arXiv:2601.03325v1 [stat.ML] 6 Jan 2026

具有多滞后依赖的状态切换模型的可识别性

Carles Balsells-Rodas
CB221@IMPERIAL.AC.UK伦敦帝国理工学院计算系，英国伦敦**ToshikoMatsui**
T.MATSUI19@IMPERIAL.AC.UK 伦敦帝国理工学院计算系，英国伦敦 **PedroA.M. Mediano**
P.MEDIANO@IMPERIAL.AC.UK伦敦帝国理工学院计算系，英国伦敦

Yixin Wang
YIXINW@UMICH.EDU美国密歇根州安娜堡密歇根大学统计学系，邮编 *48109*

Yingzhen Li yingzhen.li@imperial.ac.uk
英国伦敦伦敦帝国理工学院计算机系

摘要

可识别性对于深度潜变量模型的可解释性至关重要，它确保参数化能够由数据生成分布唯一确定。然而，对于深度机制切换时间序列，这一问题仍未得到充分研究。我们针对多滞后机制切换模型（RSM）提出了一个通用理论框架，涵盖马尔可夫切换模型（MSM）和切换动力系统（SDS）。对于 MSM，我们将模型表述为具有时间结构的有限混合，并证明在非线性高斯设定下，状态数量和多滞后转移均具有可识别性。对于 SDS，我们基于时间结构确立了潜变量在置换和缩放意义下的可识别性，进而得出状态依赖的潜在因果图具有可识别性的条件（在状态/节点置换意义下）。通过可直接利用神经网络设计实现的架构和噪声假设，我们的结果在完全无监督设定下依然成立。我们还提出了一个满足这些假设的灵活变分估计器，并在合成基准测试中验证了这些结果。在神经科学、金融和气候领域的真实世界数据集上，可识别性带来了更可信的可解释性分析，这对科学发现至关重要。

关键词： 可识别性、状态切换模型、马尔可夫状态切换模型、切换动力系统、时间序列、状态空间模型、有限混合模型、变分推断、非线性独立成分分析、因果表征学习。

1 引言

深度潜变量模型的可识别性近年来一直是一项长期挑战。它形式化地刻画了何时参数能够由数据生成分布唯一确定。在实践中，精确恢复参数往往不可能，因此通常假定存在由研究目标决定的等价类。例如，因果表征学习（CRL；（Schölkopf 等人，2021年））旨在恢复潜在因果图，通常只要求节点在置换意义下相同。在独立成分分析（ICA；（Comon，1994））中，等价类的目标是恢复独立源，但允许置换和缩放，并且通常还允许混合函数的函数形式存在差异。

©2026 Carles Balsells-Rodas、Toshiko Matsui、Pedro A.M. Mediano、Yixin Wang 和 Yingzhen Li。许可协议：知识共享署名 4.0 国际许可协议，详见 https://creativecommons.org/licenses/by/4.0/。

一般而言，若无额外结构，非线性独立成分分析是一个病态问题（Hyvärinen 和 Pajunen, 1999年）。近期研究通过将深度生成模型与辅助信息结合（Khemakhem 等人，2020年），或将有限混合模型与分段线性混合结合（Kivva 等人，2022年），确立了可识别性。其他研究则利用时间结构（Hyvarinen 和 Morioka，2017年；Hyvarinen 等人，2019年；Hälvä 和 Hyvarinen, 2020；Hälvä 等人，2021年），或潜在因果过程（Yao 等人，2022a、b；Lippe 等，2023b、a）。

本文从深度序列生成模型的视角，重点研究状态切换模型（RSM）的可识别性，包括马尔可夫切换模型（MSM）和切换动力系统（SDS）。RSM 描述的是其行为取决于随时间变化的离散潜变量状态（state，或 switch）的时间序列。它们是状态空间模型（SSM）这一成熟序列模型类别的子集（Lindgren，1978年；Poritz，1982；Hamilton，1989）。随着序列潜变量模型日益普及（Chung 等人，2015；Li 和 Mandt，2018；Babaeizadeh 等人，2018；Saxena 等人，2021），状态空间模型已被纳入潜在动力学先验（Johnson 等人，2016；Linderman 等人，2016、2017；Fraccaro 等人，2017；Dong 等人，2020年；Ansari 等人，2021；Liu 等人，2023年；Geadah 等人，2024年）。然而，这些研究主要关注灵活先验和稳定训练方案的设计（Dong 等人，2020年；Ansari 等人，2021；Liu 等人，2023年；Zhao 和 Linderman，2023）。

为弥补这些不足，我们建立了一个统一理论框架，为具有多滞后依赖的 RSM 提供可识别性。首先，我们证明，多滞后马尔可夫切换模型可以表述为具有时间结构化成分的有限混合模型，并在非线性高斯设定下证明状态数量和多滞后转移函数的可识别性。该分析将经典有限混合可辨识性（Yakowitz 和 Spragins，1968 年）扩展到多滞后情形；由于存在滞后依赖，这一扩展并非易事。其次，对于 SDS，我们通过将 Kivva等（2022）扩展到时间设定，证明潜变量在置换和缩放意义下是可识别的。由此可得，状态依赖的潜在因果图在状态和节点的置换意义下是可识别的。通过对潜在噪声分布和生成模型的函数类施加限制，我们的结果适用于完全无监督设定。相比之下，其他处理状态切换的研究需要辅助信息（Yao 等人，2022a,b；Lippe 等，2023a）或时间相关性中的平稳性（Hyvarinen 和 Morioka，2017；Hälvä 等人，2021）。近期，已有研究在更灵活的假设下建立了类似结果（Song 等，2023年、2024年；Balsells-Rodas 等人，2025年）。

本文其余部分安排如下。[1] 第2节讨论相关工作，并阐明我们的贡献。第3节回顾有限混合和潜变量模型。第4节介绍我们的主要结果：（i）具有非线性高斯转移的参数化 MSM（非参数分析见附录 A）；（ii）SDS 潜变量在置换和缩放意义下的可识别性；以及（iii）依赖状态的因果图在状态/节点置换意义下的可识别性。第5节介绍一种灵活的估计方法，该方法遵循

1. 本文在会议论文版本 Balsells-Rodas 等人（2024）的基础上做出了重大扩展：(i) 以更直观的视角将马尔可夫切换模型视为有限混合，并将可识别性扩展到更高阶的时间依赖（正文中新增了大量证明和直观解释）；(ii) 不再局限于 SDS 中潜变量的仿射可识别性，而是进一步确立了置换/缩放可识别性，以及在节点/状态置换意义下的因果结构可辨识性；(iii) 新增了现实世界应用。

可识别性假设，并以 Dong et al. (2020)；Ansari et al. (2021) 为基础。第6节通过合成数据以及基线方法和消融实验验证我们的理论框架。最后，第7节展示脑活动、金融时间序列和气候数据方面的结果，说明可识别性带来的可解释性提升。

可识别性为何在实践中至关重要。 神经科学（Karniol- Tambour 等人，2024年）、气候（Saggioro 等，2020年）、人体运动（Dong 等人，2020年；Liu 等人，2023年）和动物行为（Ansari 等人，2021年）领域的应用展现了深度状态切换模型的前景。然而，如果缺乏可识别性，对学习得到的潜变量（例如状态、因果结构）的解释可能是虚假的，因为不同甚至相互矛盾的参数化可能实现相同的似然，从而使科学发现失效。更广泛地说，可识别性能够防止任意的潜变量重新参数化（Locatello 等人，2019年），并支持有效的因果主张（Schölkopf 等人，2021年）。我们的理论框架提供了灵活的条件，在这些条件下，学习得到的状态、潜变量和图能够在可接受的等价性范围内被唯一确定，从而支持可信的解释。

2 相关工作

马尔可夫切换模型（或自回归隐马尔可夫模型）最初用于语音分析（Poritz, 1982; Ephraim 和 Roberts, 2005）。在计量经济学领域，Hamilton (1989) 推广了马尔可夫切换模型在商业周期研究中的应用，并催生了许多扩展，例如马尔可夫切换广义自回归条件异方差模型（Haas et al., 2004）。关于渐近性质的综述和其他应用，请参见 Frühwirth-Schnatter and Frèuhwirth-Schnatter (2006)。切换动力系统在马尔可夫切换模型的基础上，扩展纳入了连续潜在状态和状态依赖动力学。经典进展包括切换卡尔曼滤波器（Murphy, 1998）以及用于呼吸数据的变分学习（Ghahramani and Hinton, 2000）。近来，状态切换已被用作深度潜变量模型中的先验分布，包括切换变分自编码器（Johnson et al., 2016; Zhao and Linderman, 2023）、软切换卡尔曼滤波器（Fraccaro et al., 2017），以及结合显式持续时间模型（Ansari 等人, 2021）或基于图的交互（Liu et al., 2023）的循环切换动力系统（Linderman et al., 2016; Dong 等人, 2020）。

有限状态隐马尔可夫模型在温和条件下具有很强的可识别性（Allman 等人，2009年；Gassiat等人，2016年、2020年）。然而，由于连续变量中存在显式自回归，这些策略无法直接迁移至马尔可夫切换模型。尽管针对某些离散马尔可夫切换模型已有研究结果（An 等人，2013年），但多滞后马尔可夫切换模型的一般非参数理论仍然缺失。我们的马尔可夫切换模型可识别性结果将经典有限混合可辨识性（Yakowitz 和 Spragins，1968 年）扩展至多滞后情形，从而弥补了这一空白。

现代深度生成模型的可识别性研究通常建立在带辅助变量或时间结构的非线性独立成分分析之上。iVAE 利用辅助信息中的条件先验证明了可识别性（Khemakhem 等人，2020年）；Kivva 等人（2022年）通过采用分段线性混合的混合先验，去除了这一要求。时间信息也可以充当辅助信息，从而通过对比学习实现可识别性（Hyvarinen 和 Morioka，2016年；Hyvarinen 等人，2019年）。在 CRL 中，时间可识别性已在分布偏移或潜在时间结构下得到确立（Yao 等人，2022b、2022a；Lippe 等，2023b、2023a）。

若干研究引入潜在状态结构，以放宽对平稳性的要求或对辅助信息的依赖。HM-NICA 用可识别的平稳

HMM先验（Gassiat等人，2016年）取代观测到的状态；IIA-HMM（Morioka等人，2021年）通过平稳HMM（Gassiat等人，2016年），引入具有非平稳创新的循环混合。SNICA（Hälvä 等人，2021年）研究具有弱时间依赖的线性SDS（弱平稳性（Hyvarinen 和 Morioka，2017年）），但无法识别切换次数或潜在转移。NCTRL（Song 等，2023年）为SDS建立了相关结果。该策略沿用 Balsells-Rodas 等人（2024年）提出的一阶MSM结果，并在观测层面假设高斯性。CtrlNS（Song 等，2024年）通过机制稀疏性（Lachapelle等人，2022年）和直接的观测-状态依赖实现可识别性，并假设状态数量已知。近期研究明确聚焦于状态依赖因果发现：CASTOR（Rahmani 和 Frossard，2025年）假设存在同期效应，SDCI（Balsells-Rodas 等人，2025年）则允许灵活的状态依赖；两者都依赖于不同状态之间因果结构的差异。我们的结果与这些方法互为补充，因为多滞后马尔可夫切换模型可视为一种直接针对状态依赖转移的推广。

定位。 与此前关于状态切换和时间CRL/ICA的可识别性结果相比，我们的框架具有以下主要优势：

- **多滞后MSM理论**。 据我们所知，我们首次针对多滞后马尔可夫切换模型给出了一般性的可识别性结果。我们的结果还可推广到标准MSM之外的情形（备注9）。
- **函数层面的假设**。 关于时间CRL（Yao 等人，2022b、2022a；Song 等，2023年、2024年）和ICA（Morioka等人，2021年；Hälvä 等人，2021年）的相关工作，通常依赖分布或尾部假设，而这些假设在设计模型时难以验证。相反，我们直接施加可通过神经网络参数化实现的函数约束，从而使模型设定更易于控制。
- **未知状态数**。 HMM-ICA（Hälvä 和 Hyvarinen，2020年）以及CtrlNS（Song 等，2024年）等方法假设状态数已知。通过扩展经典的有限混合理论（Yakowitz 和 Spragins，1968 年），我们的结果不要求预先知道状态数。这在实践中十分重要，因为我们可以通过模型选择（McLachlan 和 Peel，2000年）来确定状态数；而当状态数被假定为已知时，这是无法实现的。

3 背景

3.1 有限混合模型

我们在第 4.1 节提出的理论框架扩展了 Yakowitz and Spragins (1968) 的有限混合模型结果，该结果表明，混合成分的线性无关性可确保可识别性。考虑一个分布族，其函数定义在 $\boldsymbol{z} \in \mathbb{R}^m$ 上，

$$\mathcal{F}_{\mathcal{A}} := \{F_a(\boldsymbol{z}) \mid a \in \mathcal{A}\}, \tag{1}$$

其中， $F_a(\boldsymbol{z})$ 是一个 m 维累积分布函数。假设指标集 $\mathcal{A}$ 满足： $F_a(\boldsymbol{z})$ 作为 $(\boldsymbol{z},\ a)$ 的函数，在 $\mathbb{R}^m \times \mathcal{A}$ 上可测。我们引入族 $\mathcal{F}_{\mathcal{A}}$ 的有限混合下的线性无关性这一概念。

定义 1 若函数族 $\mathcal{F}_{\mathcal{A}}$ （式 （1））在有限混合下包含线性无关函数，则称其具有有限混合下的线性无关性；也就是说，对于任意满足 $|\mathcal{A}_0| < +\infty$ 的 $\mathcal{A}_0 \subset \mathcal{A}$，$\{F_a(\boldsymbol{z}) \mid a \in \mathcal{A}_0\}$ 中的函数均线性无关。

上述定义对函数族的线性无关性要求较弱，因为它允许由无穷多个其他函数构成的线性组合产生线性相关。考虑以下有限混合分布族：

$$\mathcal{H}_{\mathcal{A}} := \left\{ H(\boldsymbol{z}) = \sum_{i=1}^{K} c_i F_{a_i}(\boldsymbol{z}) \mid c_i > 0,\ a_i \in \mathcal{A},\ a_i \neq a_j, \forall i \neq j,\ \sum_{i=1}^{K} c_i = 1,\ K < +\infty \right\}, \tag{2}$$

该分布族定义于凸包 $\mathcal{F}_{\mathcal{A}}$ （conv($\mathcal{F}_{\mathcal{A}}$)）之上，或者说，是由$\mathcal{F}_{\mathcal{A}}$中的累积分布函数通过具有正系数的线性组合构成的。现在，我们按照 Yakowitz and Spragins (1968) 的定义，定义可识别有限混合族 。

定义 *2* 若满足以下条件，则称有限混合族 $\mathcal{H}_{\mathcal{A}}$ 在置换意义下可识别：~对于任意两个有限混合 $H(\boldsymbol{z}) = \sum_{i=1}^{K} c_i F_{a_i}(\boldsymbol{z})$和 $\tilde{H}(\boldsymbol{z}) = \sum_{i=1}^{K} \tilde{c}_i F_{\tilde{a}_i}(\boldsymbol{z})$， $H(\boldsymbol{z}) = \tilde{H}(\boldsymbol{z})$ ~对所有 $\boldsymbol{z} \in \mathbb{R}^m$成立，当且仅当 $K = K$，并且对于每个 $1 \leq i \leq K$，都存在某个 $1 \leq j \leq K$，使得 $c_i = \tilde{c}_j$ 且 $F_{a_i}(\boldsymbol{z}) = F_{\tilde{a}_j}(\boldsymbol{z})$对所有 $\boldsymbol{z} \in \mathbb{R}^m$成立。

接下来，有限混合模型的可识别性表述如下。

命题 3 （*Yakowitz* 和 *Spragins*，*1968* 年）有限混合分布族 $\mathcal{H}_{\mathcal{A}}$ 在置换意义下可辨识，当且仅当 $\mathcal{F}_{\mathcal{A}}$ 中的函数在有限混合下线性无关。

对于本文的目的，上述结果可以推广到定义在 $\boldsymbol{z} \in \mathbb{R}^m$ 上的概率密度函数。考虑一个等价的概率密度函数族：

$$\tilde{\mathcal{F}}_{\mathcal{A}} = \left\{ f_a(\boldsymbol{z}) = \frac{\partial^m}{\partial z_1 \dots \partial z_m} F_a(\boldsymbol{z}) \mid a \in \mathcal{A} \right\}. \tag{3}$$

那么，在假设所有 F_a 都绝对连续且几乎处处可微的条件下，族 $\mathcal{F}_{\mathcal{A}}$ 在有限混合下线性无关，当且仅当相应的 ~族 $\mathcal{F}_{\mathcal{A}}$ 在有限混合下线性无关。

~**命题 4** 考虑式 (3) 中的族 $\mathcal{F}_{\mathcal{A}}$ 。那么，有限混合分布~族在置换意义下可辨识，当且仅当 $\mathcal{F}_{\mathcal{A}}$ 中的函数在有限混合下线性无关。

3.2 可识别潜变量模型

在非时间情形下，许多研究探讨了潜变量模型的可识别性（Khemakhem 等人，2020年；Kivva等（2022））。具体而言，考虑一个生成模型，其中潜变量 $\boldsymbol{z} \in \mathbb{R}^m$ 取自包含 K 个成分的高斯混合先验（$K < +\infty$）。然后通过（有噪声的）分段线性映射对 $\boldsymbol{z}$ 进行变换，从而得到观测 $\boldsymbol{x} \in \mathbb{R}^n$， $n \geq m$：

$$\boldsymbol{x} = f(\boldsymbol{z}) + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{z} \sim p(\boldsymbol{z}) := \sum_{k=1}^{K} p(s=k) \mathcal{N}(\boldsymbol{z} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k). \tag{4}$$

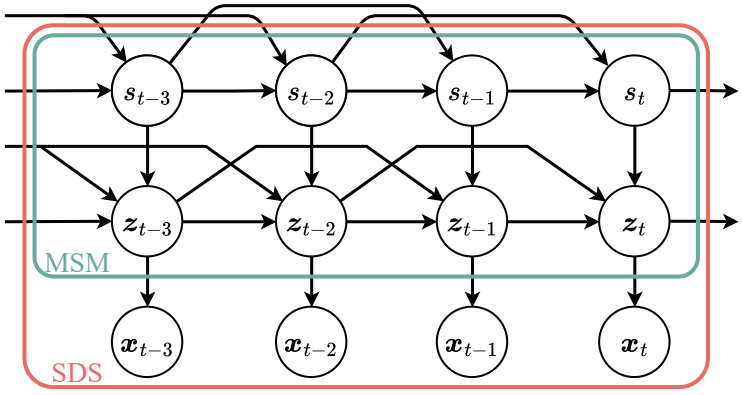


图 1：本文所考虑的生成模型。MSM（绿色）将z_t 直接视为观测，而 SDS（红色）则将z_t 转换为观测到的 x_t。

Kivva等（2022）证明，如果变换 f 是弱单射（见下文定义），那么先验分布 $p(z)$ 可根据观测识别到仿射变换[2] 。如果进一步假设 f 连续且单射，那么先验分布 $p(z)$ 和非线性映射 f 均可识别到仿射变换。在第4.2节中，我们采用类似的证明策略，将这些结果扩展到 SDS 的可识别性。

记号。对于 $(x_1^*, \ldots, x_M^*) \in \mathbb{R}^{nM}$和 $M < +\infty$，我们定义 $B(x_1^*, \ldots, x_M^*, \delta) := \{(x_1, \ldots, x_M) \in \mathbb{R}^{nM} : ||(x_1, \ldots, x_M) - (x_1^*, \ldots, x_M^*)|| < \delta\}$，其中 $||\cdot||$表示 $\mathbb{R}^{nM}$上的欧几里得范数。此外，我们将函数记号扩展到集合，即对于集合 $\mathcal{S}$，$f(\mathcal{S}) := \{f(x) : x \in \mathcal{S}\}$，以及 $f^{-1}(\mathcal{S}) := \{x : f(x) \in \mathcal{S}\}$。

定义 5 若映射 $f : \mathbb{R}^m \to \mathbb{R}^n$ 满足以下条件，则称其为弱单射：

(i) there exists $x_0 \in \mathbb{R}^n$ *and* $\delta > 0$ *s.t.* $|f^{-1}(\{x\})| = 1$ *for every* $x \in B(x_0, \delta) \cap f(\mathbb{R}^m)$*, and*

(ii) $\{x \in \mathbb{R}^n : |f^{-1}(\{x\})| > 1\} \subseteq f(\mathbb{R}^m)$ 相对于 $f(\mathbb{R}^m)$上的勒贝格测度为零。

定义 6 f是单射，如果 $|f^{-1}(\{x\})| = 1, \forall x \in f(\mathbb{R}^m)$。

3.3 切换动力系统

切换动力系统（SDS）的示例图模型如图 1 所示，它是一种序贯潜变量模型，其动力学由离散和连续潜状态共同决定，分别为 $s_t \in \{1, \ldots, K\}$， $z_t \in \mathbb{R}^m$。在每个时间步 t，离散状态 s_t 决定当前连续潜变量 z_t 应遵循的动力学先验分布的状态，而观测 x_t 则由 z_t 经过（带噪声的）非线性变换生成。因此，状态和观测变量的概率模型如下：

$$p_{\theta}(x_{1:T}, z_{1:T}, s_{1:T}) = p_{\theta}(s_{1:T}) p_{\theta}(z_{1:T} \mid s_{1:T}) \prod_{t=1}^{T} p_{\theta}(x_t \mid z_t), \tag{5}$$

2. 参见 Kivva 等（2022）中的定义 2.2，或适用于 SDS 的等价定义（定义 11）。

其中，我们用 $\boldsymbol{\theta}$ 表示模型参数。由于 $p_{\boldsymbol{\theta}}(\boldsymbol{x}_{1:T})$ 难以处理，近期研究（Dong 等人，2020年；Ansari 等人，2021年）基于变分推断（Kingma 和 Welling，2014年）开发了推断技术。这些研究考虑了从 $\boldsymbol{x}_t$ 到切换变量 s_t 的额外依赖关系，以实现更具表现力的分段；为简化起见，我们的理论分析未考虑这一点。

潜在动态先验 $p_{\boldsymbol{\theta}}(\boldsymbol{z}_{1:T})$ 考虑了一种马尔可夫切换模型（MSM），该模型也被称为自回归隐马尔可夫模型（Hamilton，1989；Ephraim 和 Roberts，2005）。此类先验使用潜变量“切换” s_t 来对每个时间步的 $\boldsymbol{z}_t$ 分布进行条件化，而给定 $\boldsymbol{s}_{1:T}$ 的 $\boldsymbol{z}_{1:T}$ 条件动态模型遵循自回归过程，其中似然可通过边缘化 $\boldsymbol{s}_{1:T}$ 进行计算。为简单起见，我们考虑 $\boldsymbol{s}_{1:M} := s_M$ 和 $p_{\boldsymbol{\theta}}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_M \mid s_M)$ 作为初始 M 个观测的分布，并将滞后 M 阶的 MSM 分解如下：

$$p_{\boldsymbol{\theta}}(\boldsymbol{z}_{1:T}) = \sum_{\boldsymbol{s}_{M:T}} p_{\boldsymbol{\theta}}(\boldsymbol{s}_{M:T}) p_{\boldsymbol{\theta}}(\boldsymbol{z}_{1:T} \mid \boldsymbol{s}_{M:T}),$$

$$p_{\boldsymbol{\theta}}(\boldsymbol{z}_{1:T} \mid \boldsymbol{s}_{M:T}) = p_{\boldsymbol{\theta}}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_M \mid s_M) \prod_{t=M+1}^{T} p_{\boldsymbol{\theta}}(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, s_t). \quad (6)$$

请注意，离散潜在状态先验 $p_{\boldsymbol{\theta}}(\boldsymbol{s}_{M:T})$ 的结构并未指定，下一节呈现的可识别性结果也不要求在此作出进一步假设。在实验中，我们对 $p_{\boldsymbol{\theta}}(\boldsymbol{s}_{M:T})$ 使用一阶马尔可夫过程，并由转移矩阵 $Q \in \mathbb{R}^{K \times K}$ 描述，使得 $p_{\boldsymbol{\theta}}(s_t = j \mid s_{t-1} = i) = Q_{ij}$，且初始分布$p_{\boldsymbol{\theta}}(s_M = i) = \pi_i$。

4 可识别性分析

在本节中，我们将在适当的假设下，证明第 3.3 节所述 SDS（式（5））的可识别性。首先，我们将第 3.1 节介绍的有限混合模型可识别性结果扩展到具有有限滞后阶数 M 的马尔可夫切换模型。随后，我们借鉴 Kivva 等（2022）的思路；他们将有限混合先验用于非时序生成模型。重要的是，该理论依赖于使用可识别的有限混合先验（在混合分量置换意义下）。

可识别性： 形式上，可识别性建立了概率分布 $p_{\boldsymbol{\theta}}(\boldsymbol{x})$ 与模型参数 $\boldsymbol{\theta}$之间的单射对应关系：

$$p_{\boldsymbol{\theta}}(\boldsymbol{x}) = p_{\tilde{\boldsymbol{\theta}}}(\boldsymbol{x}) \iff \boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}. \quad (7)$$

鉴于我们的模型由深度神经网络参数化，由于网络对称性（即不同的参数映射到同一函数），参数的显式可识别性是不可能实现的。因此，本文从函数形式和模型混合成分的角度讨论可识别性。

4.1 可识别的高阶马尔可夫切换模型

符号说明。对于集合的重复笛卡尔积和函数族的逐点乘积，我们使用以下符号：

$$\times^N \mathcal{A} := \underbrace{\mathcal{A} \times \cdots \times \mathcal{A}}_{N \text{ times}}, \quad \otimes^N \mathcal{P} := \underbrace{\mathcal{P} \otimes \cdots \otimes \mathcal{P}}_{N \text{ times}}, \tag{8}$$

其中， $\mathcal{A}$ 是一个指标集， $\mathcal{P}$ 是一个函数族。 $\mathcal{P}$ 中的函数可能共享重叠变量，从而产生定义在联合定义域中的函数。例如，考虑两个函数族

$$\mathcal{P}_{1,\mathcal{A}_1} = \{p_{1,a_1}(\boldsymbol{z}_1, \boldsymbol{z}_2) : a_1 \in \mathcal{A}_1\}, \quad \mathcal{P}_{2,\mathcal{A}_2} = \{p_{2,a_2}(\boldsymbol{z}_2, \boldsymbol{z}_3) : a_2 \in \mathcal{A}_2\}, \tag{9}$$

其中 $(\boldsymbol{z}_1, \boldsymbol{z}_2), (\boldsymbol{z}_2, \boldsymbol{z}_3) \in \mathbb{R}^{2m}$。然后，它们的逐点乘积定义为：

$$\mathcal{P}_{1,\mathcal{A}_1} \otimes \mathcal{P}_{2,\mathcal{A}_2} := \Big\{ p_{1,a_1}(\boldsymbol{z}_1, \boldsymbol{z}_2) p_{2,a_2}(\boldsymbol{z}_2, \boldsymbol{z}_3) \mid (a_1, a_2) \in \mathcal{A}_1 \times \mathcal{A}_2 \Big\}, \tag{10}$$

其中，函数定义在 $(\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3) \in \mathbb{R}^{3m}$中。对于从 1 到 K 的指标集，我们也记作 $[K] := \{1, \ldots, K\}$。

MSM 的有限混合模型视角。 为简便起见，我们省略下标 $\boldsymbol{\theta}$。MSM $p(\boldsymbol{z}_{1:T})$具有有限混合模型的等价表述，其中根据公式 (6)，给定切换序列 $p(\boldsymbol{z}_{1:T} \mid \boldsymbol{s}_{M:T})$ 时的联合分布由滞后阶数为 M 的初始分布和转移分布定义： $p(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_M \mid s_M)$ 和 $p(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, s_t)$，分别对应于 $t > M$。与 Yakowitz 和 Spragins (1968) 中的做法等价，我们首先定义对应于初始分布和转移分布的函数族。

$$\Pi_{\mathcal{A}}^M := \Big\{ p_a(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_M) \mid a \in \mathcal{A} \Big\}, \quad \mathcal{P}_{\mathcal{B}}^M := \Big\{ p_b(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}) \mid b \in \mathcal{B} \Big\}, \tag{11}$$

其中 $\mathcal{A}$ 和 $\mathcal{B}$ 是满足温和测度论条件的指标集（与第 3.1 节类似）。举例来说，我们可以将这些族视为无穷多个参数化函数，其中每个 $a \in \mathcal{A}$ 对应一种不同的参数配置 $\boldsymbol{\theta}_a$。我们使用不同的花体符号（例如， $\mathcal{A}$、 $\mathcal{B}$），因为它们分别索引具有不同结构和输出域的函数。例如，如果初始分布族是高斯族，那么集合 $\mathcal{A}$ 索引有限维参数空间中的元素（均值和协方差）。相比之下，转移族中使用的集合 $\mathcal{B}$ 索引函数空间中的元素，其输出定义了条件分布的参数。我们将在下文详细介绍参数化。

我们假设一般情形，并考虑 MSM 的初始分布有 K_0 个状态，而转移分布有 K 个状态。因此，MSM 的离散状态满足$s_t \in [K]$，其中 $K < +\infty$，且 $M < t \leq T$，初始状态为 $s_M \in [K_0]$， $K_0 < +\infty$。根据上述表述，马尔可夫切换模型考虑 K_0 初始分布$\Pi_{\mathcal{A}_0}^M$ 和 K 转移分布 $\mathcal{P}_{\mathcal{B}_0}^M$，其中 $|\mathcal{A}_0| = K_0$、 $|\mathcal{B}_0| = K$，且 $\mathcal{A}_0 \subset \mathcal{A}$、 $\mathcal{B}_0 \subset \mathcal{B}$分别成立。因此，在这些函数族下，对于每个 $k \in [K_0]$： $p(\boldsymbol{z}_{1:M} | s_M = k) = p_a(\boldsymbol{z}_{1:M})$，其中某个 $a \in \mathcal{A}_0$；对于每个 $k \in [K]$： $p(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, \ldots\ldots, \boldsymbol{z}_{t-M}, s_t = k) = p_{b_t}(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1:t-M})$，其中某个 $b_t \in \mathcal{B}_0$，且 $M < t \leq T$。换言之，这些族 $\Pi_{\mathcal{A}}^M$

以及 $\mathcal{P}_{\mathcal{B}}^{M}$包含无限多个函数，MSM 中的每个分布都与其中一个这样的函数相对应。

我们可以将联合分布 $p(\boldsymbol{z}_{1:T})$ 视为路径混合，其中每条路径由 $\boldsymbol{s}_{M:T} \in [K_0] \times [K]^{T-M}$ 的一个实现定义。因此，可以定义双射路径索引函数 $\varphi\colon [C] \to [K_0] \times [K]^{T-M}$，其混合路径总数为 $C = K_0 K^{T-M}$，使得每个 $i \in [C]$ 都能唯一检索一个状态集合 $\boldsymbol{s}_{M:T} = \varphi(i)$。然后，我们可以用 $c_i = p(\boldsymbol{s}_{M:T} = \varphi(i))$ 表示 p 下状态 $\boldsymbol{s}_{M:T} = \varphi(i)$ 的联合概率。于是，对于任意 M，T，只要满足 $M < T < +\infty$，我们就可以从有限混合模型的角度构造 M 阶 MSM 族:

$$\begin{aligned}\mathcal{M}^{T}(\Pi_{\mathcal{A}}^{M}, \mathcal{P}_{\mathcal{B}}^{M}) := \Big\{ &\sum_{i=1}^{C} c_i p_{a^i}(\boldsymbol{z}_{1:M}) \prod_{t=M+1}^{T} p_{b_t^i}(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}) \mid c_i > 0, \\ &p_{a^i} \in \Pi_{\mathcal{A}_0}^{M}, \quad p_{b_t^i} \in \mathcal{P}_{\mathcal{B}_0}^{M}, \quad M < t \le T, \quad a^i \in \mathcal{A}_0 \subset \mathcal{A}, \quad b_t^i \in \mathcal{B}_0 \subset \mathcal{B}, \quad K_0, K < +\infty, \\ &|\mathcal{A}_0| = K_0, \quad |\mathcal{B}_0| = K, \quad (a^i, b_{M+1:T}^i) \neq (a^j, b_{M+1:T}^j), \forall i \neq j, \quad \sum_{i=1}^{C} c_i = 1 \Big\}. \quad (12)\end{aligned}$$

由于式 (12) 要求对任意 $i \neq j$，均有 $(a^i,\ b_{M+1:T}^i) \neq (a^j,\ b_{M+1:\,T}^j)$，因此可构造单射映射 $\phi\colon [C] \to \mathcal{A}_0 \times (\times^{T-M}\mathcal{B}_0)$，使得 $\phi(i) = (a^i,\ b_{M+1}^i,\ \ldots,\ b_T^i)$。结合路径索引函数，我们可以用 $\phi \circ \varphi^{-1}$ 将一个状态集合 $\boldsymbol{s}_{M:T}$ 唯一映射到索引。

上述记法说明了 MSM 如何将有限混合模型扩展到时间情境，将 $p(\boldsymbol{z}_{1:T})$ 表示为 $\Pi_{\mathcal{A}_0}^{M}$ 中的一个初始分布与来自 $\mathcal{P}_{\mathcal{B}_0}^{M}$的条件分布所构成的 $K_0 K^{T-M}$轨迹的有限混合。不过，这一概念与 Yakowitz and Spragins (1968) 中的独立同分布混合模型设置有着根本区别。为形式化这种区别，我们定义以下轨迹族:

$$\begin{aligned}\mathcal{P}_{\mathcal{A},\mathcal{B}}^{T,M} := \Pi_{\mathcal{A}}^{M} \otimes (\otimes^{T-M}\mathcal{P}_{\mathcal{B}}^{M}) = \Big\{ &p_a(\boldsymbol{z}_{1:M}) \prod_{t=M+1}^{T} p_{b_t}(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}) \mid \\ &a \in \mathcal{A},\ b_t \in \mathcal{B},\ M < t \le T \Big\}. \quad (13)\end{aligned}$$

在第 3.1 节中，$\mathcal{H}_{\mathcal{A}}$ 被定义为 $\mathcal{F}_{\mathcal{A}}$的凸包 conv($\mathcal{F}_{\mathcal{A}}$)。相比之下，式 (12) 中定义的 MSM族构成 conv($\mathcal{P}^{T}{}_{,\mathcal{A}}{}^{M}_{\mathcal{B}}$,) 的真子集，因为混合分量仅限于有限子集 $\mathcal{A}_0 \subset \mathcal{A}$ 和 $\mathcal{B}_0 \subset \mathcal{B}$。另一方面，对 $\mathcal{P}^{T}{}_{,\mathcal{A}}{}^{M}_{\mathcal{B}}$,进行混合时，可以在每个时间步从 $\mathcal{B}$ 中任意选取，从而得到表达能力更强的模型族。这种推广对于下文的理论分析至关重要；我们将通过证明 $\mathcal{P}^{T}{}_{,\mathcal{A}}{}^{M}_{\mathcal{B}}$,的有限混合下具有线性无关性，来证明 MSM 的可识别性。

可辨识的马尔可夫切换模型。在确立了马尔可夫切换模型的有限混合模型视角后，本节第 4.1 节余下部分将采用 MSM 这一记号，因为我们将使用有限混合建模技术来确立其可识别性。具体而言，我们首先如下定义 $\mathcal{M}^{T}(\Pi_{\mathcal{A}}^{M},\ \mathcal{P}_{\mathcal{B}}^{M})$ 的可识别性。

定义 7 若对于任意 $p, \tilde{p} \in \mathcal{M}^T(\Pi_{\mathcal{A}}^M, \mathcal{P}_{\mathcal{B}}^M)$ 其中，$M, T \in \mathbb{Z}^+$ 和 $M < T$，阶数为 M 、长度为 T 的马尔可夫切换模型族在置换意义下可辨识，

$$p(\boldsymbol{z}_{1:T}) = \sum_{i=1}^{C} c_i p_{a^i}(\boldsymbol{z}_{1:M}) \prod_{t=M+1}^{T} p_{b_t^i}(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}), \quad C = K_0 K^{T-M};\ and$$
$$\tilde{p}(\boldsymbol{z}_{1:T}) = \sum_{i=1}^{\tilde{C}} \tilde{c}_i p_{\tilde{a}^i}(\boldsymbol{z}_{1:M}) \prod_{t=M+1}^{T} p_{\tilde{b}_t^i}(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}), \quad \tilde{C} = \tilde{K}_0 \tilde{K}^{T-M},$$

给定 $p(\boldsymbol{z}_{1:T}) = \tilde{p}(\boldsymbol{z}_{1:T})\ \forall \boldsymbol{z}_{1:T} \in \mathbb{R}^{mT}$，我们有 $C = \tilde{C}$、$K_0 = \tilde{K}_0$、$K = \tilde{K}$，并且对于每个 $1 \leq i \leq C$，都存在某个 $1 \leq j \leq C$，使得：

1. $c_i = \tilde{c}_j$;
2. *if* $b_{t_1}^i = b_{t_2}^i$ *for* $M < t_1, t_2 \leq T$, $t_1 \neq t_2$, *then* $\tilde{b}_{t_1}^j = \tilde{b}_{t_2}^j$;
3. $p_{a^i}(\boldsymbol{z}_{1:M}) = p_{\tilde{a}^j}(\boldsymbol{z}_{1:M}),\ \forall \boldsymbol{z}_{1:M} \in \mathbb{R}^{mM}$;
4. $p_{b_t^i}(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}) = p_{\tilde{b}_t^j}(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}),\ \forall (\boldsymbol{z}_t, \ldots, \boldsymbol{z}_{t-M}) \in \mathbb{R}^{m(M+1)}$.

