(\tau)$，其中 $\boldsymbol{\Omega}_*(\tau) = [\omega_{ij,*}(\tau)]_{2\times 2}$ 满足 $\omega_{11,\,*}(\tau) = \omega_{22,\,*}(\tau) \equiv 1$，且 $\omega_{12,\,*}(\tau) = \omega_{21,\,*}(\tau) = 1.4\Phi(5(\tau - 1/2)) - 0.7$。$\mathbf{A}_1(\tau)$ 的对角结构意味着变量之间不存在格兰杰因果关系，而 $\boldsymbol{\Omega}(\tau)$ 的分块对角结构会导致 X_t的各分量之间存在较弱的截面依赖。

表1报告了时变转移矩阵和格兰杰网络的估计结果。对于所提出的 tv-wgLASSO，与d^2（潜在有向格兰杰因果关系的总数或转移矩阵的条目数）相比，其假阳性和假阴性值非常小。因此，TPR、TNR、PPV、NPV、F1分数和马修斯相关系数等度量值都很高，且接近 1。我们还可以看到，当 d 从 50 增加到 100 时，假阳性和假阴性值翻倍；但当 n从 200 增至 400 时，这些值显著下降。这些结果清楚地表明，只要样本量适度大，tv-wgLASSO 就能准确恢复时变格兰杰网络。tv-wgLASSO 的平均 R^2 与 tv-Oracle 的接近，而朴素的 tv-Full 方法由于模型过拟合，往往会有较大的 R^2。尽管当$_A$ 时，tv-wgLASSO 的能量效率值高于 tv-Oracle，但当 $n = 200$时，这些值会显著下降，甚至

略低于 tv-Oracle 的相应值。当 $n = 400$ 时，可以观察到 $RMSE_e$ 中类似的模式，这表明所提出的 tv-wgLASSO 能够很好地近似 VAR 误差，而这些误差将用于后续的时变精度矩阵估计。不出所料，当 $d = 100$ 和 $n = 200$ 时，tv-Full 方法无法估计时变转移矩阵。

表2 报告了时变精度矩阵和偏相关网络的估计结果。当 $n = 200$时，tv-CLIME 和 tv-GLASSO 的假阳性值均为零，而 tv-CLIME 的假阴性值低于 tv-GLASSO。因此，就 F1分数和马修斯相关系数这两项度量而言，所提出的 tv-CLIME 优于 tv-GLASSO。当 $n = 400$时，tv-CLIME 和 tv-GLASSO 均能正确恢复时变偏相关网络。就精度矩阵估计准确度（能量效率$_\Omega$）而言，tv-GLASSO 的表现略优于 tv-CLIME。此外，通过比较 tv-CLIME 与不可行时变 CLIME，我们可以得出结论：VAR 误差近似对精度矩阵和偏相关网络估计的影响可以忽略不计。

示例 2。数据由时变 VAR(1) 模型生成，其中 $\mathbf{A}_1(\tau)$ 对所有 $\tau \in [0,1]$ 均为上三角矩阵。$\mathbf{A}_1(\tau)$ 的每个对角元素取值为 $0.7\Phi(5(\tau - 1/2))$，每个超对角元素取值为 $0.7 - 0.7\Phi(5(\tau - 1/2))$，其余元素取值为 0。我们设定 $\mathbf{\Omega}(\tau) = [\omega_{ij}(\tau)]_{d\times d}$ 对所有 $\tau \in [0,1]$ 均为带状对称矩阵，其中 $\omega_{ii}(\tau) \equiv 1$，$\omega_{i,\,(i+1)}(\tau) = 0.7\Phi(5(\tau - 1/2)) - 0.7$，$\omega_{i,\,(i+2)}(\tau) = 0.7 - 0.7\Phi(5(\tau - 1/2))$，且$\omega_{i,\,j}(\tau) \equiv 0$ 当 $|i - j| > 2$ 时。

表3报告了时变转移矩阵和格兰杰网络的估计结果。请注意，本例中的时间序列变量比例1中的变量相关性更强，这会影响网络估计的准确性。当 $d = 100$ 且 $n = 200$时，tv-wgLASSO 的 FP 和 FN 值分别达到最大值 20.73 和 37.55，而 F1分数和马修斯相关系数约为 0.85。与例1一样，当 n从 200 增至 400 时，F1分数和马修斯相关系数均有所提高，并且 tv-wgLASSO 的平均 R^2 与 tv-Oracle 的接近。然而，tv-wgLASSO 的能量效率$_A$ 和均方根误差$_e$ 远高于 tv-Oracle。

表4 报告了时变精度矩阵和偏相关网络的估计结果。从表$_A$ 和均方根误差$_e$ 结果可知，VAR误差近似的效果不如例1。因此，所提出的 tv-CLIME 的表现不如直接将真实VAR误差用于估计的不可行时变 CLIME。特别是，当 $n = 200$时，tv-CLIME 的 FN 远高于不可行时变 CLIME。出于同样的原因，不可行时变 CLIME 的表现也优于 tv-GLASSO。此外，我们发现，当 $n = 200$时，tv-CLIME 在恢复时变精度网络方面优于 tv-GLASSO；而当 $n = 400$时，两者表现相当。

例3。数据由 VAR(1) 模型生成，其中 $\mathbf{A}_1(\tau) = [a_{ij}(\tau)]_{d\times d}$ 为托普利兹

表 1：例1中的转移矩阵和格兰杰网络估计。

		tv-wgLASSO		tv-Oracle		tv-Full	
度量	维度 $n = 200$	$n = 400$		$n = 200$	$n = 400$	$n = 200$	$n = 400$
FP	$d = 50$	0.97	0.04	0	0	2450	2450
	$d = 100$	1.73	0.08	0	0	-	9900
FN	$d = 50$	3.53	0.08	0	0	0	0
	$d = 100$	8.55	0.15	0	0	-	0
TPR	$d = 50$	0.929	0.998	1	1	1	1
	$d = 100$	0.915	0.999	1	1	-	1
TNR	$d = 50$	1.000	1.000	1	1	0	0
	$d = 100$	1.000	1.000	1	1	-	0
PPV	$d = 50$	0.980	0.999	1	1	0.02	0.02
	$d = 100$	0.982	0.999	1	1	-	0.01
NPV	$d = 50$	0.999	1.000	1	1	1	1
	$d = 100$	0.999	1.000	1	1	-	1
F1	$d = 50$	0.953	0.999	1	1	0.039	0.039
	$d = 100$	0.947	0.999	1	1	-	0.020
MCC	$d = 50$	0.953	0.999	1	1	0	0
	$d = 100$	0.947	0.999	1	1	-	0
平均值 R^2	$d = 50$	0.289	0.296	0.296	0.297	0.933	0.721
	$d = 100$	0.296	0.306	0.305	0.307	-	0.959
EE_A	$d = 50$	0.214	0.160	0.185	0.163	54.29	1.410
	$d = 100$	0.224	0.163	0.189	0.166	-	112.8
$RMSE_e$	$d = 50$	0.203	0.115	0.162	0.120	1.119	0.876
	$d = 100$	0.213	0.113	0.159	0.119	-	1.145

所有表格中，除精确值 *0* 和 *1* 外，假阳性和假阴性度量均四舍五入到小数点后 *2* 位，其余数值均四舍五入到小数点后 *3* 位。

Tabl例 2：精度矩阵与偏相关网络估计 e 1.

		tv-CLIME		不可行时变 CLIME		tv-GLASSO	
测量维度 $n = 200$		$n = 400$		$n = 200$	$n = 400$	$n = 200$	$n = 400$
FP	$d = 50$	0	0.02	0	0.02	0	0
	$d = 100$	0	0.03	0	0.01	0	0
FN	$d = 50$	5.06	0	3.49	0	9.24	0
	$d = 100$	13.25	0	9.01	0	28.31	0
TPR	$d = 50$	0.798	1	0.860	1	0.630	0
	$d = 100$	0.735	1	0.820	1	0.434	0
TNR	$d = 50$	1	1.000	1	1.000	1	1
	$d = 100$	1	1.000	1	1.000	1	1
PPV	$d = 50$	1	0.999	1	0.999	1	1
	$d = 100$	1	0.999	1	1.000	1	1
NPV	$d = 50$	0.996	1	0.097	1	0.992	1
	$d = 100$	0.997	1	0.998	1	0.994	1
F1	$d = 50$	0.884	1.000	0.922	1.000	0.768	1
	$d = 100$	0.845	1.000	0.899	1.000	0.600	1
MCC	$d = 50$	0.889	1.000	0.925	1.000	0.788	1
	$d = 100$	0.855	1.000	0.904	1.000	0.653	1
EE_Ω	$d = 50$	0.510	0.436	0.503	0.435	0.451	0.407
	$d = 100$	0.481	0.421	0.473	0.419	0.433	0.397

矩阵和 $a_{ij}(\tau) = (0.4 - 0.1\tau)^{|i-j|+1}$。我们还设定 $\mathbf{\Omega}(\tau) = [\omega_{ij}(\tau)]_{d\times d}$ 为一个托普利茨矩阵，其 $\omega_{ij}(\tau) = (0.8-0.1\tau)^{|i-j|}$。在这个例子中，转移矩阵和精度矩阵均为非稀疏矩阵；我们旨在考察当（精确）稀疏性假设不成立时，所提出的方法表现如何。

表 5 报告了所考察各种方法的估计误差。在本例中，tv-Oracle 等价于 tv-Full，并且在对时变转移矩阵采用传统局部线性估计程序时，两者都受到维数灾难的影响（尤其是在$d = 100$ 和 $n = 200$时）。因此，tv-wgLASSO 的能量效率$_A$ 和均方根误差$_e$ 远小于 tv-Oracle。tv-Oracle 的能量效率$_\Omega$ 结果与不可行时变 CLIME 的结果非常接近，这表明正如例1中所讨论的，VAR误差近似对 tv-CLIME 的性能影响很小。此外，tv-CLIME 和 Oracle tv-CLIME 的能量效率$_\Omega$ 结果通常与 tv-GLASSO 的结果接近。模拟结果表明，当转移矩阵和精度矩阵不满足稀疏性假设时，所提出的 tv-wgLASSO 和 tv-CLIME 仍具有合理的表现。

示例4。数据由形式为（5.2）的因子调整时变 VAR 模型生成。时变因子模型的特质误差由示例 2 中的 VAR(1)模型生成。中的两个因子由两个单变量

$^\intercal$

表3：示例 2 中的转移矩阵和格兰杰网络估计。

		tv-wgLASSO		tv-Oracle		tv-Full	
度量	维度 $n = 200$	$n = 400$		$n = 200$	$n = 400$	$n = 200$	$n = 400$
FP	$d = 50$	13.53	12.75	0	0	2401	2401
	$d = 100$	20.73	7.73	0	0	-	9801
FN	$d = 50$	18.56	11.11	0	0	0	0
	$d = 100$	37.55	13.90	0	0	-	0
TPR	$d = 50$	0.813	0.888	1	1	1	1
	$d = 100$	0.811	0.930	1	1	-	1
TNR	$d = 50$	0.994	0.995	1	1	0	0
	$d = 100$	0.998	0.999	1	1	-	0
PPV	$d = 50$	0.859	0.875	1	1	0.040	0.040
	$d = 100$	0.888	0.960	1	1	-	0.020
NPV	$d = 50$	0.992	0.995	1	1	0	0
	$d = 100$	0.996	0.999	1	1	-	0
F1	$d = 50$	0.834	0.881	1	1	0.076	0.076
	$d = 100$	0.847	0.945	1	1	-	0.039
MCC	$d = 50$	0.828	0.876	1	1	0	0
	$d = 100$	0.846	0.943	1	1	-	0
平均值 R^2	$d = 50$	0.465	0.448	0.477	0.462	0.963	0.829
	$d = 100$	0.473	0.467	0.483	0.471	-	0.978
EE_A	$d = 50$	0.328	0.250	0.171	0.122	58.44	1.510
	$d = 100$	0.323	0.204	0.168	0.122	-	82.60
$RMSE_e$	$d = 50$	0.631	0.476	0.417	0.305	1.673	1.414
	$d = 100$	0.613	0.390	0.414	0.309	-	1.720

表 4：示例 2 中的精度矩阵和偏相关网络估计。

		tv-CLIME		不可行时变 CLIME		tv-GLASSO	
测量维度	$n = 200$	$n = 400$		$n = 200$	$n = 400$	$n = 200$	$n = 400$
FP	$d = 50$	0.03	0.04	0.02	0.03	0	0.01
	$d = 100$	0.01	0	0	0.01	0	0.01
FN	$d = 50$	12.62	0.82	2.34	0	20.84	0.06
	$d = 100$	24.71	0.23	6.21	0.01	49.73	0.43
TPR	$d = 50$	0.742	0.983	0.952	1	0.575	0.997
	$d = 100$	0.750	0.998	0.937	1.000	0.498	0.996
TNR	$d = 50$	1.000	1.000	1.000	1.000	1	1.000
	$d = 100$	1.000	1	1	1.000	1	1.000
PPV	$d = 50$	0.999	0.999	1.000	0.999	1	1.000
	$d = 100$	1.000	1	1	1.000	1	1.000
NPV	$d = 50$	0.989	0.999	0.998	1	0.983	1.000
	$d = 100$	0.995	1.000	0.999	1.000	0.990	1.000
F1	$d = 50$	0.850	0.991	0.975	1.000	0.725	0.998
	$d = 100$	0.857	0.999	0.967	1.000	0.662	0.998
MCC	$d = 50$	0.856	0.991	0.975	1.000	0.749	0.998
	$d = 100$	0.864	0.999	0.967	1.000	0.701	0.998
EE_Ω	$d = 50$	0.598	0.533	0.526	0.485	0.560	0.514
	$d = 100$	0.560	0.489	0.486	0.458	0.536	0.496

表5：例3中双网络的估计准确度。

		tv-wgLASSO		tv-Oracle		tv-Full	
度量	维度 $n = 200$	$n = 400$		$n = 200$	$n = 400$	$n = 200$	$n = 400$
平均值 R^2	$d = 50$	0.009	0.029	0.891	0.588	0.891	0.588
	$d = 100$	0.005	0.020	-	0.930	-	0.930
EE_A	$d = 50$	0.383	0.348	56.66	1.927	56.66	1.927
	$d = 100$	0.388	0.364	-	97.60	-	97.60
$RMSE_e$	$d = 50$	0.515	0.463	1.716	1.300	1.716	1.300
	$d = 100$	0.523	0.486	-	1.776	-	1.776
		tv-CLIME		不可行时变 CLIME		tv-GLASSO	
		$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$
EE_Ω	$d = 50$	1.669	1.601	1.613	1.572	1.584	1.570
	$d = 100$	1.674	1.615	1.616	1.580	1.587	1.588

AR(1)过程：$F_{t,1} = 0.6F_{t-1,1} + \sqrt{1-0.6^2}u^F_{t,1}$ 和 $F_{t,2} = 0.3F_{t-1,2} + \sqrt{1-0.3^2}u^F_{t,2}$，其中 $u^F_{t,1}$和$u^F_{t,2}$ 独立地取自标准正态分布。因子载荷矩阵定义为 $\Lambda_t = (\Lambda_{t,1}, \Lambda_{t,2})$，其中 $\Lambda_{t,1} \equiv \Lambda_1$ 是从 d维标准多元正态分布中抽取的时不变向量，且$\Lambda_{t,2} = (\Lambda_{1t,2}, \ldots, \Lambda_{dt,2})$ 满足$\Lambda_{it,2} = 2/\{1 + \exp\{-2[10(t/n) - 5(i/d) - 2]\}\}$，其中 $i = 1, \ldots, d$。$^{\intercal}$

表 6报告了特质误差的时变转移矩阵和格兰杰网络的估计结果，表 7 报告了时变精度矩阵和偏相关网络的估计结果。与表 3 和4中的结果相比，我们可以观察到，因子调整估计引入了额外的估计误差，导致 F1分数和马修斯相关系数的值较低。当 $n = 200$ 时，影响更为明显，但当 $n = 400$ 时，影响则大幅减弱。与前面的例子一样，随着n从200增加到400，F1分数和马修斯相关系数的值也随之增加。因此，我们可以得出结论：尽管因子模型估计误差会传递到三阶段估计程序中，但当样本量适中且较大（$n = 400$）时，其对网络估计的影响并不显著。

7 实证应用

在本节中，我们使用 FRED-MD 宏观经济数据集，应用所提出的方法估计格兰杰因果关系网络和偏相关网络。该数据集可从 FRED-MD网站[1], 获取，包含 127 个美国宏观经济变量，观测频率为月度，时间跨度为 1959年1月至 2022年7月。这些宏观经济变量可分为八类：消费、订单与库存；住房；利率与汇率；劳动力市场；货币与信贷；产出与收入；价格；以及股票市场。更详细的描述见McCracken 和 Ng (2016)。

我们遵循McCracken 和 Ng (2016) 和 McCracken 和 Ng(2020) 的方法，剔除离群值并填补缺失值。将每个变量标准化，使其具有零均值和单位方差。为处理较强的截面依赖，我们考虑第5 节中的两种因子建模方法：具有恒定因子载荷的近似因子模型（5.1），以及具有动态因子载荷的时变因子模型（5.2）。我们使用Bai 和 Ng (2002) 和 Su andWang (2017) 提出的信息准则，确定这两个模型中的因子数量（准则说明见补充材料中的附录 E）。具有恒定载荷的因子模型选取了七个因子，而时变因子模型仅选取了四个因子。由于后者的模型设定更为简约，以下我们仅报告该模型的网络估计结果。估计的特质误差记为$\widehat{x}_{t,\ i}$,

[1]https://research.stlouisfed.org/econ/mccracken/fred-databases/

表6：示例4中的因子调整转移矩阵和格兰杰网络估计。

		tv-wgLASSO	
度量	维度 $n = 200$	$n = 400$	
FP	$d = 50$	11.35	10.60
	$d = 100$	20.40	10.41
FN	$d = 50$	35.97	14.77
	$d = 100$	65.45	20.68
TPR	$d = 50$	0.637	0.851
	$d = 100$	0.671	0.896
TNR	$d = 50$	0.995	0.996
	$d = 100$	0.998	0.999
PPV	$d = 50$	0.852	0.890
	$d = 100$	0.869	0.945
NPV	$d = 50$	0.985	0.994
	$d = 100$	0.993	0.998
F1	$d = 50$	0.725	0.869
	$d = 100$	0.756	0.920
MCC	$d = 50$	0.725	0.865
	$d = 100$	0.759	0.919
平均值 R^2	$d = 50$	0.298	0.350
	$d = 100$	0.339	0.389
EE_A	$d = 50$	0.413	0.283
	$d = 100$	0.396	0.241
$RMSE_e$	$d = 50$	1.319	1.025
	$d = 100$	1.230	0.856

表7：示例4中的因子调整精度矩阵和偏相关网络估计。

		tv-CLIME	
度量	维度 $n = 200$		$n = 400$
FP	$d = 50$	0.01	0.01
	$d = 100$	0	0.02
FN	$d = 50$	38.22	5.36
	$d = 100$	65.99	2.21
TPR	$d = 50$	0.220	0.891
	$d = 100$	0.333	0.978
TNR	$d = 50$	1.000	1.000
	$d = 100$	1	1.000
PPV	$d = 50$	0.999	1.000
	$d = 100$	1	1.000
NPV	$d = 50$	0.969	0.995
	$d = 100$	0.987	1.000
F1	$d = 50$	0.349	0.941
	$d = 100$	0.496	0.989
MCC	$d = 50$	0.448	0.941
	$d = 100$	0.570	0.988
EE_Ω	$d = 50$	0.670	0.585
	$d = 100$	0.628	0.534

$i = 1$、……、127、 $t = 1$……、763，随后用于我们的实证分析。Miao、Phillips 和 Su (2022) 提出通过比率准则比较不同滞后阶数下估计的转移矩阵的弗罗贝尼乌斯范数，以确定高维 VAR 模型的最优阶数。我们将他们的准则扩展到时变 VAR 模型的情境（详情见补充材料中的附录 $\hat{E}$），随后为 $X_t = (\hat{x}_{t,1}, \ldots\ldots \hat{x}_{t,\ 127})^{\intercal}$ 选择时变 VAR(1) 模型。

图1 展示了由静态 VAR(1) 模型和时变 VAR(1) 模型估计得到的格兰杰网络。根据估计得到的时变转移矩阵，我们在格兰杰因果网络中发现了 190 条有向连接，其中 78 条为自连接，143 条是同一类别内的连接。具体而言，自连接对应于转移矩阵中显著的对角元素，这表明以下四类宏观经济变量：消费、订单与库存；利率与汇率；货币与信贷；以及价格，比其他变量更具持续性，尽管在初步分析中所有变量均已转换为平稳序列。相比之下，通过静态 VAR(1) 估计的格兰杰网络包含 155 条有向连接，因此，我们的时变 VAR(1) 模型在网络估计中捕捉到了更多连接。图2展示了未经因子调整估计得到的格兰杰网络。与因子调整后的版本相比，通过时变 VAR(1) 得到的格兰杰网络更加稠密，包含 1118 条有向连接，其中 104 条为自连接，432 条是类别内连接。正如McCracken 和 Ng（2016）所指出的，共同因子（可解释为商业周期）是宏观经济变量之间格兰杰因果关系的主要来源，从而形成相当稠密的网络结构。另一方面，未经因子调整、通过静态 VAR(1) 估计得到的格兰杰网络仅有 450 条连接。

