arXiv:2603.01468v1 [stat.ME] 2 Mar 2026

含随机性的非负矩阵分解的野生自助法推断

效应

Kenichi Satoh 滋
贺大学数据科学学院

作者说明

Kenichi Satoh https://orcid.org/0000-0003-4436-9347

有关本文的通信请联系 Kenichi Satoh，滋贺大学数据科学学院，地址：日本滋贺县彦根市番场 1-1-1，邮编 522-8522。电子邮箱：kenichi-satoh@biwako.shiga-u.ac.jp

作者声明不存在利益冲突。

摘要

非负矩阵分解（NMF）广泛用于基于成分的表示，但当基在非负性约束下学习时，协变量效应的正式推断却很少可用。我们提出带随机效应的非负矩阵分解（NMF-RE），这是一种均值结构潜变量模型 $Y = X(\Theta A + U) + \mathcal{E}$，将协变量驱动的得分与个体特异性偏差相结合。随机效应作为一种工作性工具，用于建模异质性并控制复杂度；我们监控其有效自由度，并强制设置基于自由度的上限，以防止近饱和拟合。估计过程交替进行闭式岭回归（类似最佳线性无偏预测的）更新，用于 U，以及乘法式非负$\widehat{\ }$ $\widehat{\ }$更新，用于 X 和 Θ。对于 Θ的推断，我们以（X，U）为条件，并通过渐近线性化、单步牛顿更新和乘子（wild）自助法实现快速不确定性量化；这避免了重复的约束再优化。模拟包括一项有针对性的压力测试，结果表明，如果不控制自由度，随机效应惩罚项可能会崩溃，而 Θ 的推断也会变得退化；相比之下，基于自由度的上限可以防止这种失效模式。非负约束会产生稀疏的、基于成分的载荷，即测量侧变量选择；而对 Θ 的推断则能识别哪些协变量影响哪些成分，从而实现协变量侧选择。纵向、心理测量、空间流和文本示例进一步展示了稳定且易于解释的协变量效应推断。

关键词：非负矩阵分解（NMF），随机效应，变量选择，有效自由度，正则化后推断，野生自助法

含随机的非负矩阵分解的野生自助法推断

效应

1 引言

非负矩阵分解（NMF）可为多元非负数据生成可解释的、基于部分的表示，目前已常规应用于文本分析、图像处理、主题建模和基因组学（Cemgil，2009年；D. Lee & Seung，2000年；D. D. Lee & Seung，1999年；Schmidt 等人，2009年）。除算法上的吸引力外，NMF 还可作为通过重构（均值）结构加以设定的潜变量模型获得统计学解释：每个观测值由非负基底轮廓与潜在得分的加性组合来近似。这种对均值结构的强调，与通常通过协方差结构来表述的经典因子分析和结构方程模型（SEM）形成对照（Bollen，1989年；Jöreskog & Goldberger，1975年）。

均值结构建模在纵向研究和心理测量研究中有着悠久传统；在这类研究中，增长曲线模型通过对预先指定的设计矩阵进行回归来表示重复测量（Laird & Ware，1982年；Potthoff & Roy，1964年；Satoh & Yanagihara，2010年；Vonesh 和 Carter，1987年）。

从这个角度看，非负矩阵分解可被视为均值结构模型的一种数据自适应 类比：基矩阵扮演设计矩阵的角色，但它是在非负约束下从数据中学习得到的，而非事先固定。

贝叶斯表述与概率表述通过将非负矩阵分解刻画为带有工作似然和先验分布的潜变量模型，使这种数据自适应观点得以明确（Cemgil，2009年；Schmidt等人，2009年），并且统计非负矩阵分解已针对基因组学及相关场景中的计数数据与模型选择得到发展（Rosales等人，2017）。