第二个要求消除了置换等价性，例如 $\boldsymbol{s}_{1:4} = (1,2,3,2)$；$\tilde{\boldsymbol{s}}_{1:4} = (3,1,2,3)$，这种情况在独立同分布有限混合情形下是成立的。

为了构建深度生成模型，我们寻求定义可辨识的参数族，并将非参数情形的研究留待附录 A。特别是，我们采用如下非线性高斯转移族：

$$\mathcal{G}_{\mathcal{B}}^M = \Big\{ p_b(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}) = \mathcal{N}(\boldsymbol{z}_t \mid \boldsymbol{m}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, b), \boldsymbol{\Sigma}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, b) \mid b \in \mathcal{B} \Big\}, \quad (14)$$

其中 $\boldsymbol{m}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, b)$：$\mathbb{R}^{mM} \to \mathbb{R}^m$，且 $\boldsymbol{\Sigma}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, b)$：$\mathbb{R}^{mM} \to \mathbb{R}^{m \times m}$，它们关于 $(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M})$ 是非线性的，并分别表示均值和协方差矩阵。为使每个 p_b 都是密度，我们假设对所有 $b \in \mathcal{B}$，$\boldsymbol{\Sigma}(\cdot, b) \succ 0$。我们还要求满足以下**唯一索引** 假设：

$$\forall b \neq b' \in \mathcal{B},\ \exists \mathcal{Z} \subset \mathbb{R}^{mM}, \mu(\mathcal{Z}) > 0,\ s.t.\ \forall (\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}) \in \mathcal{Z},\ \boldsymbol{m}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, b) \neq \boldsymbol{m}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, b') \text{ or } \boldsymbol{\Sigma}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, b) \neq \boldsymbol{\Sigma}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, b'), \quad (15)$$

换言之，对于这样的 $(\boldsymbol{z}_{t-1}, \ldots..., \boldsymbol{z}_{t-M})$，$p_b(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, \ldots..., \boldsymbol{z}_{t-M})$ 和 $p_{b'}(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, \ldots..., \boldsymbol{z}_{t-M})$ 在至少一个非零测度集上是两个不同的高斯分布。我们还引入一个具有唯一索引的初始分布族：

$$\mathcal{I}_{\mathcal{A}}^M := \Big\{ p_a(\boldsymbol{z}_{1:M}) = \mathcal{N}(\boldsymbol{z}_{1:M} \mid \boldsymbol{\mu}(a), \boldsymbol{\Sigma}_1(a)) \ \mid \ a \in \mathcal{A} \Big\}, \quad (16)$$

$$\forall a \neq a' \in \mathcal{A} \Leftrightarrow \boldsymbol{\mu}(a) \neq \boldsymbol{\mu}(a') \text{ or } \boldsymbol{\Sigma}_1(a) \neq \boldsymbol{\Sigma}_1(a'), \quad (17)$$

其中，同理，对于所有 $a \in \mathcal{A}$，$\boldsymbol{\Sigma}_1(a) \succ 0$。上述高斯族与唯一索引假设相结合，满足有利于 MSM 可识别性的条件。

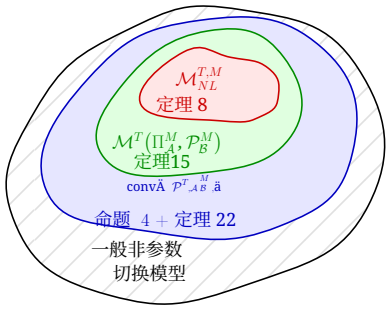


图 2：红色区域展示了我们关于具有高斯转移的参数化 MSM 的主要结果（第 4.1 节）。绿色区域表示我们针对非参数 MSM 得出的更一般的可辨识性结果（附录 A）。蓝色区域强调，我们的证明技巧适用于族包含在 conv($\mathcal{P}^T{}_{,\mathcal{A}}{}_{\mathcal{B}}^{M}$,) 中的任意切换模型。

定理8（参数化马尔可夫切换模型的可辨识性）考虑公式（*12*）中定义的阶数为 M、具有非线性高斯族 $\mathcal{M}^T{}_{,}{}_{NL}^{M}$: $=\mathcal{M}^T$（$\mathcal{I}_{\mathcal{A}}^M$、$\mathcal{G}_{\mathcal{B}}^M$）的马尔可夫切换模型族，其中 $\mathcal{G}_{\mathcal{B}}^M$、 $\mathcal{I}_{\mathcal{A}}^M$分别由公式（*14*）、（*16*）定义。则，根据定义*7*，在以下假设下，该 *MSM* 是可辨识的：

（*m1*） $\mathcal{G}_{\mathcal{B}}^M$和 $\mathcal{I}_{\mathcal{A}}^M$具有唯一索引：公式（*15*）和（*17*）成立；

（*m2*） $\mathcal{G}_{\mathcal{B}}^M$、 $\boldsymbol{m}$（·、 b）： $\mathbb{R}^{mM} \to \mathbb{R}^m$ 和 $\boldsymbol{\Sigma}$（·、 b）： $\mathbb{R}^{mM} \to \mathbb{R}^{m\times m}$中的均值和协方差对于任意 $b \in \mathcal{B}$均为解析函数。

证明概要：完整证明请参见附录 A。其策略可概括为以下几个步骤。更多直观解释和细节将在一系列备注之后、本节末尾处给出。

1. 从有限混合模型的视角来看，只需证明轨迹族 $\mathcal{P}^T{}_{,\mathcal{A}}{}_{\mathcal{B}}^{M}$, 在有限混合下线性无关（Yakowitz 和 Spragins，1968 年）。由于 $\mathcal{M}^T(\Pi_{\mathcal{A}}^M, \mathcal{P}_{\mathcal{B}}^M)$ 是 conv($\mathcal{P}^T{}_{,\mathcal{A}}{}_{\mathcal{B}}^{M}$,) 的子集，conv($\mathcal{P}^T{}_{,\mathcal{A}}{}_{\mathcal{B}}^{M}$,) 的可识别性意味着 $\mathcal{M}^T(\Pi_{\mathcal{A}}^M, \mathcal{P}_{\mathcal{B}}^M)$ 的可识别性。

2. 为了证明 $\mathcal{P}^T{}_{,\mathcal{A}}{}_{\mathcal{B}}^{M}$, 在有限混合下线性无关，我们从 $T = 2M$ 出发，并利用非参数分量 $\mathcal{P}_{\mathcal{B}}^M$ 和 $\Pi_{\mathcal{A}}^M$ 构造一族联合密度：

$$\left\{ p_a(\boldsymbol{z}_{1:M}) \prod_{t=M+1}^{2M} p_{b_t}(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}) \mid a \in \mathcal{A},\ b_t \in \mathcal{B},\ M < t \le 2M \right\}.$$

我们在引理 21 中证明，这一族在有限混合下线性无关。

3. 对于 $T > 2M$，我们采用归纳法（定理 22）。在归纳假设下，我们重复使用引理 21 的结果，将线性无关性推广到更长的轨迹。

4. 我们证明，在假设 (m1-m2) 下，高斯族 $\mathcal{I}_{\mathcal{A}}^M$ 和 $\mathcal{G}_{\mathcal{B}}^M$ 分别满足为 $\Pi_{\mathcal{A}}^M$ 和 $\mathcal{P}_{\mathcal{B}}^M$ 定义的非参数 情形。

备注9 我们的策略通过证明有限混合下轨迹族 $\mathcal{P}^T{}_{,\mathcal{A}}{}_{\mathcal{B}}^{M}$,的线性无关性（定理 *22*）来确立可识别性；这一结果与命题 *4* 相结合，确保了对于其族包含于conv($\mathcal{P}^T{}_{,\mathcal{A}}{}_{\mathcal{B}}^{M}$,)的任意切换模型，都能根据定义 *2* 确认可识别性。图 *2* 展示了由此形成的层级：

- 红色区域： 主要结果（定理 *8*），涵盖具有高斯转移的参数化 *MSM*。
- 绿色区域： 推广至非参数 *MSM*（定理 *15*），其中也包括非高斯转移族。

- 蓝色区域： 我们方法的完整理论适用范围，涵盖族包含于$\mathrm{conv}(\mathcal{P}^{T}_{\mathcal{A}}{}^{M}_{\mathcal{B}},)$中的任何切换模型，例如具有时变分量的 *MSM* $K(t)$。

直观而言，“切换”可视为动态模型中的不连续性，并完全由离散状态控制。假设 (m2) 意味着 $\boldsymbol{m}(\cdot, b)$ 和 $\boldsymbol{m}(\cdot, b')$ 与$b \neq b'$ 只能在勒贝格测度为零的点集上相交（ $\boldsymbol{\Sigma}(\cdot, b)$也同样如此），而这一光滑性性质有助于连续状态转移 $p_b(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M})$ $b \in \mathcal{B}$的可识别性。定理 8 的以下推论结合起来，展现了可辨识 MSM 的强大表达能力：

- 假设 (m2) 使得可以用多项式和具有解析激活函数（例如 Softplus）的神经网络（NN）等来参数化 $p_b(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, \ldots \boldsymbol{z}_{t-M})$。如前所述，对于神经网络，可识别性仅适用于函数形式，因为神经网络权重无法唯一地确定神经网络函数。
- 可识别性结果不受 $p(\boldsymbol{s}_{M:\,T})$ 选择的影响，因此允许非平稳性和高阶马尔可夫依赖等情形。

我们的实验对 $p(\boldsymbol{s}_{M:T})$ 使用平稳一阶马尔可夫先验。根据 Balsells-Rodas等人（2024年，推论 C.1），其转移参数在置换意义下可识别。其他情形留待未来研究。

非参数可识别性。我们证明技术的关键在于确立证明概要中的步骤 (2)，该步骤通过一项新的理论结果得以证明：具有 M 个连续重叠变量的乘积函数的线性无关性（LIPO-M；引理 21），这一结果本身也可能具有独立的研究价值。与隐马尔可夫模型（HMMs）不同，在马尔可夫切换模型（MSMs）中， $\boldsymbol{z}_t$和 $\boldsymbol{z}_{t+1}$ 在给定离散状态的条件下并非条件独立（本例中假设 $M = 1$ ）。因此，经典隐马尔可夫模型可识别性结果（Allman等人，2009年；Gassiat等人，2016年）利用 HMMs 中 $\boldsymbol{z}_t$ 和 $\boldsymbol{z}_{t+1}$ 的条件独立性，并采用 Kruskal 的“三维数组”技术（Kruskal，1977 年），因而不适用于非参数 MSM，这使得发展新理论成为必要。

线性无关性（情形 $M = 1$）。为了解如何在包含 M 个连续变量的乘积上实现有限混合下的线性无关性，我们从 Balsells-Rodas 等人（2024）提出的 $M = 1$(LIPO-1) 入手。为完整起见，我们在附录 A.2.1（引理 16）中给出该结果。考虑步骤 (2) 上的一阶轨迹族，即 $T = 2$:

$$\Pi^1_{\mathcal{A}} \otimes \mathcal{P}^1_{\mathcal{B}} = \left\{ p_a(\boldsymbol{z}_1) p_{b_2}(\boldsymbol{z}_2 \mid \boldsymbol{z}_1) \mid (a, b_2) \in \mathcal{A} \times \mathcal{B} \right\}, \tag{18}$$

其中，重叠变量 $\boldsymbol{z}_1$ 对有限混合下的线性无关性构成挑战。鉴于族 $\Pi^1_{\mathcal{A}}$ 上满足有限混合下的线性无关性（a3 – 引理 16），重叠变量 $\boldsymbol{z}_1$ 会使 $\mathcal{P}^1_{\mathcal{B}}$ 与 $\Pi^1_{\mathcal{A}}$ 相乘后的有限混合下的线性无关性变得复杂。为解决这一问题，需要在 $\mathcal{P}^1_{\mathcal{B}}$ 的特定非零测度子集上施加有限混合下的线性无关性（a4 – 引理 16），并要求线性相关意味着函数重复（a5– 引理 16）；此外，还需要 $\boldsymbol{z}_1$ 的连续性（a6– 引理 16）。图3展示了这一

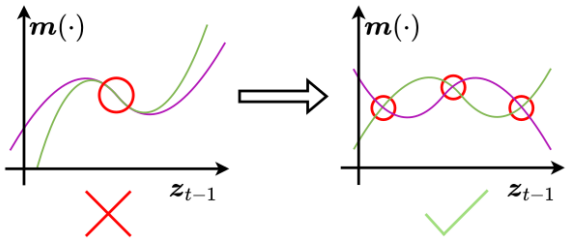


图3：引理 16 背后的直觉：如果对于任意一对函数（以绿色和紫色显示），它们在条件变量 $\boldsymbol{z}_{t-1}$ 的定义域上的交集测度为零，则线性无关性成立。

结果背后的直观理解：只要线性相关仅发生在零测度集合中，乘积族就可以满足有限混合下的线性无关性。对于情形 $M = 1$，上述条件需要在重叠变量空间中至少一个非零测度集上成立。

线性无关性（一般情形）。我们将结果推广到任意滞后阶数 $M < T$ 的策略，是将先前的结果扩展到涵盖多个重叠变量。挑战在于，自回归阶数 M 增大时，连续分布之间的重叠变量数量也会增加。对于 $M = 1$，每个新因子只与前一个因子共享一个变量，因此重叠始终是“局部的”，可以直接加以控制。然而，对于 $M > 1$，单个变量可能同时与多个不同因子重叠。这使有限混合下线性无关性的证明更加复杂，因为我们不再像式（18）那样只有一个重叠。举例来说，对于 $M = 3$，$T = 6$ 处的轨迹族具有如下结构：

$$\Pi^3_{\mathcal{A}} \otimes (\otimes^3 \mathcal{P}^3_{\mathcal{B}}) = \Big\{ p_a(\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3) p_{b_4}(\boldsymbol{z}_4 \mid \boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3) p_{b_5}(\boldsymbol{z}_5 \mid \boldsymbol{z}_4, \boldsymbol{z}_3, \boldsymbol{z}_2) p_{b_6}(\boldsymbol{z}_6 \mid \boldsymbol{z}_5, \boldsymbol{z}_4, \boldsymbol{z}_3) \Big\}, \tag{19}$$

对于每个 $a \in \mathcal{A}$ 以及 $b_4, b_5, b_6 \in \mathcal{B}$，我们用不同颜色标示重叠变量。与式（18）中的 $M = 1$ 相比，在相乘初始分布和转移分布时，$p_a(\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3) p_{b_4}(\boldsymbol{z}_4 \mid \boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3)$，重叠部分（$\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3$）的结构与 $M = 1$ 情形相同，因此可通过 LIPO-1 建立 $T = 4$ 的线性无关性。然而在 $T = 5$处，下一个因子 $p_{b_5}(\boldsymbol{z}_5 \mid \boldsymbol{z}_4, \boldsymbol{z}_3, \boldsymbol{z}_2)$引入了重叠部分（$\boldsymbol{z}_4, \boldsymbol{z}_3, \boldsymbol{z}_2$），它不再局限于一对相邻项中，因为（$\boldsymbol{z}_3, \boldsymbol{z}_2$）出现在$p_a(\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3)$中。这种分支重叠违反了 LIPO-1 的假设（a3），因为乘积族无法直接化简为式（18）中的单重叠结构。对于 $T = 6$，也有类似的论证，其中 $\boldsymbol{z}_3$ 与所有函数都重叠。

为确保轨迹族在任意滞后阶数M下具有有限混合下的线性无关性，我们对变量进行分组，以使重叠仅出现在连续分布之间（类似于式（18））。例如，$M = 3$ 的示例如下：

$$\Pi^3_{\mathcal{A}} \otimes (\otimes^3 \mathcal{P}^3_{\mathcal{B}}) = \Big\{ p_a(\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3) p_{b_{4:6}}(\boldsymbol{z}_4, \boldsymbol{z}_5, \boldsymbol{z}_6 \mid \boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3) \mid a \in \mathcal{A}, b_{4:6} \in (\times^3 \mathcal{B}) \Big\}. \tag{20}$$

与情形 $M = 1$ 一样，只要乘积（$\otimes^3 \mathcal{P}^3_{\mathcal{B}}$）满足 LIPO-1 中的假设（a4–a6），有限混合下的线性无关性便成立。因此，我们的策略是在 $\mathcal{P}^M_{\mathcal{B}}$ 上施加这些假设，并探究它们是否能扩展到所有 $1 < r \leq M$ 对应的乘积族（$\otimes^r \mathcal{P}^M_{\mathcal{B}}$）。在附录A.2.2中，我们证明，这种扩展要求将非参数假设从（a4–a6）加强为（b4–b6）（引理 19 和 20）。值得注意的是，假设（b4）要求族 $\mathcal{P}^M_{\mathcal{B}}$ 在

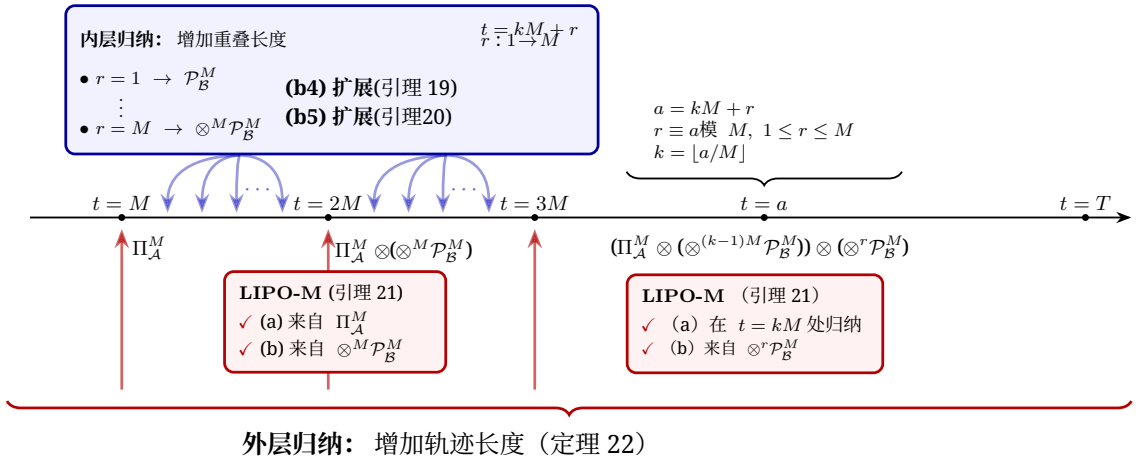


图 4：通过嵌套归纳策略证明 $\mathcal{P}^{T}$、$\mathcal{A}\mathcal{B}^{M}$、在有限混合下的线性无关性。内层归纳（蓝色方框）构造乘积族（$\otimes^{r}\mathcal{P}_{\mathcal{B}}^{M}$、 $1 \le r \le M$），而外层归纳（红色方框）则增加轨迹族的长度。

$\times^{M}\mathbb{R}^{m}$ 的全测度子集的每个子集上满足有限混合下的线性无关性；而如前文针对 $M = 1$ 所述，只需 $\mathcal{Y} \subset \mathbb{R}^{m}$ 是非零测度子集即可。尽管这些假设更强，但它们通过定理 27 中的（d3）与非线性高斯相关联。它们也与（m2）相关，因为（m2）确保不同函数几乎处处不同。

通过嵌套归纳证明线性无关性。为了证明 LIPO-M 并将假设推广到乘积族（引理 19 和 20），我们采用归纳法，基例是连续变量的单个乘积（$\mathcal{P}_{\mathcal{B}}^{M} \otimes \mathcal{P}_{\mathcal{B}}^{M}$）。因此，要完整证明轨迹族在有限混合下的线性无关性，需要使用嵌套归纳（见附录 A.2.4 中的定理 22）。图 4 阐释了这一思路：

- 内层归纳通过增加连续转移分布 r 的数量，构造具有 $1 \le r \le M$ 的乘积族（$\otimes^{r}\mathcal{P}_{\mathcal{B}}^{M}$）。
- 外层归纳增加轨迹族 $\mathcal{P}^{M}$、$\mathcal{A}\mathcal{B}^{t}$、的长度，其中 $M < t \le T$，确保在引入更多时间步时处理重叠。

定理 22 中的一个关键技术点是，归纳假设仅在轨迹长度 t 是 M 的倍数时适用。因此我们区分两种情形：

- 如果 t 是 M 的倍数，我们在 $t - M$ 处应用归纳假设，然后使用 LIPO-M 将时间 $t - M$ 处的轨迹族与来自 $\mathcal{P}_{\mathcal{B}}^{M}$ 的 M 个连续乘积结合起来。
- 如果 t 不是 M 的倍数，计算 $r \equiv t$ 除以 M 的余数。我们在 $t - r$ 处应用归纳假设（它是 M 的倍数），然后使用 LIPO-M 将时间 $t - r$ 处的轨迹族与额外的 r 个乘积（$\otimes^{r}\mathcal{P}_{\mathcal{B}}^{M}$）结合起来。

同样的逻辑也适用于基例 $M < t \le 2M$：通过将初始族 $\Pi_{\mathcal{A}}^{M}$与来自 $\mathcal{P}_{\mathcal{B}}^{M}$的 $t - M$ 个连续乘积相结合来应用 LIPO-M。

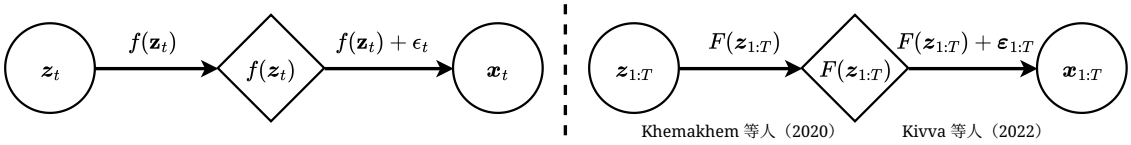


图 5：我们假设 z_t 在每个时间步 t 都独立地经由 f 转换，并受到噪声 ε_t 的影响。我们将其视为通过带噪声 $\varepsilon_{1:T}$ 的分解 F 对 $z_{1:T}$ 进行的变换。

4.2 可辨识的切换动力系统

基于可辨识的 MSM，我们对切换动力系统的分析将 Kivva等（2022）（第3节.2）中的设定扩展到了时间情形。假设先验动态模型 $p(z_{1:T})$ 属于由定理 8 指定的非线性高斯 MSM 族 $\mathcal{M}^{T,}{}_{NL}^{M}$。在每个时间步 t，潜变量 $z_t \in \mathbb{R}^m$ 通过分段线性变换 f 生成观测数据 $x_t \in \mathbb{R}^n$。此类 SDS 模型的生成过程可表示如下：

$$x_t = f(z_t) + \varepsilon_t, \quad \varepsilon_t \sim p_\varepsilon, \quad z_{1:T} \sim p(z_{1:T}) \in \mathcal{M}_{NL}^{T,M}. \tag{21}$$

请注意，我们也可以将解码过程写作 $x_{1:T} = F(z_{1:T}) + \varepsilon_{1:T}$，其中 F 是由下文定义的 f 组成的分解变换。重要的是，这一概念使 F 能够继承 f 的性质，例如分段线性和弱单射性。

定义 10我们称函数 $F: \mathbb{R}^{mT} \to \mathbb{R}^{nT}$ 是可分解的，如果它由$f: \mathbb{R}^m \to \mathbb{R}^n$构成，使得对于任意 $z_{1:T} \in \mathbb{R}^{mT}$, $F(z_{1:T}) = (f(z_1) \ldots, f(z_T))^\top$。

SDS 模型（式 (21)）的可识别性分析分为两个步骤（见图 5）。(i) 按照 Kivva等（2022）的方法，我们证明在无噪声情形下，可以从 $(F_\# p)(x_{1:T})$ 中识别先验分布 $p(z_{1:T}) \in \mathcal{M}^{T,}{}_{NL}^{M}$。这里，$F_\# p$ 表示由 F 对 p 诱导的推前测度。(ii) 当噪声 $\varepsilon_{1:T}$ 的分布已知时，利用 Khemakhem 等人（2020）的卷积技巧[3]，可以从 $((F_\# p) * p_{\varepsilon_{1:T}})(x_{1:T})$ 中识别 $F_\# p$。具体而言，我们首先根据无噪声观测定义 $p \in \mathcal{M}^T(\Pi_{\mathcal{A}}^M, \mathcal{P}_{\mathcal{B}}^M)$ 的可识别性概念。

定义 11 给定一个分解变换族 $\mathbb{F}$，对于 $F \in \mathbb{F}$ 先验分布 $p \in \mathcal{M}^T(\Pi_{\mathcal{A}}^M, \mathcal{P}_{\mathcal{B}}^M)$，如果对于任意 $\tilde{F} \in \mathbb{F}$和 $\tilde{p} \in \mathcal{M}^T(\Pi_{\mathcal{A}}^M, \mathcal{P}_{\mathcal{B}}^M)$ 满足 $F_\# p = \tilde{F}_\# \tilde{p}$，都存在一个由 $h: \mathbb{R}^m \to \mathbb{R}^m$组成的可逆分解仿射映射$H: \mathbb{R}^{mT} \to \mathbb{R}^{mT}$，其中 $p = H_\# \tilde{p}$，则称其在仿射变换下可识别。等价地，在置换、缩放和平移下可识别，是指具有 $h(x) = PDx + b, x \in \mathbb{R}^m$，其中 P 和 D 分别为置换矩阵和对角矩阵。

我们针对非线性SDS陈述如下可识别性结果，并将证明置于附录B中，该证明主要基于 Kivva 等人（2022年）的工作。

定理 12（参数化SDS的可辨识性）假设观测 $x_{1:T}$ 由遵循式 *(21)* 的潜变量模型生成，其先验分布 $p(z_{1:T})$ 属于 $\mathcal{M}^{T,}{}_{NL}^{M} = \mathcal{M}^T(\mathcal{I}_{\mathcal{A}}^M, \mathcal{G}_{\mathcal{B}}^M)$，并满足 *(m1 - m2)*。假设：

3. 特别地，在含噪情形下，若 p_ε 的特征函数的零点集是零测度，则可识别性成立。我们使用 $p_{\varepsilon_{1:T}}$ 表示 $\varepsilon_{1:T}$的分布。

(f1) f 是弱单射（定义 *5*）且为分段线性；

以及关于先验分布 $p(\boldsymbol{z}_{1:T})$的以下假设。

(s1) 在时间 t，给定 $\boldsymbol{z}_{t-1}$ 和s_t时，连续潜变量 $\boldsymbol{z}_t$ 彼此独立。对于任意 $i \neq j$, $z_t^{(i)} \perp\!\!\!\perp z_t^{(j)} \mid \boldsymbol{z}_{t-1}, s_t$；*(s2)* 存在 $k_1, k_2 \in [K]$ 和一个非零测度集 $\mathcal{Z} \subset \mathbb{R}^{mM}$，使得对于任意$(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}) \in \mathcal{Z}$， $\{\boldsymbol{\Sigma}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, k_1)_{ii} / \boldsymbol{\Sigma}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, k_2)_{ii} \mid i \in [m]\}$ 中的元素互不相同。

在上述假设下，先验分布 $p(\boldsymbol{z}_{1:T})$ 是可辨识的，具体如下。

(i) 若满足 *(f1)*，则先验分布 $p(\boldsymbol{z}_{1:T})$ 在仿射变换下可识别（定义 *11*）。*(ii)* 若满足 *(f1)*、*(s1)* 和 *(s2)*，则先验分布 $p(\boldsymbol{z}_{1:T})$ 在置换、缩放和平移下可识别。

上述结果使可辨识的 SDS 能够采用例如 Softplus网络来表示 $p_{\boldsymbol{\theta}}(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, s_t)$，并适用于具有历史依赖性或分段平稳噪声的情形；还可采用某些（Leaky）ReLU 网络来表示 f（参见 Kivva等（2022）的附录 H）。同样，对于神经网络，可识别性仅指其函数形式。

放宽分段线性条件。 虽然我们的方法放宽了对发射函数的单射性假设，但仍限制 f 为分段线性。Yao 等人的其他研究（如 Yao等（2022a））假设函数为单射，其结果可以调整后用于我们的框架。具体而言，Yao等（2022a）在以下条件下建立了可识别性结果：(1) 潜变量独立性（定理 12 中的 s1）；(2) 存在解混函数；(3) 可以获取满足特定线性无关条件的 $K = 2m + 1$ 分量，这些条件施加于由 $\log p(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, k)$ $k \in [K]$的导数构成的向量上。然而，这一策略要么假设可以获取真实的分量标签 k，要么要求直接对观测条件分布 $p(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}, \ldots, \boldsymbol{x}_{t-M}, k)$ $k \in [K]$施加假设。例如，Song 等（2023）直接对观测分布提出假设 (m1-m2)，从而隐式地约束潜在动力学和发射函数。潜在结构与观测模型之间缺乏区分，使得这些假设在实践中难以验证或控制（例如，若观测分布和潜变量分布均为高斯分布，则 f 必须是仿射变换）。此外，若没有额外的结构约束，这些策略无法充分利用我们针对 MSMs 得出的可识别性结果，因为这些结果是在潜在动力学层面推导的。为作说明，考虑联合概率分布；一个等价的 SDS 分布族可定义如下：

$$\left\{ \sum_i^C c_i (F_{\#} p_i)(\boldsymbol{x}_{1:T}) \mid \; F \in \mathbb{F},\; \sum_i^C c_i = 1,\; c_i > 0,\; C = K_0 K^{T-M},\; K_0, K < +\infty, \right.$$
$$\left. p_i \in \mathcal{P}^{T,M}_{\mathcal{A}_0, \mathcal{B}_0} \subset \mathcal{P}^{T,M}_{\mathcal{A}, \mathcal{B}},\; \mathcal{A}_0 \subset \mathcal{A},\; \mathcal{B}_0 \subset \mathcal{B},\; |\mathcal{A}_0| = K_0,\; |\mathcal{B}_0| = K \right\}, \quad (22)$$

其中每个元素都由应用于 MSM族 $\mathcal{M}^T$中混合分量 p_i的分解变换 $F \cdot {}^{M}_{NL}$ 构成。这里的可识别性需要进一步分析，因为上述族中具有 $F \neq F$的两个不同元素可能对应于具有不同分量的不同 MSM，$(K_0,\ K)$ $\neq$ $(\tilde{K}_0,\ \tilde{K})$，它们通过逆累积分布函数映射相关联：$p = F^{-1} \circ \tilde{F} \# \tilde{p}$。

4.3 制度依赖的因果表示学习

在制度依赖时间序列（Saggioro 等，2020年）中，SEM 依赖于时间。然而，在每个时间步 t以离散潜变量 $s_t \in [K]$为条件时，它是因果平稳的 （Assaad 等人，2022年）。在潜变量模型的语境下，SEM 作用于潜在空间；因此，我们的框架与因果表征学习（ Schölkopf 等人，2021年）相关。在考虑 SDS 时，相应的潜在因果结构成为制度依赖因果图 ，我们将其记为$\mathbf{G}_{1:K} := \{\mathbf{G}_k\}_{k=1}^{K} \in \mathbb{G}^K$。它可以定义为一组因果图，使得对于每个 $k \in [K]$，$\mathbf{G}_k \in \mathbb{G}$ 在 $s_t = k$时编码时间 t 的潜在 SEM。为建立与前述 MSMs 的联系，我们假设存在一个加性噪声模型 （ANM；（Hoyer 等人，2008年）），且不存在瞬时效应。因此，变量 $j \in [m]$ 在潜在空间中的相应制度依赖 SEM 表示如下：

$$z_t^{(j)} = f_{s_t}^{(j)}\left(\{z_{t-\tau}^{(i)} \mid z_{t-\tau}^{(i)} \in \mathbf{Pa}_{s_t}^{(j)}(\tau), 1 \le \tau \le M\}\right) + \varepsilon_t^{(j)}, \quad \varepsilon_t^{(j)} \sim p_{\varepsilon^{(j)}}, \tag{23}$$

其中 $\mathbf{Pa}_{s_t}^{(j)}(\tau)$ 表示变量 j在滞后阶数 τ时的父节点，与 $s_t = k$对应的 $\mathbf{G}_k$相关。

定义 *13* 若对于任意 p、$\tilde{p} \in \mathcal{M}^T(\Pi_{\mathcal{A}}^M, \mathcal{P}_{\mathcal{B}}^M)$ 和 F、$\tilde{F} \in \mathbb{F}$，其对应的依赖机制的图分别为 $G_{1:K} \in (\times^K \mathbb{G})$ 和 $\tilde{G}_{1:\tilde{K}} \in (\times^{\tilde{K}} \mathbb{G})$，且满足 $(F_{\#}p)(\boldsymbol{x}_{1:T}) = (\tilde{F}_{\#}\tilde{p})(\boldsymbol{x}_{1:T})$，$K = \tilde{K}$ ，并且存在节点置换 $\pi \in S_m$ (定义 *35*) 和图置换 $\sigma \in S_K$ ，使得对于 $k \in [K]$，$G(k) = \pi(\tilde{G}\sigma(k))$，则称阶数为 M 的 *SDS* 的依赖机制的图在置换意义下可辨识。

上述节点置换（定义 35）会置换集合 $\mathbf{G}_{1:K}$ 中每个图的节点。我们的定义旨在恢复固有的图结构，使其在连续和离散潜变量的置换下保持不变。我们如下建立制度依赖时间序列的潜在结构可识别性。

推论 14（潜在结构可识别性）考虑公式（*12*）中定义的阶数为 M 的马尔可夫切换模型族，其中采用非线性高斯族 $\mathcal{M}^{T,\ M}_{NL}$ 。假设满足定理 *8* 中的 *(m1)*；

(m2)* 异质噪声： $\boldsymbol{m}(\cdot, b) : \mathbb{R}^{mM} \to \mathbb{R}^m$ 中的转移分布均值是解析的，并且 $\boldsymbol{\Sigma}(\cdot, b) : \mathbb{R}^{mM} \to \mathbb{R}^{m \times m}$ 是对角的、解析的，且满足 *(s2)*；

(m3) 最小性：若 $z^{(i)} \notin \mathbf{Pa}_k^{(j)}(\tau)\ 1 \le \tau \le M$，则 $\boldsymbol{m}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, k)$ 的第 i 维相对于 $z_{t-\tau}^{(j)}$ 是常数。

因此，依赖机制的图在置换意义下可辨识（定义 *13*）。

证明见附录 C。鉴于定理 8 和 12 给出了可辨识的转移导数，只需证明依赖机制的结构在置换意义下可以恢复。在假设 (m1-m3) 下，我们注意到，传统因果时间序列分析中的其他结构，例如完整时间图 （Peters 等人，2017），在高阶 MSM 的情境中不可辨识；因为我们无法获取离散潜变量 $\boldsymbol{s}_{M:T}$，只能获取其在置换意义下的分布。

5 估计

我们将结合下文理论框架与推断方案的完整方法称为可识别切换动力系统 （iSDS）。对于现实应用（第7节），我们还考虑马尔可夫切换模型。结合我们的假设，我们将相应方法称为可识别马尔可夫切换模型 （iMSM）。

我们采用 Dong et al. (2020) 的折叠变分推断框架。通过最大化证据下界（ELBO）（Kingma 和 Welling，2014年）来学习参数。潜变量 $\{\boldsymbol{z}_{1:T}, \boldsymbol{s}_{M:T}\}$ 的近似后验采用一种因子分解形式，其中纳入了在连续潜变量条件下离散潜状态的精确后验:

$$q_{\boldsymbol{\phi},\boldsymbol{\theta}}(\boldsymbol{z}_{1:T}, \boldsymbol{s}_{M:T} \mid \boldsymbol{x}_{1:T}) = \prod_{t=1}^{T} q_{\boldsymbol{\phi}}(\boldsymbol{z}_t \mid \boldsymbol{x}_t) p_{\boldsymbol{\theta}}(\boldsymbol{s}_{M:T} \mid \boldsymbol{z}_{1:T}). \tag{24}$$

关于 $\boldsymbol{z}_{1:T}$ 的变分后验按时间进行因子分解。这与我们对潜变量变换的单射假设相呼应，也简化了估计。先前的研究（Ansari 等人，2021年；Dong et al. (2020)）通过对 $\boldsymbol{x}_{1:T}$ 使用双向循环神经网络，再接一个前向循环神经网络，构造 $q_{\boldsymbol{\phi},\ \boldsymbol{\theta}}(\boldsymbol{z}_{1:T} \mid \boldsymbol{x}_{1:T})$。有了精确后验 $p_{\boldsymbol{\theta}}(\boldsymbol{s}_{M:T} \mid \boldsymbol{z}_{1:T})$，我们可以从证据下界中边缘化 $\boldsymbol{s}_{M:T}$ （详见附录 D）：

$$p_{\boldsymbol{\theta}}(\boldsymbol{x}_{1:T}) \geq \mathcal{L}_{\text{ELBO}} := \mathbb{E}_{q_{\boldsymbol{\phi}}(\boldsymbol{z}_{1:T} \mid \boldsymbol{x}_{1:T})}\Big[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}_{1:T} \mid \boldsymbol{z}_{1:T})\Big] - KL\Big(q_{\boldsymbol{\phi}}(\boldsymbol{z}_{1:T} \mid \boldsymbol{x}_{1:T}) || p_{\boldsymbol{\theta}}(\boldsymbol{z}_{1:T})\Big). \tag{25}$$

我们使用蒙特卡洛样本 $\tilde{\boldsymbol{z}}_{1:T} \sim q_{\boldsymbol{\phi}}(\boldsymbol{z}_{1:T} \mid \boldsymbol{x}_{1:T})$ 估计期望，并应用重参数化技巧进行反向传播（Kingma 和 Welling，2014年），从而实现随机梯度上升。梯度结果如下:

$$\nabla \mathcal{L}_{\text{ELBO}} \approx \nabla \log p_{\boldsymbol{\theta}}(\boldsymbol{x}_{1:T} \mid \tilde{\boldsymbol{z}}_{1:T}) - \nabla \log q_{\boldsymbol{\phi}}(\tilde{\boldsymbol{z}}_{1:T} \mid \boldsymbol{x}_{1:T}) + \nabla \log p_{\boldsymbol{\theta}}(\tilde{\boldsymbol{z}}_{1:T}). \tag{26}$$

遵循 Dong 等人（2020年）的做法，先验梯度可以进行高效分解，类似于标准的隐马尔可夫模型推断（Bishop，2006年）：

$$\begin{aligned}
\nabla \log p_{\boldsymbol{\theta}}(\tilde{\boldsymbol{z}}_{1:T}) &= \mathbb{E}_{p_{\boldsymbol{\theta}}(\boldsymbol{s}_{M:T} \mid \tilde{\boldsymbol{z}}_{1:T})}\Big[\nabla \log p_{\boldsymbol{\theta}}(\tilde{\boldsymbol{z}}_{1:T}, \boldsymbol{s}_{M:T})\Big] && (27)\\
&= \sum_{t=M+1}^{T} \sum_{k=1}^{K} \gamma_{t,k}(\tilde{\boldsymbol{z}}_{1:T}) \nabla \log p_{\boldsymbol{\theta}}(\tilde{\boldsymbol{z}}_t \mid \tilde{\boldsymbol{z}}_{t-1}, \ldots, \tilde{\boldsymbol{z}}_{t-M}, s_t = k) && (28)\\
&\quad + \sum_{k=1}^{K} \gamma_{M,k}(\tilde{\boldsymbol{z}}_{1:T}) \nabla \log \pi_k + \sum_{t=M+2}^{T} \sum_{k=1}^{K} \sum_{k'=1}^{K} \xi_{t,k,k'}(\tilde{\boldsymbol{z}}_{1:T}) \nabla \log Q_{k',k}, && (29)
\end{aligned}$$

其中，$\gamma_{t,\ k}(\tilde{\boldsymbol{z}}_{1:T}) = p_{\boldsymbol{\theta}}(s_t = k \mid \tilde{\boldsymbol{z}}_{1:T})$ 和 $\xi_{t,\ k,\ k'}(\tilde{\boldsymbol{z}}_{1:T}) = p_{\boldsymbol{\theta}}(s_t = k,\ \ s_{t-1} = k' \mid \tilde{\boldsymbol{z}}_{1:T})$ 是在 $\boldsymbol{s}_{M:T}$ 上的一阶马尔可夫链假设下，使用前向后向更新规则计算的。详情请参阅附录 D。近期的一种替代方法（Ansari et al., 2021）仅使用前向后验来计算 $p(\boldsymbol{z}_{1:T})$，但这需要通过整个序列进行反向传播，从而大幅增加计算成本。

初始化与 iMSM。Dong et al. (2020) 指出，上述方法容易导致状态坍塌，并引入了采用退火调度的额外损失；Ansari et al. (2021) 则通过对离散逻辑值进行温度退火并结合前向后验，缓解状态坍塌。这类调度取决于数据，需要谨慎进行超参数调优。根据我们的经验，对 $p_\theta(\boldsymbol{z}_{1:T})$ 进行强初始化是稳定训练并缓解状态坍塌的关键。我们首先通过线性主成分分析（PCA）将 $\boldsymbol{x}_{1:T}$ 投影到低维空间，再按照公式（29）对投影时间序列拟合标准 MSM。学习得到的 π， Q 和制度特定参数用于初始化 $p_{\boldsymbol{\theta}}(\boldsymbol{z}_{1:T})$。类似的初始化策略已在先前工作中证明有效（Linderman et al., 2016），为提升离散潜变量模型的训练稳定性奠定了基础。iMSM 采用相同的基于似然的优化，并使用公式（29）中的梯度分解。

稀疏正则化。为恢复潜在因果结构，我们通过对雅可比矩阵 $J_{\boldsymbol{m}_k}$（针对 $k \in [K]$）施加逐元素 ℓ_1 惩罚，促使状态特定均值转移 $\boldsymbol{m}(\cdot, k)$ 具有稀疏性。正则化后的目标函数为

$$\mathcal{L}_{\mathrm{REG}} := \mathcal{L}_{\mathrm{ELBO}} + \eta \sum_{k=1}^{K} \sum_{i,j}^{m} \left| \left(J_{\boldsymbol{m}_k}\right)_{ij} \right|, \tag{30}$$

其中， m 表示潜在维度。尽管最大化证据下界是一种实用的估计方法，但通常无法保证统计一致性；相关讨论可参见 Khemakhem 等人（2020）。一致性分析留待未来工作。

6 合成实验

本节介绍合成数据实验。由于可以访问真实结构，我们能够通过实证验证理论结果。具体而言，我们（i）比较 iSDS 与近期基线方法在恢复可辨识潜变量方面的能力；（ii）考察样本效率和序列长度效率；（iii）检验定理 8 中的假设对于 MSM 和 SDS 的可识别性是否必要。

数据生成。我们主要按照 Balsells-Rodas 等人（2024）生成数据，并额外设置了若干噪声分布情形。各状态遵循一阶马尔可夫链，其转移矩阵 Q 设定为：留在同一状态的概率为 0.9，转移到下一个状态的概率为 0.1。连续潜变量 $\boldsymbol{z}_{1:M}$ 从具有均匀类别先验的 K 分量高斯混合模型中初始化。对于 $t > M$，它们依据具有 K 个状态和状态特定参数的高斯转移模型演化。对于每个 $k \in [K]$，均值函数 $\boldsymbol{m}(\cdot, k)$ 由具有余弦激活函数的双层多层感知机构成，其中每一层都是局部线性映射（Zheng 等人，2018），并受随机生成的因果图 $\mathbf{G}_k$ 约束。协方差矩阵 $\boldsymbol{\Sigma}(\cdot, k)$ 为对角矩阵，我们考虑三种情形：

- 恒定噪声：固定的， $\boldsymbol{\Sigma}(\cdot, k) = \boldsymbol{\Sigma}$。
- 异质噪声：依赖于状态， $\boldsymbol{\Sigma}(\cdot, k) = \boldsymbol{\Sigma}(k)$。
- 历史依赖噪声：对角项与均值的缩放后的绝对值成正比。

后两者满足假设（s2），即至少有一对状态的对角线比值不同。观测是通过使用带泄漏的ReLU激活函数的双层多层感知机对潜变量进行变换生成的。除非另有说明，否则我们生成长度为 $T = 100$ 的 $N = 10000$ 个样本用于训练，并生成1000个样本用于评估。

指标。 为将学习得到的解与真实值进行比较，我们使用以下指标（详见附录 E.3）：

- **机制估计：** 选择后验概率最高的分量，并在进行机制排列对齐后计算 F_1 得分。

- **连续潜变量对齐：** 我们报告弱MCC和强MCC的平均相关系数。强MCC（Khemakhem 等人（2020）也对其进行了定义）是 ICA 中的标准测度，用于评估在缩放和置换意义下的对齐情况。相比之下，弱MCC（Kivva 等人，2022年）允许进行仿射变换。

- **因果结构恢复：** 真实因果图与估计因果图之间的 F_1 得分。该得分通过在完成状态和连续潜变量置换对齐后，对雅可比矩阵进行阈值化获得。

- **转移函数：** 在进行状态置换和潜变量仿射对齐后，真实均值 $\boldsymbol{m}(\cdot,\ k)$ 与估计均值之间的平均 L_2 距离和 R^2 得分。

基线比较。 我们将 iSDS 与五种具有代表性的可辨识表示学习基线方法进行比较：

- **iVAE*** (Khemakhem 等人，2020年)，将真实的离散状态作为辅助信息提供（一个更强的基线方法）。
- **SNICA** (Hälvä 等人，2021年)，基于线性切换动力系统，其中每个潜变量都遵循自身的 SDS。

- **TDLR** (Yao 等人，2022a)，假设时序潜变量具有非高斯性，但不存在切换。

- **NCTRL** (Song 等，2023年)，基于 SDS，但假设观测具有高斯性，且潜变量具有一定的非高斯性。
- **CtrlNS** (Song 等，2024年)，假设切换次数已知，且直接依赖于观测。

为将 iSDS 与其他基线方法进行比较，我们探讨了数据集设置 A 至 F；所有设置均具有 $n = 10$ 个观测维度，并具备以下特征：

A) $m = 3$, $K = 3$, $M = 1$，恒定噪声。B) $m = 3$, $K = 3$, $M = 1$，异质噪声。C) $m = 3$, $K = 3$, $M = 1$，历史依赖噪声。D) $m = 5$, $K = 3$, $M = 2$，历史依赖噪声。E) $m = 5$, $K = 5$, $M = 2$，异质噪声。F) $m = 5$, $K = 5$, $M = 5$，异质噪声。

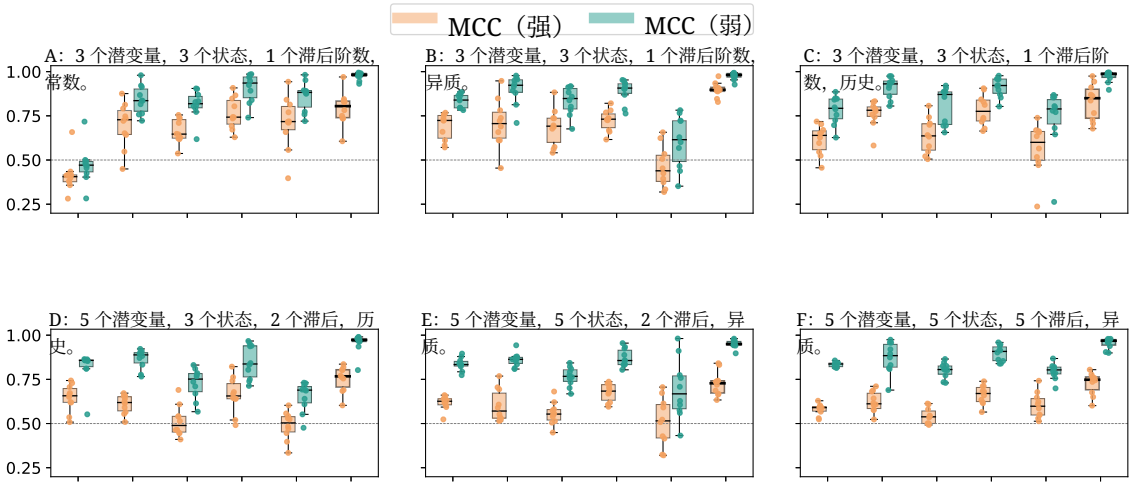
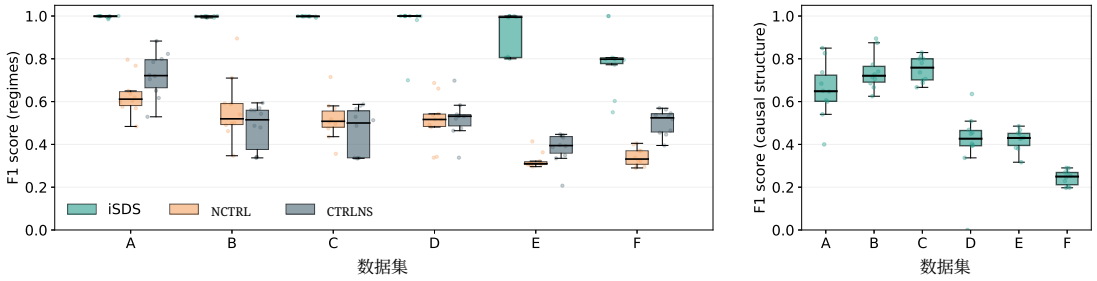


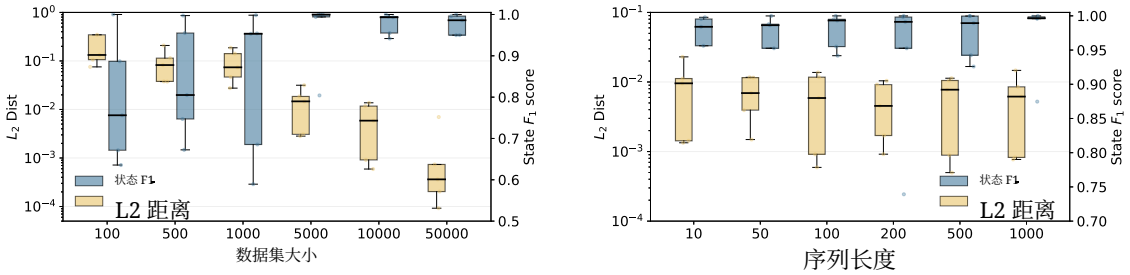
图 6: 数据集设置 A–F 中的弱MCC 分数和强MCC 分数（每个设置包含 10 个数据种子）.