我们进一步探讨高斯因果联系的动态平滑结构变化。以标准普尔市盈率（S&P PE ratio）的对数增长率为例[2]，该变量存在四条有向连接关系：对数货币基础（BOGMBASE）的加速度、标普500指数（S&P 500）的对数收益率、标准普尔500工业指数（S&P: indust）的对数收益率，以及标准普尔市盈率的对数增长率（这是一条自连接关系）。我们仅使用这四个选定的预测变量，通过非参数自回归模型重新估计相应的时变系数，并使用 R软件包“tvReg”绘制90%置信带。图3绘出了四个系数函数的估计曲线。我们发现，在最近二十年中，标准普尔市盈率的对数增长率总体上具有持续性，并与 BOGMBASE 呈正相关。标普500工业指数收益率的估计时变系数显著，但接近于零。因此，静态 VAR(1) 模型结合经典 LASSO 惩罚未能检测出从该变量出发的格兰杰因果关系连接，也就不足为奇了

[2]括号中列出了 FRED-MD 数据集中使用的变量名称。变量变换按照该数据集中的指南进行。

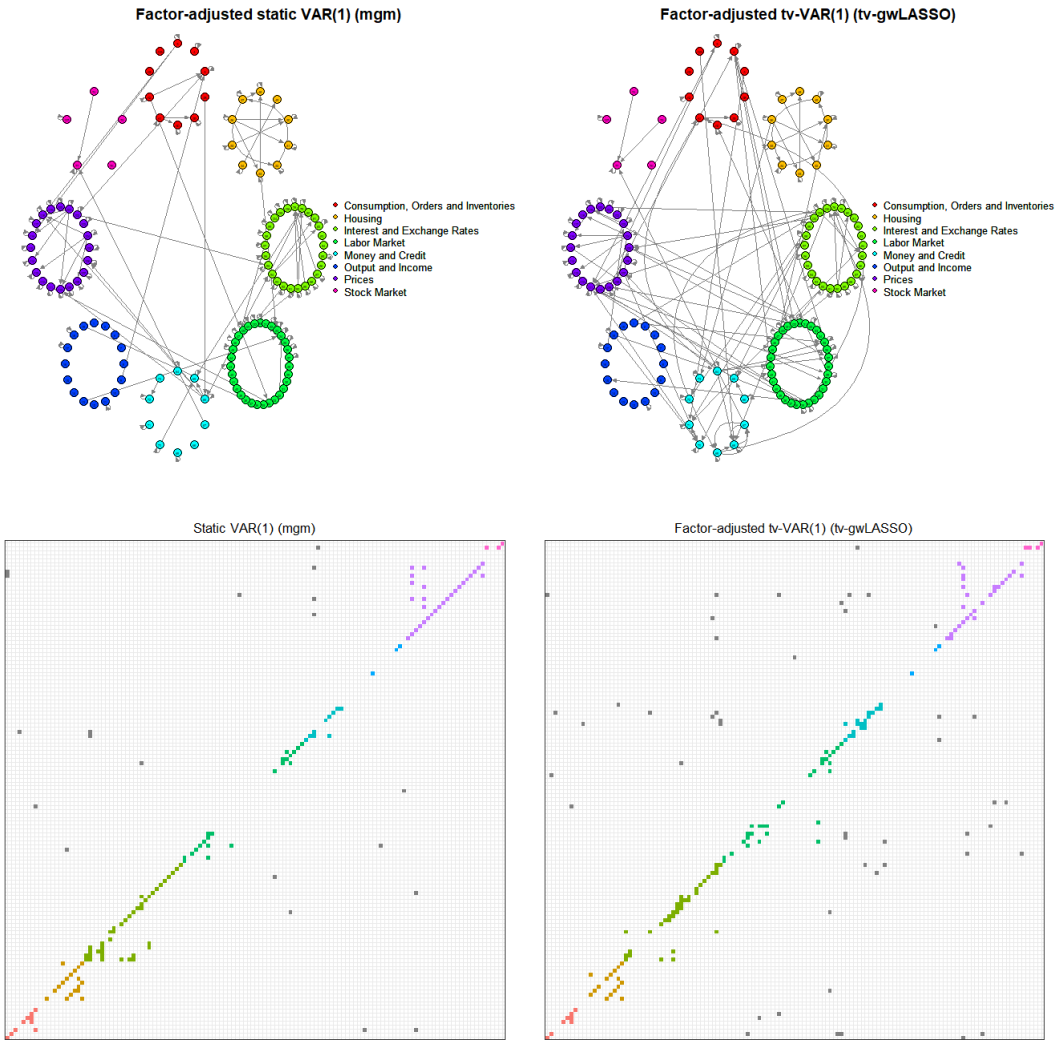


图1：使用因子调整静态 VAR(1) 模型（左）和时变 VAR(1) 模型（右）估计的格兰杰因果网络。

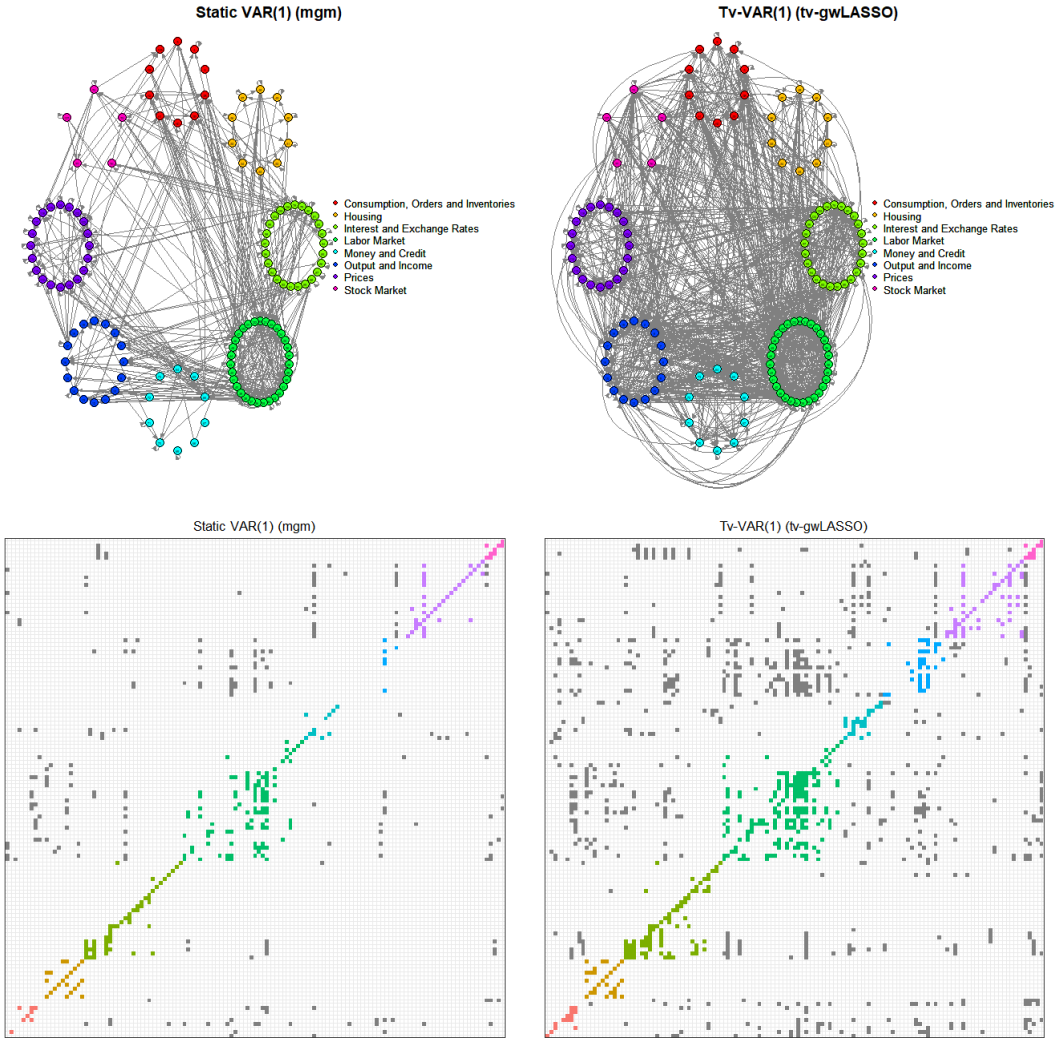


图2：未进行因子调整时，使用静态 VAR(1) 模型（左）和时变 VAR(1) 模型（右）估计的格兰杰因果网络。

变量。事实上，LASSO 往往只会从一组高度相关的预测变量中选出一个变量。由于两个指数收益率之间存在高相关性，静态 VAR(1) 模型仅选中了标普500指数收益率。相比之下，所提出的时变 LASSO 会在不同时间段选中这两个指数收益率，而第二阶段的加权组 LASSO 则汇总了各时段的信息，并选中这两个指数收益率。

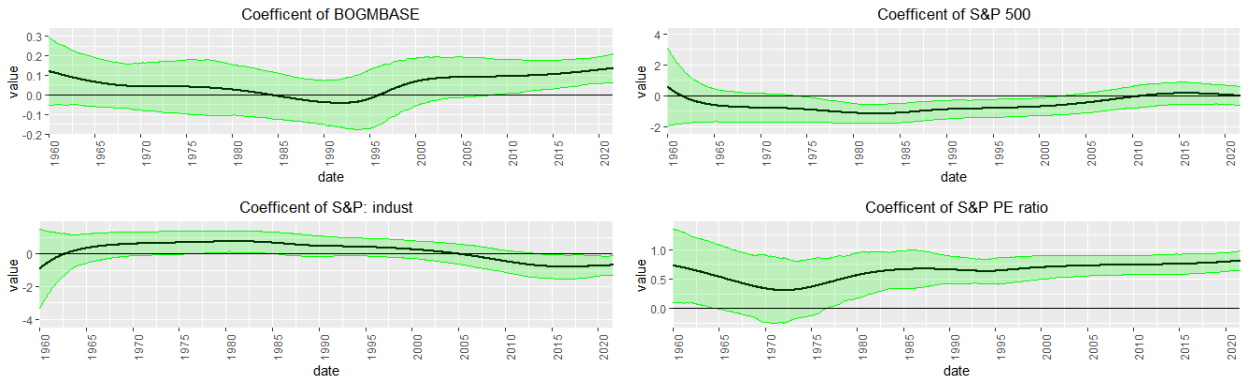


图3: 与标准普尔市盈率相连的时变系数估计值及其90%置信区间 s.

我们在图4中绘制了估计得到的偏相关网络，这些网络总体上较为稀疏。采用因子调整时变 CLIME 后，估计网络中检测到234条无向连接，其中205条位于同一类别内。相比之下，未经因子调整的估计网络包含236条连接，其中211条位于同一类别内。与格兰杰网络估计不同，是否进行因子调整似乎对偏相关网络估计影响不大。

接下来，我们考察标准普尔市盈率与其他四个变量之间偏相关连接关系的时变模式：标普500、标普工业指数、标普股息收益率（即标普综合普通股股息收益率的增量）以及 BAAFFM（穆迪评级 Baa 的长期公司债券收益率与有效联邦基金利率之间的利差）。我们使用 200个月移动窗口重新估计相关时变函数（Jankova 和 van de Geer，2015），并使用 R软件包“SILGGM”绘制图5中的 90%置信带。请注意，偏相关的符号与精度矩阵中对应元素的符号相反。我们发现，标准普尔市盈率与标普500和标普工业指数呈正（偏）相关，而与标普股息收益率呈负（偏）相关。图 5 中的置信带表明，时不变的偏相关连接关系并不适用于描述 FRED-MD 数据的网络结构。

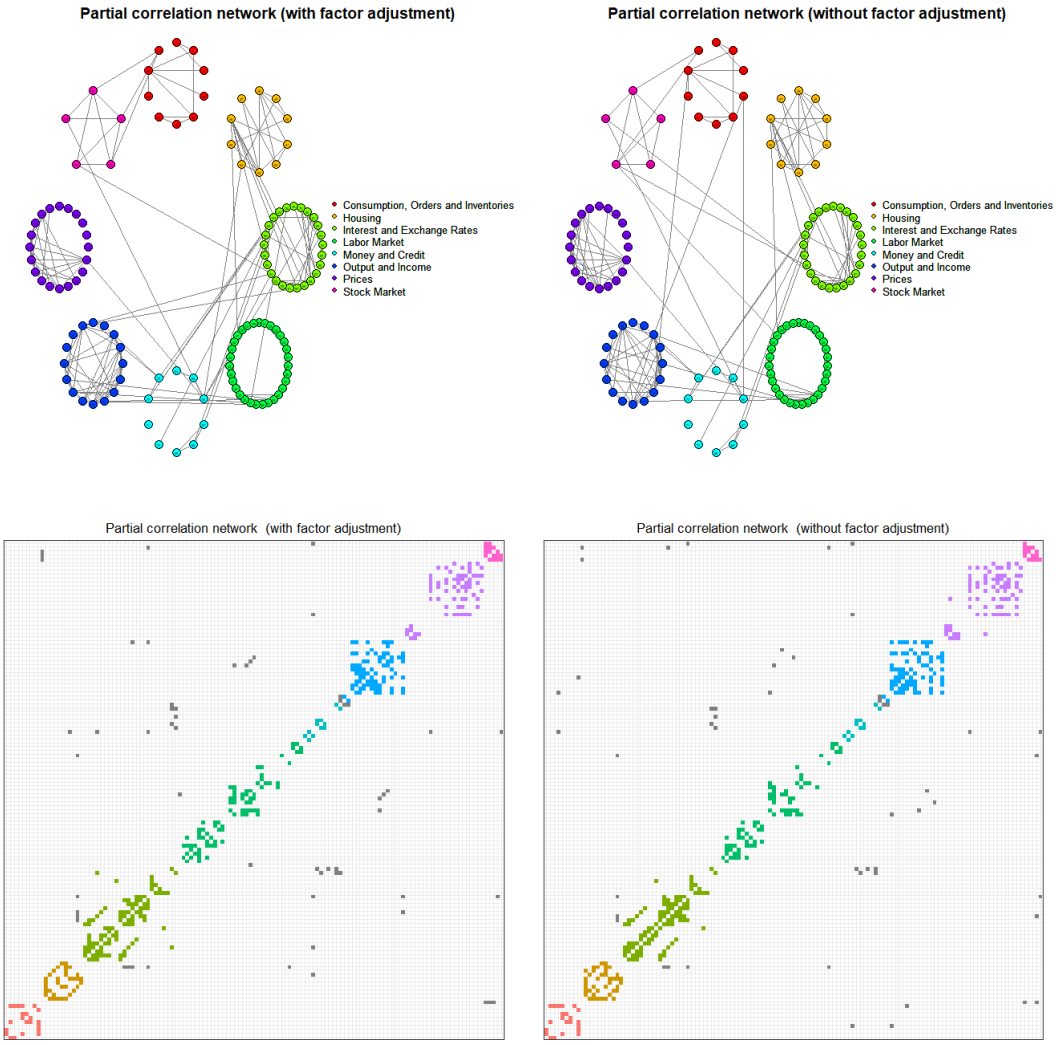


图 4：估　　计得到的偏相关网络，包含（左）和不包含（右）因子调整　　.

8 结论

本文针对高维局部平稳时间序列估计一般的时变 VAR 模型。我们提出一种结合时变 LASSO、加权组 LASSO 和时变 CLIME 的三阶段估计程序，用于估计转移矩阵和误差精度矩阵，从而允许结构随时间平滑变化。随后，我们利用估计得到的转移矩阵和精度矩阵，分别构建包含有向格兰杰因果关系和无向偏相关关系的双重网络结构。在稀疏结构假设及其他技术条件下，我们推导出所提出估计量的一致一致性和 Oracle性质。为了适应大规模时间序列之间的高相关性

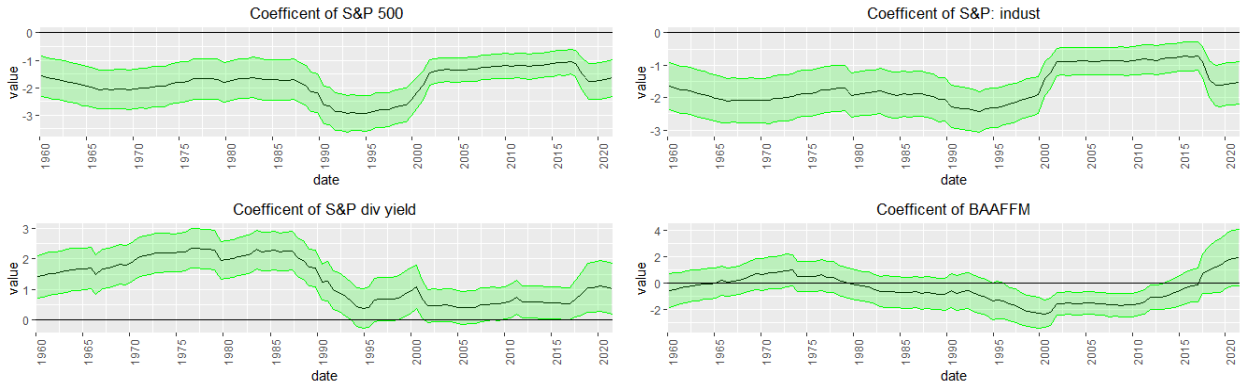


图5：与标准普尔市盈率相关的精度矩阵中时变元素的估计值及其90%置信带。

序列，并避免直接施加稀疏性假设，我们还将方法和理论扩展至更一般的因子调整时变向量自回归和网络结构。模拟和实证研究均表明，所提出的网络模型和方法在有限样本中具有可靠的数值表现。

补充材料

补充材料包含主要渐近定理的证明、若干附有证明的技术引理、对假设 3(ii) 的验证以及关于调优参数选择的讨论。

参考文献

BAI, J. AND NG, S.（2002）。确定近似因子模型中的因子数量。《计量经济学》90，191–221。

BARIGOZZI, M. 和 BROWNLEES, C. (2019). NETS：时间序列网络估计。《应用计量经济学杂志》34, 347–364。

BARIGOZZI, M., CHO, H. 和OWENS, D. (2022). FNETS：面向高维时间序列的因子调整网络估计与预测。工作论文可见于https://arxiv.org/pdf/2201.06110.pdf。

BASU, S. 和 MICHAILIDIS, G. (2015). 稀疏高维时间序列模型中的正则化估计。《统计年鉴》43, 1535–1567。

BASU, S. SHOJAIE, A. AND MICHAILIDIS, G. (2015). 具有内在分组结构的网络格兰杰因果关系。《机器学习研究杂志》16, 417–453.

BICKEL, P., RITOV, Y. AND TSYBAKOV, A. (2009). Lasso与Dantzig选择器的联合分析。《统计年鉴》, 37, 1705–1732. BURT, R. S., KILDUFF, M. AND TASSELLI, S. (2013). 社会网络分析：优势的基础与前沿。《心理学年度评论》64, 527–547. CAI, T. T., LIU, W. AND LUO, X. (2011). 一种用于稀疏精度矩阵估计的约束 ℓ_1 最小化方法。《美国统计学会会刊》106, 594–607. CAI, Z. (2007). 具有序列相关误差的趋势型时变系数时间序列模型。《计量经济学杂志》136, 163–188. CHAMBERLAIN, G. AND ROTHSCHILD, M. (1983). 大型资产市场中的套利、因子结构与均值—方差分析。《计量经济学》51, 1305–1324. CHEN, E., FAN, J. AND ZHU, X. (2020). 面向高维时间序列的社区网络自回归。工作论文，见 https://arxiv.org/abs/2007.05521. CHEN, J., LI, D., WEI, L. AND ZHANG, W. (2021). 函数系数模型中的非参数同质性追寻。《非参数统计学杂志》33, 387–416. CHENG, M., HONDA, T., LI, J. AND PENG, H. (2014). 超高维纵向数据的非参数独立性筛选与结构识别。《统计年鉴》, 42, 1819–1849. DAHLHAUS, R. (1997). 将时间序列模型拟合到非平稳过程。《统计年鉴》25, 1–37. DAHLHAUS, R. AND SUBBA RAO, S. (2006). 时变 ARCH 过程的统计推断。《统计年鉴》34, 1075–1114. DAVIS, R., ZANG, P. AND ZHENG, T. (2016). 稀疏向量自回归建模。《计算与图形统计学杂志》25, 1077–1096. DEMPSTER, A.P. (1972). 协方差选择。《生物统计学》28, 157–175. DIEBOLD, F. AND YILMAZ, K. (2014). 方差分解的网络拓扑：度量金融公司的关联性。《计量经济学杂志》182, 119–134. DIEBOLD, F. AND YILMAZ, K. (2015).《金融与宏观经济关联性：测量与监测的网络方法》。牛津大学出版社。DING, X., QIU, Z. AND CHEN, X. (2017). 高维局部平稳向量自回归模型的稀疏转移矩阵估计。《电子统计学杂志》11, 3871–3902.