针对特定似然设定，研究者已开发出协变量感知贝叶斯非负矩阵分解扩展，尤其是基于计数的突变特征模型（Grabski et al., 2025; Robinson et al., 2019）。在频率学派研究方向，Phelan 和 Campbell（2025）近期将可分（凸）非负矩阵分解与事后回归及非参数自助法相结合

用于主题模型；他们的两阶段方法固定基矩阵，并将主题比例对协变量进行回归，但既未纳入随机效应，也未对协变量效应矩阵施加非负约束。尽管相关统计学文献日益增多，但相较之下，将协变量效应、单位层级随机效应和非负约束整合到单一估计与自助法框架中的联合频率学派推断，受到的关注仍然很少。我们关注的正是这一空白：针对随机效应，采用基于优化的推断并进行明确的复杂度诊断。

一种自然的纳入协变量的方法是用回归$B = \Theta A$替代自由得分，从而得到协变量驱动的近似 $Y \approx X\Theta A$ ，其中“设计”矩阵X 在非负性约束下进行学习（Ding 等人，2006年；Satoh，2023）。这种表述隐含了一种变量选择形式：对 X施加的非负性和低秩约束共同诱导出稀疏的基于部分的载荷，使每个成分仅保留对其有贡献的观测变量——无需显式的稀疏性惩罚，即可实现测量侧选择。当对 Θ 中协变量对潜在得分的影响进行正式推断时，所得显著性检验还能进一步识别哪些协变量会影响哪些成分，从而实现互补的协变量侧选择。

当 Θ 是科学研究关注的参数时，正式推断至关重要；然而，当 P 为中等或较大时，以协方差为中心的建模会变得脆弱。因此，我们通过引入单位特异性随机效应来扩展协变量驱动的非负矩阵分解，得到带随机效应的非负矩阵分解（NMF-RE）：

$$Y = X(\Theta A + U) + \mathcal{E}, \qquad \boldsymbol{y}_n = X(\Theta \boldsymbol{a}_n + \boldsymbol{u}_n) + \boldsymbol{\varepsilon}_n.$$

这里， X 是共享的非负基矩阵， Θ 刻画了潜在得分上的系统性协变量效应， U 包含潜在得分空间中的个体特异性偏差，而$\mathcal{E}$ 是误差矩阵。这与混合效应建模（Breslow & Clayton，1993年；Kenward & Roger，1997年；Laird & Ware，1982年）相似，但根本区别在于，基 X 是未知的，并且是在非负约束下估计的。因此，我们将 NMF-RE 视为一种定义在重构（均值）结构上、而非

协方差结构上的混合效应潜变量模型。

正则化较弱的随机效应可能会吸收特异性变异，并使针对 Θ 的推断不稳定（Efron，2004；Tibshirani与Taylor，2012）。因此，我们将 $\lambda = \sigma^2/\tau^2$ 视为复杂度控制装置，并明确监测 U所消耗的有效自由度。

估计采用分块方案：随机效应通过岭型最佳线性无偏预测（BLUP）闭式解进行更新，而 X 和 Θ 则通过非负乘法规则进行更新。为避免进入近饱和状态，我们对 U的有效自由度设定诊断上限；该上限是一种数值保护措施，在常规运行条件下不会触发。

统计推断针对 Θ 进行，并在给定估计得到的$\widehat{\ }$ $\widehat{\ }$正则化解（X，U）的条件下开展。采用正则化后视角（Javanmard与Montanari，2014；van de Geer等人，2014），我们通过结合（i）渐近线性化、（ii）单步牛顿更新，以及（iii）基于加权得分贡献的乘子（wild）自助法，避免反复进行约束再优化。由此可实现计算高效的不确定性量化，且与非负矩阵分解的算法结构相兼容。