(a) (b)

图 7: （a）数据集设置 A–F 中的机制估计，以及（b）因果结构恢复（每个设置包含 10 个数据种子）。

图 6 展示了 10 个数据集随机种子下，各种配置（A-F）的弱MCC和强MCC分数。总体而言，在弱MCC方面，iSDS 优于基线方法，在所有设置中都实现了近乎完美的仿射对齐。其次是 SNICA 和 NCTRL，这符合预期，因为二者都属于基于SDS的方法。在强MCC方面，iSDS 同样优于所有基线方法，不过在设置 A 和 D-F 中，性能有所下降。A-C 之间唯一的区别在于噪声分布。A 中较低的强MCC符合预期，因为在恒定协方差矩阵下，无法保证可识别性达到缩放和置换意义下的程度。D-F 中较低的 MCC 分数可归因于更高阶滞后、维度、状态数量以及非平稳噪声带来的复杂性增加。对于状态估计，图 7a 将 iSDS 与 NCTRL 和 CtrlNS 进行了比较，因为它们以相近的方式提取状态。iSDS 始终取得更优的 F_1 分数，而状态数量增加时（E 和 F），性能会下降。尽管 NCTRL 基于 SDS，其表现仍显著较差。这可以解释为 NCTRL 直接从观测中估计状态



(a) (b)

图 8：通过改变 (a) 数据集大小或 (b) 序列长度，对 iSDS 进行数据效率分析。

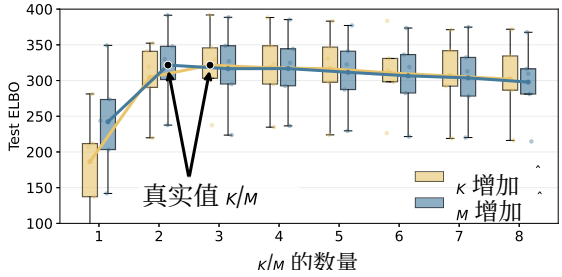


图 9：使用 iSDS 对 5 个数据集随机种子进行模型选择的结果。随着切换数 (K；黄色) 和滞后阶数 (M；蓝色) 增加，模型的测试证据下界。黑点表示真实值中的切换数和滞后阶数（分别为 $K = 3$ 和 $M = 2$）。

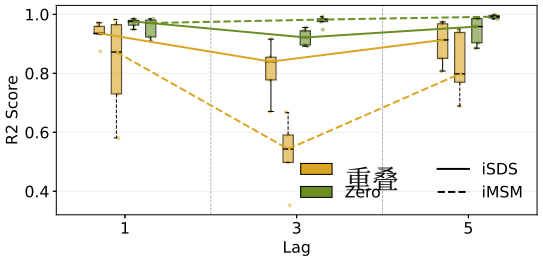


图 10：使用 iMSM 和 iSDS 对 5 个数据集随机种子进行交集假设消融实验。跨滞后阶数的均值函数平均 R^2，其测度为零的交集（零）和非零测度的交集（重叠）。

在高斯性假设下，这会造成模型失配，因为生成的观测通常并非高斯分布。最后，图 7b 报告了因果结构恢复结果。在 A-C 中，随着噪声非平稳性的增加，F_1 分数有所提高，这与我们的理论结果一致。相比之下，D-F 中的表现有所下降，与先前报告的 MCC 分数一致。

数据效率分析。为了更好地了解 iSDS 在样本和序列方面的效率，我们生成了包含 $n = m = 5$ 个潜变量和观测维度、$M = 2$ 个滞后阶数、$K = 3$ 个状态以及恒定噪声的数据。我们评估估计状态的 F_1 分数，以及仿射对齐后均值函数的平均 L_2 距离 $\boldsymbol{m}(\cdot)$。结果如图 8a 和 8b 所示，分别展示了数据集大小和序列长度变化时的结果。图 8a 中，序列长度固定为 $T = 100$；图 8b 中，样本数固定为$N = 10000$。

我们观察到，较大的样本量显著提升了状态恢复和均值函数估计的表现。然而，当样本数量充足时，增加序列长度几乎没有影响。结果表明，训练 iSDS 时，通常多收集一些序列比增加序列长度更有益。

选择 K 和 M。我们的理论框架证明，对于 iMSM 和 iSDS，状态数量 K 是可辨识的。假设 iSDS 满足一致性，我们的结果有助于通过模型选择学习真实值 K （McLachlan 和 Peel，2000）。类似的论证也适用于选择 M：滞后阶数较小时，模型会出现设定错误；而滞后阶数 M较大时，可能会有多余的滞后阶数未被使用。我们通过实验展示了这一点，更深入的理论分析留待未来工作。选择 K 和 M 时，我们采用肘部法（Thorndike，1953），这是有限混合模型的一种标准启发式方法。对于 iSDS，我们使用留出数据集上的证据下界，而非 iMSM 可能采用的对数似然。图 9 展示了一个 iSDS 模型选择示例，其中包含 $M = 2$ 个滞后ˆ和 $K = 3$ ，我们增加估计的状态数量 K ，同时保持真实值ˆ滞后阶数 M 不变（黄色线）；增加估计的滞后阶数 M ，同时保持真实值状态数量不变（蓝色线）。两种情况下，我们都观察到，最高测试证据下界出现在真实值ˆˆ M 和 K处，而当选择更大的 M 和 K时，性能趋于平稳。

消融实验。 第 4.1 节证明了具有解析函数和唯一索引的参数化 MSM 的可识别性。对于 $M = 1$, 一个充分条件是存在一个开集，使得特定于各状态的均值（和协方差）函数不相交。然而，对于一般的M，我们要求该集合具有全测度。这些条件是充分条件，但尚未证明其必要性。我们通过实验评估其必要性，针对 iMSM 和 iSDS 分别设计了两种设置：(i) 零，其中真实值均值函数至多在一个零测度集合上相交；(ii) 重叠，其中函数在一个非零测度集合上重合，但在其他区域有所不同。我们考虑 $n = m = 5$、 $K = 3$、恒定噪声以及滞后 $M \in \{1, 3, 5\}$。

图 10 报告了在状态排列对齐后（对 iSDS 还进行了额外的仿射对齐），均值函数 $\boldsymbol{m}$（·）的平均 R^2 。使用 R^2 而非 L_2 ，便于比较具有不同输入/输出尺度的滞后阶数。正如预期，零优于重叠。不过，重叠仍表现出较强的拟合效果，且 R^2 得分较高，远不能说明模型失效。对于 iSDS，仿射对齐可能会使 R^2 得分更高。然而，对于 iMSM，我们观察到 $M = 5$的拟合效果较强。这表明，要求具有更强的全测度不相交性在实践中是充分但非必要的；否则，我们会得到低得多的 R^2 得分（例如 $R^2 < 0.5$）。因此，重叠设置下性能下降，可归因于重叠区域带来的估计难题——在该区域追踪状态更加困难，而非模型在根本上缺乏可识别性。

7 真实世界数据的应用

在受控条件下确立了理论结果的有效性后，我们现在转向真实世界数据，在这类数据中，真实结构通常不可得。因此，本节着重评估第4节所提出理论在完全无监督设定下的适用性。我们研究来自不同科学领域的数据集，并探究关于平滑性和非平稳性的假设（即假设 (m2) 和 (s2)）是否有助于学习结构化潜在表示。重要的是，我们基于可辨识性的框架表明，学习到的表示与观测之间是一一对应的，且仅存在可接受的等价性（例如缩放、置换）。这一唯一性使我们能够更可靠地解读恢复出的结构，并将其与既有领域知识联系起来。总之，本节旨在展示所提出的 iMSM 和 iSDS 如何助力真实世界数据中的科学发现。

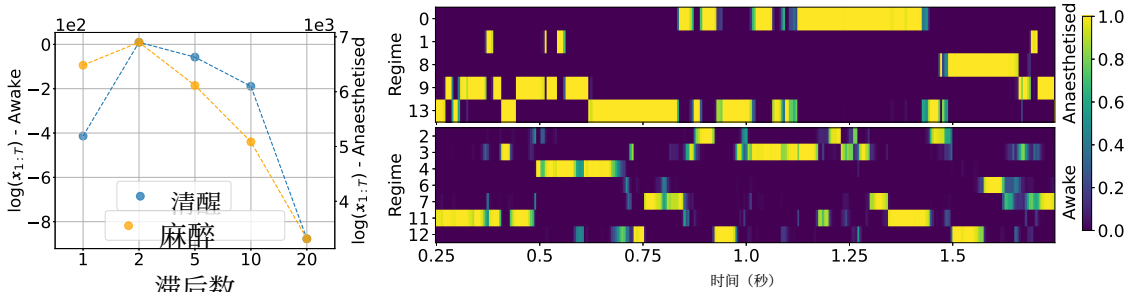


图11：（a）使用不同滞后和 $K = 15$ 个状态时，皮层脑电图数据的测试对数似然。（b）清醒和麻醉状态下的状态后验概率示例。使用长度为 3 的均匀核对后验概率进行平滑，以抑制帧级伪影。

7.1 脑活动数据

我们首先通过 NeuroTycho数据库[4], 中的皮层脑电图（ECoG）记录展示 iMSM；这些记录最初由 Yanagawa等人（2013年）发表。在两种条件下记录了一只猕猴的128个电极信号：正常清醒状态（清醒）和丙泊酚诱导的意识丧失（麻醉）。根据先前的研究（Mediano 等，2023），我们预期清醒条件下会呈现更复杂的动力学，表现为潜在状态之间更频繁的状态转移。

原始皮层脑电图以 1 kHz 采样，包含清醒和麻醉片段，时长分别为 929.8 秒和 650.65 秒。我们选取了 21 个电极，大致覆盖大脑视觉皮层。首先，我们使用 50 Hz 二阶巴特沃斯陷波滤波器消除工频噪声。随后，将数据降采样至 200 Hz，对每个通道分别进行标准化，并将每个序列切分为 2 秒的时间段（$T = 400$）。由此得到 464 个清醒片段和 325 个麻醉片段，其中每种条件各留出 50 个片段用于测试。

在两种数据设置中，我们设定 $K = 15$，并使用余弦激活函数参数化转移函数（模型选择详情见附录 E.4）。图 11a 展示了 iMSM 的测试对数似然随滞后阶数 M 变化的情况。正如预期，清醒状态始终更难建模（对数似然更低），表明其动力学更加复杂。在两种设置中，滞后阶数为 $M = 2$ 的 iMSM 均取得最佳表现。值得注意的是，麻醉状态在 $M = 1$ 时仍有竞争力，但当 $M > 2$ 时性能迅速下降，这与其动力学较为简单的情况一致。相比之下，清醒状态从更高的滞后阶数中显著获益。

图 11b 展示了各条件下以 2 秒为时段的后验状态占用率（经长度为 3 的均匀核平滑）。我们仅显示该时段内占用率超过 1% 的状态。麻醉状态下，活动会长时间停留在较小的一组状态中；而清醒状态下，涉及的状态比例更大，状态转移也更频繁。表1量化了这些差异：与麻醉状态相比，清醒状态总体上使用的状态更多、每个样本使用的状态也更多，状态转换频率（赫兹）更高，平均状态持续时间则更短。

至关重要的是，iMSM 提供状态可辨识性保证，以确保状态使用上的差异反映出有数据支持的条件特异性结构。这凸显了

4. http://neurotycho.org/

表1：带有 $M = 2$ 和 $K = 15$ 的测试视觉皮层脑电图数据的状态统计。“总使用率”是测试集中至少占用过一次的状态所占的比例。“每个样本的使用率”是单个 2 秒时段内所占用状态比例的均值。“最长持续时间”是每个时段最大驻留时间的均值。

条件	转移频率（赫兹）	总使用率	每个样本的使用率	最大持续时间（秒）
清醒	22.05 ± 1.06	53.33%	44% ± 1.47%	0.22 ± 0.02
麻醉	17.71 ± 1.29	33。33%	26.40% ± 1.48%	0.35 ± 0.04

对于 ECoG 等神经科学时间序列，考虑采用 iMSM 或 iSDS 的重要性，并可与现有方法（Mediano 等，2023）相互补充。

7.2 金融数据

金融时间序列是状态切换模型的一个直接应用领域。早期关于马尔可夫切换模型（Hamilton，1989）的研究侧重于对市场周期状态进行建模，为金融领域的状态切换模型奠定了基础。本节研究美国基于行业划分的股票价格数据，并使用 iSDS 提取结构化潜在表示和状态。相关应用既涉及马尔可夫切换模型（Alizadeh 和 Nomikos，2004；Cavaliere 等，2014；Mahmoudi 和 Ghaneei，2022 ），也涉及 SDS （ Azzouzi 和 Nabney，1999 ；Carvalho 和 Lopes，2007；Fox 等，2011）。在实证研究中，研究人员通常使用对数收益率或对数价格（减去股息）。

对数收益率 $r_t = \log\text{Ä}\frac{p_t}{p_{t-1}}\text{ä}$ 通常被视为平稳序列（Cont，2001），而状态切换主要关注依赖于状态的均值和/或波动率。这包括结合条件异方差的马尔可夫切换模型（例如广义自回归条件异方差（Bollerslev，1986）），其中条件方差取决于潜变量和/或过去的收益率（Ang 和 Bekaert，2002；Carvalho 和 Lopes，2007）。这类模型通常将状态解释为牛市/熊市或平静/动荡，并广泛用于风险评估和资产配置（Guidolin 和 Timmermann，2007）。

由于价格通常是非平稳的，因此，对数价格数据的状态切换通常通过现值框架进行建模，将公司基本面（例如收益）与投机成分分离开来（Balke 和 Wohar，2009年；Chan 和 Santi，2021年）。这些模型能够捕捉较长期的状态（持续熊市/牛市阶段）以及投机时期。

数据模态与可识别性。 我们分析了两种数据模态（对数收益率和对数价格），以在真实场景中阐释我们的理论。我们的 SDS 结果表明，在适当的建模条件下，状态是可辨识的（定理 12）。对于连续潜变量，若要实现置换和缩放意义下的可识别性，则需要某些异质噪声或依赖历史的噪声（定理 12；推论 14）。更具体地说，一对状态在不同维度上的噪声比率必须各不相同。实践中，非平稳的对数价格更可能满足上述条件，因为其尺度具有状态特异性和行业特异性。相比之下，对数收益率中的噪声尺度经过了归一化，因为它们是相邻价格的比率。这意味着，即使是依赖历史的波动率，也可能不满足置换和缩放意义下潜变量可识别性所需的条件

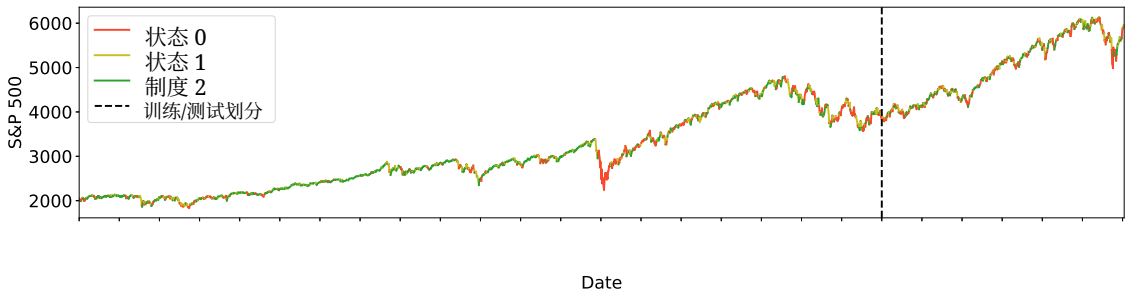
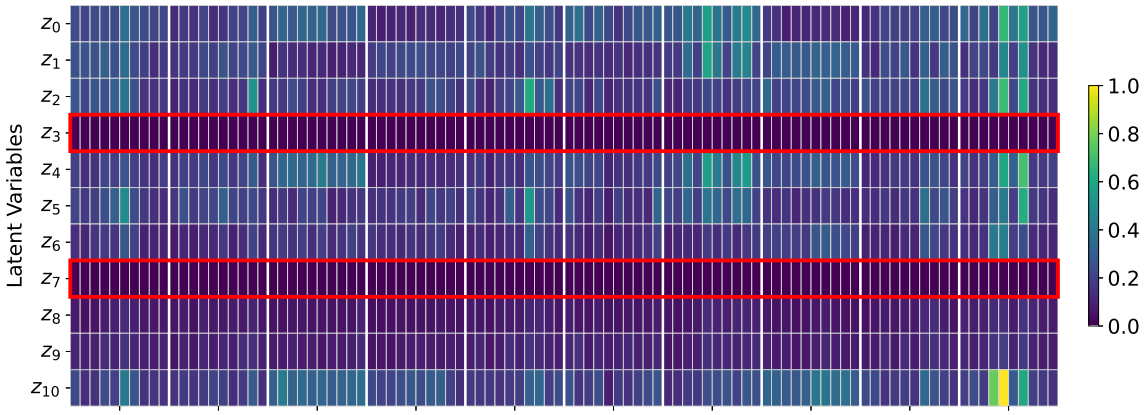


图 12：与标准普尔 500 指数叠加显示的对数收益率状态分割；虚线分隔训练数据与测试数据。



股票板块

Fi图 13：按行业分组的对数收益率解码器雅可比矩阵（潜变量 $\rightarrow$ 资产） .

。因此，我们评估这两种设置，以判断哪一种能在 iSDS 中捕捉到更具结构化的表征。

数据集和模型设置。 我们从彭博终端（彭博有限合伙公司，2025年）获取了100只美国股票的股票价格，并将其分为10个行业，保留过去十年（2015年1月–2025年5月）的每日价格数据。由此得到一个包含2630个时间步的序列。公司列表和行业映射见附录 E.5。训练时，我们从前2000个时间步中随机抽取长度为$T = 200$的子序列，并将剩余的630个时间步留作测试。我们使用11个潜变量和3个状态训练iSDS：10个行业潜变量（每个行业一个）加上一个全局潜变量，以及3个市场状态，旨在捕捉例如牛市/熊市/横盘或高/中/低波动率，这与金融时间序列相关工作中的做法类似（Chan 和 Santi，2021）。我们设置 $M = 5$ 并使用异质协方差，即，对于 $k \in [K]$，$\boldsymbol{\Sigma}(\cdot,\ k) = \boldsymbol{\Sigma}(k)$。其他训练细节见附录 E.2。

对数收益率的结果。图 12 展示了叠加在标准普尔 500 指数上的训练/测试分割，以供参考。使用对数收益率进行训练时，iSDS 展现出的状态能够捕捉

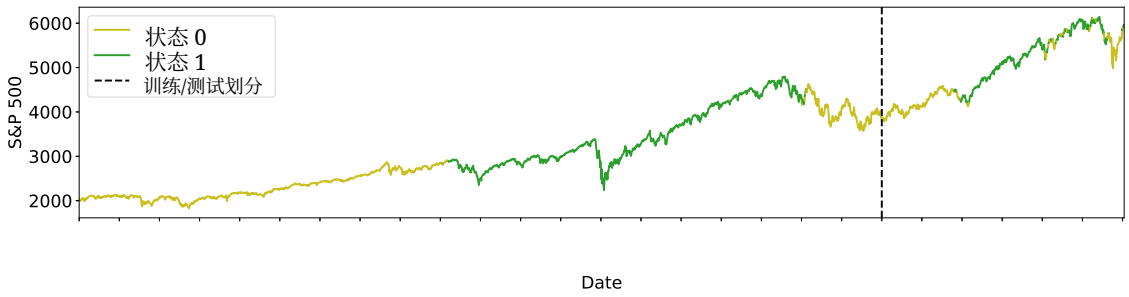


图 14：叠加标准普尔 500 指数的对数价格状态分割；虚线表示训练集与测试集的分界。

短期波动率适应，其中绿色、黄色和红色分别对应低、中、高波动率。我们观察到，从 2016 年末到 2018 年初，低波动率时期持续了较长时间；而在 2020 年 3 月（COVID-19）则出现了持续时间显著较长的高波动状态。自 2020 年起，分割结果主要由中、高波动率状态构成，这与标准普尔 500 指数波动性增加的情况一致。

为评估潜在结构，图 13 展示了解码器雅可比矩阵（$\rightarrow$ 个资产对应的潜变量），并按行业分组。该模式显示，行业特定结构并不明显：多个潜变量映射到同一行业（例如，在“材料”行业附近显著集中），而某些潜变量（例如，z_3 和 z_7）的使用较弱。这与我们的理论一致：对数收益率中的比率会对不同维度的噪声尺度进行归一化，从理论上阻碍解缠。

对数价格的结果。图 14 展示了对数价格对应的分割结果。在这里，iSDS 提取出时间跨度更长的状态：绿色状态与牛市时期相对应，黄色状态则与横盘/熊市阶段相对应。与收益率模型相比，推断出的状态持续性更强，并能追踪宏观经济阶段，这与现值框架相关文献中的结论一致（Chan 和 Santi，2021年）。

解码器雅可比矩阵（图 15）呈现出的结构明显比收益率情形下更清晰（参见图 13）。若干潜变量与行业或行业组合相对应（例如，z_5：原材料/必需消费品；z_5、z_7：必需消费品；z_8：房地产；z_4、z_6：医疗保健）。其他潜变量（z_2 和 z_10）则表现为全局因子。一些行业（例如，信息技术和工业）分布在多个潜变量中，考虑到这些行业的周期性及其对市场状况的敏感性，这种情况是合理的。总体而言，对数价格模型得到了结构更清晰、能够感知行业差异的表示，这符合我们的可识别性预期。行业特异性解缠未能进一步增强，也与合成数据结果一致：iSDS 未能取得高强MCC分数。

总结。比较两种数据模态可知，在对数收益率上应用 iSDS 能发现短时波动率状态，但行业解缠较弱；而对数价格则呈现出更持久的宏观状态，以及更清晰的行业对齐潜变量。对数价格数据模态中更强的结构化表示，与数据的非平稳性以及我们对可识别表示提出的异质性要求相一致。更多细节请参见附录 E.5，其中我们还考察了 iSDS 在两种数据模态中学习到的潜在结构。

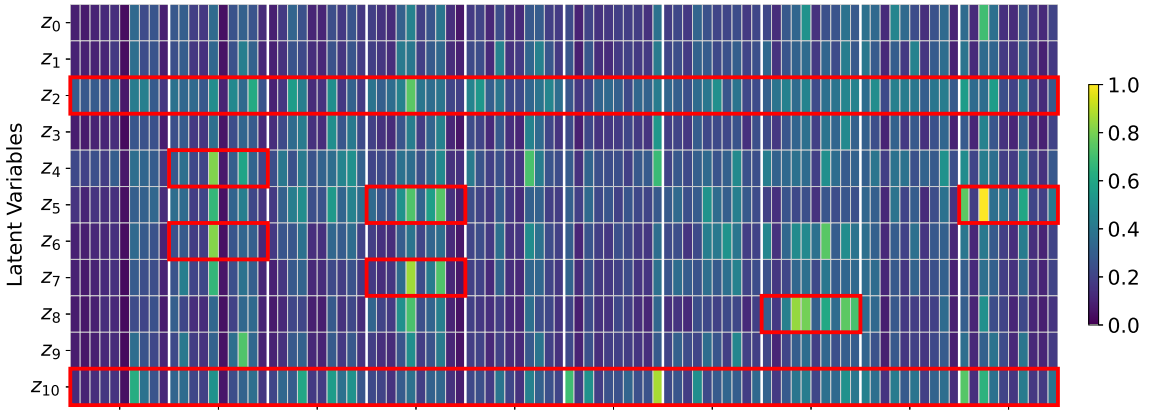


图 15：按行业分组的对数价格解码器雅可比矩阵（潜变量 → 资产）。

7.3 气候数据

气候数据具有显著的季节性模式，近期已有研究在这一背景下探索依赖状态的方法（Saggioro 等，2020年）。对地球系统进行建模，对于理解气候动态和检测极端事件至关重要。本节将研究高维时空气候数据，以展示我们的理论框架在这一科学领域中的适用性。

我们研究非洲的归一化植被指数（NDVI），采用与 Varando 等人（2022）类似的设置。NDVI 通过近红外与红光反射率的归一化比值衡量植被绿度，广泛用作植被生产力和陆地表面物候的替代指标（Pettorelli，2013）。我们使用 MODIS/Terra 16天植被指数 L3 全球0.05度 CMG V006 产品（Didan，2015）以及 ENSO 3.4 气候指数；Varando 等人（2022）提供了这两者的预处理形式。该数据集采用与其研究类似的预处理方法：将 NDVI 重投影至 25 km 等面积可扩展地球（EASE）网格[5]，，并保留对应非洲的像素（边界框 $[-20,\ 55]$、 $[-37,\ 38]$ ，纬度、经度]分别为 []）。每年，该产品提供从第 1 天到第 353 天、间隔为 16 天的 NDVI 值。我们选取 2001—2020 年这一时段。缺失值通过线性插值按时间（逐像素）填补。由此得到的数据集是一个包含 460 个时间步和 289×356 个像素的视频张量。ENSO 3.4 指数经过采样，以匹配 16天时间分辨率。由于该指数按日提供，我们采用 16 天滑动窗口，随后进行 $16\times$ 下采样。基于这些数据集，我们预期 iMSM 和 iSDS 能够提取对应季节动态的状态，并与 Granger-PCA（Varando 等人，2022）类似，考察学习到的潜变量是否与 ENSO 3.4 相吻合。

5. 关于 EASE 网格重投影的信息，请参见此处 https://nsidc.org/data/user-resources/help-center/guide-ease-grids。

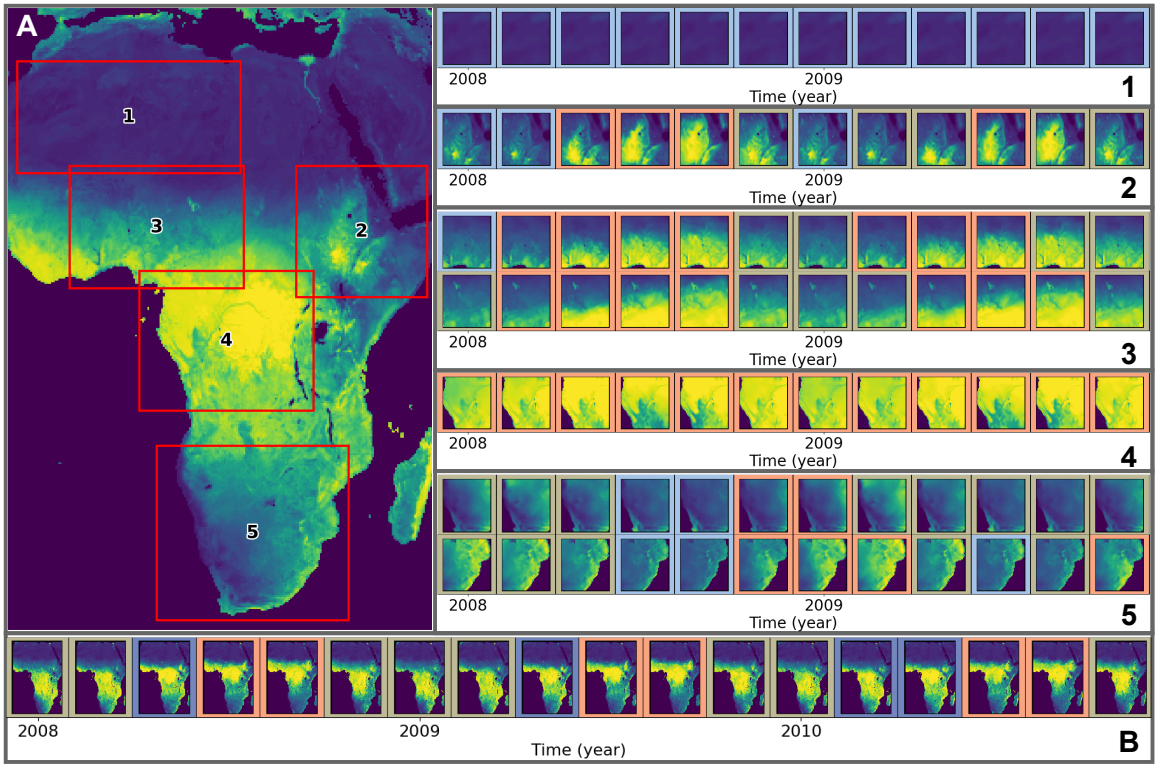


图16：A. 非洲大陆上 iSDS 的分割示例：(1) 撒哈拉，(2) 埃塞俄比亚，(3) 萨赫勒，(4) 刚果，以及 (5) 南非。B. iMSM 的分割结果。

模型设置。 我们选取最近的 360 个时间步，并将最初的 100 个留作测试和模型选择。我们按照 Varando 等人（2022）的方法训练 iMSM：对陆地像素（共 56196 个像素）进行主成分分析，并保留 10 个主成分。对于 iSDS，我们从陆地像素中随机选取 64×64 个补丁，并使用 CNN-MLP 编码器-解码器。这里我们考虑 20 个潜变量，以纳入有关大陆形状的额外空间特征。鉴于气候固有的季节性，我们假设 $K = 4$ 已知（每个季节对应一个状态）。通过模型选择，我们为 iMSM 设置异质协方差、在状态转移中使用 GELU 激活函数，并设置 $M = 1$；为 iSDS 设置 $M = 2$。更多训练和模型选择细节分别见附录 E.2 和 E.6。

分割结果。图16展示了两个模型的分割结果。由于 iMSM 基于 10 个主成分进行训练，提取出的状态会在整个大陆范围内发挥作用。我们观察到三个连贯的季节状态，每年都会重复出现：3月至5月、6月至8月以及9月至次年2月。相比之下，iSDS 得到的状态结构具有区域依赖性：撒哈拉和刚果以一种状态为主（图16A中的区块1和4）；埃塞俄比亚、萨赫勒和南非则在两种或三种状态之间交替（图16B中的区块2、3和5）。尽管不同地区的空间状态结构各不相同，iSDS 中的状态效应却是共享的：例如，南非的分割与萨赫勒/埃塞俄比亚的分割大致相反。附录 E.6 提供了全球和区域基于年份的结构的可视化结果。

表 2：厄尔尼诺-南方涛动与 iMSM主成分的季节性及整体对齐情况。厄尔尼诺-南方涛动领先时的峰值相关性及滞后阶数。

季节	相关性	滞后天数
所有时间	0.38	80
3月至5月	0.49	208
6月至9月	0.37	80
10月至次年2月	**0.53**	**80**

表3：各季节 ENSO 与 iSDS潜变量 的对齐情况。对于每个区域，我们报告 ENSO 领先时的平均峰值相关性和滞后阶数。

区域	季节	相关性	滞后天数
撒哈拉	所有时间	0.15	48
埃塞俄比亚	1月至3月	**0.56**	**144**
	4月至8月	0.27	100
	8月至12月	0.26	0
萨赫勒	3月至8月	0.18	112
	9月至2月	0.19	0
刚果	所有时间	0.12	144
南非	3月至6月和9月至11月	0.25	64
	6月至9月	0.22	128
	12月至次年2月	0.28	192

与厄尔尼诺-南方涛动的相关性。为了定量评估学习得到的状态依赖动力学，我们将学习得到的潜变量与预处理后的厄尔尼诺-南方涛动 3.4 时间序列联系起来。对于 iMSM（全球），我们根据状态确定的季节切片划分时间序列，并计算每个主成分与厄尔尼诺-南方涛动 3.4 在0至240天滞后范围内的滞后皮尔逊相关（步长为16天；厄尔尼诺-南方涛动领先），同时报告“所有时间”基线（即不作划分）。对于 iSDS（区域），我们对每个区域抽取 $10\ 64 \times 64$ 个陆地区块，计算潜变量与厄尔尼诺-南方涛动的相关矩阵（潜变量 $\times$ 滞后），并使用 Fisher z 平均法将其合并。对于每个季节/区域，我们报告峰值相关强度及其滞后阶数。完整的相关矩阵及其他可视化结果见附录 E.6。

结果汇总于表2和表3中。对于 iMSM（参见表2），我们发现，按状态划分得到的峰值比“所有时间”这一基准更清晰。这表明，跨季节汇总会削弱厄尔尼诺-南方涛动对时间序列的滞后效应。结果显示，iMSM 将时间序列划分为不同状态，在这些状态下，厄尔尼诺-南方涛动的影响会以不同的滞后阶数显现。对于 iSDS（参见表3），各区域的响应存在差异。我们发现，撒哈拉、刚果和萨赫勒与厄尔尼诺-南方涛动并无显著关联。另一方面，与冬季相比，南非在南非夏季与厄尔尼诺-南方涛动的相关性略高。此外，埃塞俄比亚在1月至3月期间与厄尔尼诺-南方涛动的相关性非常高。在气候分析中，研究人员会考察厄尔尼诺-南方涛动等现象在不同滞后尺度上的影响（Rouault 和 Tomety，2022年），以及其在不同区域的影响（例如萨赫勒（Nicholson，2013年）和埃塞俄比亚（Bekele-Biratu 等人，2018年））。总体而言，iMSM 和 iSDS 能够揭示厄尔尼诺-南方涛动等外部驱动因素的重要信息，因为它们将季节效应和区域效应分别解耦。

总结。 非洲的气候数据呈现出明显的季节性归一化植被指数（NDVI）结构。iMSM 捕捉全局尺度，而 iSDS 则揭示区域特定的状态结构。尽管在气候时间序列中，潜变量的可解释性较为复杂，我们的状态可识别性结果仍能支持可解释的状态条件分析。将 iSDS 扩展至近期提出的可识别时空框架，或可进一步区分区域效应，并提升潜在空间的可解释性（Wang 等人，2025年）。

8 结论

我们为具有多滞后依赖的状态切换模型构建了统一的可识别性框架。对于马尔可夫切换模型，我们将模型表述为时间有限混合模型，并证明在非线性高斯设定下，状态数量和状态特定的多滞后转移均具有可识别性；附录 A 中还提供了全面的非参数分析。对于切换动力系统，我们确立了潜变量在置换和缩放意义下的可识别性，并推导出状态依赖的潜在因果图在状态置换和节点置换意义下可识别的条件。重要的是，通过灵活的架构（函数类）和噪声假设，这些保证在完全无监督设定下依然成立。在实证方面，我们在合成数据上验证了理论，并展示了 iMSM 和 iSDS 在各科学领域的应用，表明可识别性通过确保表示在可接受的等价关系下唯一，从而带来可信的解释。

虽然我们为 iMSM 提供了强有力的可识别性结果，但 iSDS 的保证依赖于分段线性生成函数以及特定的噪声和函数类，而这些条件未必普遍成立。在估计方面，我们依赖变分推断，且未证明统计一致性。相关文献中已有一致性结果，其假设是变分族具有通用逼近性质（Khemakhem 等人，2020年）。从实证结果来看，我们观察到模型在合成数据上表现出色；随着维数增加，恢复机制依赖的因果结构会变得更具挑战性。

我们的框架与现有方法互为补充，并为未来研究指明了若干方向。例如，可以纳入同期效应、连续时间动力学和无限混合过程，也可以超越分段线性这一限制。这些扩展将拓宽可辨识状态切换模型在科学发现中的适用范围。此外，我们的可识别性结果不仅适用于采用时间齐次马尔可夫切换的标准 MSM，也适用于具有时变状态分布的模型，为实验探索提供了空间。

致谢与经费披露

我们感谢 Gherardo Varando 和 Gustau Camps-Valls（瓦伦西亚大学图像处理实验室）就气候应用进行的早期讨论，并分享了第7.3节中使用的预处理数据集。C. B.-R. 也感谢伦敦帝国理工学院的同事就本文的理论结果和内容展开了有益的讨论。Y.W 获得海军研究办公室（资助编号 N00014-23-1-2590）、美国国家科学基金会（资助编号 2231174、2310831、2428059、2435696、2440954）以及密歇根大学数据科学研究所“推进原创数据科学”（PODS）项目资助的部分支持。

参考文献

Amir Alizadeh 和 Nikos Nomikos。用于对冲股票指数的马尔可夫状态切换方法。《期货市场杂志：期货、期权及其他衍生产品》，24 (7):649–674，2004年。

Elizabeth S Allman、Catherine Matias 和 John A Rhodes。具有多个观测变量的潜在结构模型中的参数可识别性。《统计年刊》，37(6A): 3099–3132，2009年。Yonghong An、Yingyao Hu、约翰斯·霍普金斯大学和 Matt Shum。隐马尔可夫模型的可识别性与推断。技术报告，技术报告，2013年。Andrew Ang 和 Geert Bekaert。体制转换下的国际资产配置。《金融研究评论》，15(4):1137–1187，2002年。Abdul Fatir Ansari、Konstantinos Benidis、Richard Kurle、Ali Caner Turkmen、Harold Soh、Alexander J Smola、Bernie Wang 和 Tim Januschowski。用于时间序列的深度显式持续时间切换模型。载于神经信息处理系统进展，34: 29949–29961，2021年。Charles K Assaad、Emilie Devijver 和 Eric Gaussier。时间序列因果发现方法综述与评估。《人工智能研究杂志》，73:767–819，2022年。Mehdi Azzouzi 和 Ian T Nabney。使用切换状态空间模型对金融时间序列建模。载于《*IEEE/IAFE 1999*年金融工程计算智能会议》*(CIFEr)(IEEE*目录号：*99TH8408)*，第240–249页。IEEE，1999年。Mohammad Babaeizadeh、Chelsea Finn、Dumitru Erhan、Roy H. Campbell 和 Sergey Levine。随机变分视频预测。载于国际学习表征会议，2018年。URL https://openreview.net/forum?id=rk49Mg-CW。Nathan S Balke 和 Mark E Wohar。股票价格中的市场基本面与理性泡沫：贝叶斯视角。《应用计量经济学杂志》，24(1):35–75，2009年。Carles Balsells-Rodas、Yixin Wang 和 Yingzhen Li。论切换动力系统的可识别性。载于第四十一届国际机器学习会议，2024年。URL https://openreview.net/forum?id=Eew3yUQQtE。Carles Balsells-Rodas、Xavier Sumba、Tanmayee Narendra、Ruibo Tu、Gabriele Schweikert、Hedvig Kjellstrom 和 Yingzhen Li。基于条件平稳时间序列的因果发现。载于第*42*届国际机器学习会议，2025年。URL https://openreview.net/forum?id=j88QAtutwW。Endalkachew Bekele-Biratu、Wassila M. Thiaw 和 Diriba Korecha。埃塞俄比亚贝尔格雨季降雨的季节内变率。《国际气候学杂志》，38(7):2940–2953，2018年。Christopher M. Bishop。《模式识别与机器学习》（信息科学与统计）。施普林格出版社，柏林，海德堡，2006年。ISBN 0387310738。彭博有限合伙公司。美国股票的历史收盘价（px_last）——股票代码见本文所列。彭博终端，伦敦帝国理工学院，2025年。数据范围：2015年1月1日至2025年5月31日；访问日期：2025年11月9日。