FAN, J., FENG, Y. 和WU, Y. (2009). 通过自适应 Lasso 和 SCAD 惩罚进行网络探索。《应用统计年鉴》3, 521–541.FAN, J. 和GIJBELS, I. (1996). 局部多项式建模及其应用.Chapman &Hall.FAN, J., MASINI, R. 和MEDEIROS, M. (2021). 融合因子模型与稀疏模型。工作论文见https://arxiv.org/abs/2102.11341。FAN, J. 和 LI, R. (2001). 通过非凹惩罚似然进行变量选择及其 Oracle 性质。《美国统计学会会刊》96, 1348–1360. FAN, J., LIAO, Y. 和 MINCHEVA, M. (2013). 通过对主正交补进行阈值化来估计大规模协方差（附讨论）。《英国皇家统计学会杂志》*B*辑 75, 603–680. FAN, J., MA, Y. 和 DAI, W. (2014). 稀疏超高维变系数模型中的非参数独立性筛选。《美国统计学会会刊》109, 1270–1284. FAN, J., XUE, L. 和 ZOU, H. (2014). 折叠凹惩罚估计的强预言机最优性。《统计年鉴》42, 819–849. GRANGER, C. W. (1969). 利用计量经济模型和互谱方法考察因果关系。《计量经济学》37, 424–438.

HAFNER, C. AND LINTON, O. (2010). 多变量乘法波动率模型的有效估计。《计量经济学杂志》159, 55–73.

HAN, F., LU, H. AND LIU H. (2015). 高维平稳向量自回归的直接估计。《机器学习研究杂志》16, 3115–3150.

HAUTSCH, N., SCHAUMBURG, J. 和 SCHIENLE, M.（2014）。金融网络中的系统性影响预测。《国际预测杂志》30，781–794。JANKOVA, J. 和 VAN DE GEER S.（2015）。高维逆协方差估计的置信区间。《电子统计学杂志》9，1205–1229。KOCK, A.B. 和 CALLOT, L.（2015）。高维向量自回归的预言机不等式。《计量经济学杂志》186，325–344。KOLAR, M.、SONG, L.、AHMED, A. 和 XING, E.（2010）。时变网络估计。《应用统计年鉴》4，94–123。KOO, B. 和 LINTON O.（2012）。半参数局部平稳扩散模型的估计。《计量经济学杂志》170，210–233。

Krampe, J. 和 Margaritella, L.（2022）。具有稀疏VAR特质成分的因子模型。工作论文见 https://arxiv.org/pdf/2112.07149.pdf。LAM, C. 和 FAN, J.（2009）。大规模协方差矩阵估计中的稀疏性与收敛速率。《统计年鉴》37，4254–4278。LI, D.、CHEN, J. 和 GAO, J.（2011）。带固定效应的非参数时变系数面板数据模型。《计量经济学杂志》14，387–408。LI, D.、KE, Y. 和 ZHANG, W.（2015）。超高维广义半变系数模型中的模型选择与结构设定。《统计年鉴》43，2676–2705。LIAN, H.（2012）。高维广义变系数模型的变量选择。《中华统计学志》，22，1563–1588。LIU, J.、LI, R. 和 WU, R.（2014）。具有超高维协变量的变系数模型的特征选择。《美国统计学会会刊》，109，266–274。LOH, P. 和 WAINWRIGHT, M.（2013）。离散图模型的结构估计：广义协方差矩阵及其逆矩阵。《统计年鉴》41，3022–3049。LIU, L. 和 ZHANG, D.（2021）。高维向量自回归模型的稳健估计。工作论文见https://arxiv.org/abs/2109.10354。L ̈UTKEPOHL, H.（2006）。《多元时间序列分析新导论》。施普林格。MCCRACKEN, M.W. 和 NG, S.（2016）。FRED-MD：用于宏观经济研究的月度数据库。《商业与经济统计杂志》34，574–589。MCCRACKEN, M.W. 和 NG, S.（2020）。FRED-QD：用于宏观经济研究的季度数据库。工作论文见https://www.nber.org/papers/w26872。MIAO, K.、PHILLIPS, P.C.B. 和 SU, L.（2022）。具有共同因子的高维VAR。即将发表于《计量经济学杂志》。Motta, G.、Hafner, C. 和 von Sachs, R.（2011）。局部平稳因子模型：识别与非参数估计。《计量经济学理论》27，1279–1319。NEWMAN, M. E. J.（2002）。网络中流行病的传播。《物理评论*E*辑》66，016128。SAFIKHANI, A. 和 SHOJAIE, A.（2022）。高维非平稳向量自回归模型中的联合结构突变检测与参数估计。《美国统计学会会刊》117，251–264。SCOTT, J.（2017）。《社会网络分析》（第4版）。Sage，伦敦。

Stock, J. H. 和 Watson, M. W.（2002）。利用大量预测变量的主成分进行预测。《美国统计学会会刊》97，1167–1179。SU, L. AND WANG, X.（2017）。关于时变因子模型：估计与检验。《计量经济学杂志》198，84–101。TIBSHIRANI, R. J.（1996）。通过 LASSO 进行回归收缩与变量选择。英国皇家统计学会会刊B辑 58，267–288。VOGT, M.（2012）。局部平稳时间序列的非参数回归。《统计年鉴》40，2601–2633。WAINWRIGHT, M. J.（2019）。《高维统计：非渐近视角》。剑桥统计与概率数学丛书。WANG, L., LI, H. AND HUANG, J.（2008）。用于重复测量分析的非参数变系数模型中的变量选择。《美国统计学会会刊》103，1556–1569。WANG, H. AND XIA, Y.（2009）。变系数模型的收缩估计。《美国统计学会会刊》104，747–757。WANG, D., YU, Y. AND RINALDO, A.（2021）。稀疏动态网络中的最优变点检测与定位。《统计年鉴》49，203–232。XU, M., CHEN, X. AND WU, W.（2020）。高维非平稳时间序列的动态网络估计。熵 22，55。YAN, Y., GAO, J. AND PENG, B.（2020）。一类时变向量移动平均 (∞) 模型。工作论文，见 https://arxiv.org/abs/2010.01492。YUAN, M.（2010）。通过线性规划进行高维逆协方差矩阵估计。《机器学习研究杂志》11，2261–2286。YUAN, M. AND LIN, Y.（2007）。高斯图模型中的模型选择与估计。《生物统计学》94，19–35。ZHANG, T. AND WU, W. B.（2012）。时变回归模型的推断。《统计年鉴》40，1376–1402。ZHANG, D. AND WU, W.（2021）。高维局部平稳过程的协方差与谱密度估计量的收敛性。《统计年鉴》49，233–254。ZHAO, J., LIU, X., WANG, H. AND LENG, C.（2022）。网络数据中协变量的降维。《生物统计学》109，85–102。ZHOU, S., LAFFERTY, J. AND WASSERMAN, L.（2010）。时变无向图。机器学习 80，295–319。

ZHU, X., CHANG, X., LI, R. AND WANG, H.（2019）。大规模社交网络的门户节点筛选。《计量经济学杂志》209，145–157。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。ZOU, H. AND LI, R.（2008）。非凹惩罚似然模型中的单步稀疏估计（附讨论）。《统计年鉴》，36，1509–1566。

《高维时间序列的时变网络估计》补充材料

附录 A：定理4.1–4.3的证明

定理 4.1 的证明。本证明所采用的主要思路与Bickel、Ritov 和 Tsybakov (2009)、Lian (2012) 以及 Li, Ke and Zhang (2015) 的研究类似，他们研究的是经典独立性假设下的高维数据。在以下证明中，我们需要利用时变 VAR 的核加权量的一致收敛性质（例如附录 B 中的引理 B.3 ）。事实上，接下来我们将证明（4.4）的一个强化版本，其中还包含导数函数估计的一致一致性：

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant t\leqslant n}\left(\|\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau_t)-\boldsymbol{\alpha}_{i\bullet}(\tau_t)\|+h\left\|\widetilde{\boldsymbol{\alpha}}'_{i\bullet}(\tau_t)-\boldsymbol{\alpha}'_{i\bullet}(\tau_t)\right\|\right)=O_P\left(\sqrt{s}\lambda_1\right). \tag{A.1}$$

由于我们仅考虑时变 VAR(1) 模型，

$$\boldsymbol{\alpha}_{i\bullet}(\tau_t)=[\alpha_{i,1}(\tau_t),\alpha_{i,2}(\tau_t),\cdots,\alpha_{i,d}(\tau_t)]^\intercal \text{ and } \boldsymbol{\alpha}'_{i\bullet}(\tau_t)=\left[\alpha'_{i,1}(\tau_t),\alpha'_{i,2}(\tau_t),\cdots,\alpha'_{i,d}(\tau_t)\right]^\intercal.$$

回顾 $\mathcal{J}_i(\tau_t)=\{j:\ \alpha_{i,j}(\tau_t)\neq 0\}$ 并定义 $\mathcal{J}'_i(\tau_t)=\left\{j:\ \alpha'_{i,j}(\tau_t)\neq 0\right\}$。我们首先证明，对于任意 $i=1、\cdots、d$ 和 $t=1、\cdots、n$,

$$\sum_{j\notin\mathcal{J}_i(\tau_t)}|\delta_{i,j}(\tau_t)|+\sum_{j\notin\mathcal{J}'_i(\tau_t)}|\delta'_{i,j}(\tau_t)|\leqslant 2\left(\sum_{j\in\mathcal{J}_i(\tau_t)}|\delta_{i,j}(\tau_t)|+\sum_{j\in\mathcal{J}'_i(\tau_t)}|\delta'_{i,j}(\tau_t)|\right), \tag{A.2}$$

其中 $\delta_{i,\ j}(\tau_t)=\widetilde{\alpha}_{i,\ j}(\tau_t)-\alpha_{i,\ j}(\tau_t)$ 以及 $\delta'_{i,\ j}(\tau_t)=h\left[\widetilde{\alpha}'_{i,j}(\tau_t)-\alpha'_{i,j}(\tau_t)\right]$。

根据初步时变LASSO的定义，我们有

$$\mathcal{L}_i^*\left(\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau_t),\widetilde{\boldsymbol{\alpha}}'_{i\bullet}(\tau_t)\mid\tau_t\right)\leqslant\mathcal{L}_i^*\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t),\boldsymbol{\alpha}'_{i\bullet}(\tau_t)\mid\tau_t\right)$$

对于任意的 $i=1、\cdots、d$ 和 $t=1、\cdots、n$，其中 $\mathcal{L}_i^*(\boldsymbol{\alpha},\boldsymbol{\beta}\mid\tau_t)$ 定义于式 (3.4) 中。于是，我们很容易得到

$$\begin{aligned}&\mathcal{L}_i\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t),\boldsymbol{\alpha}'_{i\bullet}(\tau_t)\mid\tau_t\right)-\mathcal{L}_i\left(\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau_t),\widetilde{\boldsymbol{\alpha}}'_{i\bullet}(\tau_t)\mid\tau_t\right)\\&\geqslant\lambda_1\left[\sum_{j=1}^d|\widetilde{\alpha}_{i,j}(\tau_t)|+h\sum_{j=1}^d|\widetilde{\alpha}'_{i,j}(\tau_t)|-\sum_{j=1}^d|\alpha_{i,j}(\tau_t)|-h\sum_{j=1}^d|\alpha'_{i,j}(\tau_t)|\right].\end{aligned} \tag{A.3}$$

Let

$$\delta_i(\tau_t) = [\delta_{i,1}(\tau_t), \cdots, \delta_{i,d}(\tau_t)]^\tau \quad \text{and} \quad \delta_i'(\tau_t) = \left[\delta_{i,1}'(\tau_t), \cdots, \delta_{i,d}'(\tau_t)\right]^\tau.$$

注意，

$$\begin{aligned}
& \mathcal{L}_i\left(\alpha_{i\bullet}(\tau_t), \alpha_{i\bullet}'(\tau_t) \mid \tau_t\right) - \mathcal{L}_i\left(\widetilde{\alpha}_{i\bullet}(\tau_t), \widetilde{\alpha}_{i\bullet}'(\tau_t) \mid \tau_t\right) \\
= \; & 2\left[L_{i,0}^\tau(\tau_t)\delta_i(\tau_t) + L_{i,1}^\tau(\tau_t)\delta_i'(\tau_t)\right] - \frac{1}{n}\sum_{s=1}^{n}\left\{\left[\delta_i(\tau_t) + \delta_i'(\tau_t)\left(\frac{\tau_s - \tau_t}{h}\right)\right]^\tau X_{s-1}\right\}^2 K_h(\tau_s - \tau_t) \\
\leqslant \; & 2\left[L_{i,0}^\tau(\tau_t)\delta_i(\tau_t) + L_{i,1}^\tau(\tau_t)\delta_i'(\tau_t)\right], \qquad (A.4)
\end{aligned}$$

其中$L_{i,\,0}(\tau_t)$和$L_{i,\,1}(\tau_t)$在附录 B 中定义。根据引理B.3，我们可以证明

$$\left|L_{i,0}^\tau(\tau_t)\delta_i(\tau_t) + L_{i,1}^\tau(\tau_t)\delta_i'(\tau_t)\right| \leqslant O_P\left(\zeta_{n,d}\right) \cdot \left(\sum_{j=1}^{d}|\delta_{i,j}(\tau_t)| + \sum_{j=1}^{d}|\delta_{i,j}'(\tau_t)|\right) \qquad (A.5)$$

对 $i = 1$、$\cdots$、d 以及 $t = 1$、$\cdots$、n 一致成立。

另一方面，根据三角不等式，我们可以证明

$$\begin{aligned}
& \lambda_1\left[\sum_{j=1}^{d}|\widetilde{\alpha}_{i,j}(\tau_t)| + h\sum_{j=1}^{d}|\widetilde{\alpha}_{i,j}'(\tau_t)| - \sum_{j=1}^{d}|\alpha_{i,j}(\tau_t)| - h\sum_{j=1}^{d}|\alpha_{i,j}'(\tau_t)|\right] \\
= \; & \lambda_1\left[\sum_{j\in\mathcal{J}_i(\tau_t)}\left(|\widetilde{\alpha}_{i,j}(\tau_t)| - |\alpha_{i,j}(\tau_t)|\right) + h\sum_{j\in\mathcal{J}_i'(\tau_t)}\left(|\widetilde{\alpha}_{i,j}'(\tau_t)| - |\alpha_{i,j}'(\tau_t)|\right)\right] + \\
& \lambda_1\left[\sum_{j\notin\mathcal{J}_i(\tau_t)}|\widetilde{\alpha}_{i,j}(\tau_t)| + h\sum_{j\notin\mathcal{J}_i'(\tau_t)}|\widetilde{\alpha}_{i,j}'(\tau_t)|\right] \\
\geqslant \; & -\lambda_1\left(\sum_{j\in\mathcal{J}_i(\tau_t)}|\delta_{i,j}(\tau_t)| + \sum_{j\in\mathcal{J}_i'(\tau_t)}|\delta_{i,j}'(\tau_t)|\right) + \lambda_1\left(\sum_{j\notin\mathcal{J}_i(\tau_t)}|\delta_{i,j}(\tau_t)| + \sum_{j\notin\mathcal{J}_i'(\tau_t)}|\delta_{i,j}'(\tau_t)|\right) \qquad (A.6)
\end{aligned}$$

由（A.3）–（A.6）以及假设$\zeta_{n,\,d} = o(\lambda_1)$中的假设 3(i)，我们完成了（A.2）的证明。令 $u_1 = (u_{1,1}, \cdots, u_{1,d})$和$u_2 =^\tau (u_{2,1}, \cdots, u_{2,d})$为两个 d维向量，并且

$$\mathcal{B}_i(\tau_t; M) = \left\{u = \left(u_1^\tau, u_2^\tau\right)^\tau : \|u_1\|^2 + \|u_2\|^2 = M, \sum_{j=1}^{d}\left(|u_{1,j}| + |u_{2,j}|\right) \leqslant 3\left(\sum_{j\in\mathcal{J}_i(\tau_t)}|u_{1,j}| + \sum_{j\in\mathcal{J}_i'(\tau_t)}|u_{2,j}|\right)\right\},$$

其中 M 是一个可以充分大的正常数。注意，对于任意的 $i = 1$、$\cdots$、d,

$t=1$、$\cdots$、n 和 $u\in\mathcal{B}_i(\tau_t;\ M)$,

$$\mathcal{L}_i^*\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t)+\sqrt{s}\lambda_1 u_1,\boldsymbol{\alpha}'_{i\bullet}(\tau_t)+\sqrt{s}\lambda_1 u_2/h\mid\tau_t\right)-\mathcal{L}_i^*\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t),\boldsymbol{\alpha}'_{i\bullet}(\tau_t)\mid\tau_t\right)=\sum_{k=1}^{3}\Xi_{i,k}(\tau_t),\quad(A.7)$$

其中

$$\begin{aligned}
\Xi_{i,1}(\tau_t) &= \mathcal{L}_i\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t)+\sqrt{s}\lambda_1 u_1,\boldsymbol{\alpha}_{i\bullet}(\tau_t)+\sqrt{s}\lambda_1 u_2/h\mid\tau_t\right)-\mathcal{L}_i\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t),\boldsymbol{\alpha}'_{i\bullet}(\tau_t)\mid\tau_t\right),\\
\Xi_{i,2}(\tau_t) &= \lambda_1\left(\sum_{j=1}^{d}|\alpha_{i,j}(\tau_t)+\sqrt{s}\lambda_1 u_{1,j}|-\sum_{j=1}^{d}|\alpha_{i,j}(\tau_t)|\right),\\
\Xi_{i,3}(\tau_t) &= \lambda_1\left(\sum_{j=1}^{d}|h\alpha'_{i,j}(\tau_t)+\sqrt{s}\lambda_1 u_{2,j}|-\sum_{j=1}^{d}|h\alpha'_{i,j}(\tau_t)|\right).
\end{aligned}$$

对于$\Xi_{i,\ 1}(\tau_t)$，它可写为

$$\Xi_{i,1}(\tau_t)=-2\sqrt{s}\lambda_1 u^{\intercal}L_i(\tau_t)+s\lambda_1^2 u^{\intercal}\boldsymbol{\Psi}(\tau_t)u,\quad(A.8)$$

其中 $L_i(\tau)=\left[L_{i,\theta}(\tau),L_{i,\gamma}(\tau)\right]^{\intercal}$ 且 $\boldsymbol{\Psi}(\tau)$ 在 (4.2) 中定义。根据 $\mathcal{B}_i(\tau_t;\ M)$，引理B.3 和柯西-施瓦茨不等式，我们有

$$\max_{1\leqslant i\leqslant d}\left|\sqrt{s}\lambda_1 u^{\intercal}L_i(\tau_t)\right|=o_P\left(s\lambda_1^2\right)\cdot\|u\|.\quad(A.9)$$

根据 (A.8)、(A.9以及一致受限特征值条件 (4.3)，当 n足够大且 M选得足够大时，我们有

$$\min_{1\leqslant i\leqslant d}\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}_i(\tau_t;M)}u^{\intercal}\Xi_{i,1}(\tau_t)=s\lambda_1^2 u^{\intercal}\boldsymbol{\Psi}(\tau_t)u(1+o_P(1))>\frac{1}{2}\kappa_0 s\lambda_1^2\|u\|^2,\ w.p.a.1.\quad(A.10)$$

接下来，我们考虑$\Xi_{i,2}(\tau_t)$ 和 $\Xi_{i,3}(\tau_t)$。容易证明,

$$\begin{aligned}
\Xi_{i,2}(\tau_t) &= \lambda_1\left(\sum_{j=1}^{d}|\alpha_{i,j}(\tau_t)+\sqrt{s}\lambda_1 u_{1,j}|-\sum_{j=1}^{d}|\alpha_{i,j}(\tau_t)|\right)\\
&= \lambda_1\sum_{j\in\mathcal{J}_i(\tau_t)}\left[|\alpha_{i,j}(\tau_t)+\sqrt{s}\lambda_1 u_{1,j}|-|\alpha_{i,j}(\tau_t)|\right]+\lambda_1\sum_{j\notin\mathcal{J}_i(\tau_t)}|\sqrt{s}\lambda_1 u_{1,j}|\\
&= O\left(s\lambda_1^2\right)\cdot\|u_1\|+\lambda_1\sum_{j\notin\mathcal{J}_i(\tau_t)}|\sqrt{s}\lambda_1 u_{1,j}|=O\left(s\lambda_1^2\right)\cdot\|u_1\|,\quad(A.11)
\end{aligned}$$

类似地，

$$\Xi_{i,3}(\tau_t) = O\left(s\lambda_1^2\right) \cdot \|u_2\| + \lambda_1 \sum_{j \notin \mathcal{J}_i'(\tau_t)} |\sqrt{s}\lambda_1 u_{2,j}| = O\left(s\lambda_1^2\right) \cdot \|u_2\|, \tag{A.12}$$