Tim Bollerslev。广义自回归条件异方差。《计量经济学》，31(3):307–327，1986年。Carlos M Carvalho 和 Hedibert F Lopes。基于模拟的马尔可夫切换随机波动率模型序贯分析。《计算统计与数据分析》，51(9):4526–4542，2007年。Giuseppe Cavaliere、Michele Costa 和 Luca De Angelis。运用多元马尔可夫切换方法研究股票市场行为。载于《潜变量进展：方法、模型与应用》，第185–196页。施普林格，2014年。Joshua CC Chan 和 Caterina Santi。现值模型中的投机泡沫：一种贝叶斯马尔可夫切换状态空间方法。*Journal of Economic Dynamics and Control*，127:104101，2021年。Junyoung Chung、Kyle Kastner、Laurent Dinh、Kratarth Goel、Aaron C Courville 和 Yoshua Bengio。用于序列数据的循环潜变量模型。载于神经信息处理系统进展，28，2015年。Pierre Comon。独立成分分析：一个新概念？《信号处理》，36(3):287–314，1994年。Rama Cont。资产收益的经验特性：典型事实与统计问题。《定量金融》，1(2):223，2001年。Kamel Didan。MOD13C1 MODIS/Terra 植被指数16天L3全球0.05度CMG V006，2015年。Zhe Dong、Bryan Seybold、Kevin Murphy 和 Hung Bui。用于切换非线性动力系统的折叠摊销变分推断。载于国际机器学习会议，第2638–2647页。PMLR，2020年。Yariv Ephraim 和 William JJ Roberts。重新审视语音信号的自回归隐马尔可夫建模。《*IEEE*信号处理快报》，12(2):166–169，2005年。Ronald A Fisher。来自无限大总体的样本中相关系数值的频率分布。《生物统计学》，10(4):507–521，1915年。Emily Fox、Erik B Sudderth、Michael I Jordan 和 Alan S Willsky。切换动态线性模型的贝叶斯非参数推断。《*IEEE*信号处理汇刊》，59 (4):1569–1585，2011年。Marco Fraccaro、Simon Kamronn、Ulrich Paquet 和 Ole Winther。用于无监督学习的解耦识别与非线性动力学模型。载于神经信息处理系统进展，30，2017年。Sylvia Frühwirth-Schnatter 和 Sylvia Frèuhwirth-Schnatter。有限混合与马尔可夫切换模型，第425卷。施普林格，2006年。

Élisabeth Gassiat、Alice Cleynen 和 Stephane Robin。有限状态空间非参数隐马尔可夫模型中的推断及其应用。*Statistics and Computing*，26(1):61–71，2016年。Elisabeth Gassiat、Sylvain Le Corff 和 Luc Lehéricy。一般状态空间下非参数平移隐马尔可夫模型的可识别性与一致估计。机器学习研究杂志，21(1):4589–4628，2020年。Victor Geadah、International Brain Laboratory 和 Jonathan W. Pillow。利用无限递归切换线性动力系统解析神经动力学。载于第十二届国际学习表征会议，2024年。URL https://openreview. net/forum?id=YIls9HEa52。Zoubin Ghahramani 和 Geoffrey E Hinton。切换状态空间模型的变分学习。*Neural computation*，12(4):831–864，2000年。Massimo Guidolin 和 Allan Timmermann。多变量状态转换下的资产配置。*Journal of Economic Dynamics and Control*，31(11):3503–3544，2007年。Markus Haas、Stefan Mittnik 和 Marc S Paolella。马尔可夫转换 GARCH 模型的一种新方法。*Journal of financial Econometrics*，2(4):493–530，2004年。Hermanni Hälvä 和 Aapo Hyvarinen。隐马尔可夫非线性独立成分分析：从非平稳时间序列中进行无监督学习。载于人工智能不确定性会议，第939–948页。PMLR，2020年。Hermanni Hälvä、Sylvain Le Corff、Luc Lehéricy、Jonathan So、Yongjie Zhu、Elisabeth Gassiat 和 Aapo Hyvarinen。利用结构化非线性独立成分分析，从噪声数据中分离可辨识特征。神经信息处理系统进展，34:1624–1633，2021年。James D Hamilton。非平稳时间序列与商业周期经济分析的一种新方法。*Econometrica: Journal of the econometric society*，第357–384页，1989年。Patrik Hoyer、Dominik Janzing、Joris M Mooij、Jonas Peters 和 Bernhard Schölkopf。利用加性噪声模型进行非线性因果发现。载于神经信息处理系统进展，21，2008年。Aapo Hyvarinen 和 Hiroshi Morioka。通过时间对比学习和非线性独立成分分析进行无监督特征提取。载于神经信息处理系统进展，29，2016年。Aapo Hyvarinen 和 Hiroshi Morioka。时间依赖的平稳源的非线性独立成分分析。载于人工智能与统计学，第460–469页。PMLR，2017年。Aapo Hyvärinen 和 Petteri Pajunen。非线性独立成分分析：存在性与唯一性结果。神经网络，12(3):429–439，1999年。Aapo Hyvarinen、Hiroaki Sasaki 和 Richard Turner。利用辅助变量和广义对比学习进行非线性独立成分分析。载于第二十二届人工智能与统计学国际会议，第859–868页。PMLR，2019年。

Matthew J Johnson、David K Duvenaud、Alex Wiltschko、、Sandeep R Datta 和 Ryan P Adams。将图模型与神经网络相结合，实现结构化表征和快速推断。神经信息处理系统进展，29，2016年。Orren Karniol-Tambour、David M. Zoltowski、E. Mika Diamanti、Lucas Pinto、Carlos D Brody、David W. Tank 和 Jonathan W. Pillow。利用切换非线性动力系统对脑区之间的状态依赖通信进行建模。载于第十二届国际学习表征会议，2024年。网址：https://openreview. net/forum?id=WQwV7Y8qwa。Ilyes Khemakhem、Diederik Kingma、Ricardo Monti 和 Aapo Hyvarinen。变分自编码器与非线性独立成分分析：一个统一框架。载于国际人工智能与统计会议，第2207–2217页。PMLR，2020年。Diederik P. Kingma 和 Jimmy Ba。Adam：一种随机优化方法。载于 Yoshua Bengio 和 Yann LeCun（编），第三届国际学习表征会议，*ICLR 2015*，圣迭戈，美国加利福尼亚州，*2015*年*5*月*7*日至*9*日，会议论文集，2015年。Diederik P. Kingma 和 Max Welling。自编码变分贝叶斯。载于 Yoshua Bengio 和 Yann LeCun （编），第二届国际学习表征会议，*ICLR 2014*，班夫，艾伯塔省，加拿大，*2014*年*4*月*14*日至*16*日，会议论文集，2014年。网址：http://arxiv.org/abs/1312.6114。Bohdan Kivva、Goutham Rajendran、Pradeep Ravikumar 和 Bryon Aragam。无辅助信息时深度生成模型的可识别性。神经信息处理系统进展，35：15687–15701，2022年。Joseph B Kruskal。三阶数组：三线性分解的秩与唯一性及其在算术复杂度和统计学中的应用。线性代数及其应用，18（2）：95–138，1977年。Sébastien Lachapelle、Pau Rodriguez、Yash Sharma、Katie E Everett、Rémi Le Priol、Alexan- dre Lacoste 和 Simon Lacoste-Julien。通过机制稀疏性正则化实现解缠：非线性独立成分分析的新原则。载于因果学习与推理会议，第428–484页。PMLR，2022年。Yingzhen Li 和 Stephan Mandt。解缠序列自编码器。载于国际机器学习会议，2018年。Scott Linderman、Matthew Johnson、Andrew Miller、Ryan Adams、David Blei 和 Liam Paninski。循环切换线性动力系统中的贝叶斯学习与推断。载于人工智能与统计学，第914–922页。PMLR，2017年。Scott W Linderman、Andrew C Miller、Ryan P Adams、David M Blei、Liam Paninski 和 Matthew J Johnson。循环切换线性动力系统。*arXiv*预印本 *arXiv:1610.08466*，2016年。Georg Lindgren。混合分布的马尔可夫状态模型与切换回归。斯堪的纳维亚统计学杂志，第81–91页，1978年。

Phillip Lippe、Sara Magliacane、Sindy Löwe、Yuki M Asano、Taco Cohen 和 Efstratios Gavves。BISCUIT：从二元交互中进行因果表征学习。载于Robin J. Evans 和 Ilya Shpitser（编），第三十九届人工智能不确定性会议论文集，第216卷机器学习研究论文集，第1263–1273页。PMLR，2023年7月31日至8月4日。网址：https://proceedings.mlr.press/v216/ lippe23a.html。

Phillip Lippe、Sara Magliacane、Sindy Löwe、Yuki M Asano、Taco Cohen 和 Efstratios Gavves。交互系统中瞬时和时间效应的因果表征学习。载于第十一届国际学习表征会议，2023b。网址：https://openreview.net/forum?id=itZ6ggvMnzS。

Yongtuo Liu、Sara Magliacane、Miltiadis Kofinas 和 Efstratios Gavves。图切换动力系统。载于国际机器学习会议，第 21867–21883 页。PMLR，2023年。

Francesco Locatello、Stefan Bauer、Mario Lucic、Gunnar Raetsch、Sylvain Gelly、Bernhard Schölkopf 和 Olivier Bachem。对无监督学习解耦表征中的常见假设提出挑战。载于国际机器学习会议，第 4114–4124 页。PMLR，2019年。

Mohammadreza Mahmoudi 和 Hana Ghaneei. 结构体制的检测及原油市场对加拿大股票市场影响的分析：马尔可夫体制转换方法. 《经济与金融研究》, 39(4):722–734, 2022年。

Geoffrey J McLachlan 和 David Peel. 有限混合模型. John Wiley & Sons, 2000年。

Pedro A.M. Mediano、Fernando E. Rosas、Andrea I. Luppi、Valdas Noreika、Anil K. Seth、Robin L. Carhart-Harris、Lionel Barnett 和 Daniel Bor. 神经信号多样性的频谱和时间分辨估计. *eLife*, 2023年3月。doi: doi.org/10.7554/ eLife.88683.1. URL http://dx.doi.org/10.7554/eLife.88683.1.

Boris Mityagin. 实解析函数的零点集. *arXiv*预印本 *arXiv:1512.07276*, 2015年。

Hiroshi Morioka、Hermanni Hälvä 和 Aapo Hyvarinen。《非线性向量自回归过程的独立创新分析》。载于国际人工智能与统计会议，第 1549–1557 页。PMLR，2021年。

Kevin P Murphy。《切换卡尔曼滤波器》，1998年。

Sharon E Nicholson。《西非萨赫勒地区：关于降雨体制及其年际变率的近期研究综述》。《style id='1'>International Scholarly Research Notices，2013年(1):453521，2013年。

Adam Paszke、Sam Gross、Francisco Massa、Adam Lerer、James Bradbury、Gregory Chanan、Trevor Killeen、Zeming Lin、Natalia Gimelshein、Luca Antiga、Alban Desmaison、Andreas Kopf、Edward Yang、Zachary DeVito、Martin Raison、Alykhan Tejani、Sasank Chilamkurthy、Benoit Steiner、Lu Fang、Junjie Bai 和 Soumith Chintala。PyTorch：一种命令式风格、

高性能深度学习库。载于神经信息处理系统进展 *32*，第8024–8035页。Curran Associates, Inc.，2019年。Jonas Peters、Dominik Janzing 和 Bernhard Schölkopf。《因果推断要义：基础与学习算法》。麻省理工学院出版社，2017年。Nathalie Pettorelli。归一化差异植被指数。《归一化差异植被指数》，第1–208页，2013年10月。doi：10.1093/acprof:osobl/9780199693160.001.0001。A Poritz。线性预测隐马尔可夫模型与语音信号。载于*ICASSP'82*。*IEEE*声学、语音与信号处理国际会议，第7卷，第1291–1294页。IEEE，1982年。Abdellah Rahmani 和 Pascal Frossard。因果时序机制结构学习。载于第*28*届人工智能与统计国际会议，2025年。M Rouault 和 FS Tomety。厄尔尼诺—南方涛动对本格拉上升流的影响。《物理海洋学杂志》，52(10)：2573–2587，2022年。Elena Saggioro、Jana de Wiljes、Marlene Kretschmer 和 Jakob Runge。从观测时间序列中重构机制依赖的因果关系。《混沌：非线性科学跨学科期刊》，30(11)：113115，2020年。Vaibhav Saxena、Jimmy Ba 和 Danijar Hafner。发条变分自编码器。载于神经信息处理系统，2021年。Bernhard Schölkopf、Francesco Locatello、Stefan Bauer、Nan Rosemary Ke、Nal Kalchbrenner、Anirudh Goyal 和 Yoshua Bengio。迈向因果表征学习。《*IEEE*会刊》，109(5)：612–634，2021年。Xiangchen Song、Weiran Yao、Yewen Fan、Xinshuai Dong、Guangyi Chen、Juan Carlos Niebles、Eric Xing 和 Kun Zhang。未知非平稳性下的时间解耦表示学习。神经信息处理系统进展，36：8092–8113，2023年。Xiangchen Song、Zijian Li、Guangyi Chen、Yujia Zheng、Yewen Fan、Xinshuai Dong 和 Kun Zhang。具有非平稳稀疏转移的因果时序表征学习。神经信息处理系统进展，37：77098–77131，2024年。Robert L Thorndike。谁属于这一族？《心理计量学》，18(4)：267–276，1953年。Gherardo Varando、Miguel-Ángel Fernández-Torres、Jordi Muñoz-Marí 和 Gustau Camps-Valls。利用格兰杰主成分分析学习因果表征。载于*UAI 2022*因果表征学习研讨会，2022年。Kun Wang、Sumanth Varambally、Duncan Watson-Parris、Yian Ma 和 Rose Yu。从时空数据中发现潜在因果图。载于第*42*届国际机器学习会议，2025年。Sidney J Yakowitz 和 John D Spragins。有限混合的可识别性。《数理统计年刊》，39(1)：209–214，1968年。

Toru Yanagawa、Zenas C. Chao、Naomi Hasegawa 和 Naotaka Fujii。在意识状态和无意识状态下的大规模信息流：一项针对猴子的皮层脑电图研究。*PLOS ONE*，8(11):null，2013年11月。doi: 10.1371/journal.pone.0080845。URL https://doi.org/10.1371/journal. pone.0080845。

Weiran Yao、Guangyi Chen 和 Kun Zhang。时间解耦表征学习。载于 Alice H. Oh、Alekh Agarwal、Danielle Belgrave 和 Kyunghyun Cho 编，《神经信息处理系统进展》，2022a。URL https://openreview.net/forum? id=Vi-sZWNA_Ue。

Weiran Yao、Yuewen Sun、Alex Ho、Changyin Sun 和 Kun Zhang。从一般时间数据中学习时间因果潜在过程。载于《国际学习表征会议》，2022b。URL https://openreview.net/forum?id=RDlLMjLJXdq。

Yixiu Zhao 和 Scott Linderman。重新审视结构化变分自编码器。载于国际机器学习会议，第42046–42057 页。PMLR，2023 年。

Xun Zheng、Bryon Aragam、Pradeep K Ravikumar 和 Eric P Xing。《无撕裂有向无环图：用于结构学习的连续优化》。神经信息处理系统进展，31，2018年。

附录 A. 定理 8 的证明

证明概要：我们将证明策略分为 4 个步骤。

1. 我们证明，可识别性的要求是轨迹族线性无关（附录 A.1）。2. 我们给出非参数函数乘积的线性无关性结果。首先从 $M = 1$ 开始（附录 A.2.1），然后加强假设（附录 A.2.2），以允许 M 个函数的乘积（附录 A.2.3）。3. 我们通过归纳法证明，阶数为 M 的非参数转移所对应的轨迹族线性无关（附录 A.2.4）。4. 我们证明参数假设 (m1)、(m2) 满足轨迹族线性无关的条件（附录 A.3）。

A.1 线性无关性确保可识别性

定理15 假设 $\mathcal{P}^T_{\mathcal{A},\mathcal{B}}$ 中的函数在有限混合下线性无关，则分布族 $\mathcal{M}^T(\Pi^M_{\mathcal{A}}, \mathcal{P}^M_{\mathcal{B}})$ 如定义 7 所述是可辨识的。

证明 根据命题 4，在有限混合下，$\mathcal{P}^T_{\mathcal{A},\mathcal{B}}$ 的线性无关性意味着根据定义 2，在置换意义下具有可辨识性。然后，对于来自 $\tilde{}$ 定义 7 的 $p(\boldsymbol{z}_{1:T})$ 和 $\tilde{p}(\boldsymbol{z}_{1:T})$，我们有 $C = C$，并且对于每个 $1 \leq i \leq C$，都存在 $1 \leq j \leq C$，使得 $c_i = \tilde{c}_j$ 并且：

$$p_{a^i}(\boldsymbol{z}_{1:M}) \prod_{t=M+1}^{T} p_{b^i_t}(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}) = p_{\tilde{a}^j}(\boldsymbol{z}_{1:M}) \prod_{t=M+1}^{T} p_{\tilde{b}^j_t}(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}),\ \forall \boldsymbol{z}_{1:T} \in \mathbb{R}^{Tm}. \quad (31)$$

鉴于我们拥有条件概率密度函数，如果联合分布在 $\boldsymbol{z}_{1:T}$ 上相等，那么在 $\boldsymbol{z}_{1:T-1}$ 上的分布也相等：

$$p_{a^i}(\boldsymbol{z}_{1:M}) \prod_{t=M+1}^{T-1} p_{b^i_t}(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}) = p_{\tilde{a}^j}(\boldsymbol{z}_{1:M}) \prod_{t=M+1}^{T-1} p_{\tilde{b}^j_t}(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}),\ \forall \boldsymbol{z}_{1:T-1} \in \mathbb{R}^{(T-1)m}. \quad (32)$$

因此，对于所有 $\boldsymbol{z}_t, \ldots, \boldsymbol{z}_{t-M} \in \mathbb{R}^m$，我们有 $p_{b^i_T}(\boldsymbol{z}_T|\boldsymbol{z}_{T-1}, \ldots, \boldsymbol{z}_{T-M}) = p_{\tilde{b}^j_T}(\boldsymbol{z}_T|\boldsymbol{z}_{T-1}, \ldots, \boldsymbol{z}_{T-M})$。对于其他时间索引，我们也可以依照相同的推理，得到对于所有 $t > M$ $\boldsymbol{z}_t \ldots \boldsymbol{z}_{t-M} \in \mathbb{R}^m$ $p_{b^i_t}(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}) = p_{\tilde{b}^j_t}(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M})$。类似的逻辑也适用于初始分布：对于所有 $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_M \in \mathbb{R}^m$，我们有 $p_{a^i}(\boldsymbol{z}_{1:M}) = p_{\tilde{a}^j}(\boldsymbol{z}_{1:M})$；并且有 $\tilde{K}_0 = K_0$，$K = K$。

最后，若存在 $t_1 \neq t_2$ 使得 $b^i_{t_1} = b^i_{t_2}$，但 $\tilde{b}^j_{t_1} \neq \tilde{b}^j_{t_2}$，则对于任意 $\boldsymbol{\alpha} \in \mathbb{R}^{Mm}$，$\boldsymbol{\beta} \in \mathbb{R}^m$，我们有：

$$\begin{aligned} p_{\tilde{b}^j_{t_1}}(\boldsymbol{z}_{t_1} = \boldsymbol{\beta} | \boldsymbol{z}_{t_1-1:t_1-M} = \boldsymbol{\alpha}) &= p_{b^i_{t_1}}(\boldsymbol{z}_{t_1} = \boldsymbol{\beta} | \boldsymbol{z}_{t_1-1:t_1-M} = \boldsymbol{\alpha}) \\ &= p_{b^i_{t_2}}(\boldsymbol{z}_{t_2} = \boldsymbol{\beta} | \boldsymbol{z}_{t_2-1:t_2-M} = \boldsymbol{\alpha}) \\ &= p_{\tilde{b}^j_{t_2}}(\boldsymbol{z}_{t_2} = \boldsymbol{\beta} | \boldsymbol{z}_{t_2-1:t_2-M} = \boldsymbol{\alpha}), \end{aligned}$$

这意味着 $\mathcal{P}^M_{\mathcal{B}}$ 的有限混合之间存在线性相关。我们注意到，这与轨迹族 $\mathcal{P}^{T,M}_{\mathcal{A},\mathcal{B}}$ 在有限混合下线性无关的假设相矛盾：该假设意味着我们应当有

$$\sum_i \sum_j \gamma_{ij} p_{b^i_{1:T-1}}(\boldsymbol{z}_{1:T-1}) p_{b^j_T}(\boldsymbol{z}_T | \boldsymbol{z}_{T-1:T-M}) = 0, \quad \forall \boldsymbol{z}_{1:T} \in \mathbb{R}^{Tm}, \tag{33}$$

当且仅当对所有 $1 \leq i \leq C/K$，$1 \leq j \leq K$，$\gamma_{ij} = 0$。然而，鉴于 $\mathcal{P}^M_{\mathcal{B}}$ 上存在线性相关，我们可以重新排列各项，并设定 $\gamma_{i,\,j} = \gamma_j \neq 0$，$1 \leq j \leq K$，使上述等式成立。因此，这就产生了矛盾，由此可得，对所有 $M < t_1$，$t_2 \leq T$，$b^i_{t_1} = b^i_{t_2} \Leftrightarrow \tilde{b}^j_{t_1} = \tilde{b}^j_{t_2}$。 ■

下一步是说明轨迹族 $\mathcal{P}^{T,M}_{\mathcal{A},\mathcal{B}}$ 在有限混合下线性无关的条件。

A.2 非参数乘积族的线性无关性

为了确立非参数轨迹族的线性无关性，我们分析条件分布的连续乘积。例如，当 $M = 1$ 时，考察两个连续变量可得到乘积

$$\{p(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}, s_t) p(\boldsymbol{z}_{t+1} \mid \boldsymbol{z}_t, s_{t+1})\}$$

具有一个重叠变量 $\boldsymbol{z}_t$。增加 M 会增加考虑 $M+1$ 个连续变量时的重叠数量。下面的乘积

$$\left\{ \prod_{\ell=0}^{M} p(\boldsymbol{z}_{t+\ell} \mid \boldsymbol{z}_{t-1+\ell}, \ldots, \boldsymbol{z}_{t-M+\ell}, s_{t+\ell}) : (s_t, \ldots, s_{t+M}) \in [K]^{M+1} \right\}, \tag{34}$$

包含两类重叠变量，用于 $M > 1$：

1. M 个变量，既作为观测变量出现，也作为条件变量出现，即式（34）中的 $\boldsymbol{z}_{t+M-1}$、……、$\boldsymbol{z}_t$。2. $M-1$ 个变量，仅作为多个因子中的条件变量出现，即式（34）中的 $\boldsymbol{z}_{t-1}$、……、$\boldsymbol{z}_{t-M+1}$。

对于 $M > 1$，条件变量 ($\boldsymbol{z}_{t-1}$、……、$\boldsymbol{z}_{t-M+1}$) 之间额外的重叠使分析变得复杂，因为这些重叠妨碍了对 Balsells-Rodas 等人（2024）中所述情形 $M = 1$ 的直接扩展。

A.2.1 预备知识

我们回顾 Balsells-Rodas 等人（2024）中情形 $M=1$ 的核心结果，并将其命名为积函数的线性无关：带***1***个连续重叠变量（*LIPO-1*）。该结果给出了在只有一个重叠变量的情况下，乘积函数保持线性无关的条件。

引理 16（LIPO-1）考虑两个实值标量函数族 $\mathcal{U}_I := \{u_i(\mathbf{y}\ \boldsymbol{x}) | i \in I\}$和 $\mathcal{V}_J := \{v_j(\boldsymbol{z}, \mathbf{y}) | j \in J\}$，其中 $\boldsymbol{x} \in \mathcal{X}$、$\mathbf{y} \in \mathbb{R}^{d_y}$和 $\boldsymbol{z} \in \mathbb{R}^{d_z}$。我们进一步作出以下假设：

(a1) 正函数值：对所有 $i \in I$，$(\mathbf{y}\ \boldsymbol{x}) \in \mathbb{R}^{d_y} \times \mathcal{X}$，$u_i(\mathbf{y}\ \boldsymbol{x}) > 0$。类似的正函数值假设也适用于 $\mathcal{V}_J$：对所有 $j \in J$，$(\boldsymbol{z}, \mathbf{y}) \in \mathbb{R}^{d_z} \times \mathbb{R}^{d_y}$，$v_j(\boldsymbol{z}, \mathbf{y}) > 0$。*(a2)* 唯一索引：对于 $\mathcal{U}_I$，$i \neq i' \in I \Leftrightarrow \exists\, \boldsymbol{x}$，若 $\mathbf{y}$ 满足 $u_i(\mathbf{y}\ \boldsymbol{x}) \neq u_{i'}(\mathbf{y}\ \boldsymbol{x})$。类似的唯一索引假设也适用于 $\mathcal{V}_J$；*(a3)* $\mathcal{U}_I$在特定非零测度子集上的有限混合线性无关性：对于任意非零测度子集 $\mathcal{Y} \subset \mathbb{R}^{d_y}$，$\mathcal{U}_I$ 在 $(\mathbf{y}\ \boldsymbol{x}) \in \mathcal{Y} \times \mathcal{X}$上包含有限混合下线性无关的函数。*(a4)* $\mathcal{V}_J$在特定非零测度子集上的有限混合线性无关性：存在一个非零测度子集 $\mathcal{Y} \subset \mathbb{R}^{d_y}$，使得对于任意非零测度子集 $\mathcal{Y}' \subset \mathcal{Y}$和 $\mathcal{Z} \subset \mathbb{R}^{d_z}$，$\mathcal{V}_J$ 在 $(\boldsymbol{z}, \mathbf{y}) \in \mathcal{Z} \times \mathcal{Y}'$上包含有限混合下线性无关的函数；

(a5) $\mathcal{V}_J$ 中函数子集在有限混合下线性相关意味着函数重复：对于任意 $\boldsymbol{\beta} \in \mathbb{R}^{d_y}$、任意非零测度子集 $\mathcal{Z} \subset \mathbb{R}^{d_z}$ 以及满足 $|J_0| < +\infty$，$\{v_j(\boldsymbol{z}, \mathbf{y} = \boldsymbol{\beta}) | j \in J_0\}$ 仅在存在满足 $v_j(\boldsymbol{z}, \boldsymbol{\beta}) = v_{j'}(\boldsymbol{z}, \boldsymbol{\beta})$ 对所有 $\boldsymbol{z} \in \mathbb{R}^{d_z}$成立的 $\exists\, j \neq j' \in J_0$时，才包含定义在 $\boldsymbol{z} \in \mathcal{Z}$ 上的线性相关函数的任意子集 $J_0 \subset J$ 。

(a6) $\mathcal{V}_J$的连续性：对于任意 $j \in J$，$v_j(\boldsymbol{z}, \mathbf{y})$关于$\mathbf{y} \in \mathbb{R}^{d_y}$连续。

那么，对于任意非零测度子集 $\mathcal{Z} \subset \mathbb{R}^{d_z}$，$\mathcal{U}_I \otimes \mathcal{V}_J := \{v_j(\boldsymbol{z}, \mathbf{y}) u_i(\mathbf{y}\ \boldsymbol{x}) | i \in I$在$(\boldsymbol{x}, \mathbf{y}\ \boldsymbol{z}) \in \mathcal{X} \times \mathbb{R}^{d_y} \times \mathcal{Z}$上定义的有限混合下包含线性无关函数。

证明 假设该充分性陈述不成立，则存在一个非零测度子集$\mathcal{Z} \subset \mathbb{R}^{d_z}$, $S_0 \subset I \times J$，满足 $|S_0| < +\infty$，以及一组非零值 $\{\gamma_{ij} \in \mathbb{R} | (i, j) \in S_0\}$，使得

$$\sum_{(i,j) \in S_0} \gamma_{ij} v_j(\boldsymbol{z}, \mathbf{y}) u_i(\mathbf{y}, \boldsymbol{x}) = 0, \quad \forall (\boldsymbol{x}, \mathbf{y}, \boldsymbol{z}) \in \mathcal{X} \times \mathbb{R}^{d_y} \times \mathcal{Z}. \tag{35}$$

注意，S_0和γ_{ij}的取值独立于任何 $\boldsymbol{x}$，$\mathbf{y}$ $\boldsymbol{z}$值，但可能依赖于 $\mathcal{Z}$。根据假设 (a1)，指标集 S_0 至少包含两个不同的索引（i，j）和（i'，j'）。特别地，S_0至少包含两个不同的索引（i，j）和（i'，j'），且$j \neq j'$；否则，我们可以将公共项$v_j(\boldsymbol{z}$，$\mathbf{y})$提取出来：

$$\sum_{(i,j) \in S_0} \gamma_{ij} v_j(\boldsymbol{z}, \mathbf{y}) u_i(\mathbf{y}, \boldsymbol{x}) = v_j(\boldsymbol{z}, \mathbf{y}) \left(\sum_{i:(i,j) \in S_0} \gamma_{ij} u_i(\mathbf{y}, \boldsymbol{x}) \right) = 0, \quad \forall (\boldsymbol{x}, \mathbf{y}, \boldsymbol{z}) \in \mathcal{X} \times \mathbb{R}^{d_y} \times \mathcal{Z},$$

由于 S_0中至少存在两个不同的索引（i'，j）和（i，j），因此至少有一个$i' \neq i$，而上面的等式与假设 (a1)–(a3) 矛盾。

现在定义 $J_0 = \{j \in J | \exists (i,\ j) \in S_0\}$ 为出现在 S_0 中的所有可能 j索引构成的集合，并且由 $|S_0| < +\infty$，我们也有 $|J_0| < +\infty$。对于任意 $\boldsymbol{\beta} \in \mathbb{R}^{d_y}$，我们将线性组合方程（式 (35)）改写为

$$\sum_{j \in J_0} \left(\sum_{i:(i,j) \in S_0} \gamma_{ij} u_i(\mathbf{y} = \boldsymbol{\beta}, \boldsymbol{x}) \right) v_j(\boldsymbol{z}, \mathbf{y} = \boldsymbol{\beta}) = 0, \quad \forall (\boldsymbol{x}, \boldsymbol{z}) \in \mathcal{X} \times \mathcal{Z}. \tag{36}$$

由假设（a3）可知，集合 $\mathcal{Y}_0 := \{\boldsymbol{\beta} \in \mathbb{R}^{d_y} | \sum_{i:(i,j) \in S_0} \gamma_{ij} u_i(\mathbf{y} = \boldsymbol{\beta}, \boldsymbol{x}) = 0, \forall \boldsymbol{x} \in \mathcal{X}\}$ 在 $\mathbb{R}^{d_y}$ 中只能是零测度。设 $\mathcal{Y} \subset \mathbb{R}^{d_y}$ 为由假设 (a4) 定义的非零测度子集，则 $\mathcal{Y}_1 := \mathcal{Y} \backslash \mathcal{Y}_0 \subset \mathcal{Y}$ 也具有非零测度，并满足假设 (a4)。结合假设（a1），对于每个 $\boldsymbol{\beta} \in \mathcal{Y}_1$，存在 $\boldsymbol{x} \in \mathcal{X}$，使得 $\sum_{i:(i,j) \in S_0} \gamma_{ij} u_i(\mathbf{y} = \boldsymbol{\beta}, \boldsymbol{x}) \neq 0$ 对于 J_0 中至少两个 j索引成立。对于每个 $\boldsymbol{\beta} \in \mathcal{Y}_1$，$\{v_j(\boldsymbol{z}, \mathbf{y} = \boldsymbol{\beta}) | j \in J_0\}$ 在 $\boldsymbol{z} \in \mathcal{Z}$ 上包含线性相关函数。现在，在假设 (a5) 下，我们可以将指标集 J_0 划分为由 $k \in K(\boldsymbol{\beta})$ 索引的子集，如下所示，使得在每个指标子集 $J_k(\boldsymbol{\beta})$ 内，对应索引的函数相等：

$$\begin{aligned} J_0 = \cup_{k \in K(\boldsymbol{\beta})} J_k(\boldsymbol{\beta}), \quad J_k(\boldsymbol{\beta}) \cap J_{k'}(\boldsymbol{\beta}) = \emptyset, \forall k \neq k' \in K(\boldsymbol{\beta}), \\ j \neq j' \in J_k(\boldsymbol{\beta}) \quad \Leftrightarrow \quad v_j(\boldsymbol{z}, \mathbf{y} = \boldsymbol{\beta}) = v_{j'}(\boldsymbol{z}, \mathbf{y} = \boldsymbol{\beta}), \quad \forall \boldsymbol{z} \in \mathcal{Z}. \end{aligned} \tag{37}$$

那么，对于任意 $\boldsymbol{\beta} \in \mathcal{Y}_1$，我们可以将公式（36）改写为

$$\sum_{k \in K(\boldsymbol{\beta})} \left(\sum_{j \in J_k(\boldsymbol{\beta})} \sum_{i:(i,j) \in S_0} \gamma_{ij} u_i(\mathbf{y} = \boldsymbol{\beta}, \boldsymbol{x}) v_j(\boldsymbol{z}, \mathbf{y} = \boldsymbol{\beta}) \right) = 0, \quad \forall (\boldsymbol{x}, \boldsymbol{z}) \in \mathcal{X} \times \mathcal{Z}. \tag{38}$$

回顾公式（37），$v_j(\boldsymbol{z}, \boldsymbol{y} = \boldsymbol{\beta})$ 和 $v_{j'}(\boldsymbol{z}, \boldsymbol{y} = \boldsymbol{\beta})$ 在 $\boldsymbol{z} \in \mathcal{Z}$ 上是相同的函数，当且仅当 $j \neq j'$ 属于同一指标集 $J_k(\boldsymbol{\beta})$。这意味着，若公式（35）成立，那么对于任意 $\boldsymbol{\beta} \in \mathcal{Y}_1$，在假设 (a1) 和 (a5) 下，

$$\sum_{j \in J_k(\boldsymbol{\beta})} \sum_{i:(i,j) \in S_0} \gamma_{ij} u_i(\mathbf{y} = \boldsymbol{\beta}, \boldsymbol{x}) = 0, \quad \forall \boldsymbol{x} \in \mathbb{R}^d, \quad k \in K(\boldsymbol{\beta}). \tag{39}$$

定义 $C(\boldsymbol{\beta}) = \min_k |J_k(\boldsymbol{\beta})|$ 为 $J_k(\boldsymbol{\beta})$ 子集中的 j 个索引的最小基数。选择 $\boldsymbol{\beta}^* \in$ 极小值点$_{\boldsymbol{\beta} \in \mathcal{Y}_1} C(\boldsymbol{\beta})$：

1. 我们有 $C(\boldsymbol{\beta}^*) < |J_0|$ 和 $|K(\boldsymbol{\beta}^*)| \geq 2$。否则，对于所有 $j \neq j' \in J_0$，我们都有 $v_j(\boldsymbol{z}, \mathbf{y} = \boldsymbol{\beta}) = v_{j'}(\boldsymbol{z}, \mathbf{y} = \boldsymbol{\beta})$，对所有 $\boldsymbol{z} \in \mathcal{Z}$ 和 $\boldsymbol{\beta} \in \mathcal{Y}_1$ 均成立，因此它们在 $(\boldsymbol{z}, \mathbf{y}) \in \mathcal{Z} \times \mathcal{Y}_1$ 上线性相关，这与假设 (a4) 矛盾（令 $\mathcal{Y}' = \mathcal{Y}_1$）。
2. 现在不失一般性地假设 $|J_1(\boldsymbol{\beta}^*)| = C(\boldsymbol{\beta}^*)$。根据假设 (a5)，我们知道，对于任意 $j \in J_1(\boldsymbol{\beta}^*)$ 和 $j' \in J_0 \backslash J_1(\boldsymbol{\beta}^*)$，$v_j(\boldsymbol{z}, \mathbf{y} = \boldsymbol{\beta})$ 与 $= v_{j'}(\boldsymbol{z}, \mathbf{y} = \boldsymbol{\beta})$ 至多仅在 $\mathcal{Z}$ 的零测度子集上。由于 $|J_0| < +\infty$ 和 $\mathcal{Z} \subset \mathbb{R}^{d_z}$ 均具有非零测度，因此存在 $\boldsymbol{z}_0 \in \mathcal{Z}$ 和 $\delta > 0$，使得

$$|v_j(\boldsymbol{z} = \boldsymbol{z}_0, \mathbf{y} = \boldsymbol{\beta}^*) - v_{j'}(\boldsymbol{z} = \boldsymbol{z}_0, \mathbf{y} = \boldsymbol{\beta}^*)| \geq \delta, \quad \forall j \in J_1(\boldsymbol{\beta}^*), \forall j' \in J_0 \backslash J_1(\boldsymbol{\beta}^*).$$

根据假设 (a6)，存在 $\epsilon_j > 0$ ，使得我们可以利用 ℓ_2-范数构造一个 ϵ-球 $B(\boldsymbol{\beta}^*,\epsilon_j)$，满足

$$|v_j(z = \boldsymbol{z}_0, \mathbf{y} = \boldsymbol{\beta}^*) - v_j(\boldsymbol{z} = \boldsymbol{z}_0, \mathbf{y} = \boldsymbol{\beta})| \leq \delta/3, \quad \forall \boldsymbol{\beta} \in B(\boldsymbol{\beta}^*, \epsilon_j).$$

选择合适的 $0 < \epsilon \leq \min^{j\in J_0} \epsilon_j$ （注意，$\min_{j\in J_0} \epsilon_j > 0$ as $|J_0| < +\infty$）并构造一个基于 ℓ_2-范数的 ϵ-球 $B(\boldsymbol{\beta}^*, \epsilon) \subset \mathcal{Y}_1$，则对所有 $j \in J_1(\boldsymbol{\beta}^*)$ $j' \in J_0 \backslash J_1(\boldsymbol{\beta}^*)$，由于 $j' \notin J_1(\boldsymbol{\beta})$，对所有 $\boldsymbol{\beta} \in B(\boldsymbol{\beta}^*, \epsilon)$ 均有

$$|v_j(\boldsymbol{z} = \boldsymbol{z}_0, \mathbf{y} = \boldsymbol{\beta}) - v_{j'}(\boldsymbol{z} = \boldsymbol{z}_0, \mathbf{y} = \boldsymbol{\beta})| \geq \delta/3, \quad \forall \boldsymbol{\beta} \in B(\boldsymbol{\beta}^*, \epsilon).$$

因此，对于任意 $\boldsymbol{\beta} \in B(\boldsymbol{\beta}^*, \epsilon)$ 的划分 $\{J_k(\boldsymbol{\beta})\}$ ，我们有 $J_1(\boldsymbol{\beta}) \subset J_1(\boldsymbol{\beta}^*)$，因此 $|J_1(\boldsymbol{\beta})| \leq |J_1(\boldsymbol{\beta}^*)|$。根据 $\boldsymbol{\beta}^* \in$ 极小值点$_{\boldsymbol{\beta}\in\mathcal{Y}} C(\boldsymbol{\beta})$ 和$|J_1(\boldsymbol{\beta}^*)| = C(\boldsymbol{\beta}^*)$ 的定义，对所有 $\boldsymbol{\beta} \in B(\boldsymbol{\beta}^*, \epsilon)$，我们有 $J_1(\boldsymbol{\beta}) = J_1(\boldsymbol{\beta}^*)$。

3. 可以证明 $|J_1(\boldsymbol{\beta}^*)| = 1$，否则根据划分的定义（公式（37））及上述结论，存在 $j \neq j' \in J_1(\boldsymbol{\beta}^*)$，使得对于所有$\boldsymbol{z} \in \mathcal{Z}$，都有 $v_j(\boldsymbol{z},\ \mathbf{y} = \boldsymbol{\beta}) = v_{j'}(\boldsymbol{z},\ \mathbf{y} = \boldsymbol{\beta})$，且 $\boldsymbol{\beta} \in B(\boldsymbol{\beta}^*, \epsilon)$，这与假设（a4）矛盾，只需令 $\mathcal{Y}' = B(\boldsymbol{\beta}^*,\ \epsilon)$。现在假设 $j \in J_1(\boldsymbol{\beta}^*)$ 是该子集中的唯一指标，那么上述结论中已证明的事实，即对于所有 $\boldsymbol{\beta} \in B(\boldsymbol{\beta}^*,\ \epsilon)$，$J_1(\boldsymbol{\beta}) = J_1(\boldsymbol{\beta}^*)$，意味着

$$\sum_{i:(i,j)\in S_0} \gamma_{ij} u_i(\mathbf{y} = \boldsymbol{\beta}, \boldsymbol{x}) = 0, \quad \forall \boldsymbol{x} \in \mathcal{X}, \quad \forall \boldsymbol{\beta} \in B(\boldsymbol{\beta}^*, \epsilon),$$

只需令 $\mathcal{Y} = B(\boldsymbol{\beta}^*,\ \epsilon)$，便再次与假设（a3）矛盾。

上述三点表明，在假设（a3）–（a6）下，公式（39）不可能对所有 $\boldsymbol{\beta} \in \mathcal{Y}_1$ （因此也不可能对所有$\boldsymbol{\beta} \in \mathcal{Y}$）成立，故得出矛盾。 ■

证明表明，在假设 (a1)–(a4) 下，原则上，对于重叠变量 $\mathbf{y}$ 的每个取值，都可能出现线性相关，因此还需要额外的假设 (a5)–(a6)。

为什么 LIPO-1 不能解决一般情形 $M > 1$**?** 如第 4.1 节所述，当 $M = 1$ 时，以及更一般地，当 $T = M + 1$ 时，LIPO-1 可直接应用。对于给定的初始族和转移族，我们令

$$u_i(\mathbf{y}, \boldsymbol{x}) := p_a(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_M), \quad \boldsymbol{x} := \emptyset, \quad \mathbf{y} := (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_M), \quad i = a,$$

and

$$v_j(\boldsymbol{z}, \mathbf{y}) := p_b(\boldsymbol{z}_{M+1} \mid \boldsymbol{z}_M, \ldots, \boldsymbol{z}_1), \quad \boldsymbol{z} := \boldsymbol{z}_{M+1}, \quad \mathbf{y} := (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_M), \quad j = b,$$

因此，重叠变量为 $\mathbf{y}$ $=(\boldsymbol{z}_1$、.....、 $\boldsymbol{z}_M)$，且假设 (a1)–(a6) 均得到满足。对于 $T > M + 1$，这种方法会失效，因为为了应用 LIPO-1，一步中的变量 $\boldsymbol{z}$ 必须在下一步变为$\mathbf{y}$ 。例如，对于 $T = M + 2$，我们要求 $v'(\boldsymbol{z}',\ \mathbf{y}')$ 满足 (a4-a6)。因此我们定义

$$v'_j(\boldsymbol{z}', \mathbf{y}') := p_b(\boldsymbol{z}_{M+2} \mid \boldsymbol{z}_{M+1}, \ldots, \boldsymbol{z}_2), \quad \mathbf{y}' := (\boldsymbol{z}_2, \ldots, \boldsymbol{z}_{M+1}), \quad \boldsymbol{z}' := \boldsymbol{z}_{M+2}.$$

然而，根据上一步我们有 $\boldsymbol{z} := \boldsymbol{z}_{M+1}$，而唯一能够保持 (a3) 的 $u_i'(\mathbf{y}\ \boldsymbol{x})$ 定义为

$$u_i'(\mathbf{y}', \boldsymbol{x}') := p_a(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_M) p_b(\boldsymbol{z}_{M+1} \mid \boldsymbol{z}_M, \ldots, \boldsymbol{z}_1), \quad \boldsymbol{x}' := (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_M), \quad \mathbf{y}' := \boldsymbol{z}_{M+1}.$$

我们注意到，这与 $v_j'(\boldsymbol{z}', \mathbf{y}')$ 中所要求的 $(\boldsymbol{z}_2, \ldots, \boldsymbol{z}_{M+1})$ 不兼容，因此无法使用 LIPO-1。为确保 $\mathbf{y}' = \boldsymbol{z}$，我们可以将 $v_j(\boldsymbol{z}, \mathbf{y})$ 定义为跨越 M 个连续状态转移的乘积：

$$v_j(\boldsymbol{z}, \mathbf{y}) := \prod_{\ell=1}^{M} p_{b_\ell}(\boldsymbol{z}_{t+\ell} \mid \boldsymbol{z}_{t-M+\ell:t-1+\ell}),$$
$$\mathbf{y} := (\boldsymbol{z}_{t-M+1}, \ldots, \boldsymbol{z}_t),\ \boldsymbol{z} := (\boldsymbol{z}_{t+1}, \ldots, \boldsymbol{z}_{t+M}),\ j := (b_1, \ldots, b_M).$$

根据这一定义，重叠变量现在满足 $\mathbf{y}' = \boldsymbol{z}$，因此该乘积族具备按顺序应用 LIPO-1 所需的结构。剩下的任务是验证该乘积族仍满足假设 (a1)–(a2) 和 (a4)–(a6)。

A.2.2 乘积族假设的扩展

记号。 在正文及上文中，我们按升序书写坐标变量。然而，为了提高可读性，本节按降序书写变量， $\boldsymbol{z}_M, \ldots, \boldsymbol{z}_1$ 以及 $\mathbf{y}_M, \ldots \mathbf{y}_1$,, 以对应乘积族的构造。我们还使用以下字母表示证明中的不同概念： r 表示连续乘积的数量， ℓ 表示乘积因子索引， q 表示划分中的类别标签（稍后用于 LIPO-M、引理 21）。此外，我们删除 d_y, d_z （用于 LIPO-1），改用 d表示单个变量 $\mathbf{y}_i, i \in [M]$ 或 $\boldsymbol{z}$的维数。

下面我们探讨关于含有 $M+1$ 个变量的非参数族的假设（与某个转移分布 $p(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1}$, ……， $\boldsymbol{z}_{t-M}$， s_t)相对应），使得 M个连续变量的乘积满足(a1)、(a2)和(a4-a6)。这样我们便可以使用 LIPO-1（或等价的扩展）来证明轨迹族 $\mathcal{P}^T, \mathcal{A}^M \mathcal{B}$ 在有限混合下的线性无关性。我们注意到，假设(a2)是多余的，因为如果(a3)和(a4)分别成立，那么 U_I 和 V_J 也成立。因此，为简化理论分析，我们从中删除(a2)。给定一个族 $\mathcal{V}_J = \{v_j(\boldsymbol{z}, \mathbf{y}_M, \ldots\ldots, \mathbf{y}_1) | j \in J\}$，其变量$(\boldsymbol{z}, \mathbf{y}_M, \ldots\ldots, \mathbf{y}_1)$定义在$\times^{M+1}\mathbb{R}^d$上，我们按照 LIPO-1（引理 16）中的编号，对假设进行相应修改。我们保持(a3)不变，因为不会修改族 $\mathcal{U}_I$。

(a1) $\to$ (b1) $v_j(\boldsymbol{z}, \mathbf{y}_M, \ldots, \mathbf{y}_1) > 0$ for all $j \in J, (\boldsymbol{z}, \mathbf{y}_M, \ldots, \mathbf{y}_1) \in (\times^{M+1}\mathbb{R}^d)$.

(a4) $\to$ (b4) 存在一个满测度集 $\mathcal{Y} \subset (\times^M \mathbb{R}^d)$（即 $\mu((\times^M \mathbb{R}^d) \setminus \mathcal{Y}) = 0$），使得对于每个非零测度集合 $\mathcal{Y}' \subset \mathcal{Y}$ 和 $\mathcal{Z} \subset \mathbb{R}^d$， $\mathcal{V}_J$ 在 $(\boldsymbol{z}, \mathbf{y}_M, \ldots\ldots, \mathbf{y}_1) \in \mathcal{Z} \times \mathcal{Y}'$ 上的有限混合中都包含线性无关函数;

(a5) $\to$ (b5) 对任意 $(\boldsymbol{\beta}_\ell, \ldots, \boldsymbol{\beta}_1) \in (\times^\ell \mathbb{R}^d)$、任意满足 $1 \le \ell \le M$ 的 ℓ、任意非零测度子集 $\mathcal{Z} \subset \mathbb{R}^d, \mathcal{Y} \subset (\times^{M-\ell}\mathbb{R}^d)$，以及满足 $|J_0| < +\infty$ 的任意子集 $J_0 \subset J$： 仅当存在 $\exists\, j \neq j' \in J_0$ 使得对于所有 $(\boldsymbol{z}, \mathbf{y}_M, \ldots, \mathbf{y}_{\ell+1}) \in \times^{M-\ell+1}\mathbb{R}^d$， $\{v_j(\boldsymbol{z}, \mathbf{y}_M, \ldots, \mathbf{y}_{\ell+1}, \mathbf{y}_\ell = \boldsymbol{\beta}_\ell, \ldots, \mathbf{y}_1 = \boldsymbol{\beta}_1) | j \in J_0\}$ 在 $(\boldsymbol{z}, \mathbf{y}_M, \ldots, \mathbf{y}_{\ell+1})$ 上包含线性相关函数时， $v_j(\boldsymbol{z}, \mathbf{y}_M, \ldots, \mathbf{y}_{\ell+1}, \mathbf{y}_\ell = \boldsymbol{\beta}_\ell, \ldots, \mathbf{y}_1 = \boldsymbol{\beta}_1) = v_{j'}(\boldsymbol{z}, \mathbf{y}_M, \ldots, \mathbf{y}_{\ell+1}, \mathbf{y}_\ell = \boldsymbol{\beta}_\ell, \ldots, \mathbf{y}_1 = \boldsymbol{\beta}_1)$。

(a6) $\to$ (b6) 对任意 $j \in J$， $v_j(\boldsymbol{z}, \mathbf{y}_M, \ldots\ldots, \mathbf{y}_1)$ 在 $(\mathbf{y}_M, \ldots\ldots, \mathbf{y}_1) \in (\times^M \mathbb{R}^d)$ 上连续 .

与对 (a1) 和 (a6) 的修改不同，我们注意到，将 (a4) 和 (a5) 扩展到 M 连续乘积时，分别需要将其加强为 (b4) 和 (b5)。对于 (b4)，非零测度子集 $\mathcal{Y}$ 现在必须具有全测度。这与情形 $M = 1$ 不同，在该情形下， $\mathcal{Y}$ 可以具有任意非零测度；这一要求背后的直观解释可从 $M > 1$ 条件化重叠的出现中看出（参见式 (34)）。对于 (b5)，原先只要求固定$\mathbf{y}$ 对应的重复函数线性相关，现在这一要求得到加强：对于来自$\mathbf{y}_M, \ldots, \mathbf{y}_1$的某些固定变量子组所对应的重复函数，也必须满足这一条件。

下面给出将在后续结果中使用的定义。

定义 17（坐标投影）设 $\mathcal{A} \subseteq (\times^r \mathbb{R}^d)$ 且 $I = \{i_1, \ldots, i_\ell\} \subset [r]$，其中 $1 \le \ell \le r$。$\mathcal{A}$ 在 I 所含坐标上的投影为

$$\pi_I(\mathcal{A}) := \{(\boldsymbol{x}_i)_{i \in I} \in (\times^\ell \mathbb{R}^d) \mid (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r) \in \mathcal{A}\} \subseteq (\times^\ell \mathbb{R}^d).$$

定义 18（条件投影）设 $\mathcal{A} \subseteq (\times^r \mathbb{R}^d)$ 具有变量$(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r)$。给定 $I = \{i_1, \ldots, i_\ell\} \subset [r]$ 及赋值$(\tilde{\boldsymbol{x}}_{i_1}, \ldots, \tilde{\boldsymbol{x}}_{i_\ell})$，定义

$$\mathcal{A}\big|_{(\boldsymbol{x}_{i_1} = \tilde{\boldsymbol{x}}_{i_1}, \ldots, \mathbf{x}_{i_\ell} = \tilde{\boldsymbol{x}}_{i_\ell})} := \Big\{(\boldsymbol{x}_j)_{j \notin I} \in (\times^{r-\ell} \mathbb{R}^d) : (\mathbf{y}_1, \ldots, \mathbf{y}_r) \in \mathcal{A},\ y_i = \tilde{\boldsymbol{x}}_i\ \forall i \in I,\ y_j = \boldsymbol{x}_j\ \forall j \notin I\Big\}.$$

假设（b4）扩展至连续的乘积函数族，如下所示。

引理 19（（b4）的扩展）设标量函数族 $\mathcal{V}_J = \{v_j(\boldsymbol{z}, \mathbf{y}_M, \ldots, \mathbf{y}_1) | j \in J\}$ 定义在$(\boldsymbol{z}, \mathbf{y}_M, \ldots, \mathbf{y}_1) \in \times^{M+1} \mathbb{R}^d$ 上，且满足（*b1*）、（*b4*）、（*b5*）和（*b6*）。

那么，对于任意 r 满足 $2 \le r \le M$，存在一个满测度集 $\mathcal{Y} \subset (\times^M \mathbb{R}^d)$，使得对于任意非零测度集 $\mathcal{Y}' \subset \mathcal{Y}$ 和 $\mathcal{Z}_r \subset (\times^r \mathbb{R}^d)$，乘积族

$$\otimes^r \mathcal{V}_J := \{v_{j_r}(\boldsymbol{z}_r, \ldots, \boldsymbol{z}_1, \mathbf{y}_M, \ldots, \mathbf{y}_r) \ldots v_{j_1}(\boldsymbol{z}_1, \mathbf{y}_M, \ldots, \mathbf{y}_1) | (j_r, \ldots, j_1) \in \times^r J\},$$

是线性 在有限混合下关于 $(\boldsymbol{z}_r$、……、 $\boldsymbol{z}_1$、$\mathbf{y}_M$、……、$\mathbf{y}_1) \in \mathcal{Z}_r \times \mathcal{Y}'$ 几乎独立。

证明 我们通过归纳法证明上述陈述，首先从基例（$r = 2$）开始，具体如下。基例的证明策略大体遵循 LIPO-1 证明。

情形 $r = 2$。假设该陈述不成立。那么，对于每个满足$\mu(\mathcal{Y}) = \mu(\times^M \mathbb{R}^d)$的集合 $\mathcal{Y} \subset (\times^M \mathbb{R}^d)$，都存在非零测度集 $\mathcal{Z}_2 \subset (\mathbb{R}^d \times \mathbb{R}^d)$和 $\mathcal{Y}' \subset \mathcal{Y}$，使得$\mathcal{V}_J \otimes \mathcal{V}_J$ 包含线性相关函数。我们固定来自 (b4) 的 $\mathcal{Y}$，并考虑 $\mathcal{Y}'$和 $\mathcal{Z}_2$的任意取值。这意味着存在一个指标集 $S_0 \subset (J \times J)$ $|S_0| < +\infty$ 以及相应的非零系数 $\{\gamma_{j_1 j_2} \in \mathbb{R}, (j_1, j_2) \in S_0\}$，使得以下线性相关条件成立：

$$\sum_{(j_1, j_2) \in S_0} \gamma_{j_1 j_2} v_{j_2}(\boldsymbol{z}_2, \boldsymbol{z}_1, \mathbf{y}_M, \ldots, \mathbf{y}_2) v_{j_1}(\boldsymbol{z}_1, \mathbf{y}_M, \ldots, \mathbf{y}_1) = 0, \quad \forall (\boldsymbol{z}_2, \boldsymbol{z}_1, \mathbf{y}_M, \ldots, \mathbf{y}_1) \in (\mathcal{Z}_2 \times \mathcal{Y}'), \quad (40)$$

S_0 和$\gamma_{j_1 j_2}$ 的取值可能取决于 $\mathcal{Z}_2$ 和 $\mathcal{Y}'$的选取，而 $\mathcal{Y}'$ 则取决于满测度 $\mathcal{Y}$的选取。

由 (b1) 可知，集合 S_0 至少包含两个不同的索引（j_1，j_2）和（j_1'，j_2'），其中$j_2 \neq j_2'$。为说明这一点，假设对于所有可能的（j_1，j_2），（j_1'，j_2'）$\in S_0$，我们都有 $j_2 = j_2'$。那么，我们可以提出公因子 v_{j_2}（$\boldsymbol{z}_2$，$\boldsymbol{z}_1$，$\mathbf{y}_M$，.....，$\mathbf{y}_2$），并将线性相关方程改写为

$$v_{j_2}(\boldsymbol{z}_2, \boldsymbol{z}_1, \mathbf{y}_M, \ldots, \mathbf{y}_2) \sum_{j_1:(j_1,j_2)\in S_0} \gamma_{j_1 j_2} v_{j_1}(\boldsymbol{z}_1, \mathbf{y}_M, \ldots, \mathbf{y}_1) = 0,$$
$$\forall (\boldsymbol{z}_2, \boldsymbol{z}_1, \mathbf{y}_M, \ldots, \mathbf{y}_1) \in (\mathcal{Z}_2 \times \mathcal{Y}'), \quad (41)$$

这与 (b4) 矛盾，因为 $\mathcal{Y}' \subset \mathcal{Y}$。

现在我们定义 $J_0 := \{j_2 \in J | \exists$（$j_1$，$j_2$）$\in S_0\}$，$|J_0| < +\infty$。于是，对于任意（$\boldsymbol{\beta}_M$，.....，$\boldsymbol{\beta}_1$）$\in$（$\pi_{\{1\}}$（$\mathcal{Z}_2$）$\times \pi_{\{M,\ldots,2\}}$（$\mathcal{Y}'$））$\times$，都可能出现线性相关，其中 π 表示定义 17 中定义的集合投影。

$$\sum_{j_2 \in J_0} \left(\sum_{j_1:(j_1,j_2)\in S_0} \gamma_{j_1 j_2} v_{j_1}(\boldsymbol{z}_1 = \boldsymbol{\beta}_M, \mathbf{y}_M = \boldsymbol{\beta}_{M-1}, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1, \mathbf{y}_1) \right)$$
$$v_{j_2}(\boldsymbol{z}_2, \boldsymbol{z}_1 = \boldsymbol{\beta}_M, \mathbf{y}_M = \boldsymbol{\beta}_{M-1}, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1) = 0,$$
$$\forall (\boldsymbol{z}_2, \mathbf{y}_1) \in (\mathcal{Z}' \times \mathcal{Y}_1'). \quad (42)$$

其中，$\mathcal{Z}' = \mathcal{Z}_2|_{(\boldsymbol{z}_1 = \boldsymbol{\beta}_M)}$和 $\mathcal{Y}_1' = \mathcal{Y}'|_{(\mathbf{y}_M = \boldsymbol{\beta}_{M-1}, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1)}$分别是 $\mathcal{Z}_2$和 $\mathcal{Y}'$的条件投影（定义 18），并且依赖于 $(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1)$。根据构造，这些条件投影非空，因为 $(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) \in (\pi_{\{1\}}(\mathcal{Z}_2) \times \pi_{\{M,\ldots,2\}}(\mathcal{Y}'))$。现在，我们定义如下集合。

$$\mathcal{D}_0 := \Bigg\{ (\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) \in \Big(\pi_{\{1\}}(\mathcal{Z}_2) \times \pi_{\{M,\ldots,2\}}(\mathcal{Y}') \Big) \Bigg|$$
$$\sum_{j_1:(j_1,j_2)\in S_0} \gamma_{j_1,j_2} v_{j_1}(\boldsymbol{z}_1 = \boldsymbol{\beta}_M, \mathbf{y}_M = \boldsymbol{\beta}_{M-1}, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1, \mathbf{y}_1) = 0, \ \forall \mathbf{y}_1 \in \mathcal{Y}_1' \Bigg\}. \quad (43)$$

由 (b4) 可知，对于每个 $\mathcal{Y}' \subset \mathcal{Y}$，其中 $\mathcal{Y}$ 具有全测度，以及每个 $\mathcal{Z}' \subset \mathbb{R}^d$，我们都要求线性无关。因此，如上所述的线性相关至多可能发生在一个零测度集合中，因为 $\mathcal{Y}_0 := (\times^M \mathbb{R}^d) \setminus \mathcal{Y}$ 在 $(\times^M \mathbb{R}^d)$ 中的测度为零。因此，该集合

$$\mathcal{D}_0 \subset \pi_{\{1\}}(\mathcal{Z}_2) \times \pi_{\{M,\ldots,2\}}(\mathcal{Y}_0)$$

也具有零测度，因为 $\mathcal{Y}_0$ 的测度为零。

现在定义 $\mathcal{D} := (\pi_{\{1\}}(\mathcal{Z}_2) \times \pi_{\{M,\ldots,2\}}(\mathcal{Y}')) \setminus \mathcal{D}_0$，它是非零测度。根据假设（b1），我们有 $\forall (\boldsymbol{z}_1 = \boldsymbol{\beta}_M$，$\mathbf{y}_M = \boldsymbol{\beta}_{M-1}$，.....，$\mathbf{y}_2 = \boldsymbol{\beta}_1) \in \mathcal{D}$，存在 $\mathbf{y}_1 \in \mathcal{Y}'|_{(\mathbf{y}_M = \boldsymbol{\beta}_{M-1}, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1)}$ 使得

$$\tilde{\gamma}_{j_2}(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) := \sum_{j_1:(j_1,j_2)\in S_0} \gamma_{j_1 j_2} v_{j_1}(\boldsymbol{z}_1 = \boldsymbol{\beta}_M, \mathbf{y}_M = \boldsymbol{\beta}_{M-1}, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1, \mathbf{y}_1) \neq 0$$

如公式（41）所示，至少有两个不同的 $j_2 \in J_0$ 。根据公式（40），这意味着 $\{v_{j_2}(\boldsymbol{z}_2, \boldsymbol{z}_1 = \boldsymbol{\beta}_M, \mathbf{y}_M = \boldsymbol{\beta}_{M-1}, \ldots\ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1)\}$ 与 $\boldsymbol{z}_2 \in \mathcal{Z}_2|_{(\boldsymbol{z}_1=\boldsymbol{\beta}_M)}$ 、 $\forall(\boldsymbol{z}_1 = \boldsymbol{\beta}_M, \mathbf{y}_M = \boldsymbol{\beta}_{M-1}, \ldots\ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1) \in \mathcal{D}$ 线性相关。

根据假设 (b5)，重复函数之间存在线性相关。因此，我们可以将 J_0 划分为由 $k \in K$（$\boldsymbol{\beta}_M$、……、$\boldsymbol{\beta}_1$）索引的子集，使得每个子集内的 v_j 函数 J_k（$\boldsymbol{\beta}_M$、……、$\boldsymbol{\beta}_1$）都相等，如式 (45) 所述：

$$J_0 = \cup_{k\in K(\boldsymbol{\beta}_M,\ldots,\boldsymbol{\beta}_1)} J_k(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1),\ J_k(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) \cap J_{k'}(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) = \emptyset,$$
$$\forall k \neq k' \in K(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1), \quad (44)$$

$$j \neq j' \in J_k(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) \quad \Leftrightarrow$$
$$v_j(\boldsymbol{z}_2, \boldsymbol{z}_1 = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1) = v_{j'}(\boldsymbol{z}_2, \boldsymbol{z}_1 = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1), \quad \forall \boldsymbol{z}_2 \in \mathcal{Z}_2|_{(\boldsymbol{z}_1=\boldsymbol{\beta}_M)}. \quad (45)$$

于是，对于任意 $(\boldsymbol{\beta}_M$、……、$\boldsymbol{\beta}_1) \in \mathcal{D}$，我们可以重写线性相关条件 as

$$\sum_{k\in K(\boldsymbol{\beta}_M,\ldots,\boldsymbol{\beta}_1)} \left(\sum_{j_2\in J_k(\boldsymbol{\beta}_M,\ldots,\boldsymbol{\beta}_1)} \sum_{j_1:(j_1,j_2)\in S_0} \gamma_{j_1 j_2} v_{j_1}(\boldsymbol{z}_1 = \boldsymbol{\beta}_M, \mathbf{y}_M = \boldsymbol{\beta}_{M-1}, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1, \mathbf{y}_1) \right.$$
$$\left. v_{j_2}(\boldsymbol{z}_2, \boldsymbol{z}_1 = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1) \right) = 0, \quad \forall(\boldsymbol{z}_2, \mathbf{y}_1) \in \mathcal{Z}_2|_{(\boldsymbol{z}_1=\boldsymbol{\beta}_M)} \times \mathcal{Y}'|_{(\mathbf{y}_M=\boldsymbol{\beta}_{M-1},\ldots,\mathbf{y}_2=\boldsymbol{\beta}_1)}$$
$$(46)$$

回顾可知，v_j（$\boldsymbol{z}_2$，$\boldsymbol{z}_1 = \boldsymbol{\beta}_M$，……，$\mathbf{y}_2 = \boldsymbol{\beta}_1$）和 $v_{j'}$（$\boldsymbol{z}_2$，$\boldsymbol{z}_1 = \boldsymbol{\beta}_M$，……，$\mathbf{y}_2 = \boldsymbol{\beta}_1$）在 $\boldsymbol{z}_2 \in \mathcal{Z}_2|_{(\boldsymbol{z}_1=\boldsymbol{\beta}_M)}$ 上是相同的函数，当且仅当 $j \neq j'$ 属于同一个指标集 J_k（$\boldsymbol{\beta}_M$、……、$\boldsymbol{\beta}_1$）。这意味着，若线性无关成立，则对于任意 $(\boldsymbol{\beta}_M$、……、$\boldsymbol{\beta}_1) \in \mathcal{D}$，在假设 (b1) 和 (b5) 下，

$$\sum_{j_2\in J_k(\boldsymbol{\beta}_M,\ldots,\boldsymbol{\beta}_1)} \sum_{j_1:(j_1,j_2)\in S_0} \gamma_{j_1 j_2} v_{j_1}(\boldsymbol{z}_1 = \boldsymbol{\beta}_M, \mathbf{y}_M = \boldsymbol{\beta}_{M-1}, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1, \mathbf{y}_1) = 0,$$
$$\forall \mathbf{y}_1 \in \mathcal{Y}'|_{(\mathbf{y}_M=\boldsymbol{\beta}_{M-1},\ldots,\mathbf{y}_2=\boldsymbol{\beta}_1)}, \quad k \in K(\boldsymbol{\beta}_{M-1}, \ldots, \boldsymbol{\beta}_1). \quad (47)$$

定义 $C(\boldsymbol{\beta}_M$，……，$\boldsymbol{\beta}_1) = \min_k |J_k(\boldsymbol{\beta}_M$，……，$\boldsymbol{\beta}_1)|$ 为 j_2 索引在 $J_k(\boldsymbol{\beta}_M$，……，$\boldsymbol{\beta}_1)$ 子集中计数的最小基数。

选取 $(\boldsymbol{\beta}_M^*$，……，$\boldsymbol{\beta}_1^*) \in$ 极小值点$_{(\boldsymbol{\beta}_M,\ldots,\boldsymbol{\beta}_1)\in\mathcal{D}} C(\boldsymbol{\beta}_M$，……，$\boldsymbol{\beta}_1)$：

1. 我们有 $C(\boldsymbol{\beta}_M^*$，……，$\boldsymbol{\beta}_1^*) < |J_0|$ 以及 $|K(\boldsymbol{\beta}_M^*$，……，$\boldsymbol{\beta}_1^*)| \geq 2$。否则，对于所有 $j \neq j' \in J_0$，我们都有 $v_j(\boldsymbol{z}_2$，$\boldsymbol{z}_1 = \boldsymbol{\beta}_M$，……，$\mathbf{y}_2 = \boldsymbol{\beta}_1) = v_{j'}(\boldsymbol{z}_2$，$\boldsymbol{z}_1 = \boldsymbol{\beta}_M$，……，$\mathbf{y}_2 = \boldsymbol{\beta}_1)$，对于所有 $\boldsymbol{z}_2 \in \mathcal{Z}_2|_{(\boldsymbol{z}_1=\boldsymbol{\beta}_M)}$ 和 $(\boldsymbol{\beta}_M$，……，$\boldsymbol{\beta}_1) \in \mathcal{D})$，因此它们关于 $(\boldsymbol{z}_2$，$\boldsymbol{z}_1$，$\mathbf{y}_M$，……，$\mathbf{y}_2) \in \mathcal{Z}' \times \mathcal{D}$ 线性相关，对某个非零测度 $\mathcal{Z}' \subseteq \pi_{\{2\}}(\mathcal{Z}_2)$ 都成立；这与假设（b4）矛盾。令 $\mathcal{Y}' = \mathcal{D}$，无论选择何种 $\mathcal{Z}_2$，这一结论都成立，因为 $\mathcal{Y}' \subset \mathcal{Y}$，其中 $\mathcal{Y}$ 具有全测度。

2. 现在不失一般性地假设 $|J_1(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*)| = C(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*)$。根据假设 (b5)，我们知道，对于任意的 $j \in J_1(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*)$ 和 $j' \in J_0 \backslash J_1(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*)$，$v_j(\boldsymbol{z}_2, \boldsymbol{z}_1 =$

$\boldsymbol{\beta}_M, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1) = v_{j'}(\boldsymbol{z}_2, \boldsymbol{z}_1 = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1)$ 至多只在 $\mathcal{Z}_2|_{(\boldsymbol{z}_1=\boldsymbol{\beta}_M)}$ 的一个零测度子集上成立。然后，由于 $|J_0| < +\infty$ 和 $\mathcal{Z}_2|_{(\boldsymbol{z}_1=\boldsymbol{\beta}_M)} \subset \mathbb{R}^d$ 具有非零测度，因此存在 $\boldsymbol{z}_0 \in \mathcal{Z}_2|_{(\boldsymbol{z}_1=\boldsymbol{\beta}_M)}$ 和 $\delta > 0$，使得

$$|v_j(\boldsymbol{z}_2 = \boldsymbol{z}_0, \boldsymbol{z}_1 = \boldsymbol{\beta}_M^*, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1^*) - v_{j'}(\boldsymbol{z}_2 = \boldsymbol{z}_0, \boldsymbol{z}_1 = \boldsymbol{\beta}_M^*, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1^*)| \geq \delta, \\ \forall j \in J_1(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*), \forall j' \in J_0 \backslash J_1(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*) \quad (48)$$

在假设（b6）下，存在 $\epsilon_j > 0$，使得我们可以使用 ℓ_2-范数构造一个 ϵ-球 $B(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*, \epsilon_j)$，使得

$$|v_j(\boldsymbol{z}_2 = \boldsymbol{z}_0, \boldsymbol{z}_1 = \boldsymbol{\beta}_M^*, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1^*) - v_j(\boldsymbol{z}_2 = \boldsymbol{z}_0, \boldsymbol{z}_1 = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1)| \leq \delta/3, \\ \forall (\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) \in B(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*, \epsilon_j). \quad (49)$$

选择合适的 $0 < \epsilon \leq \min^{j \in J_0} \epsilon_j$（注意，由于 $|J_0| < +\infty$，$\min_{j \in J_0} \epsilon_j > 0$）并构造一个基于 ℓ_2-范数的 ϵ-球 $B(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*, \epsilon) \subset \mathcal{D}$，对于所有 $j \in J_1(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*)$ $j' \in J_0 \backslash J_1(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*)$，$j' \notin J_1(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1)$ 对所有 $(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1)$，由于

$$|v_j(\boldsymbol{z}_2 = \boldsymbol{z}_0, \boldsymbol{z}_1 = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1) - v_{j'}(\boldsymbol{z}_2 = \boldsymbol{z}_0, \boldsymbol{z}_1 = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1)| \geq \delta/3, \\ \forall (\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) \in B(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*, \epsilon). \quad (50)$$

因此，这意味着对于任意 $(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) \in B(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*, \epsilon)$ 的划分 $\{J_k(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1)\}$，我们有 $J_1(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) \subset J_1(\boldsymbol{\beta}^*)$，从而 $|J_1(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1)| \leq |J_1(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*)|$。现在根据 $(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*) \in$ 极小值点$_{(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) \in \mathcal{D}} C(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1)$ 和 $|J_1(\boldsymbol{\beta}_M^*, \ldots \boldsymbol{\beta}_1^*)| = C(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*)$ 的定义，对于所有 $(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) \in B(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*, \epsilon)$，我们有 $J_1(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) = J_1(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*)$。

3. 可以证明，$|J_1(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*)| = 1$；否则，根据划分的定义及上述结论，存在 $j \neq j' \in J_1(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*)$，使得对所有 $\boldsymbol{z}_2 \in \mathcal{Z}_2|_{(\boldsymbol{z}_1=\boldsymbol{\beta}_M)}$ 和 $(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1)$，$v_j(\boldsymbol{z}_2, \boldsymbol{z}_1 = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1) = v_{j'}(\boldsymbol{z}_2, \boldsymbol{z}_1 = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1) \in B(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*, \epsilon)$。令 $\mathcal{Y}^* := B(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*, \epsilon) \cap \mathcal{Y}$，这与假设（b4）矛盾。现在假设 $j \in J_1(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*)$ 是该子集中的唯一指标，那么上述结论已证明的事实，即对所有 $(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1)$，$J_1(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) = J_1(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*) \in \mathcal{Y}^*$，意味着

$$\sum_{j_1:(j_1,j_2) \in S_0} \gamma_{j_1 j_2} v_{j_1}(\boldsymbol{z}_1 = \boldsymbol{\beta}_M, \mathbf{y}_M = \boldsymbol{\beta}_{M-1}, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1, \mathbf{y}_1) = 0, \\ \forall \mathbf{y}_1 \in \mathcal{Y}'|_{(\mathbf{y}_M = \boldsymbol{\beta}_{M-1}, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_1)}, \quad \forall (\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) \in \mathcal{Y}^*, \quad (51)$$

这再次与 (b4) 矛盾，因为我们发现至少在一个非零测度集上存在线性相关。

上述三点表明，对于任意选择的 $\mathcal{Z}_2$ 和 $\mathcal{Y}$，我们发现 $\mathcal{V}_J$ 至少在一个非零测度集上线性相关。这直接与假设（b4）矛盾：在 $\mathcal{Y}$ 满测度的 $\mathcal{Y} \times \mathbb{R}$ 的任意非零测度子集上，有限混合下的线性无关性都必须成立。

我们注意到，此处假设 $\mathcal{Y}$ 在（b4）中具有全测度，而非在（a4）中具有全测度，这一点至关重要。否则，由 $\mathcal{Z}_2 \subset （\mathbb{R}^d \times \mathbb{R}^d）$ 和 $\mathcal{Y}' \subset \mathcal{Y}$的投影构造出的 ϵ-球可能落在 $\mathcal{Y}$的选择范围之外。因此，我们将有 $\mu（B（\boldsymbol{\beta}_M^*, \ldots\ldots, \boldsymbol{\beta}_1^*, \epsilon）\cap \mathcal{Y}）= 0$，上述矛盾也就不成立。

情形 $r > 2$。现在假设该陈述对 $r-1$ 成立，并再次用反证法证明这一情形。假设该陈述对 $r > 2$ 不成立。那么，对于每个满足 $\mu(\mathcal{Y}) = \mu(\times^M \mathbb{R}^d)$ 的集合 $\mathcal{Y} \subset (\times^M \mathbb{R}^d)$，都存在非零测度集 $\mathcal{Z}_r \subset (\times^r \mathbb{R}^d)$ 和 $\mathcal{Y}' \subset \mathcal{Y}$，使得族 $\otimes^r \mathcal{V}_J$ 中包含线性相关函数。与之前一样，我们固定 (b4) 中的 $\mathcal{Y}$，并考虑 $\mathcal{Y}'$ 和 $\mathcal{Z}_r$ 的任意取值。于是，存在 $S_0 \subset (\times^r J)$ $|S_0| < +\infty$，使得

$$\sum_{(j_r,\ldots j_1)\in S_0} \gamma_{j_r,\ldots,j_1} v_{j_r}(\boldsymbol{z}_r, \ldots, \boldsymbol{z}_1, \mathbf{y}_M, \ldots, \mathbf{y}_r) \ldots v_{j_1}(\boldsymbol{z}_1, \mathbf{y}_M, \ldots, \mathbf{y}_1) = 0,$$
$$\forall (\boldsymbol{z}_r, \ldots, \boldsymbol{z}_1, \mathbf{y}_M, \ldots, \mathbf{y}_1) \in (\mathcal{Z}_r \times \mathcal{Y}'), \quad (52)$$

其中 $\{\gamma_{j_r,\ldots,j_1} \in \mathbb{R}, (j_r, \ldots\ldots, j_1) \in S_0\}$ 是一个非零值集合，与之前一样，它可能取决于 $\mathcal{Z}_r$ 和 $\mathcal{Y}' \subset \mathcal{Y}$ 的选择。前一个等式可整理如下

$$\sum_{(j_r,\ldots j_1)\in S_0} \gamma_{j_r,\ldots,j_1} v_{j_r}(\boldsymbol{z}_r, \ldots, \boldsymbol{z}_1, \mathbf{y}_M, \ldots, \mathbf{y}_r) v_{j_{r-1},\ldots,j_1}(\boldsymbol{z}_{r-1}, \ldots, \boldsymbol{z}_1, \mathbf{y}_M, \ldots, \mathbf{y}_1) = 0,$$
$$\forall (\boldsymbol{z}_r, \ldots, \boldsymbol{z}_1, \mathbf{y}_M, \ldots, \mathbf{y}_1) \in (\mathcal{Z}_r \times \mathcal{Y}'), \quad (53)$$

其中 $v_{j_{r-1},\ \ldots,\ j_1} := v_{j_{r-1}}(\boldsymbol{z}_{r-1},\ \ldots,\ \boldsymbol{z}_1,\ \mathbf{y}_M,\ \ldots,\ \mathbf{y}_{r-1}) \cdots v_{j_1}(\boldsymbol{z}_1,\ \mathbf{y}_M,\ \ldots,\ \mathbf{y}_1)$ 对于 $j_i \in J$, $1 \leq i < r$ 表示族 $\otimes^{r-1} \mathcal{V}_J$ 上的函数，并满足以下条件：

(b1) Positive function values: $v_{j_{r-1},\ldots,j_1}(\boldsymbol{z}_{r-1}, \ldots, \boldsymbol{z}_1, \mathbf{y}_M, \ldots, \mathbf{y}_1) > 0$ for all $(j_{r-1}, \ldots, j_1) \in (\times^{r-1} J), (\boldsymbol{z}_{r-1}, \ldots, \boldsymbol{z}_1, \mathbf{y}_M, \ldots, \mathbf{y}_1) \in (\times^{M+r-1} \mathbb{R}^d)$;

并根据归纳假设：

（b4*）存在一个全测度集 $\mathcal{Y} \subset (\times^M \mathbb{R}^d)$，使得对于任意非零测度集 $\mathcal{Y}' \subset \mathcal{Y}$ 和 $\mathcal{Z}_{r-1} \subset (\times^{r-1} \mathbb{R}^d)$，族 $\otimes^{r-1} \mathcal{V}_J$ 在 $(\boldsymbol{z}_{r-1},\ \ldots,\ \boldsymbol{z}_1,\ \mathbf{y}_M,\ \ldots,\ \mathbf{y}_1) \in \mathcal{Z}_{r-1} \times \mathcal{Y}'$ 中的有限混合下包含线性无关函数。

这里的策略是将此情形归约为上述基例（$n = 2$），其中重要的是我们需要证明：

(1) S_0 至少包含两个 $(j_r, \ldots\ldots, j_1)$ 和 $(j_r', \ldots\ldots, j_1')$，且 $j_r \neq j_r'$；并且

(2) 在任意选择 $\mathcal{Z}_r$ 和 $\mathcal{Y}' \subset \mathcal{Y}$ 时，族的乘积中的重叠变量 $(\boldsymbol{z}_{r-1}, \ldots\ldots, \boldsymbol{z}_1, \mathbf{y}_M, \ldots\ldots, \mathbf{y}_r) \in (\pi_{\{r-1,\ldots,1\}}(\mathcal{Z}_r) \times \pi_{\{M,\ldots,r\}}(\mathcal{Y}'))$ 都不会在非零测度集合上导致线性相关。

按照公式（41）中的相同逻辑，我们可以直接从 (b1) 和 (b4) 得出 (1) 成立；如果有 $j_r = j_r'$，线性相关关系就会与 (b4) 矛盾。因此，我们定义$J_0 := \{j_r \in J | \exists (j_r, \ldots\ldots, j_1) \in S_0\}$, $|J_0| < +\infty$。然后，对于任意

$(\boldsymbol{\beta}_M, \ldots\ldots, \boldsymbol{\beta}_1) \in (\pi_{\{r-1,\ldots,1\}}(\mathcal{Z}_r) \times \pi_{\{M,\ldots,r\}}(\mathcal{Y}')) \subset (\times^M \mathbb{R}^d)$，其中 π 表示定义 17 中定义的集合投影。

$$\sum_{j_r \in J_0} v_{j_r}(\boldsymbol{z}_r, \boldsymbol{z}_{r-1} = \boldsymbol{\beta}_M, \ldots, \boldsymbol{z}_1 = \boldsymbol{\beta}_{M-r+2}, \mathbf{y}_M = \boldsymbol{\beta}_{M-r+1} \ldots, \mathbf{y}_r = \boldsymbol{\beta}_1)$$
$$\left(\sum_{(j_{r-1},\ldots,j_1):(j_r,\ldots,j_1)\in S_0} \gamma_{j_r,\ldots,j_1} v_{j_{r-1},\ldots,j_1}(\boldsymbol{z}_{r-1} = \boldsymbol{\beta}_M, \ldots, \boldsymbol{z}_1 = \boldsymbol{\beta}_{M-r+2}, \mathbf{y}_M = \boldsymbol{\beta}_{M-r+1}, \right.$$
$$\left. \ldots, \mathbf{y}_r = \boldsymbol{\beta}_1, \mathbf{y}_{r-1}, \ldots, \mathbf{y}_1) \right) = 0,$$
$$\forall (\boldsymbol{z}_r, \mathbf{y}_{r-1}, \ldots, \mathbf{y}_1) \in \left(\mathcal{Z}_r|_{(\boldsymbol{z}_{r-1}=\boldsymbol{\beta}_M,\ldots,\boldsymbol{z}_1=\boldsymbol{\beta}_{M-r+2})} \times \mathcal{Y}'|_{(\mathbf{y}_M=\boldsymbol{\beta}_{M-r+1},\ldots,\mathbf{y}_r=\boldsymbol{\beta}_1)} \right). \quad (54)$$

与基例一样，$\mathcal{Z}_r|_{(\boldsymbol{z}_{r-1}=\boldsymbol{\beta}_M,\ldots,\boldsymbol{z}_1=\boldsymbol{\beta}_{M-r+2})}$ 和 $\mathcal{Y}'|_{(\mathbf{y}_M=\boldsymbol{\beta}_{M-r+1},\ldots,\mathbf{y}_r=\boldsymbol{\beta}_1)}$ 分别是来自 $\mathcal{Z}_r$ 和 $\mathcal{Y}'$ 的条件投影（定义 18），并且依赖于$\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1$。与之前一样，这些集合均非空；对于基例（$r = 2$），给定 $\mathcal{Z}_r$ 和 $\mathcal{Y}' \subset \mathcal{Y}$的选择后，我们可以建立变量与集合之间的以下等价性。

- $\boldsymbol{z}_r$等价于 $\boldsymbol{z}_2$，因此 $\mathcal{Z}_r|_{(\boldsymbol{z}_{r-1}=\boldsymbol{\beta}_M,\ldots,\boldsymbol{z}_1=\boldsymbol{\beta}_{M-r+2})}$等价于 $\mathcal{Z}_2|_{(\boldsymbol{z}_1=\boldsymbol{\beta}_M)}$。
- $\boldsymbol{z}_{r-1}, \ldots, \boldsymbol{z}_1$ 等价于 $\boldsymbol{z}_1$，因此 $\pi_{\{r-1,\ldots,1\}}(\mathcal{Z}_r)$ 等价于 $\pi_{\{1\}}(\mathcal{Z}_2)$。
- $\mathbf{y}_M, \ldots, \mathbf{y}_r$等价于 $\mathbf{y}_M, \ldots, \mathbf{y}_2$，因此 $\pi_{\{M,\ldots,r\}}(\mathcal{Y}')$ 等价于$\pi_{\{M,\ldots,2\}}(\mathcal{Y}')$。
- $\mathbf{y}_{r-1}, \ldots, \mathbf{y}_1$ 等价于 $\mathbf{y}_1$，因此 $\mathcal{Y}'|_{(\mathbf{y}_M=\boldsymbol{\beta}_{M-r+1},\ldots,\mathbf{y}_r=\boldsymbol{\beta}_1)}$ 等价于$\mathcal{Y}'|_{(\mathbf{y}_M=\boldsymbol{\beta}_{M-1},\ldots,\mathbf{y}_2=\boldsymbol{\beta}_1)}$。

根据上述等价关系，对于任何 $(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) \in (\pi_{\{r-1,\ldots,1\}}(\mathcal{Z}_r) \times \pi_{\{M,\ldots,r\}}(\mathcal{Y}'))$ 的线性相关情况，导出矛盾的论证仍然成立。唯一的区别是，投影和条件投影的维数会发生变化（$\boldsymbol{z}_r \in \mathcal{Z}_r|_{(\boldsymbol{z}_{r-1}=\boldsymbol{\beta}_M,\ldots,\boldsymbol{z}_1=\boldsymbol{\beta}_{M-r+2})}$ 除外）。因此，集合

$$\mathcal{D}_0 := \left\{ (\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) \in \left(\pi_{\{r-1,\ldots,1\}}(\mathcal{Z}_r) \times \pi_{\{M,\ldots,r\}}(\mathcal{Y}') \right) \right|$$
$$\sum_{(j_{r-1},\ldots,j_1):(j_r,\ldots,j_1)\in S_0} \gamma_{j_r,\ldots,j_1} v_{j_{r-1},\ldots,j_1}(\boldsymbol{z}_{r-1} = \boldsymbol{\beta}_M, \ldots, \boldsymbol{z}_1 = \boldsymbol{\beta}_{M-r+2}, \mathbf{y}_M = \boldsymbol{\beta}_{M-r+1},$$
$$\left. \ldots, \mathbf{y}_r = \boldsymbol{\beta}_1, \mathbf{y}_{r-1}, \ldots, \mathbf{y}_1) = 0, \forall (\mathbf{y}_{r-1}, \ldots, \mathbf{y}_1) \in \mathcal{Y}'|_{(\mathbf{y}_M=\boldsymbol{\beta}_{M-r+1},\ldots,\mathbf{y}_r=\boldsymbol{\beta}_1)} \right\} \quad (55)$$