在 $i = 1, \cdots, d$ 和 $t = 1, \cdots, n$ 上一致成立。

结合（A.7）和（A.10）–（A.12），令 M 足够大，我们可以证明

$$\mathcal{L}_i^*\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t) + \sqrt{s}\lambda_1 u_1, \boldsymbol{\alpha}_{i\bullet}'(\tau_t) + \sqrt{s}\lambda_1 u_2/h \mid \tau_t\right) - \mathcal{L}_i^*\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t), \boldsymbol{\alpha}_{i\bullet}'(\tau_t) \mid \tau_t\right)$$

在 $i = 1, \cdots, d$ 和 $t = 1, \cdots, n$ 上一致为正。因此，我们可以在 $\mathcal{L}_i^*(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau_t)$ 的内部找到一个局部极小值点，记为 $\left[\widetilde{\boldsymbol{\alpha}}_{i\bullet}(\tau_t), h\widetilde{\boldsymbol{\alpha}}_{i\bullet}'(\tau_t)\right]$，

$$\left\{\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t) + \sqrt{s}\lambda_1 u_1, h\boldsymbol{\alpha}_{i\bullet}'(\tau_t) + \sqrt{s}\lambda_1 u_2\right) : u \in \mathcal{B}_i(\tau_t; M)\right\},$$

它与（A.2）一起完成了（A.1）的证明。 ■

定理 4.2 的证明。定义

$$\begin{aligned}
\mathbf{L}_{i,j}^{\alpha} &= \left[l_{i,j}^{\alpha}(\boldsymbol{\alpha}_{\bullet 1}, \boldsymbol{\beta}_{\bullet 1} \mid \tau_1), \cdots, l_{i,j}^{\alpha}(\boldsymbol{\alpha}_{\bullet n}, \boldsymbol{\beta}_{\bullet n} \mid \tau_n)\right]^{\tau}, \\
\mathbf{L}_{i,j}^{\beta} &= \left[l_{i,j}^{\beta}(\boldsymbol{\alpha}_{\bullet 1}, \boldsymbol{\beta}_{\bullet 1} \mid \tau_1), \cdots, l_{i,j}^{\beta}(\boldsymbol{\alpha}_{\bullet n}, \boldsymbol{\beta}_{\bullet n} \mid \tau_n)\right]^{\tau}, \\
\mathbf{P}_{i,j}^{\alpha} &= \left[p_{\lambda_2}'\left(\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|\right) \frac{\alpha_{j|1}}{\|\boldsymbol{\alpha}_j\|}, \cdots, p_{\lambda_2}'\left(\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|\right) \frac{\alpha_{j|n}}{\|\boldsymbol{\alpha}_j\|}\right]^{\tau}, \\
\mathbf{P}_{i,j}^{\beta} &= \left[p_{\lambda_2}'\left(\widetilde{D}_{i,j}\right) \frac{\beta_{j|1}}{\|\boldsymbol{\beta}_j\|}, \cdots, p_{\lambda_2}'\left(\widetilde{D}_{i,j}\right) \frac{\beta_{j|n}}{\|\boldsymbol{\beta}_j\|}\right]^{\tau},
\end{aligned}$$

其中

$$\begin{aligned}
l_{i,j}^{\alpha}(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau) &= \frac{1}{n}\sum_{t=1}^{n}\left\{x_{t,i} - [\boldsymbol{\alpha} + \boldsymbol{\beta}(\tau_t - \tau)]^{\tau} X_{t-1}\right\} x_{t-1,j} K_h(\tau_t - \tau), \\
l_{i,j}^{\beta}(\boldsymbol{\alpha}, \boldsymbol{\beta} \mid \tau) &= \frac{1}{n}\sum_{t=1}^{n}\left\{x_{t,i} - [\boldsymbol{\alpha} + \boldsymbol{\beta}(\tau_t - \tau)]^{\tau} X_{t-1}\right\} x_{t-1,j}\left(\frac{\tau_t - \tau}{h}\right) K_h(\tau_t - \tau).
\end{aligned}$$

根据 KKT 条件（例如，Fan and Lv，2011；Fan, Xue and Zou，2014；Li, Ke and Zhang，2015），预言估计 $\left(\widehat{\mathbf{A}}_i^o, \widehat{\mathbf{B}}_i^o\right)$ 是目标函数 $Q_i(\mathbf{A}, \mathbf{B})$ 的唯一极小值

$$\mathbf{L}_{i,j}^{\alpha} - \mathbf{P}_{i,j}^{\alpha} = \mathbf{0}_n \text{ for } j \in \mathcal{J}_i, \quad \mathbf{L}_{i,j}^{\beta} - \mathbf{P}_{i,j}^{\beta} = \mathbf{0}_n \text{ for } j \in \mathcal{J}_i', \tag{A.13}$$

$$\max_{j \in \overline{\mathcal{J}}_i}\left\|\mathbf{L}_{i,j}^{\alpha}\right\| < \min_{j \in \overline{\mathcal{J}}_i} p_{\lambda_2}'\left(\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|\right), \quad \max_{j \in \overline{\mathcal{J}}_i}\left\|\mathbf{L}_{i,j}^{\beta}\right\| < \min_{j \in \overline{\mathcal{J}}_i} p_{\lambda_2}'\left(\widetilde{D}_{i,j}\right), \tag{A.14}$$

$\widehat{\quad}$在 $\mathbf{A} = \mathbf{A}_i^o$和 $\mathbf{B} = \mathbf{B}_i^o$处成立，其中 $\mathbf{0}_n$ 是一个 n 维零向量。

$\widehat{\quad}$注意到，根据预言估计 A.13的定义，等式($\mathbf{A}_i^o\widehat{\quad}$和 $\mathbf{B}_i^o$) 自动成立。接下来只需证明(A.14)。下面仅证明(A.14)中的第一个结论，因为第二个结论的证明类似。根据定理 4.1 以及假设 4(i)中 $(ns)^{1/2}\lambda_1 = o(\lambda_2)$ 的条件，我们可以证明 $\min_{j\in\mathcal{J}_i} p'_{\lambda_2}(\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|) = \lambda_2$ 以概率趋于 *1*。与此同时，由引理 B.3和 B.4以及假设 4(i)，我们可以证明

$$\max_{j\in\overline{\mathcal{J}}_i} \left\|\mathbf{L}_{i,j}^{\alpha}\right\| = O_P\left(\sqrt{n}s\log(n\vee d)\zeta_{n,d}\right) = o_P(\lambda_2)$$

$\widehat{\quad}$当 $\mathbf{A} = \mathbf{A}_i^o$ 且 $\mathbf{B} = \mathbf{B}_i^o$时，得到(A.14)中的第一个结论。然后，由引理 B.4可得均方收敛结果(4.8)。 ■

Pro推论 4.1 的。根据定理 4.2 和假设4(ii)，我们可以证明 t

$$P\left(\min_{(i,j)\in\mathbb{E}_n^G}\sum_{t=1}^{n}\widehat{a}_{ij}^2(\tau_t) \geqslant a_0\lambda_2 > 0\right) \to 1$$

and

$$P\left(\sum_{t=1}^{n}\widehat{a}_{ij}^2(\tau_t) = 0,\ \forall\,(i,j)\notin\mathbb{E}_n^G\right) \to 1,$$

从而得到 (4.9)。 ■

定理 4.3 的证明。由引理 B.5 （见附录 B），可得

$$\sup_{0\leqslant\tau\leqslant1}\left\|\widehat{\boldsymbol{\Sigma}}(\tau) - \boldsymbol{\Sigma}(\tau)\right\|_{max} = O_P\left(\nu_{n,d}^{\diamond} + \nu_{n,d}^{*}\right). \tag{A.15}$$

根据 (A.15)、稀疏性假设 (3.7) 以及以下不等式： $\|\mathbf{W}_1\mathbf{W}_2\|_{max} \leqslant \|\mathbf{W}_1\|_1\|\mathbf{W}_2\|_{max}$ 对于任意两个大小相同的方阵$\mathbf{W}_1$ 和$\mathbf{W}_2$,

$$\begin{aligned}
\sup_{0\leqslant\tau\leqslant1}\left\|\mathbf{I}_d - \widehat{\boldsymbol{\Sigma}}(\tau)\boldsymbol{\Omega}(\tau)\right\|_{max} &= \sup_{0\leqslant\tau\leqslant1}\left\|\boldsymbol{\Sigma}(\tau)\boldsymbol{\Omega}(\tau) - \widehat{\boldsymbol{\Sigma}}(\tau)\boldsymbol{\Omega}(\tau)\right\|_{max} \\
&\leqslant \sup_{0\leqslant\tau\leqslant1}\|\boldsymbol{\Omega}(\tau)\|_1\left\|\widehat{\boldsymbol{\Sigma}}(\tau) - \boldsymbol{\Sigma}(\tau)\right\|_{max} \\
&\leqslant C_2\sup_{0\leqslant\tau\leqslant1}\left\|\widehat{\boldsymbol{\Sigma}}(\tau) - \boldsymbol{\Sigma}(\tau)\right\|_{max} \\
&= O_P\left(\nu_{n,d}^{\diamond} + \nu_{n,d}^{*}\right),
\end{aligned} \tag{A.16}$$

其中 C_2 定义见 (3.7)。根据 (A.16)、三角不等式、假设 5(ii) 以及定义

时变 CLIME 估计，我们可直接得到

$$\begin{aligned}
&\sup_{0\leqslant\tau\leqslant1}\left\|\widehat{\boldsymbol{\Sigma}}(\tau)\left[\widetilde{\boldsymbol{\Omega}}(\tau)-\boldsymbol{\Omega}(\tau)\right]\right\|_{\max}\\
\leqslant\ &\sup_{0\leqslant\tau\leqslant1}\left\|\widehat{\boldsymbol{\Sigma}}(\tau)\widetilde{\boldsymbol{\Omega}}(\tau)-\mathbf{I}_d\right\|_{\max}+\sup_{0\leqslant\tau\leqslant1}\left\|\mathbf{I}_d-\widehat{\boldsymbol{\Sigma}}(\tau)\boldsymbol{\Omega}(\tau)\right\|_{\max}\\
\leqslant\ &\lambda_3+O_P\left(\nu^{\diamond}_{n,d}+\nu^{*}_{n,d}\right)=O_P\left(\nu^{\diamond}_{n,d}+\nu^{*}_{n,d}\right). \qquad (A.17)
\end{aligned}$$

根据Cai、Liu 和 Luo （2011）中的引理 1，$\left\|\widetilde{\boldsymbol{\Omega}}(\tau)\right\|_1\leqslant\|\boldsymbol{\Omega}(\tau)\|_1\leqslant C_2$ 在 $0\leqslant\tau\leqslant1$上一致成立。然后，由（A.16）和（A.17），我们很容易得到

$$\begin{aligned}
&\sup_{0\leqslant\tau\leqslant1}\left\|\boldsymbol{\Sigma}(\tau)\left[\widetilde{\boldsymbol{\Omega}}(\tau)-\boldsymbol{\Omega}(\tau)\right]\right\|_{\max}\\
\leqslant\ &\sup_{0\leqslant\tau\leqslant1}\left\|\widehat{\boldsymbol{\Sigma}}(\tau)\left[\widetilde{\boldsymbol{\Omega}}(\tau)-\boldsymbol{\Omega}(\tau)\right]\right\|_{\max}+\sup_{0\leqslant\tau\leqslant1}\left\|\left[\widehat{\boldsymbol{\Sigma}}(\tau)-\boldsymbol{\Sigma}(\tau)\right]\left[\widetilde{\boldsymbol{\Omega}}(\tau)-\boldsymbol{\Omega}(\tau)\right]\right\|_{\max}\\
\leqslant\ &O_P\left(\nu^{\diamond}_{n,d}+\nu^{*}_{n,d}\right)+2C_2\sup_{0\leqslant\tau\leqslant1}\left\|\widehat{\boldsymbol{\Sigma}}(\tau)-\boldsymbol{\Sigma}(\tau)\right\|_{\max}=O_P\left(\nu^{\diamond}_{n,d}+\nu^{*}_{n,d}\right). \qquad (A.18)
\end{aligned}$$

再次利用假设 $\|\boldsymbol{\Omega}(\tau)\|_1\leqslant C_2$ 和（A.18），可得

$$\begin{aligned}
\sup_{0\leqslant\tau\leqslant1}\left\|\widetilde{\boldsymbol{\Omega}}(\tau)-\boldsymbol{\Omega}(\tau)\right\|_{\max}&\leqslant\sup_{0\leqslant\tau\leqslant1}\|\boldsymbol{\Omega}(\tau)\|_1\left\|\boldsymbol{\Sigma}(\tau)\left[\widetilde{\boldsymbol{\Omega}}(\tau)-\boldsymbol{\Omega}(\tau)\right]\right\|_{\max}\\
&=O_P\left(\nu^{*}_{n,d}+\nu^{\diamond}_{n,d}\right). \qquad (A.19)
\end{aligned}$$

^由（A.19）以及（3.10）中 $\boldsymbol{\Omega}(\tau)$ 的定义，我们证明了（4.10）。

We n接下来给出（4.11）的证明。根据Cai、Liu 和 Luo （2011）中的引理 1，可得

$$\sum_{i=1}^{d}|\widehat{\omega}_{ij}(\tau)|\leqslant\sum_{i=1}^{d}|\widetilde{\omega}_{ij}(\tau)|\leqslant\sum_{i=1}^{d}|\omega_{ij}(\tau)|.$$

注意到

$$\begin{aligned}
\sum_{j=1}^{d}|\widehat{\omega}_{ij}(\tau)|\,I\left(|\widehat{\omega}_{ij}(\tau)|\leqslant\lambda_3\right)&=\sum_{j=1}^{d}|\widehat{\omega}_{ij}(\tau)|-\sum_{j=1}^{d}|\widehat{\omega}_{ij}(\tau)|\,I\left(|\widehat{\omega}_{ij}(\tau)|>\lambda_3\right)\\
&\leqslant\sum_{j=1}^{d}|\widehat{\omega}_{ij}(\tau)|-\sum_{j=1}^{d}|\omega_{ij}(\tau)|+\sum_{j=1}^{d}|\widehat{\omega}_{ij}(\tau)I\left(|\widehat{\omega}_{ij}(\tau)|>\lambda_3\right)-\omega_{ij}(\tau)|\\
&\leqslant\sum_{j=1}^{d}|\widehat{\omega}_{ij}(\tau)I\left(|\widehat{\omega}_{ij}(\tau)|>\lambda_3\right)-\omega_{ij}(\tau)|,
\end{aligned}$$

我们有

$$\begin{aligned}\sup_{0\leqslant\tau\leqslant1}\left\|\widehat{\mathbf{\Omega}}(\tau)-\mathbf{\Omega}(\tau)\right\| &\leqslant \sup_{0\leqslant\tau\leqslant1}\max_{1\leqslant i\leqslant d}\sum_{j=1}^{d}|\widehat{\omega}_{ij}(\tau)-\omega_{ij}(\tau)| \\ &\leqslant 2\sup_{0\leqslant\tau\leqslant1}\max_{1\leqslant i\leqslant d}\sum_{j=1}^{d}|\widehat{\omega}_{ij}(\tau)-\omega_{ij}(\tau)|\,I\left(|\widehat{\omega}_{ij}(\tau)|>\lambda_3\right)+ \\ &\quad 2\sup_{0\leqslant\tau\leqslant1}\max_{1\leqslant i\leqslant d}\sum_{j=1}^{d}|\omega_{ij}(\tau)|\,I\left(|\widehat{\omega}_{ij}(\tau)|\leqslant\lambda_3\right) \\ &=: \Delta_1+\Delta_2. \end{aligned} \tag{A.20}$$

定义一个事件

$$\mathcal{E}_\epsilon=\left\{\sup_{0\leqslant\tau\leqslant1}\left\|\widehat{\mathbf{\Omega}}(\tau)-\mathbf{\Omega}(\tau)\right\|_{max}\leqslant c_\epsilon\left(\nu^{\diamond}_{n,d}+\nu^{*}_{n,d}\right)\right\},$$

其中 c_ϵ 是一个 正常数，使得对于任意 $\epsilon>0$，$P(\mathcal{E}_\epsilon)\geqslant1-\epsilon$。给定 $\mathcal{E}_\epsilon$ 条件下 ，

$$\Delta_1\leqslant c_\epsilon(\nu^{\diamond}_{n,d}+\nu^{*}_{n,d})\sup_{0\leqslant\tau\leqslant1}\left[\max_{1\leqslant i\leqslant d}\sum_{j=1}^{d}I\left(|\widehat{\omega}_{ij}(\tau)|>\lambda_3\right)\right]. \tag{A.21}$$

注意，在 $\mathcal{E}$上，

$$|\widehat{\omega}_{ij}(\tau)|\leqslant|\omega_{ij}(\tau)|+|\widehat{\omega}_{ij}(\tau)-\omega_{ij}(\tau)|\leqslant|\omega_{ij}(\tau)|+c_\epsilon\left(\nu^{\diamond}_{n,d}+\nu^{*}_{n,d}\right).$$

在 $C_3=2c_\epsilon$ 中取假设 5(ii)，事件 $\{|\widehat{\omega}_{ij}(\tau)|>\lambda_3\}$ 意味着 $\left\{|\omega_{ij}(\tau)|>c_\epsilon\left(\nu^{\diamond}_{n,d}+\nu^{*}_{n,d}\right)\right\}$成立。然后，根据 (3.7) 和 (A.21)，我们可以证明，在 $\mathcal{E}_\epsilon$上，

$$\begin{aligned}\Delta_1 &\leqslant c_\epsilon\left(\nu^{\diamond}_{n,d}+\nu^{*}_{n,d}\right)\left[\sup_{0\leqslant\tau\leqslant1}\max_{1\leqslant i\leqslant d}\sum_{j=1}^{d}I\left(|\omega_{ij}(\tau)|>c_\epsilon\left(\nu^{\diamond}_{n,d}+\nu^{*}_{n,d}\right)\right)\right] \\ &\leqslant c_\epsilon\left(\nu^{\diamond}_{n,d}+\nu^{*}_{n,d}\right)\left[\sup_{0\leqslant\tau\leqslant1}\max_{1\leqslant i\leqslant d}\sum_{j=1}^{d}\frac{|\omega_{ij}(\tau)|^q}{c_\epsilon^q\left(\nu^{\diamond}_{n,d}+\nu^{*}_{n,d}\right)^q}\right] \\ &= O_P\left(\xi_d\cdot\left(\nu^{\diamond}_{n,d}+\nu^{*}_{n,d}\right)^{1-q}\right). \end{aligned} \tag{A.22}$$

另一方面，根据三角不等式，

$$|\widehat{\omega}_{ij}(\tau)|\geqslant|\omega_{ij}(\tau)|-|\widehat{\omega}_{ij}(\tau)-\omega_{ij}(\tau)|\geqslant|\omega_{ij}(\tau)|-c_\epsilon\left(\nu^{\diamond}_{n,d}+\nu^{*}_{n,d}\right)$$

在 $\mathcal{E}_\epsilon$ 上。因此，我们很容易证明 $\{|\widehat{\omega}_{ij}(\tau)| \leqslant \lambda_3\}$ 表明 $\left\{|\omega_{ij}(\tau)| \leqslant 3c_\epsilon \left(\nu^{\diamond}_{n,d} + \nu^{*}_{n,d}\right)\right\}$。然后，再次由 (3.7) 可得

$$\begin{aligned}
\Delta_2 &\leqslant \sup_{0\leqslant\tau\leqslant 1} \max_{1\leqslant i\leqslant d} \sum_{j=1}^{d} |\omega_{ij}(\tau)| \, I\left(|\omega_{ij}(\tau)| \leqslant 3c_\epsilon \left(\nu^{\diamond}_{n,d} + \nu^{*}_{n,d}\right)\right) \\
&\leqslant (3c_\epsilon)^{1-q} \left(\nu^{\diamond}_{n,d} + \nu^{*}_{n,d}\right)^{1-q} \sup_{0\leqslant\tau\leqslant 1} \max_{1\leqslant i\leqslant d} \sum_{j=1}^{d} |\omega_{ij}(\tau)|^q \\
&= O_P\left(\xi_d \left(\nu^{\diamond}_{n,d} + \nu^{*}_{n,d}\right)^{1-q}\right). \qquad \text{(A.23)}
\end{aligned}$$

(4.11) 的证明可由 (A.20)、(A.22) 和 (A.23) 完成。

按照 (4.11) 的证明，我们还可得到

$$\sup_{0\leqslant\tau\leqslant 1} \left\|\widehat{\mathbf{\Omega}}(\tau) - \mathbf{\Omega}(\tau)\right\|_1 = O_P\left(\xi_d \left(\nu^{\diamond}_{n,d} + \nu^{*}_{n,d}\right)^{1-q}\right),$$

结合以下不等式：

$$\frac{1}{d}\left\|\widehat{\mathbf{\Omega}}(\tau) - \mathbf{\Omega}(\tau)\right\|_F^2 \leqslant \left\|\widehat{\mathbf{\Omega}}(\tau) - \mathbf{\Omega}(\tau)\right\|_{max} \left\|\widehat{\mathbf{\Omega}}(\tau) - \mathbf{\Omega}(\tau)\right\|_1,$$

可得 (4.12)。定理 4.3 的证明至此完成。 ■

推论 4.2 的证明。由定理 4.3 中的 (4.10) 以及 $\min_{(i,j)\in\mathbb{E}^P} \min_{1\leqslant t\leqslant n} |\omega_{ij}(\tau_t)| \gg \lambda_3$ 的条件，我们有

$$P\left(\min_{(i,j)\in\mathbb{E}_n^P} \min_{1\leqslant t\leqslant n} |\widehat{\omega}_{ij}(\tau_t)| \geqslant \lambda_3 > 0\right) \to 1. \qquad \text{(A.24)}$$

令 $\mathcal{E}_\epsilon$ 和 c_ϵ 如定理 4.3 的证明中所定义，并在假设 5(ii) 中取 $C_3 = 2c_\epsilon$，我们可以证明

$$\max_{(i,j)\notin\mathbb{E}_n^P} \max_{1\leqslant t\leqslant n} |\widehat{\omega}_{ij}(\tau_t)| \leqslant c_\epsilon(\nu^{*}_{n,d} + \nu^{\diamond}_{n,d}) < \lambda_3 \qquad \text{(A.25)}$$

在.....条件下 $\mathcal{E}_\epsilon$。根据（A.24）和（A.25），令 $\epsilon \to 0$，即可证明（4.13）。 ■

附录 B：技术引理

在本附录中，我们给出一些技术性引理，这些引理对于附录 A 中主要理论结果的证明至关重要。一般性不失，我们聚焦于时变 VAR(1) 模型

框架。在整个证明过程中，我们用 M 表示一个通用的正常数，其值可能随行而变化。

引理 B.1. 假设假设 *1* 成立。设

$$\iota_2 = \iota_1/C_*,\ \ \iota_3 = \iota_1(1-\rho)/(C_1^2 C_*),\ \ C_* = \max_{1\leqslant t\leqslant n} \|\mathbf{\Sigma}_t\| < \infty$$

其中 ι_1 和 ρ 由假设 *1* 定义， C_1 由式 *(2.4)* 定义。对于任意满足 $\|u\| = 1$的 d维向量 u，

$$\max_{1\leqslant t\leqslant n} \mathsf{E}\left[\exp\left\{\iota_2 \left(u^\tau e_t\right)^2\right\}\right] \leqslant C_0 < \infty, \tag{B.1}$$

and

$$\max_{1\leqslant t\leqslant n} \max_{1\leqslant i\leqslant d} \mathsf{E}\left[\exp\left\{\iota_3 x_{t,i}^2\right\}\right] \leqslant C_0^{1/(1-\rho)} < \infty, \tag{B.2}$$

其中 C_0 是由假设 *1(iii)*定义的正常数。

引理 B.1的证明。写作 $u_t = u\ \mathbf{\Sigma}_t^{1/2}$ 并利用假设 1(ii)(iii)，可得

$$\begin{aligned}
\max_{1\leqslant t\leqslant n} \mathsf{E}\left[\exp\left\{\iota_2 \left(u^\tau e_t\right)^2\right\}\right] &= \max_{1\leqslant t\leqslant n} \mathsf{E}\left[\exp\left\{\iota_2 \left(u^\tau \mathbf{\Sigma}_t^{1/2}\varepsilon_t\right)^2\right\}\right] \\
&= \max_{1\leqslant t\leqslant n} \mathsf{E}\left[\exp\left\{\iota_2 \|u_t\|^2 \left(u_t^\tau \varepsilon_t/\|u_t\|\right)^2\right\}\right] \\
&\leqslant \max_{1\leqslant t\leqslant n} \mathsf{E}\left[\exp\left\{\iota_2 C_* \left(u_t^\tau \varepsilon_t/\|u_t\|\right)^2\right\}\right] \\
&= \max_{1\leqslant t\leqslant n} \mathsf{E}\left[\exp\left\{\iota_1 \left(u_t^\tau \varepsilon_t/\|u_t\|\right)^2\right\}\right] \leqslant C_0,
\end{aligned}$$

至此，完成式 (B.1) 的证明。

根据时变线性过程表示 (2.3)，可得

$$x_{t,i}^2 = \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} \left(\Phi_{t,k_1,i}^\tau e_{t-k_1}\right)\left(\Phi_{t,k_2,i}^\tau e_{t-k_2}\right)$$

其中 $\Phi_{t,k,i}^\tau$ 是 i的第 i 个行向量$\mathbf{\Phi}_{t,k}$。不失一般性，假设对所有 $k \geqslant 0$，(2.4) 均成立。令$u_{t,k,i} = \Phi_{t,k,i}/\|\Phi_{t,k,i}\|$，并注意到

$$\max_{1\leqslant t\leqslant n} \max_{1\leqslant i\leqslant d} \|\Phi_{t,k,i}\| \leqslant \max_{1\leqslant t\leqslant n} \|\mathbf{\Phi}_{t,k}\| \leqslant C_1\rho^k,$$

我们可以证明

$$\begin{aligned} x_{t,i}^2 &\leqslant C_1^2 \sum_{k_1=0}^{\infty} \rho^{k_1} \sum_{k_2=0}^{\infty} \rho^{k_2} \left| \left(u_{t,k_1,i}^\intercal e_{t-k_1}\right) \left(u_{t,k_2,i}^\intercal e_{t-k_2}\right) \right| \\ &\leqslant C_1^2 \sum_{k_1=0}^{\infty} \rho^{k_1} \sum_{k_2=0}^{\infty} \rho^{k_2} \left(u_{t,k_2,i}^\intercal e_{t-k_2}\right)^2 \\ &= \frac{C_1^2}{1-\rho} \sum_{k=0}^{\infty} \rho^k \left(u_{t,k,i}^\intercal e_{t-k}\right)^2, \end{aligned}$$

其中，结合 e_t 上的独立性假设以及 (A.1)，表明 th at

$$\begin{aligned} \max_{1\leqslant t\leqslant n} \max_{1\leqslant i\leqslant d} \mathsf{E}\left[\exp\left\{\iota_3 x_{t,i}^2\right\}\right] &\leqslant \max_{1\leqslant t\leqslant n} \max_{1\leqslant i\leqslant d} \mathsf{E}\left[\exp\left\{\frac{\iota_3 C_1^2}{1-\rho} \sum_{k=0}^{\infty} \rho^k \left(u_{t,k,i}^\intercal e_{t-k}\right)^2\right\}\right] \\ &= \max_{1\leqslant t\leqslant n} \max_{1\leqslant i\leqslant d} \prod_{k=0}^{\infty} \mathsf{E}\left[\exp\left\{\frac{\iota_3 C_1^2}{1-\rho} \rho^k \left(u_{t,k,i}^\intercal e_{t-k}\right)^2\right\}\right] \\ &= \max_{1\leqslant t\leqslant n} \max_{1\leqslant i\leqslant d} \prod_{k=0}^{\infty} \mathsf{E}\left[\exp\left\{\iota_2 \rho^k \left(u_{t,k,i}^\intercal e_{t-k}\right)^2\right\}\right] \\ &\leqslant \prod_{k=0}^{\infty} \left(\max_{1\leqslant t\leqslant n} \max_{1\leqslant i\leqslant d} \mathsf{E}\left[\exp\left\{\iota_2 \left(u_{t,k,i}^\intercal e_{t-k}\right)^2\right\}\right]\right)^{\rho^k} \\ &\leqslant \prod_{k=0}^{\infty} C_0^{\rho^k} = C_0^{1/(1-\rho)}, \end{aligned}$$

从而完成了对（B.2）的证明。 ■

以下引理是关于鞅差的一个著名伯恩斯坦型不等式（例如，Freedman，1975；de la Pe˜na，1999）。

引理 B.2。 设 $(z_t, \mathcal{F}_t)_{t\geqslant 1}$ 为一个鞅差序列，且 $\sigma_n^2 = \sum_{t=1}^n \mathsf{E}(z_t^2|\mathcal{F}_{t-1})$。假设存在常数 $a > 0$，使得对所有 $t \geqslant 2$都有$\mathsf{P}(|z_t| \leqslant a|\mathcal{F}_{t-1}) = 1$。那么，对所有 x，$y > 0$,

$$\mathsf{P}\left(\sum_{t=1}^n z_t > x,\ \sigma_n^2 \leqslant y\right) \leqslant \exp\left\{-\frac{x^2}{2(y+ax)}\right\}.$$

定义

$$L_{i,0}(\tau) = \frac{1}{n}\sum_{t=1}^n e_{t,i}(\tau) X_{t-1} K_h(\tau_t - \tau) \text{ and } L_{i,1}(\tau) = \frac{1}{n}\sum_{t=1}^n e_{t,i}(\tau) X_{t-1} \left(\frac{\tau_t - \tau}{h}\right) K_h(\tau_t - \tau),$$

其中 $e_{t,\ i}(\tau) = x_{t,\ i} - [\alpha_{i\bullet}(\tau) + \alpha'_{i\bullet}(\tau)(\tau_t - \tau)]^\intercal\ X_{t-1}$。引理 B.3 给出了核加权量 $L_{i,\ k}(\cdot)$, $k = 0, 1$ 的一致渐近阶。

引理 B.3. 假设假设 *1* 和 *2* 成立。则有

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant t\leqslant n}|L_{i,k}(\tau_t)|_{max} = O_P\left(\zeta_{n,d}\right),\ \ k = 0, 1, \tag{B.3}$$

其中 $\zeta_{n,\ d} = \log(n\vee d)\left[(nh)^{-1/2} + sh^2\right]$ 如假设*3(i)*中所述。

引理 B.3 的证明。我们仅证明 B.3)，其中 $k = 0$；对于 $k = 1$，证明类似。注意到

$$e_{l,i}(\tau_t) = e_{l,i} + [\alpha_{i\bullet}(\tau_l) - \alpha_{i\bullet}(\tau_t) - \alpha'_{i\bullet}(\tau_t)(\tau_l - \tau_t)]^\intercal\ X_{l-1} =: e_{l,i} + b_{l,i}^\intercal(\tau_t)X_{l-1},$$

我们写作

$$L_{i,0}(\tau_t) = \frac{1}{n}\sum_{l=1}^{n} e_{l,i}X_{l-1}K_h(\tau_l - \tau_t) + \frac{1}{n}\sum_{l=1}^{n} b_{l,i}^\intercal(\tau_t)X_{l-1}X_{l-1}K_h(\tau_l - \tau_t).$$

为了证明 B.3)，对于 $k = 0$，只需证明

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant j\leqslant d}\max_{1\leqslant t\leqslant n}\left|\frac{1}{n}\sum_{l=1}^{n} e_{l,i}x_{l-1,j}K_h(\tau_l - \tau_t)\right| = O_P\left((nh)^{-1/2}\log(n\vee d)\right) \tag{B.4}$$

and

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant j\leqslant d}\max_{1\leqslant t\leqslant n}\left|\frac{1}{n}\sum_{l=1}^{n} b_{l,i}^\intercal(\tau_t)X_{l-1}x_{l-1,j}K_h(\tau_l - \tau_t)\right| = O_P\left(sh^2\log(n\vee d)\right). \tag{B.5}$$

定义

$$\bar{e}_{l,i} = e_{l,i}I\left(|e_{l,i}| \leqslant 2\sqrt{\iota_2^{-1}\log(n\vee d)}\right),\ \ \widetilde{e}_{l,i} = e_{l,i} - \bar{e}_{l,i},$$

and

$$\bar{x}_{l,i} = x_{l,i}I\left(|x_{l,i}| \leqslant 2\sqrt{\iota_3^{-1}\log(n\vee d)}\right),\ \ \widetilde{x}_{l,i} = x_{l,i} - \bar{x}_{l,i},$$

其中 ι_2 和 ι_3 在引理 B.1中定义。于是，我们得到如下分解：

$$\begin{aligned}\frac{1}{n}\sum_{l=1}^{n} e_{l,i}x_{l-1,j}K_h(\tau_l - \tau_t) \ =\ & \frac{1}{n}\sum_{l=1}^{n}\bar{e}_{l,i}\bar{x}_{l-1,j}K_h(\tau_l - \tau_t) + \frac{1}{n}\sum_{l=1}^{n}\bar{e}_{l,i}\widetilde{x}_{l-1,j}K_h(\tau_l - \tau_t) + \\ & \frac{1}{n}\sum_{l=1}^{n}\widetilde{e}_{l,i}\bar{x}_{l-1,j}K_h(\tau_l - \tau_t) + \frac{1}{n}\sum_{l=1}^{n}\widetilde{e}_{l,i}\widetilde{x}_{l-1,j}K_h(\tau_l - \tau_t).\end{aligned}$$

根据 th 利用 Bonferroni 不等式和 Markov 不等式，以及（B.1），对于任意 $\epsilon > 0$，我们有

$$\begin{aligned}
& P\left(\max_{1\leqslant i\leqslant d}\max_{1\leqslant j\leqslant d}\max_{1\leqslant t\leqslant n}\left|\frac{1}{n}\sum_{l=1}^{n}\widetilde{e}_{l,i}\overline{x}_{l-1,j}K_h(\tau_l-\tau_t)\right| > \epsilon(nh)^{-1/2}\log(n\vee d)\right) \\
\leqslant\ & P\left(\max_{1\leqslant i\leqslant d}\max_{1\leqslant t\leqslant n}|e_{t,i}| > 2\sqrt{\iota_2^{-1}\log(n\vee d)}\right) \\
\leqslant\ & \sum_{i=1}^{d}\sum_{t=1}^{n}P\left(|e_{t,i}| > 2\sqrt{\iota_2^{-1}\log(n\vee d)}\right) \\
\leqslant\ & \sum_{i=1}^{d}\sum_{t=1}^{n}(n\vee d)^{-4}E\left(\exp\left\{\iota_2 e_{t,i}^2\right\}\right) \\
\leqslant\ & M(n\vee d)^{-2} = o(1). \qquad \text{(B.6)}
\end{aligned}$$

因此，我们有

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant j\leqslant d}\max_{1\leqslant t\leqslant n}\left|\frac{1}{n}\sum_{l=1}^{n}\widetilde{e}_{l,i}\overline{x}_{l-1,j}K_h(\tau_t-\tau_t)\right| = o_P\left((nh)^{-1/2}\log(n\vee d)\right). \qquad \text{(B.7)}$$

沿用（B.7）的证明，我们还有

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant j\leqslant d}\max_{1\leqslant t\leqslant n}\left|\frac{1}{n}\sum_{l=1}^{n}\overline{e}_{l,i}\widetilde{x}_{l-1,j}K_h(\tau_l-\tau_t)\right| = o_P\left((nh)^{-1/2}\log(n\vee d)\right) \qquad \text{(B.8)}$$

and

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant j\leqslant d}\max_{1\leqslant t\leqslant n}\left|\frac{1}{n}\sum_{l=1}^{n}\widetilde{e}_{l,i}\widetilde{x}_{l-1,j}K_h(\tau_l-\tau_t)\right| = o_P\left((nh)^{-1/2}\log(n\vee d)\right). \qquad \text{(B.9)}$$

根据柯西–施瓦茨不等式、马尔可夫不等式以及（B.1），我们可以证明

$$\begin{aligned}
E\left(|\widetilde{e}_{l,i}|\right) &\leqslant \left[E\left(|e_{l,i}|^2\right)\right]^{1/2}\left[P\left(|e_{l,i}| > 2\sqrt{\iota_2^{-1}\log(n\vee d)}\right)\right]^{1/2} \\
&= \left[E\left(|e_{l,i}|^2\right)\right]^{1/2}\left[P\left(\exp\left\{\iota_2 e_{l,i}^2\right\} > (n\vee d)^4\right)\right]^{1/2} \\
&\leqslant \left[E\left(|e_{l,i}|^2\right)\right]^{1/2}\left[E\left(\exp\left\{\iota_2 e_{l,i}^2\right\}\right)\right]^{1/2}(n\vee d)^{-2} \\
&\leqslant M(n\vee d)^{-2},
\end{aligned}$$

结合x_{l-1}的定义，j 以及核函数的条件，可知

$$\begin{aligned}\left|\frac{1}{n}\sum_{l=1}^{n}E\left[\bar{e}_{l,i}\bar{x}_{l-1,j}K_h(\tau_l-\tau_t)\middle|\mathcal{F}_{l-1}(X)\right]\right| &= \left|\frac{1}{n}\sum_{l=1}^{n}E\left[\widetilde{e}_{l,i}\bar{x}_{l-1,j}K_h(\tau_l-\tau_t)\middle|\mathcal{F}_{l-1}(X)\right]\right| \\ &= O_P\left((n\vee d)^{-2}\sqrt{\log(n\vee d)}\right) \\ &= o_P\left((nh)^{-1/2}\log(n\vee d)\right), \end{aligned} \tag{B.10}$$

其中 $\mathcal{F}_l(X)=\sigma(X_t:t\leqslant l)$。结合（B.7）–（B.10），我们可直接得到

$$\frac{1}{n}\sum_{l=1}^{n}e_{l,i}x_{l-1,j}K_h(\tau_l-\tau_t)=\frac{1}{n}\sum_{l=1}^{n}\{\bar{e}_{l,i}-E[\bar{e}_{l,i}|\mathcal{F}_{l-1}(X)]\}\bar{x}_{l-1,j}K_h(\tau_l-\tau_t)+o_P\left((nh)^{-1/2}\log(n\vee d)\right). \tag{B.11}$$

根据引理 B.2中的 Bonferroni不等式和伯恩斯坦不等式，我们证明了

$$\begin{aligned}&P\left(\max_{1\leqslant i\leqslant d}\max_{1\leqslant j\leqslant d}\max_{1\leqslant t\leqslant n}\left|\frac{1}{n}\sum_{l=1}^{n}\{\bar{e}_{l,i}-E[\bar{e}_{l,i}|\mathcal{F}_{l-1}(X)]\}\bar{x}_{l-1,j}K_h(\tau_l-\tau_t)\right|>M_0(nh)^{-1/2}\log(n\vee d)\right)\\ \leqslant&\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{t=1}^{n}P\left(\left|\frac{1}{n}\sum_{l=1}^{n}\{\bar{e}_{l,i}-E[\bar{e}_{l,i}|\mathcal{F}_{l-1}(X)]\}\bar{x}_{l-1,j}K_h(\tau_l-\tau_t)\right|>M_0(nh)^{-1/2}\log(n\vee d)\right)\\ \leqslant&\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{t=1}^{n}\exp\{-g_0(M_0)\log(n\vee d)\}=O\left(nd^2(n\vee d)^{-g_0(M_0)}\right)=o(1),\end{aligned}$$

令 $M_0>0$ 充分大，其中 $g_0(\cdot)$ 是满足 $g_0(z)\to\infty$ 当$z\to+\infty$时的正函数。因此，我们有

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant j\leqslant d}\max_{1\leqslant t\leqslant n}\left|\frac{1}{n}\sum_{l=1}^{n}\{\bar{e}_{l,i}-E[\bar{e}_{l,i}|\mathcal{F}_{l-1}(X)]\}\bar{x}_{l-1,j}K_h(\tau_l-\tau_t)\right|=O_P\left((nh)^{-1/2}\log(n\vee d)\right). \tag{B.12}$$

利用（B.11）和（B.12），我们完成了（B.4）的证明。

令 $\bar{X}_l=(\bar{x}_{l,1},\cdots,\bar{x}_{l,d})^\intercal$ 和 $X_l=(\widetilde{x}_{l,1},\cdots,\widetilde{x}_{l,d})^\intercal$，我们得到如下分解：

$$\begin{aligned}&\frac{1}{n}\sum_{l=1}^{n}b_{l,i}^{\intercal}(\tau_t)X_{l-1}x_{l-1,j}K_h(\tau_l-\tau_t)\\ =&\frac{1}{n}\sum_{l=1}^{n}b_{l,i}^{\intercal}(\tau_t)\bar{X}_{l-1}\bar{x}_{l-1,j}K_h(\tau_l-\tau_t)+\frac{1}{n}\sum_{l=1}^{n}b_{l,i}^{\intercal}(\tau_t)\bar{X}_{l-1}\widetilde{x}_{l-1,j}K_h(\tau_l-\tau_t)\\ &\frac{1}{n}\sum_{l=1}^{n}b_{l,i}^{\intercal}(\tau_t)\widetilde{X}_{l-1}\bar{x}_{l-1,j}K_h(\tau_l-\tau_t)+\frac{1}{n}\sum_{l=1}^{n}b_{l,i}^{\intercal}(\tau_t)\widetilde{X}_{l-1}\widetilde{x}_{l-1,j}K_h(\tau_l-\tau_t).\end{aligned} \tag{B.13}$$

与式 (B.11) 的证明类似，我们可以证明，式 (B.13) 右侧最后三项的阶为 $o_P\left(h^2\log(n\vee d)\right)$。根据假设1(i)以及 $\boldsymbol{\alpha}_{i\bullet}(\cdot)$ 的泰勒展开，我们可以证明，式 (B.13) 右侧第一项的阶为 $O_P\left(sh^2\log(n\vee d)\right)$，且这一结论关于 i、j、t一致成立。式 (B.5) 的证明至此完成。■