在假设（b4）下具有零测度，这意味着对于任意满足 $\mathcal{D} := (\pi_{\{r-1,\ldots,1\}}(\mathcal{Z}_r) \times \pi_{\{M,\ldots,r\}}(\mathcal{Y}')) \setminus \mathcal{D}_0$ 的 $(\boldsymbol{z}_{r-1} = \boldsymbol{\beta}_M$、……、$\boldsymbol{z}_1 = \boldsymbol{\beta}_{M-r+2}$，$\mathbf{y}_M = \boldsymbol{\beta}_{M-r+1}$、……、$\mathbf{y}_r = \boldsymbol{\beta}_1) \in \mathcal{D}$，都存在 $(\mathbf{y}_{r-1}$，……，$\mathbf{y}_1) \in \mathcal{Y}'|_{(\mathbf{y}_M=\boldsymbol{\beta}_{M-r+1},\ldots,\mathbf{y}_r=\boldsymbol{\beta}_1)}$，使得

$$\tilde{\gamma}_{j_r}(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) := \sum_{(j_{r-1},\ldots,j_1):(j_r,\ldots,j_1)\in S_0} \gamma_{j_r,\ldots,j_1} v_{j_{r-1},\ldots,j_1}(\boldsymbol{z}_{r-1} = \boldsymbol{\beta}_M, \ldots,$$
$$\boldsymbol{z}_1 = \boldsymbol{\beta}_{M-r+2}, \mathbf{y}_M = \boldsymbol{\beta}_{M-r+1}, \ldots, \mathbf{y}_r = \boldsymbol{\beta}_1, \mathbf{y}_{r-1}, \ldots, \mathbf{y}_1) \neq 0 \quad (56)$$

对于至少两个 $j_r \in J_0$。与之前一样，这意味着 $\{v_{j_r}(\boldsymbol{z}_r, \boldsymbol{z}_{r-1} = \boldsymbol{\beta}_M, \ldots, \boldsymbol{z}_1 = \boldsymbol{\beta}_{M-r+2}, \mathbf{y}_M = \boldsymbol{\beta}_{M-r+1} \ldots, \mathbf{y}_r = \boldsymbol{\beta}_1)\}$ 关于 $\boldsymbol{z}_r \in \mathcal{Z}_r|_{(\boldsymbol{z}_{r-1}=\boldsymbol{\beta}_M,\ldots,\boldsymbol{z}_1=\boldsymbol{\beta}_{M-r+2})}$ 的线性相关，$\forall(\boldsymbol{z}_{r-1} = \boldsymbol{\beta}_M, \ldots, \boldsymbol{z}_1 = \boldsymbol{\beta}_{M-r+2}, \mathbf{y}_M = \boldsymbol{\beta}_{M-r+1} \ldots, \mathbf{y}_r = \boldsymbol{\beta}_1) \in \mathcal{D}$。因此，利用假设（b5-b6），我们可以将 J_0 划分为由 $K(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1)$ 索引的子集，使得每个子集中的函数都相等，这与方程（44-45）类似。沿用情形 $r = 2$ 中之前的论证，我们选取 $(\boldsymbol{\beta}_M^*, \ldots, \boldsymbol{\beta}_1^*) \in$极小值点$_{(\boldsymbol{\beta}_M,\ldots,\boldsymbol{\beta}_1)\in\mathcal{D}} C(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1)$，而假设（b6）则确立了元素 $v_{j_{r-1},\ldots,j_1}(\cdot)$ 在至少一个非零测度集上呈线性相关（方程（51））

在一个非零测度集上，$\otimes^{r-1}\mathcal{V}_J$ 的线性相关与归纳假设 (b4*) 矛盾，因为 $\mathcal{Y}$ 具有全测度。因此，无论如何选择 $\mathcal{Y}$和 $\mathcal{Z}_r$，等式 (53) 都不可能成立，从而得出矛盾。

■

关于 (b5)，我们注意到，对于给定的变量子集，函数乘积在 M 上的线性相关不再意味着函数重复。相反，我们证明，在有限混合下，乘积族中至少有一个分量是线性无关的。

引理20（(b5) 的扩展）设分布族 $\mathcal{V}_J = \{v_j(\boldsymbol{z}, \mathbf{y}_M, \ldots, \mathbf{y}_1)|j \in J\}$ 的变量为 $\boldsymbol{z}, \mathbf{y}_M, \ldots, \mathbf{y}_1$，定义在 $\times^{M+1}\mathbb{R}^d$上，并满足 *(b1)*、*(b4)*、*(b5)*、*(b6)*。

那么，对于任意满足 $2 \leq r \leq M$的 r，任意 $(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) \in (\times^M\mathbb{R}^d)$，任意非零测度 $\mathcal{Z}_r \subset (\times^r\mathbb{R}^d)$，以及满足 $|S_0| < +\infty$ 的 $S_0 \subset (\times^r J)$，族

$$\Big\{v_{j_r}(\boldsymbol{z}_r, \ldots, \boldsymbol{z}_1, \mathbf{y}_M = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_r = \boldsymbol{\beta}_r) \ldots v_{j_1}(\boldsymbol{z}_1, \mathbf{y}_M = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_1 = \boldsymbol{\beta}_1)|(j_r, \ldots, j_1) \in S_0\Big\}$$

仅当存在 $\exists \ell \leq r$ 使得

$$\Big\{v_{j_\ell}(\boldsymbol{z}_\ell, \ldots, \boldsymbol{z}_1, \mathbf{y}_M = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_\ell = \boldsymbol{\beta}_\ell)|j_\ell \in J_\ell\Big\}, \quad with\ J_\ell := \big\{j_\ell \in J|\exists(j_r, \ldots, j_1) \in S_0\big\}$$

在 $(\boldsymbol{z}_\ell, \ldots, \boldsymbol{z}_1) \in \pi_{\{\ell,\ldots,1\}}(\mathcal{Z}_r)$ 上包含线性相关函数。

证明 我们从基例（$r = 2$）开始，按如下方式用归纳法证明该命题。

情形 $r = 2$。固定 $\mathcal{Z}_2 \subset \mathbb{R}^d \times \mathbb{R}^d$，并假设对于任意 $(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) \in (\times^M\mathbb{R}^d)$，以及 $S_0 \subset J \times J$，该族

$$\Big\{v_{j_2}(\boldsymbol{z}_2, \boldsymbol{z}_1, \mathbf{y}_M = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_2)v_{j_1}(\boldsymbol{z}_1, \mathbf{y}_M = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_1 = \boldsymbol{\beta}_1)|(j_2, j_1) \in S_0\Big\}$$

包含线性相关函数。根据 (b1, b4-b6)，通过令 $u_i(\mathbf{y} = \boldsymbol{z}_1, \boldsymbol{x} = \emptyset) = v_{j_1}(\boldsymbol{z}_1, \mathbf{y}_M = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_1 = \boldsymbol{\beta}_1), i = j_1$，以及 $v_j(\boldsymbol{z} = \boldsymbol{z}_2, \mathbf{y} = \boldsymbol{z}_1) = v_{j_2}(\boldsymbol{z}_2, \boldsymbol{z}_1, \mathbf{y}_M = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_2 = \boldsymbol{\beta}_2), j = j_2$，我们可以将这些假设与引理 16 中定义的假设联系起来，其中 (a1)、(a5) 和 (a6) 均成立。因此，根据引理 16，我们知道这两个族中必有一个包含线性相关函数（这与引理 16 中的 (a3) 或 (a4) 矛盾）。

情形 $r > 2$。固定 $\mathcal{Z}_r \subset (\times^r \mathbb{R}^d)$，并假设对于任意的 $(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_1) \in (\times^M \mathbb{R}^d)$ 和 $S_0 \subset (\times^r J)$，该族

$$\Big\{ v_{j_r}(\boldsymbol{z}_r, \ldots, \boldsymbol{z}_1, \mathbf{y}_M = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_r = \boldsymbol{\beta}_r) \ldots \\ v_{j_1}(\boldsymbol{z}_1, \mathbf{y}_M = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_1 = \boldsymbol{\beta}_1) | (j_r, \ldots, j_1) \in S_0 \Big\}$$

co包含线性相关函数。和之前一样，根据 (b1, b4-b6) 并通过设定

$$u_i(\mathbf{y} = (\boldsymbol{z}_{r-1}, \ldots, \boldsymbol{z}_1), \boldsymbol{x} = \emptyset) = v_{j_{r-1}}(\boldsymbol{z}_{r-1}, \ldots, \boldsymbol{z}_1, \mathbf{y}_M = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_{r-1} = \boldsymbol{\beta}_{r-1}) \ldots \\ v_{j_1}(\boldsymbol{z}_1, \mathbf{y}_M = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_1 = \boldsymbol{\beta}_1), \quad i = (j_{r-1}, \ldots, j_1),$$

以及 $v_j(\boldsymbol{z} = \boldsymbol{z}_r,\ \mathbf{y} = (\boldsymbol{z}_{r-1},\ \ldots\ldots,\ \boldsymbol{z}_1)) = v_{j_r}(\boldsymbol{z}_r,\ \ldots \boldsymbol{z}_1,\ \mathbf{y}_M = \boldsymbol{\beta}_M,\ \ldots\ldots,\ \mathbf{y}_r = \boldsymbol{\beta}_r) \quad j = j_r$，我们可以建立与引理 16 中假设的对应关系。我们知道，(a1) 和 (a6) 可分别由 (b1) 和 (b6) 直接满足。通过在 (b5) 中设定 $\ell = M + 1 - r$，即可满足 (a5)。然后，由于引理 16 中 (a3) 或 (a4) 任一项所导致的矛盾，两个函数族中必有一个包含线性相关函数。因此，或者

$$\Big\{ v_{j_r}(\boldsymbol{z}_r, \ldots, \boldsymbol{z}_1, \mathbf{y}_M = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_r = \boldsymbol{\beta}_r) | j_r \in J_r \Big\}, \text{ with } J_r = \big\{ j_r \in J | \exists (j_r, \ldots, j_1) \in S_0 \big\}$$

在 $(\boldsymbol{z}_r,\ \ldots\ldots,\ \boldsymbol{z}_1) \in \pi_{\{r,\ldots,1\}}(\mathcal{Z}_r)$ 上包含线性相关函数，或者根据归纳假设，存在某个 $\ell \leq r-1$ 使得

$$\Big\{ v_{j_\ell}(\boldsymbol{z}_\ell, \ldots, \boldsymbol{z}_1, \mathbf{y}_M = \boldsymbol{\beta}_M, \ldots, \mathbf{y}_\ell = \boldsymbol{\beta}_\ell) | j_\ell \in J_\ell \Big\}, \\ \text{with } J_\ell = \big\{ j_\ell \in J | \exists (j_r, \ldots, j_1) \in S_0 \big\}$$

在 $(\boldsymbol{z}_\ell, \ldots, \boldsymbol{z}_1) \in \pi_{\{\ell,\ldots,1\}}(\mathcal{Z}_r)$ 上包含线性相关函数。 ■

A.2.3 具有 M 个重叠变量的乘积上的线性无关性

鉴于假设(b5)无法像(b4)那样推广，我们使用引理20将 LIPO-1 推广为 LIPO-M（具有 M 个连续重叠变量的乘积函数线性无关性），以证明当 V_J 定义为具有连续变量的 M 函数的乘积时，有限混合下的线性无关性。

记号： 我们使用以下分块记号 $\boldsymbol{z}_{\ell:1}$, $\boldsymbol{\beta}_{M:\ell}$, $\mathbf{y}_{M:\ell}$ 表示变量元组 $(\boldsymbol{z}_\ell, \ldots, \boldsymbol{z}_1)$, $(\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_\ell)$, $(\mathbf{y}_M, \ldots, \mathbf{y}_\ell)$。

引理 21（LIPO-M） 假设有两个族 $\mathcal{U}_I = \{u_i(\mathbf{y}\ \boldsymbol{x}) | i \in I\}$ 和 $\mathcal{V}_J = \{v_j(\boldsymbol{z}, \mathbf{y}) | j \in J\}$，其中 $\boldsymbol{x} \in \mathcal{X}$、$\mathbf{y} \in \mathbb{R}^{d_y}$ 和 $\boldsymbol{z} \in \mathbb{R}^{d_z}$，且 $d_y = d \cdot M$、$d_z = d \cdot r$ 和 $1 \leq r \leq M$。我们进一步假设：

(a) $\mathcal{U}_I$ 族满足假设 *(a1)* 和 *(a3)*。

(b) $\mathcal{V}_J$ 族中的每个元素都定义为 r 个函数的乘积，这些函数均属于同一族 $\mathcal{W}_K = \{w_k（\boldsymbol{z}, \mathbf{y}_M, ..., \mathbf{y}_1）| k \in K\}$，其中 $\boldsymbol{z}_1 \in \mathbb{R}^d$，（$\mathbf{y}_M, ..., \mathbf{y}_1$）$\in (\times^M \mathbb{R}^d)$，并表示如下

$$\mathcal{V}_J := \otimes^r \mathcal{W}_K = \{w_{k_r}(\boldsymbol{z}_r, \ldots, \boldsymbol{z}_1, \mathbf{y}_M, \ldots, \mathbf{y}_r) \ldots w_{k_1}(\boldsymbol{z}_1, \mathbf{y}_M, \ldots, \mathbf{y}_1) \mid (k_r, \ldots, k_1) \in (\times^r K)\}$$

其中 $J := (\times^r K)$，且 $\boldsymbol{z} := (\boldsymbol{z}_r, \ldots, \boldsymbol{z}_1)$，$\mathbf{y} := (\mathbf{y}_M, \ldots, \mathbf{y}_1)$。族 $\mathcal{W}_K$ 满足假设 ***(b1)***、***(b4)***、***(b5)*** 和 ***(b6)***。

那么，对于任意非零测度子集 $\mathcal{Z} \subset \mathbb{R}^{d_z}$, $\mathcal{U}_I \otimes \mathcal{V}_J := \{u_i(\mathbf{y}\ \boldsymbol{x}) v_j(\boldsymbol{z}, \mathbf{y}) | i \in I$，在定义于 $(\boldsymbol{x}, \mathbf{y}\ \boldsymbol{z}) \in \mathcal{X} \times \mathbb{R}^{d_y} \times \mathcal{Z}$ 上的有限混合中，包含线性无关函数。

证明 假设该充分性陈述不成立，则存在一个非零测度子集$\mathcal{Z} \subset \mathbb{R}^{d_z}$, $S_0 \subset I \times J$，其中 $|S_0| < +\infty$，以及一个非零值集合 $\{\gamma_{ij} \in \mathbb{R} | (i, j) \in S_0\}$，使得

$$\sum_{(i,j) \in S_0} \gamma_{ij} v_j(\boldsymbol{z}, \mathbf{y}) u_i(\mathbf{y}, \boldsymbol{x}) = 0, \quad \forall (\boldsymbol{x}, \mathbf{y}, \boldsymbol{z}) \in \mathcal{X} \times \mathbb{R}^{d_y} \times \mathcal{Z}. \tag{57}$$

为方便起见，尽管 J 是一个元组，我们仍将其元素称为索引。注意，S_0 和 γ_{ij} 的选择独立于任何 $(\boldsymbol{x}, \mathbf{y}\ \boldsymbol{z})$ 值，但可能依赖于 $\mathcal{Z}$。我们希望遵循 LIPO-1（引理 16）在 (a1-a6) 条件下提出的逻辑。下面我们建立一些联系。首先，我们知道，按 (b) 中定义的族 $\mathcal{V}_J$ 是由 $1 \leq r \leq M$中的 r 个元素构成的乘积，且所有这些元素均满足 (b1) 和 (b4-b5)。引理 19 表明，存在某个满测度 $\mathcal{Y}$，使得对于任意子集 $\mathcal{Z} \times \mathcal{Y}' \subset \mathbb{R}^{d_z} \times \mathcal{Y}$，族 $\mathcal{V}_J$ 在有限混合下线性无关。这比引理 16 中的 (a4) 更强，因为 (b4) 提供了一个满测度 $\mathcal{Y}$，而 (a4) 只要求一个非零测度 $\mathcal{Y}$。此外，它也满足 (a1)，因为它被定义为由满足 (b1) 的 r 个函数与 $1 \leq r \leq M$构成的乘积。因此，我们沿用引理 16 中的相同论证，证明指标集 S_0 至少包含两个不同的索引 (i, j) 和 (i, j')，且满足 $j \neq j'$。此外，我们将 $J_0 = \{j \in J | \exists (i, j)$ 定义为在 S_0中出现的所有可能的 j索引构成的集合，其中$|J_0| < +\infty$，而以下集合 $\mathcal{Y}_0 := \{\boldsymbol{\beta} \in \mathbb{R}^{d_y} | \sum_{i:(i,j) \in S_0} \gamma_{ij} u_i(\mathbf{y} = \boldsymbol{\beta}, \boldsymbol{x}) = 0, \forall \boldsymbol{x} \in \mathcal{X}\}$根据 (a3)，在 $\mathbb{R}^{d_y}$ 中只能具有零测度。再次沿用引理 16，我们有一个非零测度集 $\mathcal{Y}' := \mathcal{Y} \backslash \mathcal{Y}_0 \subset \mathcal{Y}$，其中我们从 (b4) 中的满测度集选取 $\mathcal{Y}$。于是，对于每个 $\boldsymbol{\beta} \in \mathcal{Y}'$，存在 $\boldsymbol{x} \in \mathcal{X}$，使得 $\sum_{i:(i,j) \in S_0} \gamma_{ij} u_i(\mathbf{y} = \boldsymbol{\beta}, \boldsymbol{x}) \neq 0$ 至少对应于 J_0中的两个 j 索引。这意味着，对于每个 $\boldsymbol{\beta} \in \mathcal{Y}'$，$\{v_j(\boldsymbol{z}, \mathbf{y} = \boldsymbol{\beta}) | j \in J_0\}$ 在 $\boldsymbol{z} \in \mathcal{Z}$上包含线性相关函数。

现在我们将以 J、S_0 和 J_0 的元组格式书写，因为我们需要展开乘积族 $\mathcal{V}_J$。因此，我们现在将 $j =$ 视为 $(k_r, \ldots, k_1)$，并将 $J_0 = \{$ 重新定义为 $(k_r, \ldots, k_1) \in J | \exists (i, k_r, \ldots, k_1) \in S_0\}$。由 (b4-b6)，引理20表明，构成族 $\mathcal{V}_J$ 的 r 个函数 $(1 \leq r \leq M)$ 的乘积意味着至少有一个分量中的函数线性相关。也就是说，对于某个 $\ell \leq r$，$\{w_{k_\ell}(\boldsymbol{z}_\ell, \ldots, \boldsymbol{z}_1, \mathbf{y}_{M:\ell} = \boldsymbol{\beta}_{M:\ell}) | k_\ell \in K_0^{(\ell)}\}$，其中 $\mathbf{y}_{M:\ell} := (\mathbf{y}_M, \ldots, \mathbf{y}_\ell)$，$\boldsymbol{\beta}_{M:\ell} := (\boldsymbol{\beta}_M, \ldots, \boldsymbol{\beta}_\ell)$，且 $K_0^{(\ell)} := \{k_\ell \in K | \exists (i, k_r, \ldots, k_1) \in S_0\}$，在 $(\boldsymbol{z}_\ell, \ldots, \boldsymbol{z}_1) \in \pi_{\{\ell, \ldots, 1\}}(\mathcal{Z})$ 上包含线性相关函数。在假设 (b5) 下，

我们可以将指标集 $K_0^{(\ell)}$ 划分为由 $q \in L^{(\ell)}(\boldsymbol{\beta}_{M:\ell})$ 索引的子集，如下所示，使得在每个指标子集 $K_q^{(\ell)}(\boldsymbol{\beta}_{M:\ell})$ 内，对应索引的函数都相等：

$$\begin{aligned} K_0^{(\ell)} = \cup_{q\in L^{(\ell)}(\boldsymbol{\beta}_{M:\ell})} K_q^{(\ell)}(\boldsymbol{\beta}_{M:\ell}), \quad K_q^{(\ell)}(\boldsymbol{\beta}_{M:\ell}) \cap K_{q'}^{(\ell)}(\boldsymbol{\beta}_{M:\ell}) = \emptyset, \forall q \neq q' \in L^{(\ell)}(\boldsymbol{\beta}_{M:\ell}), \\ k_\ell \neq k'_\ell \in K_q^{(\ell)}(\boldsymbol{\beta}_{M:\ell}) \quad \Leftrightarrow \quad w_{k_\ell}(\boldsymbol{z}_\ell, \dots, \boldsymbol{z}_1, \mathbf{y}_{M:\ell} = \boldsymbol{\beta}_{M:\ell}) = \\ w_{k'_\ell}(\boldsymbol{z}_\ell, \dots, \boldsymbol{z}_1, \mathbf{y}_{M:\ell} = \boldsymbol{\beta}_{M:\ell}), \forall (\boldsymbol{z}_\ell, \dots, \boldsymbol{z}_1) \in \pi_{\{\ell,\dots,1\}}(\mathcal{Z}). \end{aligned} \tag{58}$$

那么，对于任意 $\boldsymbol{\beta} \in \mathcal{Y}'$，我们可以将公式（57）重写为

$$\begin{aligned} \sum_{q\in L^{(\ell)}(\boldsymbol{\beta}_{M:\ell})} \Bigg(\sum_{k_\ell \in K_q^{(\ell)}(\boldsymbol{\beta}_{M:\ell})} \sum_{(i,k_r,\dots,k_{\ell+1},k_{\ell-1},\dots,k_1):(i,k_r,\dots,k_1)\in S_0} \gamma_{i,k_r,\dots,k_1} u_i(\mathbf{y} = \boldsymbol{\beta}, \boldsymbol{x}) \\ w_{k_r}(\boldsymbol{z}_r, \dots, \boldsymbol{z}_1, \mathbf{y}_M = \boldsymbol{\beta}_M) \dots w_{k_1}(\boldsymbol{z}_1, \mathbf{y}_M = \boldsymbol{\beta}_M, \dots, \mathbf{y}_1 = \boldsymbol{\beta}_1) \Bigg) = 0, \\ \forall (\boldsymbol{x}, \boldsymbol{z}_r, \dots, \boldsymbol{z}_1) \in \mathcal{X} \times \mathcal{Z}. \end{aligned} \tag{59}$$

注意，对于任意 $(i, j) \in S_0$ 和 $(k_r, \dots, k_1) \in J_0$，我们使用 $\gamma_{i,k_r,\dots,k_1} = \gamma_{ij}$，使得 $j = (k_r, \dots, k_1)$。根据公式（58），回想一下，$w_{k_\ell}(\boldsymbol{z}_\ell, \dots, \boldsymbol{z}_1, \mathbf{y}_{M:\ell} = \boldsymbol{\beta}_{M:\ell})$ 和 $w_{k'_\ell}(\boldsymbol{z}_\ell, \dots, \boldsymbol{z}_1, \mathbf{y}_{M:\ell} = \boldsymbol{\beta}_{M:\ell})$ 在 $(\boldsymbol{z}_\ell, \dots, \boldsymbol{z}_1) \in \pi_{\{\ell,\dots,1\}}(\mathcal{Z})$ 上是相同的函数，当且仅当 $k_\ell \neq k'_\ell$ 属于同一指标集 $K_l^{(\ell)}(\boldsymbol{\beta})$。这意味着，如果公式（57）成立，那么对于任意 $\boldsymbol{\beta} \in \mathcal{Y}'$，在假设 (b1) 和 (b5) 下，

$$\begin{aligned} \sum_{k_\ell \in K_q^{(\ell)}(\boldsymbol{\beta}_{M:\ell})} \sum_{(i,k_r,\dots,k_{\ell+1},k_{\ell-1},\dots,k_1):(i,k_r,\dots,k_1)\in S_0} \gamma_{i,k_r,\dots,k_1} u_i(\mathbf{y} = \boldsymbol{\beta}, \boldsymbol{x}) w_{k_r}(\boldsymbol{z}_r, \dots, \boldsymbol{z}_1, \\ \mathbf{y}_M = \boldsymbol{\beta}_M) \dots w_{k_{\ell+1}}(\boldsymbol{z}_{\ell+1}, \dots, \boldsymbol{z}_1, \mathbf{y}_{M:\ell+1} = \boldsymbol{\beta}_{M:\ell+1}) w_{k_{\ell-1}}(\boldsymbol{z}_{\ell-1}, \dots, \boldsymbol{z}_1, \\ \mathbf{y}_{M:\ell-1} = \boldsymbol{\beta}_{M:\ell-1}) \dots w_{k_1}(\boldsymbol{z}_1, \mathbf{y}_{M:1} = \boldsymbol{\beta}_{M:1}) = 0, \\ \forall (\boldsymbol{x}, \boldsymbol{z}_r, \dots, \boldsymbol{z}_1) \in \mathcal{X} \times \mathcal{Z},\ q \in L^{(\ell)}(\boldsymbol{\beta}_{M:\ell}). \end{aligned} \tag{60}$$

定义 $C^{(\ell)}(\boldsymbol{\beta}_{M:\ell}) = \min_q |K_q^{(\ell)}(\boldsymbol{\beta}_{M:\ell})|$ 为 $K_q^{(\ell)}(\boldsymbol{\beta}_{M:\ell})$ 子集中的 k_ℓ 索引的最小基数计数。选择 $\boldsymbol{\beta}^*_{M:\ell} \in$ 极小值点$_{\boldsymbol{\beta}_{M:\ell} \in \pi_{\{M,\dots,\ell\}}(\mathcal{Y}')} C^{(\ell)}(\boldsymbol{\beta}_{M:\ell})$：

1. 我们有 $C^{(\ell)}(\boldsymbol{\beta}^*_{M:\ell}) < |K_0^{(\ell)}|$ 和 $|L^{(\ell)}(\boldsymbol{\beta}^*_{M:\ell})| \geq 2$。否则，对于所有 $k_\ell \neq k'_\ell \in K_0^{(\ell)}$，我们都有 $w_{k_\ell}(\boldsymbol{z}_\ell, \dots, \boldsymbol{z}_1, \mathbf{y}_{M:\ell} = \boldsymbol{\beta}_{M:\ell}) = w_{k'_\ell}(\boldsymbol{z}_\ell, \dots, \boldsymbol{z}_1, \mathbf{y}_{M:\ell} = \boldsymbol{\beta}_{M:\ell})$ 对所有 $(\boldsymbol{z}_\ell, \dots, \boldsymbol{z}_1) \in \pi_{\{\ell,\dots,1\}}(\mathcal{Z})$ 以及 $\boldsymbol{\beta}_{M:\ell} \in \pi_{\{M,\dots,\ell\}}(\mathcal{Y}')$ 均成立，因此它们关于 $(\boldsymbol{z}_\ell, \dots, \boldsymbol{z}_1, \mathbf{y}_{M:\ell}) \in \pi_{\{\ell,\dots,1\}}(\mathcal{Z}) \times \pi_{\{M,\dots,\ell\}}(\mathcal{Y}')$ 线性相关，这与假设（b4）矛盾。

2. 现在不失一般性地假设 $|K_1^{(\ell)}(\boldsymbol{\beta}^*_{M:\ell})| = C^{(\ell)}(\boldsymbol{\beta}^*_{M:\ell})$。根据假设（b5），我们知道，对于任意 $k_\ell \in K_1^{(\ell)}(\boldsymbol{\beta}^*_{M:\ell})$ 和 $k'_\ell \in K_0^{(\ell)} \backslash K_1^{(\ell)}(\boldsymbol{\beta}^*_{M:\ell})$，$w_{k_\ell}(\boldsymbol{z}_\ell, \dots, \boldsymbol{z}_1, \mathbf{y}_{M:\ell} = \boldsymbol{\beta}_{M:\ell}) = w_{k'_\ell}(\boldsymbol{z}_\ell, \dots, \boldsymbol{z}_1, \mathbf{y}_{M:\ell} = \boldsymbol{\beta}_{M:\ell})$ 至多仅在 $\pi_{\{\ell,\dots,1\}}(\mathcal{Z})$ 的零测度子集上成立。由于 $|K_0^{(\ell)}| < +\infty$ 以及 $\pi_{\{\ell,\dots,1\}}(\mathcal{Z}) \subset (\times^\ell \mathbb{R}^d)$ 具有非零测度，因此存在

$(\boldsymbol{z}_\ell^{(0)},\ldots,\boldsymbol{z}_1^{(0)})\in\pi_{\{\ell,\ldots,1\}}(\mathcal{Z})$ and $\delta>0$ such that

$$|w_{k_\ell}(\boldsymbol{z}_{\ell:1}=\boldsymbol{z}_{\ell:1}^{(0)},\mathbf{y}_{M:\ell}=\boldsymbol{\beta}_{M:\ell}^*)-w_{k'_\ell}(\boldsymbol{z}_{\ell:1}=\boldsymbol{z}_{\ell:1}^{(0)},\mathbf{y}_{M:\ell}=\boldsymbol{\beta}_{M:\ell}^*)|\geq\delta,\\ \forall k_\ell\in K_1^{(\ell)}(\boldsymbol{\beta}_{M:\ell}^*),\forall k'_\ell\in K_0^{(\ell)}\backslash K_1^{(\ell)}(\boldsymbol{\beta}_{M:\ell}^*).\quad(61)$$

在假设（b6）下，存在 $\epsilon_{k_\ell}>0$，使得我们可以使用 ℓ_2-范数构造一个 ϵ-球$B(\boldsymbol{\beta}_{M:\ell}^*,\epsilon_{k_\ell})$，使其满足

$$|w_{k_\ell}(\boldsymbol{z}_{\ell:1}=\boldsymbol{z}_{\ell:1}^{(0)},\mathbf{y}_{M:\ell}=\boldsymbol{\beta}_{M:\ell}^*)-w_{k_\ell}(\boldsymbol{z}_{\ell:1}=\boldsymbol{z}_{\ell:1}^{(0)},\mathbf{y}_{M:\ell}=\boldsymbol{\beta}_{M:\ell})|\leq\delta/3,\\ \forall\boldsymbol{\beta}_{M:\ell}\in B(\boldsymbol{\beta}_{M:\ell}^*,\epsilon_{k_\ell}).\quad(62)$$

选择合适的 $0<\epsilon'_\ell\leq\min^{k_\ell\in K_0^{(\ell)}}\epsilon_{k_\ell}$（注意，$\min_{k_\ell\in K_0^{(\ell)}}\epsilon_{k_\ell}>0$ 满足 $|K_0^{(\ell)}|<+\infty$）并构造一个基于 ℓ_2-范数的 ϵ-球 $B(\boldsymbol{\beta}_{M:\ell}^*,\ \epsilon'_\ell)\subset\pi_{\{M,\ldots,\ell\}}(\mathcal{Y}')$，对于所有 $k_\ell\in K_1^{(\ell)}(\boldsymbol{\beta}_{M:\ell}^*)$ $k'_\ell\in K_0^{(\ell)}\backslash K_1^{(\ell)}(\boldsymbol{\beta}_{M:\ell}^*)$，我们有 $k'_\ell\notin K_1^{(\ell)}(\boldsymbol{\beta}_{M:\ell})$；对于所有 $\boldsymbol{\beta}_{M:\ell}\in B(\boldsymbol{\beta}_{M:\ell}^*,\ \epsilon'_\ell)$，这是因为

$$|w_{k_\ell}(\boldsymbol{z}_{\ell:1}=\boldsymbol{z}_{\ell:1}^{(0)},\mathbf{y}_{M:\ell}=\boldsymbol{\beta}_{M:\ell})-w_{k'_\ell}(\boldsymbol{z}_{\ell:1}=\boldsymbol{z}_{\ell:1}^{(0)},\mathbf{y}_{M:\ell}=\boldsymbol{\beta}_{M:\ell})|\geq\delta/3,\\ \forall\boldsymbol{\beta}_{M:\ell}\in B(\boldsymbol{\beta}_{M:\ell}^*,\ \epsilon'_\ell)。(63)$$

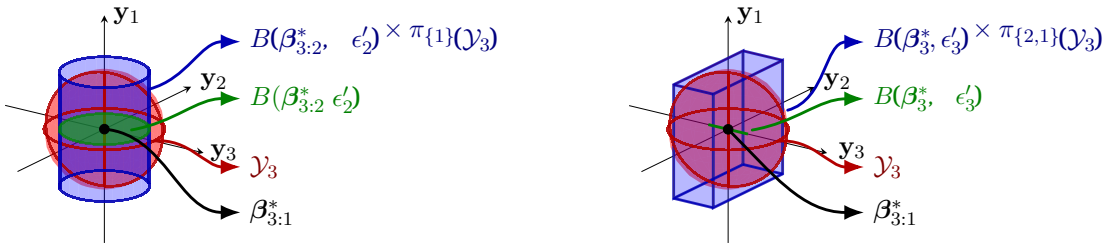
因此，这意味着对于任意 $\boldsymbol{\beta}\in B(\boldsymbol{\beta}_{M:\ell}^*,\ \epsilon)$ 的划分 $\{K_l^{(\ell)}(\boldsymbol{\beta}_{M:\ell})\}$，我们有 $K_1^{(\ell)}(\boldsymbol{\beta}_{M:\ell})\subset K_1^{(\ell)}(\boldsymbol{\beta}_{M:\ell}^*)$，因此 $|K_1^{(\ell)}(\boldsymbol{\beta}_{M:\ell})|\leq|K_1^{(\ell)}(\boldsymbol{\beta}_{M:\ell}^*)|$。根据 $\boldsymbol{\beta}_{M:\ell}^*\in$极小值点 $\boldsymbol{\beta}_{M:\ell}\in\pi_{\{M,\ldots,\ell\}}(\mathcal{Y})\,C^{(\ell)}(\boldsymbol{\beta}_{M:\ell})$ 和 $|K_1^{(\ell)}(\boldsymbol{\beta}_{M:\ell}^*)|=C^{(\ell)}(\boldsymbol{\beta}_{M:\ell}^*)$ 的定义，对于所有 $\boldsymbol{\beta}_{M:\ell}\in B(\boldsymbol{\beta}_{M:\ell}^*,\ \epsilon'_\ell)$，我们有 $K_1^{(\ell)}(\boldsymbol{\beta}_{M:\ell})=K_1^{(\ell)}(\boldsymbol{\beta}_{M:\ell}^*)$。

3. 可以证明 $|K_1^{(\ell)}(\boldsymbol{\beta}_{M:\ell}^*)|=1$；否则，根据划分的定义（公式（58））以及上述结论，存在 $k_\ell\neq k'_\ell\in K_1^{(\ell)}(\boldsymbol{\beta}_{M:\ell}^*)$，使得对所有 $(,)$ $()$ 和 $(\boldsymbol{\beta}_{M:\ell}^*,\epsilon'_\ell)$，$w_{k_\ell}(\boldsymbol{z}_\ell,\ldots,\boldsymbol{z}_1,\mathbf{y}_{M:\ell}=\boldsymbol{\beta}_{M:\ell}=w_{k'_\ell}\ \boldsymbol{z}_\ell\ \ldots\ \boldsymbol{z}_1\ \ M:\ell=\boldsymbol{\beta}_{M:\ell}\quad\boldsymbol{z}_\ell\ \ldots\ \boldsymbol{z}_1\ \in\pi_{\{\ell,\ldots,1\}}\ \mathcal{Z}\quad\boldsymbol{\beta}_{M:\ell}\in)\,(,,\mathbf{y})$；这 B 再次与假设（b）矛盾。

根据上述 3 点，并假设 $k_\ell\in K_1^{(\ell)}(\boldsymbol{\beta}_{M:\ell}^*)$ 是该子集中的唯一索引，那么上述结论中已证明的事实 $K_1^{(\ell)}(\boldsymbol{\beta}_{M:\ell})=K_1^{(\ell)}(\boldsymbol{\beta}_{M:\ell}^*)$ 意味着

$$\sum_{(i,k_r,\ldots,k_{\ell+1},k_{\ell-1},\ldots,k_1):(i,k_r,\ldots,k_1)\in S_0}\gamma_{i,k_r,\ldots,k_1}u_i(\mathbf{y}=\boldsymbol{\beta},\boldsymbol{x})w_{k_r}(\boldsymbol{z}_r,\ldots,\boldsymbol{z}_1,\mathbf{y}_M=\boldsymbol{\beta}_M)\cdot\\ \ldots\cdot w_{k_{\ell+1}}(\boldsymbol{z}_{\ell+1},\ldots,\boldsymbol{z}_1,\mathbf{y}_{M:\ell+1}=\boldsymbol{\beta}_{M:\ell+1})w_{k_{\ell-1}}(\boldsymbol{z}_{\ell-1},\ldots,\boldsymbol{z}_1,\mathbf{y}_{M:\ell-1}=\boldsymbol{\beta}_{M:\ell-1})\cdot\\ \ldots\cdot w_{k_1}(\boldsymbol{z}_1,\mathbf{y}_{M:1}=\boldsymbol{\beta}_{M:1})=0,\quad\forall(\boldsymbol{x},\boldsymbol{z}_r,\ldots,\boldsymbol{z}_1)\in\mathcal{X}\times\mathcal{Z},\forall\boldsymbol{\beta}_{M:\ell}\in B(\boldsymbol{\beta}_{M:\ell}^*,\epsilon'_\ell);\quad(64)$$

由于（b1），可以从上述方程中提取项 $w_{k_\ell}(\boldsymbol{z}_\ell,\ \ldots\ldots,\ \boldsymbol{z}_1,\ \mathbf{y}_{M:\ell}=\boldsymbol{\beta}_{M:\ell})$。鉴于我们此前已证明，对于每个 $\boldsymbol{\beta}\in\mathcal{Y}'$，$\{v_j(\boldsymbol{z},\ \mathbf{y}\ =\boldsymbol{\beta})|j\in J_0\}$ 线性相关，我们定义 $\mathcal{Y}_\ell:=(B(\boldsymbol{\beta}_{M:\ell}^*,\ \epsilon'_\ell)\times\pi_{\{\ell-1,\ldots,1\}}(\mathcal{Y}'))\setminus\mathcal{Y}_0$。根据公式（64），对于



(a) $\ell' = 2$ (b) $\ell' = 3$

图17：Il ustrations of subsets 交集非零的示意图，用于构造 $\mathcal{Y}_{\ell'}$ 与 $M = 3,\ \ell = 1$。

每个 $\boldsymbol{\beta}_{M:1} \in \mathcal{Y}_\ell$，该族

$$\Big\{ w_{k_r}(\boldsymbol{z}_r, \ldots, \boldsymbol{z}_1, \mathbf{y}_M = \boldsymbol{\beta}_M) \ldots w_{k_{\ell+1}}(\boldsymbol{z}_{\ell+1}, \ldots, \boldsymbol{z}_1, \mathbf{y}_{M:\ell+1} = \boldsymbol{\beta}_{M:\ell+1}) \cdot w_{k_{\ell-1}}(\boldsymbol{z}_{\ell-1}, \ldots, \boldsymbol{z}_1, \mathbf{y}_{M:\ell-1} = \boldsymbol{\beta}_{M:\ell-1}) \ldots w_{k_1}(\boldsymbol{z}_1, \mathbf{y}_{M:1} = \boldsymbol{\beta}_{M:1}) | , (k_r, \ldots, k_{\ell+1}, k_{\ell-1}, \ldots, k_1) \in \{k_r, \ldots, k_{\ell+1}, k_{\ell-1}, \ldots, k_1) | \exists (i, k_r, \ldots, k_1) \in S_0 \} \Big\}$$

包含线性相关函数。上述族是由 $r-1$ 个函数组成的乘积，这些函数都属于同一个族 $\mathcal{W}_K$，且该族满足 (b1)、(b4)、(b5) 和 (b6)。因此，我们可以再次使用引理20。于是我们知道，对每个 $\boldsymbol{\beta}_{M:1} \in \mathcal{Y}_\ell$，至少存在一个 $\ell' \leq r$，使得 $\{w_{k_{\ell'}}(\boldsymbol{z}_{\ell'}, \ldots\ldots, \boldsymbol{z}_1, \mathbf{y}_{M:\ell'} = \boldsymbol{\beta}_{M:\ell'}) | k'_\ell \in K^{(\ell')}{}_0 \}$，其中 $K^{(\ell')}{}_0 = \{k_{\ell'} \in K | \exists (i, k_r, \ldots\ldots, k_1) \in S_0\}$，在有限混合下包含线性相关函数。

与前面一样，我们按照公式（58）构造分割，并选择

$$\boldsymbol{\beta}^*_{M:\ell'} \in \arg \min_{\boldsymbol{\beta}_{M:\ell'} \in \pi_{\{M,\ldots,\ell'\}}(\mathcal{Y})} C^{(\ell')}(\boldsymbol{\beta}_{M:\ell'}),$$

并构造 $B(\boldsymbol{\beta}^*_{M:\ell'},\ \epsilon'_{\ell'}) \subset \pi_{\{M,\ldots,\ell'\}}(\mathcal{Y}_\ell)$，其中 $\epsilon'_\ell > \epsilon'_{\ell'} > 0$，使得根据 (b4)、(b6) 和公式（58），相应的划分满足 $|K_1^{(\ell')}(\boldsymbol{\beta}_{M:\ell'})| = 1$ 对所有 $\boldsymbol{\beta}_{M:\ell'} \in B(\boldsymbol{\beta}^*_{M:\ell'},\ \epsilon'_{\ell'})$ 都成立（我们再次假设 $|K_1^{(\ell')}(\boldsymbol{\beta}^*_{M:\ell'})| = C^{(\ell')}(\boldsymbol{\beta}^*_{M:\ell'})$，不失一般性）。因此，对于所有 $\boldsymbol{\beta}_{M:\ell'} \in B(\boldsymbol{\beta}^*_{M:\ell'},\ \epsilon'_{\ell'})$，我们可以从公式（64）中提取出因子 $\{w_{k_{\ell'}}(\boldsymbol{z}_{\ell'}, \ldots\ldots, \boldsymbol{z}_1, \mathbf{y}_{M:\ell'} = \boldsymbol{\beta}_{M:\ell'}) | k_{\ell'} \in K^{(\ell')}{}_0 \}$。

现在，我们构造一个新的非零可行 $\boldsymbol{\beta}$值集合，使得可以从乘积族 $\{v_j(\cdot)\}$ 中提取第 ℓ' 个因子。我们还需要确保新集合 $\mathcal{Y}_{\ell'}$ 是 $\mathcal{Y}_\ell$的子集，从而使所有先前提取的项仍然有效。定义

$$\mathcal{Y}_{\ell'} := \mathcal{Y}_\ell \cap \Big(B(\boldsymbol{\beta}^*_{M:\ell'}, \epsilon'_{\ell'}) \times \pi_{\{\ell'-1,\ldots,1\}}(\mathcal{Y}_\ell) \Big), \tag{65}$$

其中，根据构造，$\mathcal{Y}_{\ell'} \subseteq \mathcal{Y}_\ell$。该集合的测度也非零，因为 $B(\boldsymbol{\beta}^*_{M:\ell'},\ \epsilon'_{\ell'}) \subset \pi_{\{M,\ldots,\ell'\}}(\mathcal{Y}_\ell)$，而与$\pi_{\{\ell'-1,\ldots,1\}}(\mathcal{Y}_\ell)$的乘积会得到一个非零测度柱体，其中至少有一个非零测度子集位于 $\mathcal{Y}_\ell$ 中，因为该柱体的底面位于$\pi_{\{M,\ldots,\ell'\}}(\mathcal{Y}_\ell)$中。图17展示了 $M = 3$ 和 $\ell = 3$的这一情形，其中红色球体表示$\mathcal{Y}_3$。我们观察到，$B(\boldsymbol{\beta}^*_{M:\ell'}$ 与投影$\pi_{\{\ell'-1,\ldots,1\}}(\mathcal{Y}_\ell)$的乘积会得到一个柱体（若 $\ell' = 1$，则为一个立方体），且它们与 $\mathcal{Y}_\ell$ 的交集具有非零测度。在

这个例子中，$\boldsymbol{\beta}_{M:\ell}$ 在 $\ell \in [M]$ 上对齐：也就是说，对于 $\ell \neq \ell'$ 和 $\boldsymbol{\beta}^{*(\ell)}_{M\,\ell:}$，$\boldsymbol{\beta}^{*(\ell')}_{M:\,\ell'}$，有$\boldsymbol{\beta}^{*(\ell)}_{M\,:\,\ell'} = \boldsymbol{\beta}^{*(\ell')}_{M\,:\,\ell'}$（我们稍后会证明这种对齐）。然而，非零测度交集并不依赖于对齐：由 $B(\boldsymbol{\beta}^*_{M:\ell'}, \epsilon'_{\ell'}) \subset \pi_{\{M,\ldots,\ell'\}}(\mathcal{Y}_\ell)$即可得出。

因此，对于每个 $\boldsymbol{\beta} \in \mathcal{Y}_{\ell'}$，在有限混合$\{v_j(\boldsymbol{z},\ \mathbf{y} = \boldsymbol{\beta})|j \in J_0\}$下，该族仍然线性相关，其中乘积 v_j 中的两项各自只有一个元素（w_{k_ℓ} 和$w_{k_{\ell'}}$）。对所有剩余索引重复这一过程，我们得到一个嵌套序列：$\mathcal{Y}_{final} \subseteq \cdots \subseteq \mathcal{Y}_{\ell'} \subseteq \mathcal{Y}_\ell$其中，对于每个 $\boldsymbol{\beta} \in \mathcal{Y}_{final}$，$\{v_j(\boldsymbol{z},\ \mathbf{y} = \boldsymbol{\beta})|j \in J_0\}$中的每个因子都已缩减为单个元素。于是，公式（64）化简为

$$\sum_{i:(i,k_r,\ldots,k_1)\in S_0} \gamma_{i,k_r,\ldots,k_1} u_i(\mathbf{y} = \boldsymbol{\beta}, \boldsymbol{x}) = 0, \quad \forall (\boldsymbol{x}, \mathbf{y}) \in \mathcal{X} \times \mathcal{Y}_{final}. \tag{66}$$

由于对所有 $\boldsymbol{\beta}_{M:\ell} \in B(\boldsymbol{\beta}^*_{M:\ell},\ \epsilon)$，$|K_1^{(\ell)}(\boldsymbol{\beta}_{M:\ell})| \leq |K_1^{(\ell)}(\boldsymbol{\beta}^*_{M:\ell})|$，我们可以选择 $\boldsymbol{\beta}^*_{M:\ell}$，使重叠坐标一致。也就是说，对于任意 $\ell \neq \ell'$，不失一般性地令 $1 \leq \ell < \ell' \leq r$，我们可以得到两个 $\boldsymbol{\beta}^{*(\ell)}_{M\,\ell:}$，$\boldsymbol{\beta}^{*(\ell')}_{M:\,\ell'}$，使得 $\boldsymbol{\beta}^{*(\ell)}_{M\,:\,\ell'} = \boldsymbol{\beta}^{*(\ell')}_{M\,:\,\ell'}$。鉴于 $\boldsymbol{\beta}^*_{M:\ell}$ 已对齐，且$\epsilon'_\ell > 0$任意小，最终的嵌套域可取为 $\mathcal{Y}_{final} = B(\boldsymbol{\beta}^*,\ \epsilon'_{final})$，其中 $\min_{\ell\in[r]} \epsilon'_\ell > \epsilon'_{final} > 0$，且 $\boldsymbol{\beta}^* := \boldsymbol{\beta}^*_{M:1}$。这就得到一个非零测度集，在该集合上，族 $\{u_i(\cdot)\}$ 线性相关，这与 (a3) 矛盾。因此，得出矛盾。 ■

A.2.4 非参数轨迹族的线性无关

在 LIPO-M 中，我们针对一般的每变量维度 d 给出结果，使得 $\mathbf{y} \in \mathbb{R}^{dM}$ 和$\boldsymbol{z} \in \mathbb{R}^{dr}$ 满足 $1 \leq r \leq M$。在非参数轨迹设定中，我们令 $d = m$，因为$\boldsymbol{z}_t \in \mathbb{R}^m$。下面我们给出非参数轨迹族在有限混合下的线性无关性。

定理 22 定义如下轨迹族，对应轨迹 $(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T) \in \mathbb{R}^{Tm}$：

$$\mathcal{P}^{T,M}_{\mathcal{A},\mathcal{B}} = \Bigg\{ p_{a,b_{M+1:T}}(\boldsymbol{z}_{1:T}) = p_a(\boldsymbol{z}_{1:M}) \prod_{t=M+1}^{T} p_{b_t}(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}), \\ p_a \in \Pi^M_{\mathcal{A}}, \quad p_{b_t} \in \mathcal{P}^M_{\mathcal{B}}, \quad t = M+1, ..., T \Bigg\}, \tag{67}$$

并假设 $\Pi^M_{\mathcal{A}}$ 和 $\mathcal{P}^M_{\mathcal{B}}$满足以下假设，

(c1) $\Pi^M_{\mathcal{A}}$ *satisfies (a1), (a3), and*

(c2) $\mathcal{P}^M_{\mathcal{B}}$ 满足 *(b4-b6)*。

那么，以下结论成立：对于任意满足 $r = M$ 的 $T > M$，若 $(T \bmod M) = 0$，或 $r = (T \bmod M)$，否则；以及任意子集 $\mathcal{Z}_r \subset (\times^r \mathbb{R}^m)$，对于 $(\boldsymbol{z}_{1:T-r}, \boldsymbol{z}_{T-r+1:T}) \in \mathbb{R}^{(T-r)m} \times \mathcal{Z}_r$，轨迹族在有限混合下包含线性无关分布。

证明 我们首先通过以下归纳法，证明对于任意满足 $T \bmod M = 0$ 的 $T > M$，该结论成立。

$\underline{T = 2M}$：通过在证明中进行如下设定，可以利用引理 21 证明该结果：

$$u_i(\mathbf{y} = \boldsymbol{z}_{1:M}, \boldsymbol{x} = \boldsymbol{z}_0) = \pi_a(\boldsymbol{z}_{1:M}), i = a,$$

and

$$v_j(\boldsymbol{z} = \boldsymbol{z}_{M+1:2M}, \mathbf{y} = \boldsymbol{z}_{1:M}) = \prod_{\ell=1}^{M} p_{b_{M+\ell}}(\boldsymbol{z}_{M+\ell} | \boldsymbol{z}_{\ell:M-1+\ell}), j = b_{M+1:2M},$$

其中，j是元组索引，$\boldsymbol{z}_{\ell:M-1+\ell} := (\boldsymbol{z}_\ell, \boldsymbol{z}_{\ell+1}, \ldots\ldots, \boldsymbol{z}_{M-1+\ell})$ 是长度为 M 的滑动窗口。我们注意到，直接利用假设 (c1-c2) 即可证明该引理，因为 $v_j(\boldsymbol{z} = \boldsymbol{z}_{M+1:2M}, \mathbf{y} = \boldsymbol{z}_{1:M})$ 是由 M个函数的乘积构成的，并且满足 LIPO-M 中的 (b)（引理 21）。

$\underline{T = rM, \ r \in \mathbb{Z}^+, \ r > 2}$：假设当$T = \tau - M$时，该陈述对轨迹族成立。注意，我们可以将 $p_{a, \ b_{M+1:\tau}}(\boldsymbol{z}_{1:\tau})$ 写成

$$p_{a,b_{M+1:\tau}}(\boldsymbol{z}_{1:\tau}) = p_{a,b_{M+1:\tau-M}}(\boldsymbol{z}_{1:\tau-M}) \prod_{\ell=1}^{M} p_{b_{\tau-M+\ell}}(\boldsymbol{z}_{\tau-M+\ell} | \boldsymbol{z}_{\tau-2M+\ell:\tau-M-1+\ell}). \tag{68}$$

然后，通过设定以下内容，可以利用 LIPO-M（引理 21）证明关于 $T = \tau$ 的陈述

$$u_i(\mathbf{y} = \boldsymbol{z}_{\tau-2M+1:\tau-M}, \boldsymbol{x} = \boldsymbol{z}_{1:\tau-2M}) = p_{a,b_{M+1:\tau-M}}(\boldsymbol{z}_{1:\tau-M}), i = a, b_{M+1:\tau-M},$$

and

$$v_j(\boldsymbol{z} = \boldsymbol{z}_{\tau-M+1:\tau}, \mathbf{y} = \boldsymbol{z}_{\tau-2M+1:\tau-M}) = \prod_{\ell=1}^{M} p_{b_{\tau-M+\ell}}(\boldsymbol{z}_{\tau-M+\ell} | \boldsymbol{z}_{\tau-2M+\ell:\tau-M-1+\ell}), j = b_{\tau-M+1:\tau}. \tag{69}$$

注意，现在 i 和 j 都表示元组索引。由 $p_{a,b_{M+1:\tau-M}}(\boldsymbol{z}_{1:\tau-M})$ $i = a, b_{M+1:\tau-M}$ 张成的族根据 (c1) 满足 (a1)；根据归纳假设，该族在有限混合下线性无关，这正是 (a3)。式 (69) 中的族是由 M 个函数构成的乘积，其中 (b1) 和 (b4-b6) 根据 (c2) 成立，而这意味着引理 21 中的 (b) 成立。

根据上述结果，我们接下来证明 $T \bmod M \neq 0$ 的情形。

$\underline{T \bmod M \neq 0}$：令 r 为 T/M 除以 $M = r$ 的余数：$T \bmod M = r$。根据前面的结果，我们知道该结论对 $T - r$ 成立，即当 T 是 M 的倍数时成立。我们可以将 $p_{a,b_{M+1:T}}(\boldsymbol{z}_{1:T})$ 写成如下形式

$$p_{a,b_{M+1:T}}(\boldsymbol{z}_{1:T}) = p_{a,b_{M+1:T-r}}(\boldsymbol{z}_{1:T-r}) \prod_{\ell=1}^{r} p_{b_{T-r+\ell}}(\boldsymbol{z}_{T-r+\ell} | \boldsymbol{z}_{T-r+\ell-M:T-r+\ell-1}). \tag{70}$$

我们再次可以通过引理 21 证明该结论，只需令

$$u_i(\mathbf{y} = \boldsymbol{z}_{T-r-M+1:T-r}, \boldsymbol{x} = \boldsymbol{z}_{1:T-r-M}) = p_{a,b_{M+1:T-r}}(\boldsymbol{z}_{1:T-r}), i = a, b_{M+1:T-r},$$

and

$$v_j(\boldsymbol{z} = \boldsymbol{z}_{T-r+1:T}, \mathbf{y} = \boldsymbol{z}_{T-r-M+1:T-r}) = \prod_{\ell=1}^{r} p_{b_{T-r+\ell}}(\boldsymbol{z}_{T-r+\ell}|\boldsymbol{z}_{T-r+\ell-M:T-r+\ell-1}), j = b_{T-r+1:T}. \quad (71)$$

由 $p_{a,b_{M+1:T-r}}(\boldsymbol{z}_{1:T-r})$ $i = a, b_{M+1:T-r}$张成的族满足 (c1) 中的 (a1)。(a3) 的满足方式如下:

- 对于 $M < T < 2M$，我们有 $T - r = M$，且 (a3) 可直接由 (c1) 得出。
- 对于 $T > 2M$，我们有 $(T-r)$ 模 $M = 0$，且 (a3) 可由 $T - r$ 处的归纳假设得出。

公式（71）中的族是由 r 个函数族的乘积构成的，根据 (c2)，这些函数族满足 (b1) 和 (b4-b6)。这意味着引理 21 中的 (b) 成立。

我们证明了，对于 T 的所有可能情形，LIPO-M 中的条件均得到满足，因此根据 LIPO-M，所得轨迹族在有限混合下线性无关。

■

A.3 线性　　满足假设（m1）和（m2）时有限混合下的线性无关性

回顾假设（m1）和（m2）：

（m1） $\mathcal{G}_{\mathcal{B}}^{M}$和 $\mathcal{I}_{\mathcal{A}}^{M}$的唯一索引：公式（15）和（17）成立；

（m2） $\mathcal{G}_{\mathcal{B}}^{M}$, $\boldsymbol{m}(\cdot, b)$: $\mathbb{R}^{mM} \to \mathbb{R}^{m}$ 和 $\boldsymbol{\Sigma}(\cdot, b)$: $\mathbb{R}^{mM} \to \mathbb{R}^{m\times m}$中的均值和协方差，对于任意 $b \in \mathcal{B}$，均为解析函数。

下面我们探讨这些参数条件是否能确保轨迹族在有限混合下线性无关（从而确保可识别性）。如下所示，高斯转移族的唯一索引假设（公式（15））意味着线性无关性。

命题 23 $\mathcal{G}_{\mathcal{B}}$ 中的函数在变量$(z_t, \ldots, z_{t-M})$ 上满足有限混合下的线性无关性，前提是唯一索引假设（公式（*15*））成立。

证明 假设该陈述不成立，且不失一般性， $M = 1$。那么存在 $\mathcal{B}_0 \subset \mathcal{B}$，满足$|\mathcal{B}_0| < +\infty$，以及一个由非零值组成的集合 $\{\gamma_b | b \in \mathcal{B}_0\}$，使得

$$\sum_{b\in\mathcal{B}_0} \gamma_b \mathcal{N}(\boldsymbol{z}_t; \boldsymbol{m}(\boldsymbol{z}_{t-1}, b), \boldsymbol{\Sigma}(\boldsymbol{z}_{t-1}, b)) = 0, \quad \forall \boldsymbol{z}_t, \boldsymbol{z}_{t-1} \in \mathbb{R}^m.$$

特别地，对任意 $\boldsymbol{z}_{t-1} \in \mathbb{R}^m$，该等式都成立，这意味着高斯分布（定义在 $\boldsymbol{z}_t$ 上）的加权和等于零。请注意，Yakowitz and Spragins (1968) 证明了，均值和/或协方差不同的多元高斯分布线性无关。因此，上述等式意味着对任意 $\boldsymbol{z}_{t-1}$

$$\boldsymbol{m}(\boldsymbol{z}_{t-1}, b) = \boldsymbol{m}(\boldsymbol{z}_{t-1}, b') \quad \text{and} \quad \boldsymbol{\Sigma}(\boldsymbol{z}_{t-1}, b) = \boldsymbol{\Sigma}(\boldsymbol{z}_{t-1}, b') \quad \forall b, b' \in \mathcal{B}_0, b \neq b',$$

这与唯一索引假设相矛盾。 ■

方程 (14) 和 (16) 中分别定义的转移高斯分布族和初始高斯分布族，分别满足 LIPO-M 中的假设 (a) 和 (b)，如下所示。

命题24 在唯一索引假设（方程 *(15)*）下，条件高斯分布族 $\mathcal{G}_{\mathcal{B}}$ （方程 *(14)*）满足引理*19*、*20*和*21*中的假设 *(b1)*、*(b5)* 和 *(b6)*，若定义 $\mathcal{V}_J := \mathcal{G}_{\mathcal{B}}, \boldsymbol{z} := \boldsymbol{z}_t$ 以及 $(\mathbf{y}_M, \ldots, \mathbf{y}_1) := (\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M})$。

命题25 在唯一索引假设（方程 *(17)*）下，初始高斯分布族 $\mathcal{I}_{\mathcal{A}}$ （方程 *(16)*）满足引理*21*中的假设 *(a1)*、*(a3)*，若定义$\mathcal{U}_I := \mathcal{I}_{\mathcal{A}}, \mathbf{y} := (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_M)$ 以及 $\boldsymbol{x} = \emptyset$。

为了说明高斯转移族 $\mathcal{G}_{\mathcal{B}}$ 满足 (b5)，由命题 23 可知，只有唯一索引假设不成立时才会出现线性相关。因此，我们可以固定 $(\boldsymbol{z}_{t-1}、\ldots\ldots、\boldsymbol{z}_{t-M})$ 变量的一个子集，使所得的均值函数和协方差函数违反唯一索引假设，这意味着所得函数族线性相关。为验证 (b4)，我们需要以下矩的零测交集结果。

命题26 假设高斯族转移 $\mathcal{G}_{\mathcal{B}}$ 满足由式 *(15)* 定义的唯一索引，且矩的交集为零测集：对于任意 $b、b' \in \mathcal{B}$，若 $b \neq b'$，则集合

$$\mathcal{X}_{b,b'} := \Big\{(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}) \in \mathbb{R}^{mM} \mid \boldsymbol{m}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, b) = \boldsymbol{m}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, b'), \\ \boldsymbol{\Sigma}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, b) = \boldsymbol{\Sigma}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, b')\Big\}$$

的测度为零。因此，当 $V_J := \mathcal{G}_{\mathcal{B}}、\ \boldsymbol{z} := \boldsymbol{z}_t$，且 $(\mathbf{y}_1、\ldots\ldots、\mathbf{y}_M) := (\boldsymbol{z}_{t-1}、\ldots\ldots、\boldsymbol{z}_{t-M})$ 时，*(b4)* 成立。

证明 我们通过反证法证明该结论。根据 (b4)，定义 $\mathcal{V}_J := \mathcal{G}_{\mathcal{B}}、\ \boldsymbol{z} := \boldsymbol{z}_t$，以及 $(\mathbf{y}_1、\ldots\ldots、\mathbf{y}_M) := (\boldsymbol{z}_{t-1}、\ldots\ldots、\boldsymbol{z}_{t-M})$；不失一般性，令 $\mathcal{Y} := \mathbb{R}^{mM}$。若 (b4) 不成立，则存在非零测度集 $\mathcal{Y}' \subset \mathcal{Y}$ 和 $\mathcal{Z} \subset \mathbb{R}^m$，使得 $\mathcal{G}_{\mathcal{B}}$ 包含线性相关函数。根据命题 23，这意味着存在一个指标集 $\mathcal{B}_0 \subset \mathcal{B}$，满足 $|\mathcal{B}_0| < +\infty$，使得 $\mathcal{Y}'$ 中的唯一索引假设（式 (15)）不成立。由矩交集假设可知，该集合为$\mathcal{Y}' = \bigcup_{b \neq b' \in \mathcal{B}_0} \mathcal{X}_{b、b'}$，其测度为零。这与 (b4) 中要求 $\mathcal{Y}' \subset \mathcal{Y}$ 为非零测度集相矛盾。

■

最后，我们介绍参数情形下轨迹族在有限混合下的线性无关性。

定理 **m 27** 定义非线性高斯模式下的以下轨迹族 l

$$\Big\{p_{a,b_{M+1:T}}(\boldsymbol{z}_{1:T}) = p_a(\boldsymbol{z}_{1:M}) \prod_{t=M+1}^{T} p_{b_t}(\boldsymbol{z}_t | \boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}), \\ p_a \in \mathcal{I}_{\mathcal{A}}^M, \quad p_{b_t} \in \mathcal{G}_{\mathcal{B}}^M, \quad t = M+1, ..., T\Big\}, \quad (72)$$

其中$\mathcal{G}_{\mathcal{B}}^{M}$、$\mathcal{I}_{\mathcal{A}}^{M}$分别由公式（*14*）、（*16*）定义。假设：

(d1) $\mathcal{G}_{\mathcal{B}}^{M}$和$\mathcal{I}_{\mathcal{A}}^{M}$的唯一索引：方程 *(15)*、*(17)* 成立；

(d2) $\mathcal{G}_{\mathcal{B}}$中的函数关于（$\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{M}$） $\in \mathbb{R}^{mM}$连续；

(d3) 矩的零测交集：对于任意满足 $b \neq b'$的 b、 $b' \in \mathcal{B}$，集合

$$\mathcal{X}_{b,b'} := \Big\{(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}) \in \mathbb{R}^{mM} \mid \boldsymbol{m}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, b) = \boldsymbol{m}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, b'), \\ \boldsymbol{\Sigma}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, b) = \boldsymbol{\Sigma}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, b')\Big\}$$

具有零测度。

那么，对于（$\boldsymbol{z}_{1:T-r}, \boldsymbol{z}_{T-r+1}, \ldots, \boldsymbol{z}_{T}$） $\in \mathbb{R}^{(T-r)m} \times$ （$\times^{r}\mathbb{R}^{m}$），轨迹族在有限混合下包含线性无关的分布；其中，如果 T模 $M = 0$，则 $r = M$，否则为 $r = T$模 M 。

证明 注意到（d2）与（b6）等价。由于命题24、25，假设（a1）、（a3）、（b1）和（b5）均成立。根据假设（d3），通过命题26可知假设（b4）成立。于是，由定理 22 可得该结论成立。

■

A.4 结论

下面我们完成定理 8 的证明。

证明 我们首先说明定理 27 中假设（m1-m2）与（d1-d3）之间的对应关系。（d1）与（m1）等价。由于解析函数是$\mathcal{C}^{\infty}$，因而连续，因此（m2）满足（d2）。根据 Mityagin（2015），我们知道解析函数的零点集是零测集。因此，（m2）意味着，对于任意 b，$b' \in \mathcal{B}$，与 $b \neq b'$ 对应的均值或协方差之间的交集也都是零测集，这意味着（d3）。因此，根据定理 27，高斯转移族 $\mathcal{G}_{\mathcal{B}}$下的轨迹族在有限混合下线性无关。根据定理15，轨迹族在有限混合下线性无关意味着 MSM 在定义7意义下可识别。

■

附录B。SDS可识别性的证明

我们将 Kivva 等（2022）和 Balsells-Rodas 等人（2024）的结果扩展到以多滞后 MSM 族作为 $\boldsymbol{z}_{1:T}$ 的先验分布。该策略要求找到某个开集，使得两个同分布 SDS 的变换 F 和 G 在该开集上可逆。随后，借助 F 和 G 的分段线性性质，以及多滞后 MSM 是解析分布且在仿射变换下封闭这一特性，可以证明可识别性。

B.1 预备知识

我们需要先介绍证明中将用到的一些定义和结果。这些内容此前已由 Kivva 等（2022）定义。

定义 28 设 $D_0 \subseteq D \subseteq \mathbb{R}^n$ 是开集。设 $f_0\colon\ D_0 \to \mathbb{R}$。如果对于每个 $\boldsymbol{x} \in D_0$，都有 $f(\boldsymbol{x}) = f_0(\boldsymbol{x})$，则称解析函数 $f\colon\ D \to \mathbb{R}$ 是 f_0 在 D 上的解析延拓。

定义 29 设 $\boldsymbol{x}_0 \in \mathbb{R}^m$ 和 $\delta > 0$。设 $p\colon\ B(\boldsymbol{x}_0,\ \delta) \to \mathbb{R}$。定义

$$Ext(p) : \mathbb{R}^m \to \mathbb{R}$$

如果存在这样的延拓，则将其定义为 p 在整个空间 $\mathbb{R}^m$ 上的唯一解析延拓，否则定义为 *0*。

定义 30 设 $D_0 \subset D$ 和 $p\colon\ D \to \mathbb{R}$ 是一个函数。我们将 $p|_{D_0}\colon\ D \to \mathbb{R}$ 定义为 p 在 D_0 上的限制，即满足对于每个 $\boldsymbol{x} \in D_0$，都有 $p|_{D_0}(\boldsymbol{x}) = p(\boldsymbol{x})$ 的函数。

定义 31 设 $f\colon\ \mathbb{R}^m \to \mathbb{R}^n$ 是一个分段仿射函数。若某点$\boldsymbol{x} \in f(\mathbb{R}^m) \subseteq \mathbb{R}^n$满足原像 $f^{-1}(\{\boldsymbol{x}\})$ 为有限集，且存在 $\delta > 0$，使得对于每个 $\boldsymbol{z} \in f^{-1}(\{\boldsymbol{x}\})$，$f\colon B(\boldsymbol{z}, \delta) \to \mathbb{R}^n$ 都是仿射函数，则称该点相对于 f 是一般点。

我们还需要证明，在假设 (m1-m2) 下，可以利用多滞后 MSM 先验计算分段因式分解映射F 的原像中的点；随后将用此结果证明，存在一个非零测度集，使得 F 和 G 均可逆。我们沿用 Balsells-Rodas 等人（2024）的引理 D.6，并注意到加入额外滞后不会改变该陈述。我们还将 Balsells-Rodas 等人（2024）的推论 D.7 和定理 D.8 推广到多滞后马尔可夫切换模型，其证明需要作一些细微的澄清。

引理 32（改编自 Balsells-Rodas 等人（2024）） 设随机变量 $\boldsymbol{z}_{1:T}$ 服从多滞后 *MSM* 族 $\mathcal{M}^{T,\ M}_{N\ L}$ 并满足假设（*m1-m2*）。考虑 $f\colon\ \mathbb{R}^m \to \mathbb{R}^m$，这是一个生成随机变量 $\boldsymbol{x}_{1:T} = F(\boldsymbol{z}_{1:T})$ 的分段仿射映射，其形式为 $\boldsymbol{x}_t = f(\boldsymbol{z}_t)\ 1 \leq t \leq T$。此外，考虑相对于 f 的一般点 $\boldsymbol{x}^{(0)} \in \mathbb{R}^m$。那么，$\boldsymbol{x}_1^{(0)}\colon_T = \{\boldsymbol{x}^{(0)}, \ldots, \boldsymbol{x}^{(0)}\} \in \mathbb{R}^{Tm}$ 也是相对于 F 的一般点，且原像 $F^{-1}(\{\boldsymbol{x}_1^{(0)}\colon_T\})$ 中的点数可计算如下

$$\left|F^{-1}\left(\left\{\boldsymbol{x}_{1:T}^{(0)}\right\}\right)\right| = \lim_{\delta \to 0} \int_{\boldsymbol{x}_{1:T} \in \mathbb{R}^{Tm}} Ext\left(p|_{B(\boldsymbol{x}_{1:T}^{(0)}, \delta)}\right) d\boldsymbol{x}_{1:T}$$

证明 如果 $\boldsymbol{x}^{(0)} \in \mathbb{R}^m$ 相对于 f是一个一般点，那么 $\boldsymbol{x}_{1\ T}^{(0)}\colon$ 相对于 F也同样是一个一般点，因为原像 $F^{-1}(\{\boldsymbol{x}_1^{(0)}\colon_T\})$ 现在更大，但仍然是有限的。换言之，$F^{-1}(\{\boldsymbol{x}_1^{(0)}\colon_T\})$ 是笛卡尔积 $\times^T \mathcal{Z}$，其中 $\mathcal{Z} = \{\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_\ell\}$ 是原像 $f^{-1}(\{\boldsymbol{x}^{(0)}\})$ 中的点。鉴于此，我们定义良好的仿射映射 $G_{i_1,\ldots,i_T}$：

$B(\boldsymbol{z}_{i_1},\ldots,\boldsymbol{z}_{i_T},\epsilon) \to \mathbb{R}^m$，$i_t \in [\ell]$ 对于 $1 \le t \le T$，使得 $G_{i_1,\ldots,i_T} = F(\boldsymbol{z}_{1:T})\ \forall \boldsymbol{z}_{1:T} \in B(\boldsymbol{z}_{i_1},\ldots,\boldsymbol{z}_{i_T},\epsilon)$。该仿射映射 $G_{i_1,\ldots,i_T}$ 可分解为：$g_{i_t}(\boldsymbol{z}_t) = f(\boldsymbol{z}_t)\ \ \forall \boldsymbol{z}_t \in B(\boldsymbol{z}_i,\epsilon)$，

$$G_{i_1,\ldots,i_T} = \begin{pmatrix} A_{i_1} & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & A_{i_T} \end{pmatrix} \begin{pmatrix} \boldsymbol{z}_1 \\ \vdots \\ \boldsymbol{z}_T \end{pmatrix} + \begin{pmatrix} b_{i_1} \\ \vdots \\ b_{i_T} \end{pmatrix},$$

具有可逆的 A_{i_1}、.....、A_{i_T}。令 $\delta_0 > 0$ 使得

$$B(\boldsymbol{x}_{1:T}^{(0)},\delta_0) \subseteq \bigcap_{i_1,\ldots,i_T}^{\ell} G_{i_1,\ldots,i_T}\big(B(\boldsymbol{z}_{i_1},\ldots,\boldsymbol{z}_{i_T},\epsilon)\big).$$

我们可以使用 Balsells-Rodas 等人（2024）的命题 C.2，计算每个 $\boldsymbol{x}_{1:T} \in B(\boldsymbol{x}_{1\,T:}^{(0)}、\delta')$ 在 $0 < \delta' < \delta_0$ 下的似然；该命题表明，MSM 在因式分解仿射变换下是封闭的。不失一般性，我们可以假设 $M = 1$，因为对于滞后阶数 $M > 1$，MSM 在分解仿射变换下仍然是封闭的，

$$\begin{aligned} p|_{B(\boldsymbol{x}_{1:T}^{(0)},\delta)} = \sum_{i_1,\ldots,i_T}^{\ell} \sum_{j_1,\ldots,j_T}^{K} & p\left(s_{1:T} = \{j_1,\ldots,j_T\}\right) \mathcal{N}\left(\boldsymbol{x}_1; A_{i_1}\boldsymbol{\mu}(j_1) + \boldsymbol{b}_{i_1}, A_{i_1}\boldsymbol{\Sigma}_1(j_1)A_{i_1}^T\right) \\ & \prod_{t=2}^{T} \mathcal{N}\Big(\boldsymbol{x}_t; A_{i_t}\boldsymbol{m}\left(A_{i_{t-1}}^{-1}\left(\boldsymbol{x}_{t-1} - \boldsymbol{b}_{i_{t-1}}\right), j_t\right) + \boldsymbol{b}_{i_t}, A_{i_t}\boldsymbol{\Sigma}\Big(A_{i_{t-1}}^{-1}\left(\boldsymbol{x}_{t-1} - \boldsymbol{b}_{i_{t-1}}\right), j_t\Big)A_{i_t}^T\Big), \end{aligned}$$

其中我们使用 $\boldsymbol{x}'_{t-1} = A_{i_{t-1}}^{-1}\left(\boldsymbol{x}_{t-1} - \boldsymbol{b}_{i_{t-1}}\right)$；对于更高阶滞后 $M > 1$，在上述方程中 $\boldsymbol{x}'_{t-1},\ldots,\boldsymbol{x}'_{t-M}) = \text{Ä}A_{i_{t-1}}^{-1}\left(\boldsymbol{x}_{t-1} - \boldsymbol{b}_{i_{t-1}}\right),\ldots,A_{i_{t-M}}^{-1}\left(\boldsymbol{x}_{t-M} - \boldsymbol{b}_{i_{t-M}}\right)$ä 作用于 $\boldsymbol{m}(\cdot)$ 和 $\boldsymbol{\Sigma}(\cdot)$ 的参数。前述密度是一个解析函数，它定义在 $\boldsymbol{x}_{1\,T:}^{(0)}$ 的一个开邻域上。于是，根据解析函数恒等定理，所得密度在 $\mathbb{R}^{Tm}$ 上定义了 $p|_{B(\boldsymbol{x}_{1\,T:}^{(0)},\delta)}$ 的解析延拓。于是，我们有

$$\begin{aligned} & \int_{\boldsymbol{x}_{1:T}\in\mathbb{R}^{Tm}} \text{Ext}\left(p|_{B(\boldsymbol{x}_{1:T}^{(0)},\delta)}\right) d\boldsymbol{x}_{1:T} \\ & = \int_{\boldsymbol{x}_{1:T}\in\mathbb{R}^{Tm}} \sum_{i_1,\ldots,i_T}^{\ell} \sum_{j_1,\ldots,j_T}^{K} p\left(s_{1:T} = \{j_1,\ldots,j_T\}\right) \mathcal{N}\left(\boldsymbol{x}_1; A_{i_1}\boldsymbol{\mu}(j_1) + \boldsymbol{b}_{i_1}, A_{i_1}\boldsymbol{\Sigma}_1(j_1)A_{i_1}^T\right) \\ & \quad \prod_{t=2}^{T} \mathcal{N}\Big(\boldsymbol{x}_t; A_{i_t}\boldsymbol{m}\left(\boldsymbol{x}'_{t-1}, j_t\right) + \boldsymbol{b}_{i_t}, A_{i_t}\boldsymbol{\Sigma}\left(\boldsymbol{x}'_{t-1}, j_t\right)A_{i_t}^T\Big)\, d\boldsymbol{x}_{1:T} \\ & = \sum_{i_1,\ldots,i_T}^{\ell} \int_{\boldsymbol{x}_{1:T}\in\mathbb{R}^{Tm}} \sum_{j_1,\ldots,j_T}^{K} p\left(s_{1:T} = \{j_1,\ldots,j_T\}\right) \mathcal{N}\left(\boldsymbol{x}_1; A_{i_1}\boldsymbol{\mu}(j_1) + \boldsymbol{b}_{i_1}, A_{i_1}\boldsymbol{\Sigma}_1(j_1)A_{i_1}^T\right) \\ & \quad \prod_{t=2}^{T} \mathcal{N}\Big(\boldsymbol{x}_t; A_{i_t}\boldsymbol{m}\left(\boldsymbol{x}'_{t-1}, j_t\right) + \boldsymbol{b}_{i_t}, A_{i_t}\boldsymbol{\Sigma}\left(\boldsymbol{x}'_{t-1}, j_t\right)A_{i_t}^T\Big)\, d\boldsymbol{x}_{1:T} \\ & = \sum_{i_1,\ldots,i_T}^{\ell} 1 = \ell^T = \left|F^{-1}(\{\boldsymbol{x}_{1:T}^{(0)}\})\right|. \end{aligned}$$

■

如前所述，这导致存在一个开集合，使得 F 和 G 映射都是可逆的。

推论 *33*（改编自 *Balsells-Rodas* 等人（*2024*年））设 $F, G: \mathbb{R}^{Tm} \to \mathbb{R}^{Tn}$ 是分段仿射映射，其中 $\boldsymbol{x}_t := f(\boldsymbol{z}_t)$ 且 $\boldsymbol{x}'_t := g(\boldsymbol{z}'_t)$，对于 $1 \leq t \leq T$。假设 f 和 g 弱单射（定义 *5*）。设 $\boldsymbol{z}_{1:T}$ 和 $\boldsymbol{z}'_{1:T}$ 的分布服从可识别多滞后 *MSM* 族 $\mathcal{M}^T_{NL}$，并满足假设 *(m1-m2)*。假设 $F(\boldsymbol{z}_{1:T})$ 和 $G(\boldsymbol{z}'_{1:T})$ 同分布。假设对于 $\boldsymbol{x}_0 \in \mathbb{R}^n$ 和 $\delta > 0$，f 在 $B(\boldsymbol{x}_0, 2\delta) \cap f(\mathbb{R}^m)$ 上可逆。

那么，对于 $\boldsymbol{x}^{(0)}_{1:T} = \{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_0\} \in \mathbb{R}^{Tn}$，存在 $\boldsymbol{x}^{(1)}_{1:T} \in B(\boldsymbol{x}^{(0)}_{1:T}, \delta)$ 以及 $\delta_1 > 0$，使得 F 和 G 在 $B(\boldsymbol{x}^{(1)}_{1:T}, \delta_1) \cap F(\mathbb{R}^{Tm})$ 上均可逆。

证明 与 Balsells-Rodas 等人（2024）的证明相同，只需将其中的引理 D.6 替换为引理 32（多滞后版本）。该引理表明，在假设 (m1-m2) 下，跟随多滞后 MSM 的点，其 G 的原像中的点可借助 $\boldsymbol{z}'_{1:T}$ 计算，因为 G 的原像中几乎每个点都是一般点，从而 F 和 G 在一个非零测度集合上均可逆。■

定理 *34*（改编自 *Balsells-Rodas* 等人（*2024*））设 F、G：$\mathbb{R}^{Tm} \to \mathbb{R}^{Tn}$ 为因式分解分段仿射映射，其中 $\boldsymbol{x}_t := f(\boldsymbol{z}_t)$，$\boldsymbol{x}'_t := g(\boldsymbol{z}'_t)$，且 $1 \leq t \leq T$。设 $\boldsymbol{z}_{1:T}$ 和 $\boldsymbol{z}'_{1:T}$ 在假设 *(m1-m2)* 下服从可识别多滞后 *MSM* 族 $\mathcal{M}^T_{NL}$。假设 $F(\boldsymbol{z}_{1:T})$ 和 $G(\boldsymbol{z}'_{1:T})$ 同分布，并且存在 $\boldsymbol{x}^{(0)}_{1:T} \in \mathbb{R}^{Tn}$ 和 $\delta > 0$，使得 F 和 G 在 $B(\boldsymbol{x}^{(0)}_{1:T}, \delta) \cap F(\mathbb{R}^{Tm})$ 上可逆。那么存在一个可逆的因式分解仿射变换 H，使得 $H(\boldsymbol{z}_{1:T}) = \boldsymbol{z}'_{1:T}$。

证明 由于多滞后 MSM 在仿射变换下封闭（Balsells-Rodas 等人，2024，命题 C.2），且在假设（m1-m2）下，多滞后 MSM 族 $\mathcal{M}^T_{NL}$ 的分布是解析函数，因此证明与 Balsells-Rodas 等人（2024）的证明完全相同。■

现在，我们已具备证明定理 12 (i) 所需的全部要素。

B.2 定理 12 (i) 的证明

证明 该结论可由 Balsells-Rodas 等人（2024）推出。假设存在另一个模型，该模型从式（21）生成相同的分布，具有先验分布 $p' \in \mathcal{M}^T_{NL}$ 并满足假设（m1-m2），且其非线性发射 F'由 f' 组成，其中该函数是弱单射且分段线性的：即 $(F\#p)(\boldsymbol{x}_{1:T}) = (F'\#p')(\boldsymbol{x}_{1:T})$。

由弱单射性可知，至少对于某些 $\boldsymbol{x}_0 \in \mathbb{R}^n$ 和 $\delta > 0$，f 在$B(\boldsymbol{x}_0, 2\delta) \cap f(\mathbb{R}^m)$上可逆。这满足推论 33 的前提条件，因此可推出存在 $\boldsymbol{x}^{(1)}_{1:T} \in B(\boldsymbol{x}^{(0)}_{1:T}, \delta)$ 和 $\delta_1 > 0$，使得 F和 F'在$B(\boldsymbol{x}^{(1)}_{1:T}, \delta_1) \cap F(\mathbb{R}^{Tm})$上均可逆。因此，根据定理 34，存在一个仿射变换 H，使得 $H(\boldsymbol{z}_{1:T}) = \boldsymbol{z}'_{1:T}$，也就是说，$p \in \mathcal{M}^T_{NL}$ 在仿射变换下可识别。■

B.3 定理 12 (ii) 的证明

证明 这可由定理 12 (i) 和 Kivva等（2022）的定理 E.1 得出；该定理适用于任意 $t \in \{M+1, \ldots, T\}$，因为 $\boldsymbol{z}_t|\boldsymbol{z}_{t-1}, \ldots \boldsymbol{z}_{t-M}$是一个高斯混合模型。注意，Kivva等（2022）已经涵盖了 $t \in \{1, \ldots, M\}$，因为它可归约为初始分布情形。由定理 12 (i)，我们得到在仿射变换下的可识别性：即，存在某个 $\boldsymbol{z}'_{1:T}$，使得 $\boldsymbol{z}'_t = A\boldsymbol{z}_t + \boldsymbol{b}$，$t \in \{1, \ldots, T\}$。给定 (s1) 和 (s2)，对于任意 $t \in \{M+1, \ldots, T\}$，Kivva等（2022）中定理 E.1 的条件在 $(\boldsymbol{z}_{t-1}, \ldots \boldsymbol{z}_{t-M})$ 的空间中的一个非零测度集上得到满足。因此，我们可以找到 $\tilde{A}$，使得 $\tilde{\boldsymbol{z}}_t = \tilde{A}^{-1}\boldsymbol{z}'_t = \tilde{P}\tilde{D}\boldsymbol{z}_t + \tilde{\boldsymbol{b}}$，其中 $\tilde{P}$ 和 $\tilde{D}$ 分别是置换和缩放矩阵。 ■

附录 C. 推论 14 的证明

我们首先给出时序图结构中节点置换的定义。

定义 35 设 $\mathbf{G} = (V, E_{1:M})$ 为一个图。根据第 *4.3* 节，$\mathbf{G} \in \mathbb{G}$ 以 $V := [m]$为节点，并包含最大滞后阶数为 M 的时序边信息。对 $\mathbf{G}$ 进行节点置换 $\pi \in S_m$ 后，得到置换图 $\mathbf{G} = \pi(\mathbf{G})$。该图的顶点集合保持不变，但对于每个 $1 \le \tau \le M$，边集合经过置换，如下所示：

$$\tilde{\mathbf{G}} = (\tilde{V}, \tilde{E}_{1:M}), \quad \tilde{V} = \{\pi(1), \ldots, \pi(m)\}, \quad \tilde{E}_\tau = \left\{ \big(\pi(i), \pi(j)\big) \mid (i, j) \in E_\tau \right\} \tag{73}$$

证明 假设有两个 SDS $p(\boldsymbol{x}_{1:T})$ $\tilde{p}(\boldsymbol{x}_{1:T})$，其对应的状态依赖图为$\tilde{\mathbf{G}}_{1:K}$、$\mathbf{G}_{1:\tilde{K}}$。