引理 B.4 下文给出了不可行的预言估计量 $\widehat{\ }$ $\widehat{\ }\mathbf{A}_i^o$和$\mathbf{B}_i^o$，它们定义于第4.2节的 (4.6) 和 (4.7) 式。

引理 B.4。 假设*1–4*成立。则有

$$\max_{1\leqslant i\leqslant d}\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{d}\left\|\widehat{\alpha}_{i,j}^{o}(\tau_t)-\alpha_{i,j}(\tau_t)\right\|^2=O_P\left(s\zeta_{n,d}^2\right), \tag{B.14}$$

and

$$\max_{1\leqslant i\leqslant d}\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{d}\left\|\widehat{\alpha}_{i,j}^{\prime o}(\tau_t)-\alpha_{i,j}^{\prime}(\tau_t)\right\|^2=O_P\left(s\zeta_{n,d}^2h^{-2}\right). \tag{B.15}$$

引理B.4的证明。对任意 $1\leqslant i\leqslant d$，令

$$\mathbf{U}^o=\left[(v_1^o)^\intercal,(w_1^o)^\intercal,(v_2^o)^\intercal,(w_2^o)^\intercal,\cdots,(v_n^o)^\intercal,(w_n^o)^\intercal\right]^\intercal,$$

其中 $v_t^o=\left(v_{1|t}^o,\cdots,v_{d|t}^o\right)$ 满足 $v_{j|t}^o=0$，对 $j\in\underline{\mathcal{J}}_i$成立； $w_t^o=\left(w_{1|t}^o,\cdots,w_{d|t}^o\right)$ 满足 $w_{j|t}^o=0$，对$j\in\mathcal{J}_i^{\prime}$成立。定义

$$\mathcal{B}_i^*(M_*)=\left\{\mathbf{U}^o:\sum_{t=1}^{n}\left(\|v_t^o\|^2+\|w_t^o\|^2\right)=\|\mathbf{V}^o\|^2+\|\mathbf{W}^o\|^2=nM_*\right\},$$

其中 M_* 是一个可以充分大的正常数，

$$\mathbf{V}^o=\left[(v_1^o)^\intercal,(v_2^o)^\intercal,\cdots,(v_n^o)^\intercal\right]^\intercal\ \text{ and }\ \mathbf{W}^o=\left[(w_1^o)^\intercal,(w_2^o)^\intercal,\cdots,(w_n^o)^\intercal\right]^\intercal.$$

记

$$\begin{aligned}\mathbf{A}_i&=(\boldsymbol{\alpha}_{i,1},\cdots,\boldsymbol{\alpha}_{i,d})\ \text{ with }\ \boldsymbol{\alpha}_{i,j}=[\alpha_{i,j}(\tau_1),\cdots,\alpha_{i,j}(\tau_n)]^\intercal,\\ \mathbf{B}_i&=(\boldsymbol{\alpha}_{i,1}^{\prime},\cdots,\boldsymbol{\alpha}_{i,d}^{\prime})\ \text{ with }\ \boldsymbol{\alpha}_{i,j}^{\prime}=\left[\alpha_{i,j}^{\prime}(\tau_1),\cdots,\alpha_{i,j}^{\prime}(\tau_n)\right]^\intercal,\end{aligned}$$

作为真实时变参数矩阵。注意，

$$\mathcal{Q}_i\left(\mathbf{A}_i+\sqrt{\zeta_{n,d}^*}\mathbf{V}^o,\mathbf{B}_i+\sqrt{\zeta_{n,d}^*}\mathbf{W}^o/h\right)-\mathcal{Q}_i\left(\mathbf{A}_i,\mathbf{B}_i\right)=\Pi_{i,1}^o+\Pi_{i,2}^o+\Pi_{n,3}^o, \tag{B.16}$$

其中$\zeta^*_{n,\,d}=s\zeta^2_{n,\,d}$,

$$\begin{aligned}
\Pi^o_{i,1} &= \sum_{t=1}^{n}\left[\mathcal{L}_i\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t)+\sqrt{\zeta^*_{n,d}}v^o_t,\boldsymbol{\alpha}'_{i\bullet}(\tau_t)+\sqrt{\zeta^*_{n,d}}w^o_t/h\mid\tau_t\right)-\mathcal{L}_i\left(\boldsymbol{\alpha}_{i\bullet}(\tau_t),\boldsymbol{\alpha}'_{i\bullet}(\tau_t)\mid\tau_t\right)\right],\\
\Pi^o_{i,2} &= \sum_{j=1}^{d}p'_{\lambda_2}\left(\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|\right)\left(\left\|\boldsymbol{\alpha}_{i,j}+\sqrt{\zeta^*_{n,d}}\mathbf{v}^o_j\right\|-\|\boldsymbol{\alpha}_{i,j}\|\right),\\
\Pi^o_{i,3} &= \sum_{j=1}^{d}p'_{\lambda_2}\left(\widetilde{D}_{i,j}\right)\left(\left\|h\boldsymbol{\alpha}'_{i,j}+\sqrt{\zeta^*_{n,d}}\mathbf{w}^o_j\right\|-\left\|h\boldsymbol{\alpha}'_{i,j}\right\|\right),
\end{aligned}$$

其中 $\mathbf{v}^o_j=\left(v^o_{j|1},\cdots,v^o_{j|n}\right)^\tau$ 和 $\mathbf{w}^o_j=\left(w^o_{j|1},\cdots,w^o_{j|n}\right)^\tau$。

根据局部线性目标函数的定义，我们很容易得到

$$\Pi^o_{i,1}=-2\sqrt{\zeta^*_{n,d}}\sum_{t=1}^{n}\left[(v^o_t)^\tau,(w^o_t)^\tau\right]L_i(\tau_t)+\zeta^*_{n,d}\sum_{t=1}^{n}\left[(v^o_t)^\tau,(w^o_t)^\tau\right]\boldsymbol{\Psi}(\tau_t)\left[(v^o_t)^\tau,(w^o_t)^\tau\right]^\tau. \tag{B.17}$$

根据 $\mathcal{B}^*_i(M_*)$、引理 B.3 和柯西-施瓦茨不等式，我们证明

$$\left|\sum_{t=1}^{n}\left[(v^o_t)^\tau,(w^o_t)^\tau\right]L_i(\tau_t)\right|=O_P\left(\sqrt{\zeta^*_{n,d}}n^{1/2}\right)\cdot\|\mathbf{U}^o\| \tag{B.18}$$

一致地对 i成立。根据假设3(ii)中的一致受限特征值条件，我们有

$$\sum_{t=1}^{n}\left[(v^o_t)^\tau,(w^o_t)^\tau\right]\boldsymbol{\Psi}(\tau_t)\left[(v^o_t)^\tau,(w^o_t)^\tau\right]^\tau\geqslant\kappa_0\sum_{t=1}^{n}\left(\|v^o_t\|^2+\|w^o_t\|^2\right)=n\kappa_0M_* \tag{B.19}$$

对于 $\mathbf{U}^o\in\mathcal{B}^*_i(M_*)$。结合 (B.17)–(B.19)，并令 $M_*>0$ 充分大，我们有

$$\min_{1\leqslant i\leqslant d}\Pi^o_{i,1}\geqslant\kappa_0\zeta^*_{n,d}\|\mathbf{U}^o\|^2+O_P\left(\zeta^*_{n,d}n^{1/2}\right)\cdot\|\mathbf{U}^o\|>\frac{\kappa_0}{2}\zeta^*_{n,d}\|\mathbf{U}^o\|^2\ \ w.p.a.1. \tag{B.20}$$

另一方面，由定理 4.1 和假设4(ii)，我们有

$$P\left(\min_{1\leqslant i\leqslant d}\min_{j\in\mathcal{J}_i}\|\widetilde{\boldsymbol{\alpha}}_{i,j}\|>a_0\lambda_2\right)\to1,$$

and

$$P\left(\min_{1\leqslant i\leqslant d}\min_{j\in\mathcal{J}'_i}\widetilde{D}_{i,j}>a_0\lambda_2\right)\to1.$$

由于 $\alpha_{i,j}(\tau_t) = 0$ 和 $u^o_{1,j} = 0$ 对于 $j \in \overline{\mathcal{J}}_i(\tau_t)$，因此我们有

$$\Pi_{i,2}(\tau_t) = \sum_{j \in \mathcal{J}_i(\tau_t)} p'_{\lambda_2}\left(|\widetilde{\alpha}_{i,j}(\tau_t)|\right)\left(\left|\alpha_{i,j}(\tau_t) + \sqrt{\zeta^*_{n,d}}(\tau_t) u^o_{1,j}\right| - |\alpha_{i,j}(\tau_t)|\right) = 0 \ \ w.p.a.1, \tag{B.21}$$

同理

$$\Pi_{i,3}(\tau_t) = \sum_{j \in \mathcal{J}_i(\tau_t)} p'_{\lambda_2}\left(\left|\widetilde{\alpha}'_{i,j}(\tau_t)\right|\right)\left(\left|h\alpha'_{i,j}(\tau_t) + \sqrt{\zeta^*_{n,d}}(\tau_t) u^o_{2,j}\right| - \left|h\alpha'_{i,j}(\tau_t)\right|\right) = 0 \ \ w.p.a.1. \tag{B.22}$$

因此，根据（B.20）–（B.22），令 $M_* > 0$ 足够大，我们可以证明 $\qquad t$

$$\min_{1 \leqslant i \leqslant d}\left[\sup_{U^o \in \mathcal{B}^*_i(M_*)} Q_i\left(\mathbf{A}_i + \sqrt{\zeta^*_{n,d}}\mathbf{V}^o, \mathbf{B}_i + \sqrt{\zeta^*_{n,d}}\mathbf{W}^o/h\right) - Q_i\left(\mathbf{A}_i, \mathbf{B}_i\right)\right] > 0 \ \ w.p.a.1,$$

这表明，在 $\left(\widehat{\ }\mathbf{A}^o_i, \widehat{\ }\mathbf{B}^o_i\right)$ 的内部存在一个局部极小值点

$$\left\{\left(\mathbf{A}_i + \sqrt{\zeta^*_{n,d}}\mathbf{V}^o, \mathbf{B}_i + \sqrt{\zeta^*_{n,d}}\mathbf{W}^o/h\right) : \ \mathbf{U}^o \in \mathcal{B}^*_i(c_1)\right\}$$

对任意 $1 \leqslant i \leqslant d$。引理 B.4的证明至此完成。 ■

引理 B.5 给出了时变波动函数估计的一致收敛速率，这是证明时变 CLIME 估计一致一致性的关键结果。

引理 B.5。 假设假设*1–4*成立。则有

$$\max_{1 \leqslant i,j \leqslant d} \sup_{0 \leqslant \tau \leqslant 1} |\widehat{\sigma}_{ij}(\tau) - \sigma_{ij}(\tau)| = O_P\left(\nu^\diamond_{n,d} + \nu^*_{n,d}\right), \tag{B.23}$$

*w*其中， $\sigma_{ij}(\tau)$ 是 $\mathbf{\Sigma}(\tau)$的第 (i,j)个元素， $\nu^\diamond_{n,\ d}$ 和$\nu^*_{n,\ d}$ 在假设 *5(ii* 中定义 *).*

引理的证明B.5。根据 $\widehat{\sigma}_{ij}(\tau)$ 在 (3.8) 中的定义，我们有

$$\begin{aligned}\widehat{\sigma}_{ij}(\tau) - \sigma_{ij}(\tau) \ &= \ \left\{\frac{\sum_{t=1}^n \varpi_{n,t}(\tau) e_{t,i} e_{t,j}}{\sum_{t=1}^n \varpi_{n,t}(\tau)} - \sigma_{ij}(\tau)\right\} + \left\{\frac{\sum_{t=1}^n \varpi_{n,t}(\tau)\left(\widehat{e}_{t,i} - e_{t,i}\right) e_{t,j}}{\sum_{t=1}^n \varpi_{n,t}(\tau)} + \right. \\ &\left. \frac{\sum_{t=1}^n \varpi_{n,t}(\tau) e_{t,i}\left(\widehat{e}_{t,j} - e_{t,j}\right)}{\sum_{t=1}^n \varpi_{n,t}(\tau)} + \frac{\sum_{t=1}^n \varpi_{n,t}(\tau)\left(\widehat{e}_{t,i} - e_{t,i}\right)\left(\widehat{e}_{t,j} - e_{t,j}\right)}{\sum_{t=1}^n \varpi_{n,t}(\tau)}\right\} \\ &=: \ \chi^\diamond_{ij}(\tau) + \chi^*_{ij}(\tau).\end{aligned} \tag{B.24}$$

我们首先证明

$$\max_{1 \leqslant i,j \leqslant d} \sup_{0 \leqslant \tau \leqslant 1} \left|\chi^\diamond_{ij}(\tau)\right| = O_P\left(\nu^\diamond_{n,d}\right). \tag{B.25}$$

注意

$$\chi_{ij}^{\diamond}(\tau)=\frac{\sum_{t=1}^{n}\varpi_{n,t}(\tau)\left[e_{t,i}e_{t,j}-\sigma_{ij}(\tau_t)\right]}{\sum_{t=1}^{n}\varpi_{n,t}(\tau)}+\frac{\sum_{t=1}^{n}\varpi_{n,t}(\tau)\sigma_{ij}(\tau_t)}{\sum_{t=1}^{n}\varpi_{n,t}(\tau)}-\sigma_{ij}(\tau).$$

根据 $\sigma_{ij}(\cdot)$ 的泰勒展开以及局部线性权重的定义$\varpi_{n,\ t}(\tau)$，我们有

$$\begin{aligned}
&\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\frac{\sum_{t=1}^{n}\varpi_{n,t}(\tau)\sigma_{ij}(\tau_t)}{\sum_{t=1}^{n}\varpi_{n,t}(\tau)}-\sigma_{ij}(\tau)\right| \\
\leqslant\ &\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\sigma_{ij}''(\tau)\right|\cdot\left|\frac{\sum_{t=1}^{n}(\tau_t-\tau)^2\varpi_{n,t}(\tau)}{\sum_{t=1}^{n}\varpi_{n,t}(\tau)}\right| \\
\leqslant\ &M\sup_{0\leqslant\tau\leqslant 1}\left|\frac{\sum_{t=1}^{n}(\tau_t-\tau)^2\varpi_{n,t}(\tau)}{\sum_{t=1}^{n}\varpi_{n,t}(\tau)}\right|=O\left(b^2\right). \qquad \text{(B.26)}
\end{aligned}$$

令$e_{t,\ i}^{-}$和$\widetilde{e}_{t,\ j}$如引理B.3的证明中所定义。于是，我们有

$$\begin{aligned}
\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau}{b}\right)e_{t,i}e_{t,j}\ =\ &\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau}{b}\right)\overline{e}_{t,i}\overline{e}_{t,j}+\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau}{b}\right)\overline{e}_{t,i}\widetilde{e}_{t,j}+ \\
&\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau}{b}\right)\widetilde{e}_{t,i}\overline{e}_{t,j}+\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau}{b}\right)\widetilde{e}_{t,i}\widetilde{e}_{t,j}. \qquad \text{(B.27)}
\end{aligned}$$

沿用（B.11）的证明，（B.27）右侧的第一项是渐近主导项。考虑用若干互不相交的区间 $[0,1]$ 覆盖闭区间 $\mathcal{I}_k$，$k=1,\ \cdots,\ N$，其中心为 τ_k^*，长度为 $b^2[nb\log(n\vee d)]^{-1/2}$。根据假设 2(i)中 $K(\cdot)$ 的利普希茨连续性，我们有

$$\begin{aligned}
&\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\frac{1}{nb}\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau}{b}\right)\left[\overline{e}_{t,i}\overline{e}_{t,j}-E\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right]\right| \\
\leqslant\ &\max_{1\leqslant i,j\leqslant d}\max_{1\leqslant k\leqslant N}\left|\frac{1}{nb}\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau_k^*}{b}\right)\left[\overline{e}_{t,i}\overline{e}_{t,j}-E\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right]\right|+ \\
&\max_{1\leqslant i,j\leqslant d}\max_{1\leqslant k\leqslant N}\sup_{\tau\in\mathcal{I}_k}\left|\frac{1}{nb}\sum_{k=1}^{n}\left[K\left(\frac{\tau_t-\tau}{b}\right)-K\left(\frac{\tau_t-\tau_k^*}{b}\right)\right]\left[\overline{e}_{t,i}\overline{e}_{t,j}-E\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right]\right| \\
\leqslant\ &\max_{1\leqslant i,j\leqslant d}\max_{1\leqslant k\leqslant N}\left|\frac{1}{nb}\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau_k^*}{b}\right)\left[\overline{e}_{t,i}\overline{e}_{t,j}-E\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right]\right|+O_P\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right) \qquad \text{(B.28)}
\end{aligned}$$

根据Bonferroni不等式和引理B.2 以及条件 $nb/[\text{对数}(n\vee d)]^3\to\infty$ 中

根据假设5(i)，我们可以证明

$$\begin{aligned}
&\mathsf{P}\left(\max_{1\leqslant i,j\leqslant d}\max_{1\leqslant k\leqslant N}\left|\frac{1}{nb}\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau_k^*}{b}\right)\left[\overline{e}_{t,i}\overline{e}_{t,j}-\mathsf{E}\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right]\right|>M_1\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right)\\
\leqslant\ &\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{N}\mathsf{P}\left(\left|\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau_k^*}{b}\right)\left[\overline{e}_{t,i}\overline{e}_{t,j}-\mathsf{E}\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right]\right|>M_1\left[nb\log(n\vee d)\right]^{1/2}\right)\\
=\ &O\left(d^2N\exp\{-g_1(M_1)\log(n\vee d)\}\right)=O\left(d^2N(n\vee d)^{g_1(M_1)}\right)=o(1),
\end{aligned}$$

其中， $M_1>0$ 足够大，且 $g_1(\cdot)$ 是满足 $g_1(z)\to\infty$ 当$z\to+\infty$时的正函数。因此，我们有

$$\max_{1\leqslant i,j\leqslant d}\max_{1\leqslant k\leqslant N}\left|\frac{1}{nb}\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau_k^*}{b}\right)\left[\overline{e}_{t,i}\overline{e}_{t,j}-\mathsf{E}\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right]\right|=O_P\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right). \tag{B.29}$$

结合（B.28）和（B.29），我们可以证明

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\frac{1}{nb}\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau}{b}\right)\left[\overline{e}_{t,i}\overline{e}_{t,j}-\mathsf{E}\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)\right]\right|=O_P\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right). \tag{B.30}$$

根据$e_{t、i}$和$\widetilde{e}_{t、i}$的定义，我们有

$$\mathsf{E}\left(\overline{e}_{t,i}\overline{e}_{t,j}\right)-\sigma_{ij}(\tau_t)=\mathsf{E}\left(\widetilde{e}_{t,i}\widetilde{e}_{t,j}\right)-\mathsf{E}\left(\overline{e}_{t,i}\widetilde{e}_{t,j}\right)-\mathsf{E}\left(\widetilde{e}_{t,i}\overline{e}_{t,j}\right).$$

同时，根据柯西-施瓦茨不等式、马尔可夫不等式以及引理B.1中的（B.1），

$$\begin{aligned}
\mathsf{E}\left(|\overline{e}_{t,i}\widetilde{e}_{t,j}|\right)&\leqslant M\left[\mathsf{E}\left(\widetilde{e}_{t,j}^2\right)\right]^{1/2}\\
&\leqslant M\left[\mathsf{E}\left(|e_{t,i}|^4\right)\right]^{1/4}\left[\mathsf{P}\left(|e_{t,i}|>2\sqrt{\iota_2^{-1}\log(n\vee d)}\right)\right]^{1/4}\\
&\leqslant M\left[\mathsf{P}\left(\exp\left\{\iota_2e_{t,i}^2\right\}>(n\vee d)^4\right)\right]^{1/4}\\
&\leqslant M\left[\mathsf{E}\left(\exp\left\{\iota_2e_{t,i}^2\right\}\right)\right]^{1/4}(n\vee d)^{-1}\\
&\leqslant O\left((n\vee d)^{-1}\right)=o\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right),
\end{aligned}$$

类似地，

$$\mathsf{E}\left(|\widetilde{e}_{t,i}\overline{e}_{t,j}|\right)+\mathsf{E}\left(|\widetilde{e}_{t,i}\widetilde{e}_{t,j}|\right)=o\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right).$$

因此，我们可以证明

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\frac{1}{nb}\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau}{b}\right)\left[\mathsf{E}\left(\bar{e}_{t,i}\bar{e}_{t,j}\right)-\sigma_{ij}(\tau_t)\right]\right|=o_P\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right). \tag{B.31}$$

由 (B.27)、(B.30) 和 (B.31)，可以证明

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\frac{1}{nb}\sum_{t=1}^{n}K\left(\frac{\tau_t-\tau}{b}\right)\left[e_{t,i}e_{t,j}-\sigma_{ij}(\tau_t)\right]\right|=O_P\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right). \tag{B.32}$$

类似地，我们也有

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\frac{1}{nb}\sum_{t=1}^{n}K_1\left(\frac{\tau_t-\tau}{b}\right)\left[e_{t,i}e_{t,j}-\sigma_{ij}(\tau_t)\right]\right|=O_P\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right). \tag{B.33}$$

利用 (B.32)、(B.33) 以及$\omega_{n的t}(\tau)$定义，可以证明

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\frac{\sum_{t=1}^{n}\omega_{n,t}(\tau)\left[e_{t,i}e_{t,j}-\sigma_{ij}(\tau_t)\right]}{\sum_{t=1}^{n}\omega_{n,t}(\tau)}\right|=O_P\left(\left[\frac{\log(n\vee d)}{nb}\right]^{1/2}\right), \tag{B.34}$$

这与 (B.26) 结合，可得 (B.25)。

利用引理 B.4证明中的论证，可以证明

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant t\leqslant n}\left\|\widehat{\alpha}^{o}_{i\bullet}(\tau_t)-\alpha_{i\bullet}(\tau_t)\right\|=O_P\left(\sqrt{s}\zeta_{n,d}\right), \tag{B.35}$$

这与引理 B.1中的 (B.2) 结合，表明

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant t\leqslant n}\left|\widehat{e}_{t,i}-e_{t,i}\right|=O_P\left(s\zeta_{n,d}\sqrt{\log(n\vee d)}\right). \tag{B.36}$$

根据（B.25）、（B.36）和柯西-施瓦茨不等式，令 $\omega^{*}_{n,\ t}(\tau)=\omega_{n,\ t}(\tau)/\sum_{t=1}^{n}\omega_{n,\ t}(\tau)$，我们可以证明

$$\begin{aligned}
&\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\sum_{t=1}^{n}\omega^{*}_{n,t}(\tau)\left(\widehat{e}_{t,i}-e_{t,i}\right)e_{t,j}\right| \\
\leqslant\ &\max_{1\leqslant j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left(\sum_{t=1}^{n}\left|\omega^{*}_{n,t}(\tau)\right|e^{2}_{t,j}\right)^{1/2}\max_{1\leqslant i\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left(\sum_{t=1}^{n}\left|\omega^{*}_{n,t}(\tau)\right|\left(\widehat{e}_{t,i}-e_{t,i}\right)^{2}\right)^{1/2} \\
=\ &O_P\left(s\zeta_{n,d}\sqrt{\log(n\vee d)}\right)=O_P\left(\nu^{*}_{n,d}\right).
\end{aligned} \tag{B.37}$$

同样，我们也可以证明

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\sum_{t=1}^{n}\omega_{n,t}^{*}(\tau)e_{t,i}\left(\widehat{e}_{t,j}-e_{t,j}\right)\right|=O_{P}\left(\nu_{n,d}^{*}\right) \tag{B.38}$$

and

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\sum_{t=1}^{n}\omega_{n,t}^{*}(\tau)\left(\widehat{e}_{t,i}-e_{t,i}\right)\left(\widehat{e}_{t,j}-e_{t,j}\right)\right|=O_{P}\left(\left[\nu_{n,d}^{*}\right]^{2}\right)=o_{P}\left(\nu_{n,d}^{*}\right). \tag{B.39}$$

由（B.37）–（B.39），我们很容易得到

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\chi_{ij}^{*}(\tau)\right|=O_{P}\left(\nu_{n,d}^{*}\right),$$

结合（B.24）和（B.25），即可完成引理 B.5的证明。 ■

附录 C：命题 5.1 和 5.2 的证明

在本附录中，我们给出命题 5.1 和 5.2 中所述因子调整估计量收敛性质的证明。定义

$$\widehat{L}_{i,0}(\tau)=\frac{1}{n}\sum_{t=1}^{n}\widehat{e}_{t,i}(\tau)\widehat{X}_{t-1}K_{h}(\tau_{t}-\tau)\ \text{ and }\ \widehat{L}_{i,1}(\tau)=\frac{1}{n}\sum_{t=1}^{n}\widehat{e}_{t,i}(\tau)\widehat{X}_{t-1}\left(\frac{\tau_{t}-\tau}{h}\right)K_{h}(\tau_{t}-\tau),$$

其中 $\widehat{X}_{t}=(\widehat{x}_{t,1},\cdots,\widehat{x}_{t,d})^{\intercal}$ 定义见（5.3）或（5.4），且 $\widehat{e}_{t,\,i}(\tau)=\widehat{x}_{t,\,i}-[\alpha_{i\bullet}(\tau)+\alpha_{i\bullet}'(\tau)(\tau_{t}-\tau)]^{\intercal}X_{t-1}$。以下引理将引理 B.3推广到因子调整的核加权量。

引理 C.1。 假设 *1*、*2* 和 *6(i)* 成立。那么我们有

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant t\leqslant n}\left|\widehat{L}_{i,k}(\tau_{t})\right|_{\max}=O_{P}\left(\zeta_{n,d}^{\dagger}\right),\ k=0,1, \tag{C.1}$$

其中 $\zeta_{n,\,d}^{\dagger}=\zeta_{n,\,d}+[\text{对数}(n\vee d)]^{1/2}s\delta_{X}$ 如假设 *6(ii)* 中所述。

引理 C.1 的证明。与引理 B.3 的证明一样，我们只考虑 $k=0$。由于

$$\widehat{e}_{t,i}(\tau)=e_{t,i}(\tau)+(\widehat{x}_{t,i}-x_{t,i})+[\alpha_{i\bullet}(\tau)+\alpha_{i\bullet}'(\tau)(\tau_{t}-\tau)]^{\intercal}\left(X_{t-1}-\widehat{X}_{t-1}\right),$$

根据假设 6(i)，我们可以证明

$$\begin{aligned}\widehat{L}_{i,0}(\tau) &= L_{i,0}(\tau)+\frac{1}{n}\sum_{t=1}^{n}\left(\widehat{x}_{t,i}-x_{t,i}\right)\widehat{X}_{t-1}K_h(\tau_t-\tau)+\\ &\quad \frac{1}{n}\sum_{t=1}^{n}\left[\alpha_{i\bullet}(\tau)+\alpha'_{i\bullet}(\tau)(\tau_t-\tau)\right]^{\intercal}\left(X_{t-1}-\widehat{X}_{t-1}\right)\widehat{X}_{t-1}K_h(\tau_t-\tau)-\\ &\quad \frac{1}{n}\sum_{t=1}^{n}e_{t,i}(\tau)\left(X_{t-1}-\widehat{X}_{t-1}\right)K_h(\tau_t-\tau)\\ &= L_{i,0}(\tau)+O_P\left([\log(n\vee d)]^{1/2}s\delta_X\right).\end{aligned}$$

然后，根据引理 B.3，我们完成了关于（C.1）的证明，适用于 $k=0$。 ■

记

$$\widehat{e}_t^{\dagger}=\left(\widehat{e}_{t,1}^{\dagger},\cdots,\widehat{e}_{t,d}^{\dagger}\right)^{\intercal}=\widehat{X}_t-\widehat{A}_1^{\dagger}(\tau_t)\widehat{X}_{t-1},\ \widehat{A}_1^{\dagger}(\tau_t)=\left[\widehat{\alpha}_{ij}^{\dagger}(\tau_t)\right]_{d\times d}.$$

令 $\widehat{\sigma}_{ij}^{\dagger}(\tau)$ 为因子调整的局部线性估计量 $\sigma_{ij}(\tau)$，即在式（3.8）中用 $\widehat{e}_{t,\,i}$ 替换 $\widehat{e}_{t\,,\,i}^{\dagger}$。以下引理将引理 B.5推广到因子调整波动函数估计。

引理 C.2。 假设命题 *5.1(iii)*中的假设均成立。则有

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\widehat{\sigma}_{ij}^{\dagger}(\tau)-\sigma_{ij}(\tau)\right|=O_P\left(\nu_{n,d}^{\diamond}+\nu_{n,d}^{\dagger}\right), \tag{C.2}$$

其中 $\nu_{n,\,d}^{\diamond}$ 在假设 *5(ii)*中定义，且$\nu_{n,\,d}^{\dagger}$ 在假设 *6(iv)*中定义 .

引理 C.2的证明。与式（B.24）中一样，我们有

$$\begin{aligned}\widehat{\sigma}_{ij}^{\dagger}(\tau)-\sigma_{ij}(\tau) &= \left\{\frac{\sum_{t=1}^{n}\omega_{n,t}(\tau)e_{t,i}e_{t,j}}{\sum_{t=1}^{n}\omega_{n,t}(\tau)}-\sigma_{ij}(\tau)\right\}+\left\{\frac{\sum_{t=1}^{n}\omega_{n,t}(\tau)\left(\widehat{e}_{t,i}^{\dagger}-e_{t,i}\right)e_{t,j}}{\sum_{t=1}^{n}\omega_{n,t}(\tau)}+\right.\\ &\quad \left.\frac{\sum_{t=1}^{n}\omega_{n,t}(\tau)e_{t,i}\left(\widehat{e}_{t,j}^{\dagger}-e_{t,j}\right)}{\sum_{t=1}^{n}\omega_{n,t}(\tau)}+\frac{\sum_{t=1}^{n}\omega_{n,t}(\tau)\left(\widehat{e}_{t,i}^{\dagger}-e_{t,i}\right)\left(\widehat{e}_{t,j}^{\dagger}-e_{t,j}\right)}{\sum_{t=1}^{n}\omega_{n,t}(\tau)}\right\}\\ &=: \chi_{ij}^{\diamond}(\tau)+\chi_{ij}^{\dagger}(\tau).\end{aligned} \tag{C.3}$$

根据（B.25），我们只需证明

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\chi_{ij}^{\dagger}(\tau)\right|=O_P\left(\nu_{n,d}^{\dagger}\right). \tag{C.4}$$

根据（B.36）的证明，我们有

$$\max_{1\leqslant i\leqslant d}\max_{1\leqslant t\leqslant n}\left|\widehat{e}_{t,i}^{\dagger}-e_{t,i}\right|=O_P\left(s\zeta_{n,d}^{\dagger}\sqrt{\log(n\vee d)}\right). \tag{C.5}$$

根据（B.25）、（C.5）和柯西-施瓦茨不等式，我们可以证明

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\sum_{t=1}^{n}\varpi_{n,t}^{*}(\tau)\left(\widehat{e}_{t,i}^{\dagger}-e_{t,i}\right)e_{t,j}\right|=O_P\left(s\zeta_{n,d}^{\dagger}\sqrt{\log(n\vee d)}\right)=O_P\left(\nu_{n,d}^{\dagger}\right), \tag{C.6}$$

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\sum_{t=1}^{n}\varpi_{n,t}^{*}(\tau)e_{t,i}\left(\widehat{e}_{t,j}^{\dagger}-e_{t,j}\right)\right|=O_P\left(\nu_{n,d}^{\dagger}\right), \tag{C.7}$$

$$\max_{1\leqslant i,j\leqslant d}\sup_{0\leqslant\tau\leqslant 1}\left|\sum_{t=1}^{n}\varpi_{n,t}^{*}(\tau)\left(\widehat{e}_{t,i}^{\dagger}-e_{t,i}\right)\left(\widehat{e}_{t,j}^{\dagger}-e_{t,j}\right)\right|=o_P\left(\nu_{n,d}^{\dagger}\right). \tag{C.8}$$

结合（C.6）–（C.8），我们完成了（C.4）的证明。 ■

定义

$$\widehat{\Psi}(\tau)=\begin{bmatrix}\widehat{\Psi}_0(\tau) & \widehat{\Psi}_1(\tau)\\ \widehat{\Psi}_1(\tau) & \widehat{\Psi}_2(\tau)\end{bmatrix}\ \text{with}\ \widehat{\Psi}_k(\tau)=\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\tau_t-\tau}{h}\right)^k\widehat{X}_{t-1}\widehat{X}_{t-1}^{\intercal}K_h(\tau_t-\tau),\ \ k=0,1,2.$$

命题 5.1 的证明。我们从证明

$$\min_{1\leqslant i\leqslant d}\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}_i(\tau_t)}u^{\intercal}\widehat{\Psi}(\tau_t)u\geqslant\kappa_0/2,\ \ w.p.a.1, \tag{C.9}$$

其中，$\mathcal{B}_i(\tau)$ 的定义见（4.2）。事实上，结合假设 6(i)与引理B.3 和C.1的证明中的论证，我们可以证明

$$\max_{1\leqslant t\leqslant n}\left\|\widehat{\Psi}(\tau_t)-\Psi(\tau_t)\right\|_{\max}=O_P\left([\log(n\vee d)]^{1/2}\delta_X\right). \tag{C.10}$$

然后，利用（C.10）以及引理D.1证明中的论证，我们得到

$$\min_{1\leqslant i\leqslant d}\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}_i(\tau_t)}u^{\intercal}\widehat{\Psi}(\tau_t)u\geqslant\min_{1\leqslant i\leqslant d}\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}_i(\tau_t)}u^{\intercal}\Psi(\tau_t)u+O_P\left([\log(n\vee d)]^{1/2}s\delta_X\right),$$

这与假设 3(ii) 和 6(i) 一起完成了（C.9）的证明。

(5.6) 和 (5.7) 的证明与定理 4.1 和 4.2 的证明类似，但分别将引理B.3 和 (4.3) 替换为引理 C.1 和 (C.9)。(5.8) 的证明与定理 4.3 的证明类似，但将引理 B.5 替换为引理 C.2。为节省篇幅，此处省略细节

。 ■

命题 5.2 的证明。利用命题 5.1(ii)，式（5.9）的证明与推论 4.1 的证明类似。利用命题 5.1(iii)，式（5.10）的证明与推论 4.2 的证明类似。■

附录 D：假设3(ii)的验证

在本附录中，我们在高斯假设下验证时变向量自回归的统一受限特征值条件（4.3），即 $e_t \sim N(\mathbf{0}_d, \mathbf{\Sigma}_t)$。回顾

$$\mathbf{\Psi}(\tau) = \begin{bmatrix} \mathbf{\Psi}_0(\tau) & \mathbf{\Psi}_1(\tau) \\ \mathbf{\Psi}_1(\tau) & \mathbf{\Psi}_2(\tau) \end{bmatrix} \text{ with } \mathbf{\Psi}_k(\tau) = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\tau_t - \tau}{h}\right)^k X_{t-1}X_{t-1}^\intercal K_h(\tau_t - \tau), \ k = 0, 1, 2.$$

我们先给出一些技术引理及其证明。

引理 D.1. 在事件

$$\mathcal{E}_{\mathbf{\Psi}}(\delta) = \left\{ \max_{1\leqslant t\leqslant n} \|\mathbf{\Psi}(\tau_t) - \mathsf{E}[\mathbf{\Psi}(\tau_t)]\|_{\max} \leqslant \delta \right\},$$

我们有

$$\min_{1\leqslant i\leqslant d}\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}_i(\tau_t)} u^\intercal \mathbf{\Psi}(\tau_t)u \geqslant \min_{1\leqslant i\leqslant d}\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}_i(\tau_t)} u^\intercal \mathsf{E}[\mathbf{\Psi}(\tau_t)]u - 18\delta s,$$

其中，$\mathcal{B}_i(\tau)$ 在第 *4.1* 节中定义，s 在假设 *2(ii)* 中定义。

引理的证明 D.1。证明与Kock 和 Callot （2015）中的引理 6 类似。记 $\mathcal{J}_{i,t} = \mathcal{J}_i(\tau_t)$和 $\mathcal{J}'_{i,t} = \mathcal{J}'_i(\tau_t)$。对于 $u = (u_1、 u_2) \in \mathcal{B}_i(\tau_t)^\intercal$ 以及给定的 $\mathcal{E}_{\mathbf{\Psi}}(\delta)$，我们有

$$\begin{aligned} u^\intercal \mathsf{E}[\mathbf{\Psi}(\tau_t)]u - u^\intercal \mathbf{\Psi}(\tau_t)u &\leqslant \left|u^\intercal \mathsf{E}[\mathbf{\Psi}(\tau_t)]u - u^\intercal \mathbf{\Psi}(\tau_t)u\right| = \left|u^\intercal (\mathbf{\Psi}(\tau_t) - \mathsf{E}[\mathbf{\Psi}(\tau_t)])u\right| \\ &\leqslant \delta|u|_1^2 \leqslant 9\delta\left(|u_1(\mathcal{J}_{i,t})|_1 + \left|u_2(\mathcal{J}'_{i,t})\right|_1\right)^2 \\ &\leqslant 18\delta s\left(\|u_1(\mathcal{J}_{i,t})\|^2 + \left\|u_2(\mathcal{J}'_{i,t})\right\|^2\right) \leqslant 18\delta s, \end{aligned}$$

其中，$u(\mathcal{J})$ 表示仅由 u 中以 $\mathcal{J}$ 为索引的元素组成的向量。这表明

$$u^\intercal \mathbf{\Psi}(\tau_t)u \geqslant u^\intercal \mathsf{E}[\mathbf{\Psi}(\tau_t)]u - 18\delta s.$$

对上述不等式两边取最小值$\min_{1\leqslant i\leqslant d}\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}_i(\tau_t)}$，即可完成引理D.1的证明。

■

令

$$\overline{X}_t(\tau) = \begin{bmatrix} X_t \\ X_t\left(\frac{\tau_t-\tau}{h}\right) \end{bmatrix} \text{ and } \overline{X}_{K,t}(\tau) = K^{1/2}\left(\frac{\tau_t-\tau}{h}\right)\overline{X}_t(\tau),$$

我们可以重新写为

$$(nh)u^\intercal \Psi(\tau)u = \sum_{t=1}^{n} u^\intercal \left[\overline{X}_{K,t}(\tau)\overline{X}_{K,t}^\intercal(\tau)\right] u = \|\overline{\mathbf{X}}_u(\tau)\|^2,$$

其中 $\overline{\mathbf{X}}_u(\tau) = [u^\intercal \overline{X}_{K,1}(\tau), \cdots, u^\intercal \overline{X}_{K,n}(\tau)]^\intercal$。由于 $\overline{X}_{K,\ t}(\tau)$ 是一个高斯随机向量，我们可以采用以下引理（例如，Kock 和 Callot，2015）。

引理 D.2。令$\mathbf{Z}$ 为一个 $n \times 1$ 向量，且$\mathbf{Z} \sim N(\mathbf{0}_n, \mathbf{Q})$。那么，对于任意 δ， $m > 0$,

$$P\left(\|\mathbf{Z}\|^2 - E\|\mathbf{Z}\|^2 > \delta\right) \leqslant 2\exp\left(\frac{-\delta^2}{8n\|\mathbf{Q}\|_\infty^2 m^2}\right) + n\exp\left(-m^2/2\right).$$

引理 D.2 中的不等式对于推导事件 $\mathcal{E}_\Psi(\delta)$ 的概率至关重要，该事件在引理 D.1中定义，如以下引理所示。

引理 D.3。*S* 假设假设 *1* 和 *2(i)* 成立。那么，对于任意 δ， $m > 0$，我们有

$$P\left(\mathcal{E}_\Psi(\delta)\right) \leqslant 4nd^2\left[6\exp\left(\frac{-\delta^2 nh}{64C_\diamond^2 m^2}\right) + 6nh\exp\left(-m^2/2\right)\right], \tag{D.1}$$

其中， $C_\diamond = \frac{2C_* C_K C_1^2}{(1-\rho)(1-\rho^2)}$， C_* 定义见引理 *B.1*， C_K 是核函数 $K(\cdot)$的上界，而C_1 和 ρ定义见 *(2.4)*。

引理 D.3 的证明。设 $\Psi(\tau_t)$ 的 (i,j) 项为 $\Psi_{i,\ j}(\tau_t)$。对于任意 $\delta > 0$，注意到

$$P\left(\max_{1\leqslant t\leqslant n}\max_{1\leqslant i,j\leqslant 2d}|\Psi_{i,j}(\tau_t) - E[\Psi_{i,j}(\tau_t)]| > \delta\right) \leqslant \sum_{t=1}^{n}\sum_{i=1}^{2d}\sum_{j=1}^{2d} P\left(|\Psi_{i,j}(\tau_t) - E[\Psi_{i,j}(\tau_t)]| > \delta\right).$$

因此，只需证明

$$P\left(|\Psi_{i,j}(\tau_t) - E[\Psi_{i,j}(\tau_t)]| > \delta\right) \leqslant 6\exp\left(\frac{-\delta^2 nh}{64C_\diamond^2 m^2}\right) + 6nh\exp\left(-m^2/2\right). \tag{D.2}$$

~移除$\mathbf{X}_u(\tau)$中的零元素后，我们定义一个子向量$\mathbf{X}_u(\tau)$ ，其中仅包含
~ ~非零元素。我们将引理D.2 应用于$\mathbf{Z} = \mathbf{X}_u(\tau_t)$ 和$\mathbf{Q} = \mathbf{Q}(\tau_t) = \mathrm{Cov}(\mathbf{X}_u(\tau_t))$。考虑$\mathbf{Q}(\tau_t)$中的一个典型元素：$\mathrm{Cov}\left(u^\intercal \overline{X}_{K,l_1}(\tau_t), u^\intercal X_{K,l_2}(\tau_t)\right)$，其中 $|\tau_{l_1} - \tau_t| \leqslant h$ 且 $|\tau_{l_2} - \tau_t| \leqslant h$，这里 $u^\intercal = (u_1, u_2)$ 是一个经过适当选取的向量，其维度为 $2d$ 和 $\|u\| = 1$。令