根据假设 (m1) 和 (m2*)，由定理 12 可知，我们可以恢复 $p(\boldsymbol{z}_{1:T})$，但结果仅在置换和缩放意义下唯一。结合定理 8，根据定义7，我们可知其在置换意义下可识别，这意味着 $K = K$。我们重点关注条件 4，其中对于每个 $1 \le i \le C$，都存在某个 $1 \le j \le C$，使得：

$$p_{b_t^i}(\boldsymbol{z}_t|\boldsymbol{z}_{t-1:t-M}) = \tilde{p}_{\tilde{b}_t^j}(\boldsymbol{z}'_t|\boldsymbol{z}'_{t-1:t-M}), \quad \forall (\boldsymbol{z}'_t, \ldots, \boldsymbol{z}'_{t-M}) \in \mathbb{R}^{m(M+1)}. \tag{74}$$

于是，在每个时间步 t，都存在一个置换 σ，使得对于每个 $k \in [K]$，$p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1:t-M}, s_t = k) = \tilde{p}(\boldsymbol{z}'_t|\boldsymbol{z}'_{t-1:t-M}, s_t = \sigma(k))$ $\forall (\boldsymbol{z}_t, \ldots\ldots, \boldsymbol{z}_{t-M}) \in \mathbb{R}^{m(M+1)}$。由条件 2 可知，对于 $t > M$，σ 是常数。由于我们知道这些分布是高斯分布，根据高斯可识别性（Yakowitz 和 Spragins，1968 年），我们有

$$\boldsymbol{m}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, k) = PD\tilde{\boldsymbol{m}}(\boldsymbol{z}'_{t-1}, \ldots, \boldsymbol{z}'_{t-M}, \sigma(k)) + \boldsymbol{b} \quad k \in [K], \tag{75}$$

其中每个 $\boldsymbol{z}'_t = D^{-1}P^T\boldsymbol{z}_t + \tilde{\boldsymbol{b}}$，由于 MSM 在仿射变换下封闭（Balsells- Rodas 等，2024 年，命题 C.2）。现在，鉴于函数 $\boldsymbol{m}(\cdot, k)$ $k \in [K]$ 是解析的，雅可比矩阵将保持类似的置换等价关系。不失一般性，我们固定 $\tau \in [M]$，并且$k \in [K]$，将关于 $\boldsymbol{z}_{t-\tau}$ 的转移的雅可比矩阵记为 $J_{\boldsymbol{m}(\cdot,k),\ \tau}$。我们计算公式 (75) 中关于 $\boldsymbol{z}_{t-\tau}$ 的导数：

$$\frac{\partial \boldsymbol{m}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, k)}{\partial \boldsymbol{z}_{t-\tau}} = PD\frac{\partial \tilde{\boldsymbol{m}}(\boldsymbol{z}'_{t-1}, \ldots, \boldsymbol{z}'_{t-M}, \sigma(k))}{\partial \boldsymbol{z}'_{t-\tau}}\frac{\partial \boldsymbol{z}'_{t-\tau}}{\partial \boldsymbol{z}_{t-\tau}}, \tag{76}$$

$$J_{\boldsymbol{m}(\cdot,k),\tau} = PDJ_{\tilde{\boldsymbol{m}}(\cdot,\sigma(k)),\tau}D^{-1}P^T. \tag{77}$$

根据极小性假设 (m3)，我们可以直接将 $J_{\boldsymbol{m}(\cdot,k),\,\tau}$ 和 $J_{\tilde{\boldsymbol{m}}(\cdot,\sigma(k)),\,\tau}$ 映射为某个邻接矩阵，方法是将绝对值大于 0 的元素进行阈值化。鉴于D是对角的，$DJ_{\tilde{\boldsymbol{m}}(\cdot,\,\sigma(k))\,\tau}D^{-1}$ 的结构得以保留。由此得到的图邻接矩阵集合 $\tilde{A}_{k,\,\tau}$，$A_{k,\,\tau}$ 对于状态 k 和滞后阶数 τ 的关系如下

$$A_{k,\tau} = P\tilde{A}_{\sigma(k),\tau}P^T, \tag{78}$$

其中，置换矩阵 P 与某个置换 $\pi \in S_m$相关联，并表示对 $A_{k,\,\tau}$ 的行和列进行置换，其方式遵循 π。因此，上述操作~等价于定义 35 中定义的节点置换，且对于每个 $k \in [K]$，我们有 $\mathbf{G}_k = \pi(\mathbf{G}^{\sigma(k)})$。

■

附录 D. 估计细节

公式 (25) 中给出的证据下界推导如下：

$$\begin{aligned}
\log p_{\boldsymbol{\theta}}(\boldsymbol{x}_{1:T}) &= \log \int \sum_{\boldsymbol{s}_{M:T}} p_{\boldsymbol{\theta}}(\boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T}, \boldsymbol{s}_{M:T}) d\boldsymbol{z}_{1:T} \\
&\geq \mathbb{E}_{q_{\boldsymbol{\phi},\boldsymbol{\theta}}(\boldsymbol{z}_{1:T}, \boldsymbol{s}_{M:T}|\boldsymbol{x}_{1:T})}\left[\log \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}_{1:T}, \boldsymbol{z}_{1:T}|\boldsymbol{s}_{M:T})p_{\boldsymbol{\theta}}(\boldsymbol{s}_{M:T})}{q_{\boldsymbol{\phi}}(\boldsymbol{z}_{1:T}, \boldsymbol{s}_{M:T}|\boldsymbol{x}_{1:T})}\right] \\
&\geq \mathbb{E}_{q_{\boldsymbol{\phi}}(\boldsymbol{z}_{1:T}|\boldsymbol{x}_{1:T})}\left[\log \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}_{1:T}|\boldsymbol{z}_{1:T})}{q_{\boldsymbol{\phi}}(\boldsymbol{z}_{1:T}|\boldsymbol{x}_{1:T})} + \mathbb{E}_{p_{\boldsymbol{\theta}}(\boldsymbol{s}_{M:T}|\boldsymbol{z}_{1:T})}\left[\log \frac{p_{\boldsymbol{\theta}}(\boldsymbol{z}_{1:T}|\boldsymbol{s}_{M:T})p_{\boldsymbol{\theta}}(\boldsymbol{s}_{M:T})}{p_{\boldsymbol{\theta}}(\boldsymbol{s}_{M:T}|\boldsymbol{z}_{1:T})}\right]\right] \\
&\geq \mathbb{E}_{q_{\boldsymbol{\phi}}(\boldsymbol{z}_{1:T}|\boldsymbol{x}_{1:T})}\left[\log \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}_{1:T}|\boldsymbol{z}_{1:T})}{q_{\boldsymbol{\phi}}(\boldsymbol{z}_{1:T}|\boldsymbol{x}_{1:T})} + \log p_{\boldsymbol{\theta}}(\boldsymbol{z}_{1:T})\right] \\
&\geq \mathbb{E}_{q_{\boldsymbol{\phi}}(\boldsymbol{z}_{1:T}|\boldsymbol{x}_{1:T})}\Big[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}_{1:T}|\boldsymbol{z}_{1:T})\Big] - KL\Big(q_{\boldsymbol{\phi}}(\boldsymbol{z}_{1:T}|\boldsymbol{x}_{1:T})||p_{\boldsymbol{\theta}}(\boldsymbol{z}_{1:T})\Big).
\end{aligned}$$

给定样本 $\tilde{\boldsymbol{z}} \sim q_{\boldsymbol{\phi}}(\boldsymbol{z}_{1:T}|\boldsymbol{x}_{1:T})$，公式 (29) 中使用的后验分布 $\gamma_{t,\,k}(\tilde{\boldsymbol{z}}_{1:T})$ 和 $\xi_{t,\,k,\,k'}(\tilde{\boldsymbol{z}}_{1:T})$ 可利用前向和后向消息计算（Bishop，2006年）：$\{\alpha_{t,k}(\tilde{\boldsymbol{z}}_{1:t}) = p_{\boldsymbol{\theta}}(\tilde{\boldsymbol{z}}_{1:t},\ s_t = k)\}$，以及 $\{\beta_{t,k}(\tilde{\boldsymbol{z}}_{t+1:T}) = p_{\boldsymbol{\theta}}(\tilde{\boldsymbol{z}}_{t+1:T}|s_t = k)\}$，分别为：

$$\gamma_{t,k}(\tilde{\boldsymbol{z}}_{1:T}) = \frac{\alpha_{t,k}(\tilde{\boldsymbol{z}}_{1:t})\beta_{t,k}(\tilde{\boldsymbol{z}}_{t+1:T})}{\sum_{k=1}^{K}\alpha_{t,k}(\tilde{\boldsymbol{z}}_{1:t})\beta_{t,k}(\tilde{\boldsymbol{z}}_{t+1:T})}, \tag{79}$$

$$\xi_{t,k,k'}(\tilde{\boldsymbol{z}}_{1:T}) = \frac{\alpha_{t-1,k'}(\tilde{\boldsymbol{z}}_{1:t-1})\beta_{t,k}(\tilde{\boldsymbol{z}}_{t+1:T})p(\tilde{\boldsymbol{z}}_t|\tilde{\boldsymbol{z}}_{t-1},\ldots,\tilde{\boldsymbol{z}}_{t-M}, s_t = k)p(s_t = k|s_{t-1} = k')}{\sum_{k=1}^{K}\alpha_{t,k}(\tilde{\boldsymbol{z}}_{1:t})\beta_{t,k}(\tilde{\boldsymbol{z}}_{t+1:T})} \tag{80}$$

其中 $\alpha_{t,\,k}(\tilde{\boldsymbol{z}}_{1:t})$ 和 $\beta_{t,\,k}(\tilde{\boldsymbol{z}}_{t+1:T})$ 的计算方式如下：

$$\alpha_{t,k}(\tilde{\boldsymbol{z}}_{1:t}) = p_{\boldsymbol{\theta}}(\tilde{\boldsymbol{z}}_t|\tilde{\boldsymbol{z}}_{t-1:t-M}, s_t = k) \times \sum_{k'=1}^{K} p_{\boldsymbol{\theta}}(s_t = k|s_{t-1} = k')\alpha_{t-1,k'}(\tilde{\boldsymbol{z}}_{1:t-1}), \tag{81}$$

$$\beta_{t,k}(\tilde{\boldsymbol{z}}_{t+1:T}) = \sum_{k'=1}^{K} p_{\boldsymbol{\theta}}(\tilde{\boldsymbol{z}}_{t+1}|\tilde{\boldsymbol{z}}_{t:t+1-M}, s_{t+1} = k') \times p_{\boldsymbol{\theta}}(s_{t+1} = k'|s_t = k)\beta_{t+1,k'}(\tilde{\boldsymbol{z}}_{t+1:T}). \tag{82}$$

附录 E. 实验细节

复现我们实验的代码可在此处找到：https://github.com/ charlio23/identifiable-SDS。

E.1 合成数据生成

我们提供数据生成过程的更多细节。初始高斯分量从 $\mathcal{N}(0, 0。7_2\mathbf{I})$ 中采样。对于均值转移函数 $\boldsymbol{m}(\cdot, k)$，我们已说明使用局部连接网络（Zheng 等人，2018），其中网络依赖关系遵循机制依赖的因果结构。这些网络由具有 16 个隐藏单元和余弦激活函数的两层多层感知器组成。协方差矩阵 $\boldsymbol{\Sigma}(\cdot, k)$ 根据具体设置设计如下：

- 恒定噪声：固定为 $\boldsymbol{\Sigma} = 0.01\mathbf{I}$。
- 异质噪声：$\boldsymbol{\Sigma}(k) = \text{diag}(\boldsymbol{\sigma}_k^2)$，其中对于每个 $k \in [K]$ 和 $i \in [m]$, $\sigma_{k,\ i}^2 \sim U(0。005, 0。08)$，使得假设（s2）成立。也就是说，我们检查是否至少存在两个状态 k_1, $k_2 \in [K]$，它们在各维度上具有不同的对角线比值，否则就重新采样。
- 历史依赖噪声：$\boldsymbol{\Sigma}(\cdot, k) = \text{diag}(C_k \cdot \text{Sigmoid 函数}(\boldsymbol{m}(\cdot, k)))$，其中 $C_k \sim U(0.05, 0.1)$。

对于 iSDS，我们进一步使用具有 8 个隐藏单元的两层带泄漏的ReLU 网络生成观测。关于消融实验，我们考察的假设在均值转移中使用以下网络：

- 重叠：使用余弦激活函数的分段解析函数；当条件轨迹在时间 $t(\boldsymbol{z}_{t-1}, ..., \boldsymbol{z}_{t-M})$处的 L_2 范数介于 3 和 5 之间时，我们强制各状态使用相同的函数；以及
- 零：使用余弦激活函数的随机网络。

E.2 训练设置

实验使用 PyTorch（Paszke 等人，2019）实现，并在 NVIDIA RTX 2080 Ti GPU 上运行；气候数据上的 SDS 实验则使用 NVIDIA GeForce RTX 3090 GPU。所有实验中，我们均使用 Adam 优化器（Kingma 和 Ba，2015），iMSM 的学习率为 $7 \cdot 10^{-3}$，iSDS 的学习率为 $5 \cdot 10^{-4}$。其他具体配置取决于数据集，详见下文。与相关方法类似（Hälvä 和 Hyvarinen, 2020），我们通过随机重启获得更好的参数估计：iMSM 初始化进行 3 次随机重启（见下文），完整模型进行 5 次。

训练 iSDS 时，我们通常遵循四阶段流程。

- **初始化（iMSM）：** 我们先通过主成分分析对观测进行降维，然后在降维后的观测上训练 iMSM，以此初始化潜在动力学先验。训练 iMSM 时，我们使用 100 的批量大小，将学习率设为 $7 \cdot 10^{-3}$；当似然趋于平稳时，将学习率降低至原来的 0.5 倍，最多降低两次。我们进行 3 次随机重启，并选择训练似然最佳的模型来初始化 iSDS。

- **编码器-解码器预训练：**我们通过冻结动力学先验的梯度，为编码器和解码器找到一个初始值。
- **预热阶段：** **</style}iSDS 采用联合训练，但我们将切换参数，** Q保持冻结。这有助于我们防止状态坍塌。
- **最终阶段：** 所有网络均经过微调。

相关工作（Dong 等人，2020年；Ansari 等人，2021年）引入退火，以强制状态后验分配均匀，并确保模型在推断时使用所有切换。与此不同，我们发现，结合 iMSM 初始化和随机重启可以防止状态坍塌。下面我们将介绍各数据集的具体配置和网络架构选择。

合成数据。 训练时，我们使用批量大小为 100 的批次，按上述方法初始化潜在 iMSM，对编码器和解码器进行 40 个时间段的预训练，并额外进行 10 个时间段的预热。随后，我们解除所有梯度的冻结，并以 $5 \cdot 10^{-4}$ 的学习率训练 700 个时间段，每 200 个时间段将学习率衰减至原来的 0.8。转移均值和协方差采用与数据生成过程相同的架构。编码器和解码器均为双层多层感知机，包含 128 个隐藏单元，并采用带泄漏的ReLU激活函数。我们在损失函数公式（30）中使用 $\eta = 0.05$ 对雅可比转移进行正则化。

脑活动数据。 训练沿用 iSDS 的初始化阶段流程，状态转移由具有 32 个隐藏单元并采用 GELU 激活函数的两层多层感知机组成。

金融数据。 训练基本遵循合成数据流程，仅作少量修改；对数价格和对数收益率采用相同的设置。在初始化阶段，我们使用整个时间序列，以促进更好的状态学习；学习率设为 $7 \cdot 10^{-3}$，如果似然在 10 次迭代内停滞，则将其降低至原来的 0.5 倍。iMSM 训练在模型停滞 3 次后停止。其余阶段遵循合成数据中所述的流程，使用长度为 $T = 200$ 的子采样时间序列。我们训练 100k 次迭代，学习率为 $5 \cdot 10^{-4}$，并且每 40k 次迭代衰减至原来的 0.8 倍。预训练和预热阶段分别持续 2k 和 8k 次迭代。架构由编码器和解码器组成，二者均采用具有 256 个隐藏维度并使用带泄漏的ReLU 激活函数的两层多层感知机；均值转移采用具有 32 个隐藏维度并使用 GELU 激活函数的两层多层感知机，协方差则为依赖于状态的对角协方差。这里，我们在损失函数 Eq. (30) 中考虑 $\eta = 0$ 。

气候数据。 训练方案与金融时间序列相同：初始化时，使用包含 $T = 360$ 个时间步的完整时间序列训练 iMSM。之后，我们使用训练好的 iMSM 进行第7.3节所述的可解释性分析。为训练 iSDS 的其余部分，我们对长度为 $T = 100$ 的序列进行子采样。我们对编码器 - 解码器进行预训练，共迭代 250k 次，学习率为 $5 \cdot 10^{-4}$，并且每 100k 次迭代衰减至原来的 0.8。然后，我们对 iSDS 进行 4000 次迭代的预热，最后阶段运行 120k 次迭代，并且每 20k 次迭代衰减至原来的 0.8。网络架构方面，状态转移与金融时间序列中采用的相同，我们使用卷积神经网络进行推断和生成。推断时，我们使用一个 9 层卷积神经网络，通道数逐层增加，卷积核大小为 3，填充为 1，激活函数采用带泄漏的ReLU，并交替使用步幅 2 和 1。我们从 64 个通道开始，每两层将通道数加倍

，直到达到 512 个通道。然后，我们运行一个三层多层感知机，使用带泄漏的ReLU激活函数和 256 个隐藏维度。生成时，我们使用类似的网络：首先是一个双层多层感知机，使用带泄漏的ReLU激活函数和 256 个隐藏维度，并使用转置卷积代替卷积（配置与前述相同）。在此设置下，我们也考虑 $\eta = 0$ 。

E.3 评估指标

在第6节中，我们概述了各项指标的计算方法。此处我们进一步说明如何计算制度特定均值函数 $\boldsymbol{m}(\cdot, k)$ 的 L_2 距离和 R^2 分数，以及如何恢复潜在的制度依赖因果结构。

均值函数指标。令 $f, g: \mathbb{R}^m \to \mathbb{R}^m$ ，并定义平方 L_2 距离如下：

$$d_{L_2}^2(f,g) := \int_{\boldsymbol{x}\in\mathbb{R}^m} \left|\left|f(\boldsymbol{x}) - g(\boldsymbol{x})\right|\right|^2 d\boldsymbol{x} \approx \frac{1}{I}\sum_{i=1}^{I} \left|\left|f\left(\boldsymbol{x}^{(i)}\right) - g\left(\boldsymbol{x}^{(i)}\right)\right|\right|^2, \tag{83}$$

其中， $\boldsymbol{x}^{(i)}$ 表示来自留出数据的样本。由于模型在状态的置换以及连续潜变量的仿射变换（或置换和缩放）下可识别，我们首先通过选择置换 σ 来对齐各状态：该置换使基于 $p(\boldsymbol{s}_{M:T}|\boldsymbol{z}_{1:T})$ 的后验分配计算得到的机制估计 F_1 分数最大。请回忆，在仿射变换下，多滞后 MSM 保持以下均值函数之间的等价性：

$$\boldsymbol{m}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M}, k) = A\hat{\boldsymbol{m}}\Big(A^{-1}(\boldsymbol{z}_{t-1} - \boldsymbol{b}), \ldots, A^{-1}(\boldsymbol{z}_{t-M} - \boldsymbol{b}), \sigma(k)\Big) + \boldsymbol{b}, \tag{84}$$

其中， A, $\boldsymbol{b}$ 表示仿射变换参数。为在恰当的状态域内评估函数，我们将留出轨迹划分为 K 个集合 $\{\mathcal{Z}_k\}_{k\in[K]}$，其中每个集合 $\mathcal{Z}_k$ 包含满足以下条件的变量：对于每个 $M < t \leq T$, $\{\boldsymbol{z}_{t-1}^{(i)}, \ldots, \boldsymbol{z}_{t-M}^{(i)}\} \in \mathcal{Z}_k$，若 $k = \arg\max\ p(s_t^{(i)}|\boldsymbol{z}_{1:T}^{(i)})$，其中索引 i 表示留出数据集中的一个样本。我们使用 1000 条留出轨迹。然后按如下方式计算误差：

$$L_2 \text{ err iSDS} := \min_{A,\boldsymbol{b}} \frac{1}{K}\sum_{k=1}^{K} d_{L_2}^2\Big(\boldsymbol{m}(\cdot, k), A\hat{\boldsymbol{m}}(\cdot, k) + \boldsymbol{b}\Big), \tag{85}$$

其中， $\boldsymbol{m}(\cdot,\ k)$ 的输入取自 $\mathcal{Z}_k$，而 $\hat{\boldsymbol{m}}(\cdot,\ k)$ 的输入则根据式（84）进行映射。

对于 R^2 分数，我们采用类似的流程，并计算经方差归一化的 L_2 差异：

$$d_{R^2}(f,g) := 1 - \frac{d_{L_2}^2(f,g)}{Var(f)}, \quad Var(f) = \frac{1}{I}\sum_i \Big|\Big|f(\boldsymbol{x}^{(i)}) - \frac{1}{I}\sum_j f(\boldsymbol{x}^{(j)})\Big|\Big|^2. \tag{86}$$

在消融实验中，由于引入了函数重叠，我们观察到较低的 F_1 状态得分。因此，对于 iMSM，我们选择使平均 R^2 最大化的状态置换：

$$R^2 \text{ err iMSM} := \max_{\sigma\in S_K} \frac{1}{K}\sum_{k=1}^{K} d_{R^2}\Big(\boldsymbol{m}(\cdot, k), \hat{\boldsymbol{m}}(\cdot, \sigma(k))\Big). \tag{87}$$

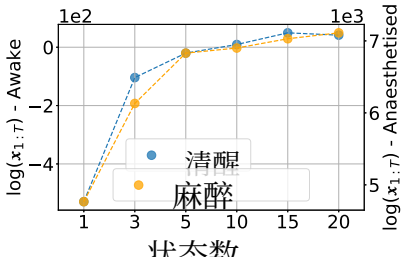


图18：清醒和麻醉状态下，使用不同状态数量和 $M = 2$ 滞后阶数时的测试对数似然。

对于 iSDS，我们采用相同的流程，但首先通过仿射映射对齐潜变量：

$$R^2 \text{ err iSDS} := \max_{\sigma \in S_K} \frac{1}{K} \sum_{k=1}^{K} d_{R^2}\Big(\boldsymbol{m}(\cdot, k), A\hat{\boldsymbol{m}}(\cdot, \sigma(k)) + \boldsymbol{b}\Big), \tag{88}$$

输入至 $\hat{\boldsymbol{m}}(\cdot)$ 的数据按式（84）进行重新缩放。此处我们在相同的 1000 条留出轨迹上进行评估；与 L_2 计算不同，在消融实验中，我们不会按状态对样本预先排序（因为 F_1 得分较低）。

因果结构恢复。令 $\boldsymbol{J}_{\hat{\boldsymbol{m}}(\cdot,k)}(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M})$ 表示 $\hat{\boldsymbol{m}}(\cdot, k)$ 在 $(\boldsymbol{z}_{t-1}, \ldots, \boldsymbol{z}_{t-M})$ 处的雅可比矩阵。我们通过对分配至状态 k 的样本的雅可比矩阵绝对值均值进行阈值化，估计依赖于状态的潜在因果图 $\mathbf{G}_{1:K}$。使用大小为 $|\mathcal{Z}_k| = N_k$ 的相同划分 $\mathcal{Z}_k$，我们定义：

$$\hat{\mathbf{G}}_k := \mathbf{1}\left(\frac{1}{N_k} \sum_{i=1}^{N_k} \left|\boldsymbol{J}_{\hat{\boldsymbol{m}}(\cdot,k)}\left(\boldsymbol{z}_{t-1}^{(i)}, \ldots, \boldsymbol{z}_{t-M}^{(i)}\right)\right| > \tau\right), \quad (\boldsymbol{z}_{t-1}^{(i)}, \ldots, \boldsymbol{z}_{t-M}^{(i)}) \in \mathcal{Z}_k, \tag{89}$$

其中，我们在实验中使用 $\tau = 0.05$，而 $\mathbf{1}(\cdot)$ 表示指示函数，当参数为真时其值为 1，否则为 0。对于滞后阶数 M，$\mathbf{G}_k$ 的形状为 $m \times (mM)$（行：当前变量；列：所有 M 个滞后阶数上的滞后变量）。根据使状态 F_1 分数最大的置换对状态进行对齐（如上所述）。对于潜变量对齐，我们将针对 L_2 距离拟合的仿射映射分解为 $A \approx DP$，其中 P 是置换矩阵，D 为对角矩阵。通过逐行选取绝对值最大的元素得到 P。对于每个滞后阶数 $\tau \in [M]$，其中 $E_{k,\tau}$ 表示 $\mathbf{G}_k$ 的滞后 -τ 邻接矩阵，我们使用 $P\hat{E}_{\sigma(k),\tau}P^T$ 进行置换，以实现潜变量对齐。最后，我们报告对齐后的估计与真实图 $\mathbf{G}_{1:K}$ 之间的平均 F_1 分数，并对各状态取平均。

E.4 脑活动数据

图18报告了随着状态数量增加 K，清醒和麻醉条件下的测试对数似然。随着 K 增大，性能有所提升，并在 $K \approx 10$ 左右趋于平稳。我们在主要实验中采用包含 $K = 15$ 的过完备设置，并预计模型在实际应用中只会使用其中一部分状态。如表1（第7.1节）所示，清醒条件下的状态总使用率为 53.33%，麻醉条件下为 33.33%，分别对应 8 个和 5 个活跃状态。允许使用更大的候选状态集合有助于防止状态坍塌，因为未使用的状态会保持不活跃，且测试似然不会下降。

表1中的指标。每个序列持续 2 秒，包含 $T = 400$ 个时间步。我们根据状态的平滑后验计算所有指标（长度为 3 的均匀核）。

- **状态转换频率（赫兹）：** 每秒状态变化次数，按时间段取平均。
- **总使用率：** 在至少一个测试周期中，占用率超过时间点总数 1% 的状态所占比例。

- **每个样本的使用率：**（跨时间段的）每周期占用率超过时间点总数 1% 的状态所占比例的均值。
- **最长持续时间（秒）：** 状态驻留时间的最大值，按时间段取平均。

E.5 金融数据

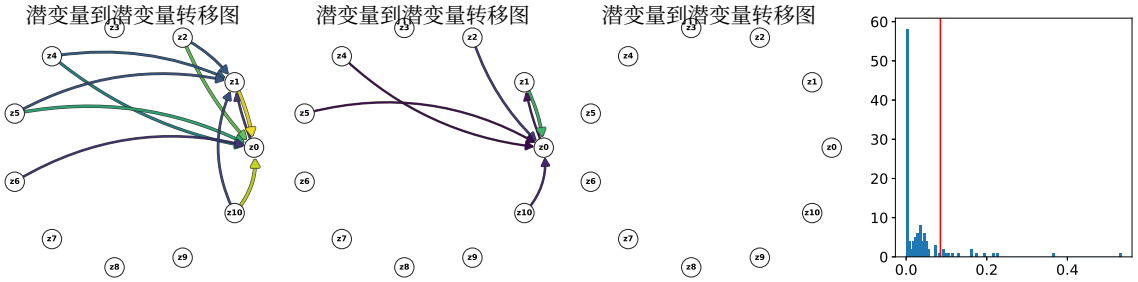
下面列出我们在实验中所考虑的按行业划分的100家美国公司的股票代码。

- **科技：** AAPL, MSFT, GOOGL, META, AMZN, NVDA, INTC, CSCO, IBM, ORCL
- **医疗保健：** JNJ, PFE, MRK, ABBV, BMY, LLY, AMGN, GILD, ILMN, UNH
- **金融：** JPM, BAC, WFC, C, GS, MS, USB, PNC, TFC, COF
- **必需消费品：** KO, PEP, PG, CL, KMB, MDLZ, GIS, K, WMT, COST
- **非必需消费品：** NKE, SBUX, MCD, DIS, BKNG, TJX, GAP, KMX, ROST, YUM

- **工业：** BA、CAT、MMM、GE、HON、LMT、RTX、UNP、CSX、DAL
- **能源：** XOM、CVX、COP、PSX、DVN、OXY、KMI、EOG、HES、VLO
- **房地产：** SPG、PLD、PSA、AVB、EQR、WELL、VTR、DLR、O、ESS
- **通信：** VZ、T、TMUS、CMCSA、CHTR、DIS、NFLX、PARA、EA、GOOGL
- **材料：** DD、SHW、LYB、NEM、FCX、NUE、AA、MLM、CRH、CE

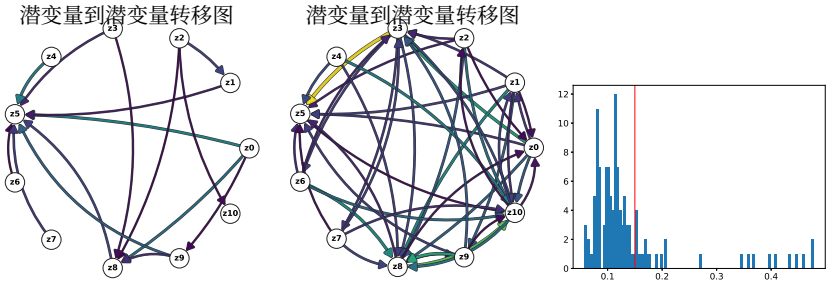
我们还研究了对数收益率在不同状态之间的潜在转移结构（图 19）以及边强度分布（图 19d）。直方图使用每个状态的雅可比矩阵非对角元素绝对值的平均值计算得出，并且集中在零附近。这表明潜在交互较弱。为便于解释，我们展示了强度较高的子集（红色阈值右侧）。在状态 0 和状态 1 中，z_0 和 z_1 主导潜在动力学（与全球市场因子一致）。此外，高波动状态（状态 0）的交互比中低波动状态（状态 2）更密集，这与动荡时期需要更强的动态调整相一致。

对于对数价格数据，图 20 展示了特定状态的潜在转移图及相应的边强度直方图（图 20c）。边分布以 0.1 附近为中心，并带有较长的强交互作用尾部。不同状态下的潜在动力学有所不同，其中横盘/熊市状态（状态 0）的动力学比牛市状态（状态 1）更简单、也更弱。两个全局因子（z_2 和 z_10）在不同状态下发挥的作用各不相同。值得注意的是，交互作用集中在 z_5 和 z_8上，它们与必需消费品、房地产和材料行业相对应。这表明这些行业需要相对



(a) 状态 0。 (b) 状态 1。 (c) 制度 2。 (d) 边直方图。

图 19 ：(a-c) 按状态划分的潜在转移图，以及 (d) 边强度（对数收益率 ）.



(a) 状态 0。 (b) 状态 1。 (c) 边直方图。

图 20：(a-b) 按状态划分的潜在转移图，以及 (c) 边强度（对数价格）。

复杂且因状态而异的调整（例如，必需消费品作为防御性行业；房地产作为利率敏感型行业）。

E.6 气候数据

模型选择。 iMSM 和 iSDS 的数据设置有所不同。对于 iMSM，我们遵循 Varando 等人（2022）的做法，将陆地像素投影到 10 个主成分上。对于 iSDS，我们在随机抽取的 64×64 陆地区块上进行训练，使用长度为 $T = 100$的子序列。我们固定$K = 4$ 个状态，并改变滞后阶数的数量 M。图 21 报告了留出集性能随 M 变化的结果：测试对数似然和证据下界（两种情形下数值越高越好）。为使量纲具有可比性，我们将两个指标都按观测进行报告（即 iMSM 中按每个主成分计算，iSDS 中按每个像素计算）。我们观察到，iMSM 在 $M = 1$ 时表现最佳，而 iSDS 受益于更高的阶数，其中 $M = 2$ 表现最佳。需要注意的是，iMSM 处理的是来自超高维数据（陆地像素）的全局混合

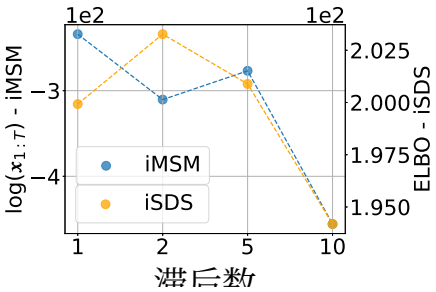
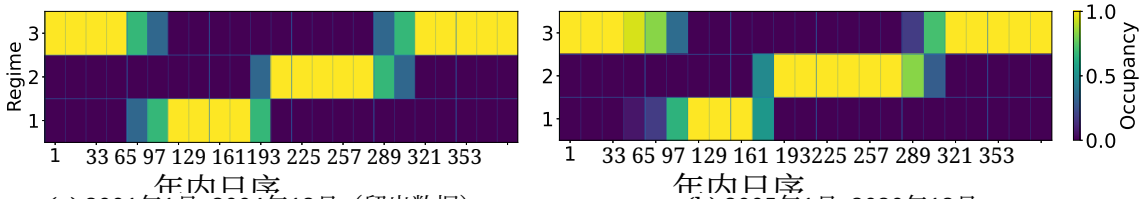
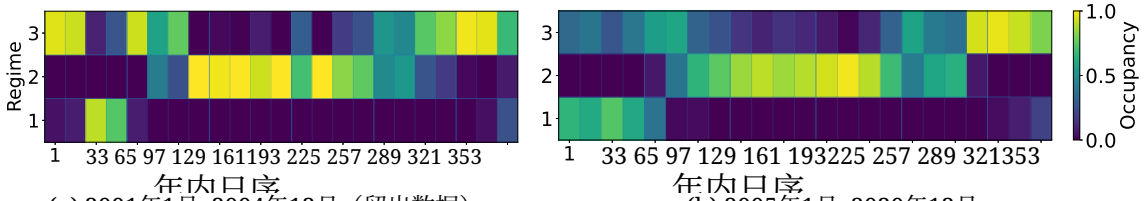


图 21：随着M增加，iMSM 和 iSDS 的测试对数似然（蓝色）与证据下界（橙色）。为便于比较，数值按观测大小报告（即 iMSM 为 10，iSDS 为 64×64）。



(a) 2001年1月–2004年12月（留出数据）。 (b) 2005年1月–2020年12月。

图22：使用 iMSM 根据陆地像素得到的基于年份的结构。



(a) 2001年1月–2004年12月（留出数据）。 (b) 2005年1月–2020年12月。

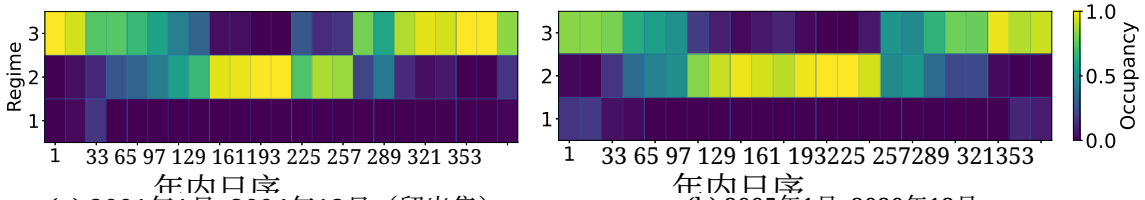
图23：iSDS揭示的埃塞俄比亚基于年份的结构。

信号，因此提高 M 可能会很快导致对未见序列的过拟合。相比之下，iSDS 使用共享状态转移对局部动态进行建模，而参数共享则对额外的滞后参数起到正则化作用。

状态日历。在第7.3节中，表2和表3分别报告了 iMSM 和 iSDS 按季节划分的 ENSO相关性。我们在图 22 中将 iMSM 的年内状态占比可视化。如前所述，其动力学按年份分为3个时段：3月至5月、6月至9月以及10月至次年2月。未观测年份也呈现出一致的结构。对于 iSDS，我们观察到5种因地而异的重复出现的年度模式。图23–25分别展示了埃塞俄比亚、萨赫勒和南非。我们省略了撒哈拉和刚果的日历，因为这些地区主要采用单一状态。各地区表3报告的月份均与可视化占比相符，而且这种一致性在未见年份中也得以保持。

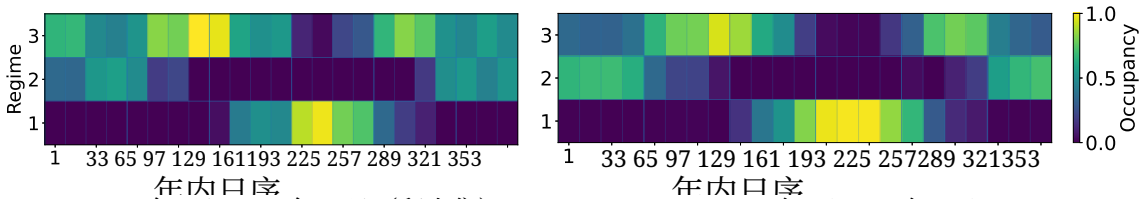
ENSO矩阵。此处我们详细说明表2和表3的计算方法。我们使用所有可用数据，并在0–240天网格上以16天步长评估与厄尔尼诺-南方涛动的滞后相关性（厄尔尼诺-南方涛动领先）。

- **iMSM：** 对于每个主成分，我们都会构建一个大小为 10×16(成分, 滞后阶数) 的相关矩阵，分别用于“所有时间”和根据 iMSM 状态后验概率得到的季节子集



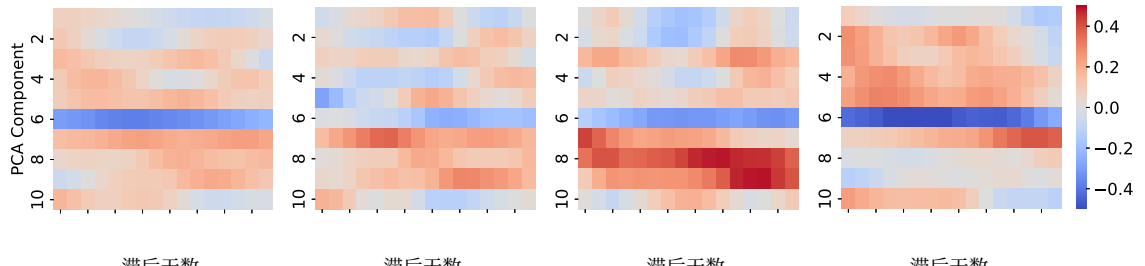
(a) 2001年1月–2004年12月（留出集）。 (b) 2005年1月–2020年12月。

图24：iSDS 揭示的萨赫勒基于年份的结构。



(a) 2001年1月–2004年12月（留出集）。 (b) 2005年1月–2020年12月。

图 25：iSDS 在南非揭示的基于年份的结构。

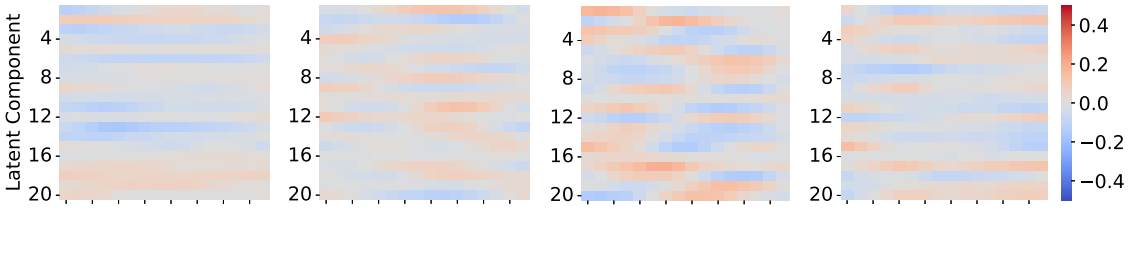


(a) 所有时间。 (b) 6月—9月。 (c) 3月—5月。 (d) 9月—2月。

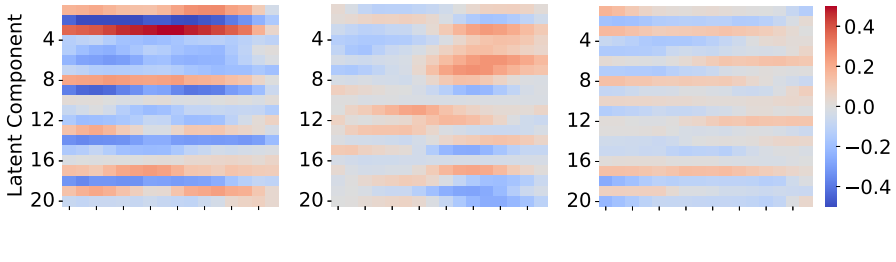
图 26：使用 (a) 所有可用数据，以及 (b-c) 按 iMSM 状态后验概率排序的基于季节的数据，计算陆地像素主成分与滞后厄尔尼诺-南方涛动（ENSO 领先）之间的相关矩阵。

。我们在图 26 中展示 iMSM 相关矩阵，表 2 则报告了每个切片的最大绝对相关性及其滞后天数。

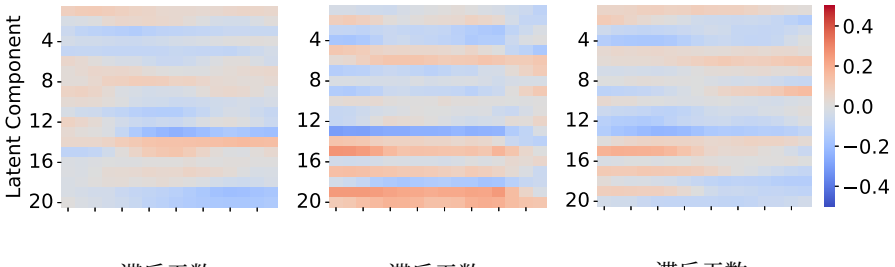
- **iSDS：**对于每个区域，我们抽取 10 个样本块，并根据 iSDS 状态后验概率对潜变量进行排序。然后，对于每个季节，我们计算大小为 20×16 (潜变量，滞后阶数) 的类似逐样本块相关矩阵，并使用 Fisher z 平均法聚合样本块（Fisher，1915年）。表 3 报告了每个区域和季节的峰值相关性及滞后天数，并在图27中展示区域相关矩阵。



(a) 撒哈拉，所有时间。 (b) 刚果，所有时间。 (c) 萨赫勒，3月至8月。 (d) 萨赫勒，9月至2月。



(e) 埃塞俄比亚，1月至3月。(f)埃塞俄比亚，4月至8月。 (g) 埃塞俄比亚，8月至12月。



(h)南非，6月至9月。(i)南非，12月至2月。(j) 南非，3月至6月和9月至11月。

图27：不同地点的 iSDS潜变量与滞后ENSO（ENSO 领先）之间的相关矩阵，分别对应 (a) 撒哈拉、(b) 刚果、(c-d) 萨赫勒、(e-g) 埃塞俄比亚和 (h-j) 南非。对于基于区域的相关矩阵，我们根据相应的状态后验分布对潜变量进行排序。