$u_{\tau,l} = (u_1 + \frac{\tau_l - \tau}{h} u_2)/\|u_1 + \frac{\tau_l - \tau}{h} u_2\|$，我们有

$$
\begin{aligned}
& \mathsf{Cov}\left(u^\intercal \overline{X}_{K,l_1}(\tau_t), u^\intercal \overline{X}_{K,l_2}(\tau_t)\right) \\
= \; & \mathsf{E}\left[\left(u_1 + \frac{\tau_{l_1} - \tau_t}{h} u_2\right)^\intercal X_{l_1} X_{l_2}^\intercal \left(u_1 + \frac{\tau_{l_2} - \tau_t}{h} u_2\right)\right] K^{1/2}\left(\frac{\tau_{l_1} - \tau_t}{h}\right) K^{1/2}\left(\frac{\tau_{l_2} - \tau_t}{h}\right) \\
\leqslant \; & \left|\mathsf{E}\left(u_{\tau_t,l_1} X_{l_1} X_{l_2}^\intercal u_{\tau_t,l_2}\right)\right| \left\|(u_1 + \frac{\tau_l - \tau}{h} u_2\right\| \left\|(u_1 + \frac{\tau_t - \tau}{h} u_2\right\| K^{1/2}\left(\frac{\tau_{l_1} - \tau_t}{h}\right) K^{1/2}\left(\frac{\tau_{l_2} - \tau_t}{h}\right) \\
\leqslant \; & \left|\mathsf{E}\left(u_{\tau_t,l_1} X_{l_1} X_{l_2}^\intercal u_{\tau_t,l_2}\right)\right| K^{1/2}\left(\frac{\tau_{l_1} - \tau_t}{h}\right) K^{1/2}\left(\frac{\tau_{l_2} - \tau_t}{h}\right).
\end{aligned}
$$

对于 $1 \leqslant l_1, l_2 \leqslant n$，当 $|\tau_{l_1} - \tau_t| \leqslant h$ 且 $|\tau_{l_2} - \tau_t| \leqslant h$ 时，由 (2.3) 和 (2.4) 可得，

$$
\begin{aligned}
\left|\mathsf{E}\left(u_{\tau_t,l_1} X_{l_1} X_{l_2}^\intercal u_{\tau_t,l_2}\right)\right| &= \left|\mathsf{E}\left[\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\left(u_{\tau_t,l_1}^\intercal \mathbf{\Phi}_{l_1,k_1} e_{l_1-k_1}\right)\left(u_{\tau_t,l_2}^\intercal \mathbf{\Phi}_{l_2,k_2} e_{l_2-k_2}\right)^\intercal\right]\right| \\
&\leqslant C_* C_1^2 \sum_{k_1=0}^{\infty} \rho^{k_1}\rho^{|l_2-l_1|+k_1} = \frac{C_* C_1^2 \rho^{|l_2-l_1|}}{1-\rho^2}.
\end{aligned}
$$

因此，

$$
\begin{aligned}
\max_{1\leqslant t\leqslant n} \|\mathbf{Q}(\tau_t)\|_\infty &\leqslant \frac{C_* C_1^2}{1-\rho^2} \max_{1\leqslant l_1\leqslant n} \sum_{l_2=1}^{n} \rho^{|l_2-l_1|}\left[\max_{1\leqslant t\leqslant n} K^{1/2}\left(\frac{\tau_{l_1} - \tau_t}{h}\right) K^{1/2}\left(\frac{\tau_{l_2} - \tau_t}{h}\right)\right] \\
&\leqslant \frac{2C^* C_1^2 C_K}{1-\rho^2} \sum_{k=0}^{\infty} \rho^k \leqslant \frac{2C^* C_1^2 C_K}{(1-\rho)(1-\rho^2)} = C_\diamond.
\end{aligned}
$$

~利用 Kock 和 Callot (2015) 中的引理 7，并注意到 $\mathbf{X}_u(\tau_t)$ 的维度为 $(2nh)$，可得对任意 δ, $m > 0$,

$$
\mathsf{P}\left(\left\|\widetilde{\mathbf{X}}_u(\tau_t)\right\|^2 - \mathsf{E}\left\|\widetilde{\mathbf{X}}_u(\tau_t)\right\|^2 > \delta\right) \leqslant 2\exp\left(\frac{-\delta^2}{16C_\diamond^2 m^2 (nh)}\right) + 2nh\exp\left(-m^2/2\right),
$$

这表明

$$
\mathsf{P}\left(u^\intercal \mathbf{\Psi}(\tau_t) u - \mathsf{E}\left[u^\intercal \mathbf{\Psi}(\tau_t) u\right] > \delta\right) \leqslant 2\exp\left(\frac{-\delta^2 (nh)}{16C_\diamond^2 m^2}\right) + 2nh\exp\left(-m^2/2\right). \tag{D.3}
$$

选择 u 为一个向量，其第 i 个元素为 1，其余元素为 0。由（D.3），我们有

$$
\mathsf{P}\left(|\Psi_{i,i}(\tau_t) - \mathsf{E}[\Psi_{i,i}(\tau_t)]| > \delta\right) \leqslant 2\exp\left(\frac{-\delta^2 (nh)}{16C_\diamond^2 m^2}\right) + 2nh\exp\left(-m^2/2\right) \tag{D.4}
$$

f或 $i=1$、$\cdots$、$2d$。类似地，我们还可进一步证明，对于 $1\leqslant i\neq j\leqslant 2d$,

$$\begin{aligned}
& P\left(|\Psi_{i,j}(\tau_t)-E[\Psi_{i,j}(\tau_t)]|>\delta\right)\\
\leqslant\ & P\left(|\Psi_{i,i}(\tau_t)-2\Psi_{i,j}(\tau_t)+\Psi_{j,j}(\tau_t)-E[\Psi_{i,i}(\tau_t)-2\Psi_{i,j}(\tau_t)+\Psi_{j,j}(\tau_t)]|/2>\delta/2\right)+\\
& P\left(|\Psi_{i,i}(\tau_t)+\Psi_{j,j}(\tau_t)-E[\Psi_{i,i}(\tau_t)+\Psi_{j,j}(\tau_t)]|/2>\delta/2\right)\\
\leqslant\ & P\left(|\Psi_{i,i}(\tau_t)+2\Psi_{i,j}(\tau_t)+\Psi_{j,j}(\tau_t)-E[\Psi_{i,i}(\tau_t)+2\Psi_{i,j}(\tau_t)+\Psi_{j,j}(\tau_t)]|>\delta\right)+\\
& P\left(|\Psi_{i,i}(\tau_t)-E[\Psi_{i,i}(\tau_t)]|>\delta/2\right)+P\left(|\Psi_{j,j}(\tau_t)-E[\Psi_{j,j}(\tau_t)]|>\delta/2\right)\\
\leqslant\ & 6\exp\left(\frac{-\delta^2nh}{64C_\diamond^2m^2}\right)+6nh\exp\left(-m^2/2\right). \qquad (D.5)
\end{aligned}$$

利用 (D.4) 和 (D.5)，我们完成了 (D.2) 的证明。 ■

下面的命题验证一致受限特征值条件。

命题 D.1. 假设假设 *1* 和 *2(i)* 成立。若

$$\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}}u^\intercal E[X_tX_t^\intercal]u\geqslant 2\kappa_0, \qquad (D.6)$$

其中 $\mathcal{B}=\{u:\|u\|=1,\ |u|_1\leqslant 3|u_{\mathcal{J}}|_1\}$, $\mathcal{J}$ 是满足 $\mathcal{J}\subset\{1,\cdots,d\}$ 且基数为

$$s=o\left((nh)^{1/2}/\log(ndh^{1/2})\right),$$

则 *(4.3)* 以概率趋于 *1* 成立。

命题的证明 D.1。在引理 $\delta=c_\circ/s$ 和 $m^2=\left(\frac{c_\circ^2nh}{32C_\diamond^2s^2}\right)^{1/2}$ D.3 中取，其中 $c_\circ$ 是一个稍后确定的适当常数，我们有

$$\begin{aligned}
P\left(\max_{1\leqslant t\leqslant n}\|\boldsymbol{\Psi}(\tau_t)-E[\boldsymbol{\Psi}(\tau_t)]\|_{\max}>\frac{c_\circ}{s}\right) &\leqslant 4nd^2\left[6\exp\left(\frac{-c_\circ^2nh}{64C_\diamond^2s^2m^2}\right)+6nh\exp\left(-m^2/2\right)\right]\\
&\leqslant 48\exp\left(\log(n^2d^2h)-\frac{c_\circ(nh)^{1/2}}{16C_\diamond s}\right),
\end{aligned}$$

当 $s=o\left((nh)^{1/2}/\log(ndh^{1/2})\right)$ 时，该式收敛于 0。根据引理 D.1，于是有

$$\min_{1\leqslant i\leqslant d}\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}_i(\tau_t)}u^\intercal\boldsymbol{\Psi}(\tau_t)u\geqslant\min_{1\leqslant i\leqslant d}\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}_i(\tau_t)}u^\intercal E[\boldsymbol{\Psi}(\tau_t)]u-18c_\circ\ w.p.a.1. \qquad (D.7)$$

接下来需要证明右侧第一项存在下界，并为D.7中的 $c_\circ$ 求出一个合适的取值。事实上，由（D.6），我们有

$$\min_{1\leqslant i\leqslant d}\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}_i(\tau_t)}u^\intercal E[\boldsymbol{\Psi}(\tau_t)]u$$

$$= \min_{1\leqslant i\leqslant d}\min_{1\leqslant t\leqslant n}\inf_{u\in\mathcal{B}_i(\tau_t)}\frac{1}{nh}\sum_{l=1}^{n}E\left[\left(u_1+\frac{\tau_l-\tau_t}{h}u_2\right)^{\tau}X_lX_l^{\tau}\left(u_1+\frac{\tau_l-\tau_t}{h}u_2\right)\right]K\left(\frac{\tau_l-\tau_t}{h}\right)$$

$$\geqslant 2\kappa_0\min_{1\leqslant t\leqslant n}\frac{1}{nh}\sum_{l=1}^{n}K\left(\frac{\tau_l-\tau_t}{h}\right)=2\kappa_0-\epsilon,$$

其中， ϵ 是任意小的数。将 $c<(\kappa_0-\epsilon)/18$ 代入（D.7），即可完成（4.3）的证明。■

附录 E：调优参数选择

所提出的三状态收缩估计程序的数值表现取决于对三个调节参数的谨慎选择： λ_1 用于初步时变 LASSO 估计， λ_2 用于时变加权组 LASSO，以及 λ_3 用于时变 CLIME。它们分别通过贝叶斯信息准则（BIC）、广义信息准则（GIC）和扩展贝叶斯信息准则（EBIC）进行选择。下面简要介绍这三个准则。

(3.3)中的局部线性回归平滑本质上是使用核权重 $K_h(\tau_t-\tau)$的加权最小二乘。因此，BIC目标函数定义为

$$BIC_i(\lambda_1;\tau)=\log\left[\frac{\mathcal{L}_i\left(\widetilde{\alpha}_{i\bullet}(\tau\mid\lambda_1),\widetilde{\alpha}'_{i\bullet}(\tau\mid\lambda_1)\right)}{\sum_{t=1}^{n}K_h(\tau_t-\tau)}\right]+\frac{\log(n_e)}{n_e}\cdot\left[|\widetilde{\alpha}_{i\bullet}(\tau\mid\lambda_1)|_0+\left|\widetilde{\alpha}'_{i\bullet}(\tau\mid\lambda_1)\right|_0\right],\quad(E.1)$$

其中 $\widetilde{\alpha}_{i\bullet}(\tau\mid\lambda_1)$ 和 $\widetilde{\alpha}'_{i\bullet}(\tau\mid\lambda_1)$ 是使用调节参数 λ_1 在点 τ处得到的局部线性估计，而有效样本量 n_e 定义为 $\sum_{t=1}^{n}K_h(\tau_t-\tau)/\max_t\{K_h(\tau_t-\tau)\}$。我们通过最小化$BIC_i(\lambda_1$来选择初步时变 LASSO 中的调节参数； τ)其定义见（E.1），并关于 λ_1进行最小化。所选调节参数同时取决于索引 i 和（缩放后的）时间点 τ。

广义信息准则由 Fan 和 Tang (2013) 在高维惩罚似然估计的背景下提出。由于我们的模型涉及未知的时变系数，且估计过程涉及局部线性平滑，因此需要按照 Li, KeandZhang (2015) 的方法对广义信息准则进行修改。例如，Cheng,Zhangand Chen (2009指出，使用 Epanechnikov 核时，每个未知函数参数相当于 $36/(35h)$ 个未知常数参数。因此，我们将广义信息准则目标函数定义为

$$GIC_i(\lambda_2)=\log\left[\frac{1}{n}\sum_{t=1}^{n}\left\{x_{t,i}-\widehat{\alpha}_{i\bullet}^{\tau}(\tau_t\mid\lambda_2)X_{t-1}\right\}^2\right]+\frac{\gamma_{n,d}}{n}\cdot\frac{36s_i(\lambda_2)}{35h},\quad(E.2)$$

其中，$\gamma_{n,d}$ 是 n 的函数，d, $\widehat{\alpha}_{i\bullet}(\tau \mid \lambda_2)$ 是使用调节参数 λ_2 得到的时变加权组 LASSO 估计，$s_i(\lambda_2)$ 是使用 λ_2选出的时变系数个数。我们选择$\gamma_{n,d} = \gamma \log(\log(n)) \log(36d/(35h))$，其中 $\gamma \in (0,1]$。我们通过最小化相对于 λ_2的 $\mathrm{GIC}_i(\lambda_2)$（定义见 (E.2)）来确定调节参数。选定的调节参数取决于指标 i。较小的 γ 会带来更稠密的网络估计。选择小于 1 的 γ的直观理由是：当某个函数参数在采样期间的大部分时间内为零、其余时间非零时，纳入相应变量对误差平方和的边际贡献较小，而较小的 γ会使信息准则更具适应性和敏感性。例如，如果我们想选出函数参数在至少 10% 的采样期间内不为零的变量，可以选择 $\gamma = 0.1$。我们在模拟中选择 $\gamma = 1$，在实证研究中选择 $\gamma = 0.1$。

扩展贝叶斯信息准则由 Chen 和 Chen (2008) 提出，并由 Foygel 和 Drton(2010) 应用于高斯图模型估计。扩展贝叶斯信息准则的目标函数定义如下：

$$\mathrm{EBIC}(\lambda_3;\tau) = -\log\left(\det(\widehat{\boldsymbol{\Omega}}(\tau \mid \lambda_3))\right) + \mathrm{Tr}(\widehat{\boldsymbol{\Omega}}(\tau \mid \lambda_3)\widehat{\boldsymbol{\Sigma}}(\tau)) + \frac{\log(n_e)}{n_e} \cdot \sum_{i<j} \mathrm{I}(|\widehat{\omega}_{ij}(\tau \mid \lambda_3)| > 0), \quad \text{(E.3)}$$

其中，$\widehat{\boldsymbol{\Omega}}(\tau \mid \lambda_3) = [\widehat{\omega}_{ij}(\tau \mid \lambda_3)]_{d\times d}$ 表示使用调节参数 λ_3得到的时变 CLIME 估计。我们通过最小化 扩展贝叶斯信息准则$(\lambda_3;\ \tau)$（E.3）中定义的函数来确定调节参数 λ_3。注意，选出的调节参数会随 τ变化。

因子调整 VAR 模型和方法的数值表现取决于对因子数的审慎选择。令 $X_t(q)$ 表示因子数设为 q时，(5.3) 或 (5.4) 中估计的特质成分，并将残差平方和定义为$V_n(q) = \sum_{t=1}^{n} |\widehat{X}_t(q)|_2^2$。对于近似因子模型 (5.1)，我们使用 Bai 和 Ng (2002) 提出的信息准则来选择因子数，即相对于 q最大化以下目标函数

$$\mathrm{IC}(q) = \log\left[V_n(q)\right] + q \cdot \left(\frac{n+d}{nd}\right) \log(n \wedge d),$$

并得到 $\hat{q}$ 作为估计因子数量。当我们考虑时变因子模型 (5.2) 时，采用 Su and Wang (2017的信息准则，即对以下目标函数关于 q

$$\mathrm{IC}(q) = \log\left[V_n(q)\right] + q \cdot \left(\frac{nh_* + d}{nh_* d}\right) \log(nh_* \wedge d),$$

并将 $\hat{q}$ 作为估计的因子数，其中 h_* 是局部主成分分析所使用的带宽。

上述两个标准　　在实证数据分析中使用 eria 来确定因子数量。

在实践中，我们需要为时变 VAR 模型选择合适的阶数。对于转移矩阵恒定的高维 VAR 模型，Miao、Phillips 和 Su (2022) 提出了一种比率准则，用于比较不同滞后阶数下估计转移矩阵的弗罗贝尼乌斯范数。接下来，我们将该准则扩展到时变 VAR 模型的情形。定义

$$R(k) = \frac{\sum_{l=k}^{2k_{max}} \sum_{t=1}^{n} (\|\widehat{\mathbf{A}}_{t,l}\|_F \vee \xi_A)}{\sum_{l=k+1}^{2k_{max}} \sum_{t=1}^{n} (\|\widehat{\mathbf{A}}_{t,k}\|_F \vee \xi_A)},$$

其中 k_{max} 和 ξ_A 由用户指定。在主文档的第 7 节中，我们设定 $k_{max} = 10$ 和$\xi_A = 0$。1，并使用时变 VAR(20) 的估计转移矩阵计算 $R(k)$。时变 VAR 的阶数选为使 $R(k), 1 \leqslant k \leqslant k_{max}$ 最大的整数。在实证分析中，我们使用上述准则选择时变 VAR(1)。

在主文档的表 1–7 中，为了评估估计的时变 VAR 和网络结构的准确性，我们报告假阳性（FP）、假阴性（FN）、真正率（TPR）、真负率（TNR）、阳性预测值（PPV）、阴性预测值（NPV）、F1 分数（F1）和马修斯相关系数（MCC）。FP 定义为被错误识别为显著变量的不显著预测变量的数量；FN 定义为被错误识别为不显著变量的显著预测变量的数量；TPR 和 TNR 定义为

$$TPR = \frac{TP}{TP + FN} \text{ and } TNR = \frac{TN}{TN + FP}$$

其中，TP 表示真正例，TN 表示真负例；阳性预测值和阴性预测值定义为

$$PPV = \frac{TP}{TP + FP} \text{ and } NPV = \frac{TN}{TN + FN};$$

F1分数是精确率和灵敏度的调和平均数，定义为

$$F_1 = 2 \times \frac{PPV \times TPR}{PPV + TPR};$$

而 MCC 定义为

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}.$$

参考文献

BAI, J. AND NG, S. (2002). 近似因子模型中因子数量的确定。《计量经济学》90, 191–221. BICKEL, P., RITOV, Y. AND TSYBAKOV, A. (2009). *Lasso*与*Dantzig*选择器的联合分析，《统计年鉴》37, 1705–1732. CAI, T. T., LIU, W. AND LUO, X. (2011). 一种用于稀疏精度矩阵估计的约束 ℓ_1 最小化方法。《美国统计学会会刊》106, 594–607. FAN, J. AND LV, J. (2011). 具有NP维数的非凹惩罚似然。*IEEE:* 信息论，57, 5467–5484. FAN, J., XUE, L. AND ZOU, H. (2014). 折叠凹惩罚估计的强预言机最优性。《统计年鉴》42, 819–849. CHEN, J. AND CHEN, Z. (2008). 大型模型空间中模型选择的扩展贝叶斯信息准则。《生物统计学》95, 759–771. CHENG, M. -Y., ZHANG, W. AND CHEN, L. -H. (2009). 广义多参数似然模型中的统计估计。《美国统计学会会刊》104, 1179–1191. FAN, Y. AND TANG, C. Y. (2013). 高维惩罚似然中的调优参数选择。《皇家统计学会杂志：*B*辑（统计方法学）》75, 531–552. Foygel, R. and Drton, M. (2010). 高斯图模型的扩展贝叶斯信息准则。《神经信息处理系统进展》23. FREEDMAN, D. A. (1975). 鞅的尾概率。《概率年刊》，3, 100–118. KOCK, A.B. AND CALLOT, L. (2015). 高维向量自回归模型的预言机不等式。《计量经济学杂志》186, 325–344. LI, D., KE, Y. AND ZHANG, W. (2015). 超高维广义半变系数模型中的模型选择与结构设定。《统计年鉴》43, 2676–2705. LIAN, H. (2012). 高维广义变系数模型的变量选择。《中华统计学志》，22, 1563–1588. MIAO, K., PHILLIPS, P.C.B. AND SU, L. (2022). 含共同因子的高维 VAR。即将发表于《计量经济学杂志》。

DE LA PE˜NA, V. H. (1999). 鞅和比率的一类一般指数不等式。《概率年刊》，27，537–564。

SU, L. AND WANG, X. (2017). 关于时变因子模型：估计与检验。《计量经济学杂志》198，84–101。