我们的主要贡献如下：

1. 我们提出 NMF-RE，即一种在非负约束下纳入单位特异性随机效应的协变量驱动非负矩阵分解均值结构表述。

2. 我们为随机效应引入有效自由度诊断，并提出基于自由度的上限，以避免进入近饱和状态。

3. 我们通过上述三步策略，为 Θ 开展正则化后推断，避免重复进行约束优化。

4. 通过模拟（包括针对 dfU 上限的专项压力测试）和四项实证研究，我们展示了该方法在纵向、心理测量、

空间流和文本分析场景下均能实现稳定推断。实证示例表明，$\widehat{\ }$所学习的 X 无需预设载荷模式，即可通过将观测变量归入可解释成分来实现测量侧变量选择；而对 Θ 的推断，则通过识别哪些协变量会显著影响哪些成分，实现协变量侧变量选择。

更广泛地说，本文倡导一种以均值结构为中心的推断方法，用于基于约束的潜变量模型；该方法结合了具有可解释性的重构结构、明确的复杂度控制，以及通过非负潜在结构实现的隐式变量选择。

2 模型

2.1 数据结构和符号表示

设 $Y \in \mathbb{R}_{\geq 0}^{P \times N}$ 为一个非负数据矩阵，其中包含 P 个观测变量（条目、词语、区域、年龄）和 N 个单位（受试者、文档、起讫节点）。令 $A \in \mathbb{R}^{K \times N}$ 收集这 N 个单位的观测协变量，其中第 n 列为 $\boldsymbol{a}_n \in \mathbb{R}^K$。我们用 $\mathbf{1}_m$ 表示一个由 m 个 1 组成的向量，并用 Q 表示目标潜在维度（秩）。

协变量编码。由于我们对协变量效应矩阵 Θ（第 2.2 节）施加逐元素非负性，因此，将协变量编码为使 A 逐元素非负是很方便的（例如截距、指示变量和非负缩放）。当某个协变量同时取正值和负值时，我们采用标准的正负部分展开 $a = a^+ - a^-$，其中 $a^+ = \max(a,\ 0)$ 和 $a^- = \max(-a,\ 0)$，并将 a^+ 和 a^- 作为单独的非负协变量纳入。这保留了 $\Theta \geq 0$，同时不限制 ΘA 所能表示的线性效应类别。

2.2 带协变量的非负矩阵分解

经典非负矩阵分解用低秩非负乘积 $Y \approx XB$来近似 Y，其中 $X \in \mathbb{R}_{\geq 0}^{P \times Q}$ 是基矩阵，$B \in \mathbb{R}_{\geq 0}^{Q \times N}$ 是得分（系数）矩阵（Gillis，2014、2020；D. Lee & Seung，2000年；D. D. Lee & Seung，1999）。为纳入观测到的

协变量，我们通过对 A的回归来参数化得分，

$$B = \Theta A, \tag{1}$$

其中， $\Theta \in \mathbb{R}_{\geq 0}^{Q \times K}$ 表示潜在得分上的协变量效应。由此得到的重构（均值）结构为

$$Y \approx X\Theta A,$$

它在形式上与增长曲线模型的均值结构相似（Laird & Ware，1982年；Potthoff & Roy，1964年；Satoh & Yanagihara，2010年；Vonesh & Carter，1987年），主要区别在于， X 是在非负约束下学习得到的，而不是预先指定的固定设计矩阵。机器学习文献中也出现了相关的三因子表示（Ding 等人，2006年），但在此我们强调回归式解释，并将 Θ 作为主要关注参数。

2.3 带随机效应的非负矩阵分解（NMF-RE）

在许多纵向和多层级研究情境中，观测到的协变量无法充分解释单位间异质性。为捕捉未解释的偏差，同时保持可解释性，我们提出带随机效应的非负矩阵分解（NMF-RE）：

$$Y = X(\Theta A + U) + \mathcal{E}, \tag{2}$$

其中， $X \in \mathbb{R}_{\geq 0}^{P \times Q}$ 是共享非负基， $\Theta \in \mathbb{R}_{\geq 0}^{Q \times K}$ 是非负协变量效应矩阵，
$U \in \mathbb{R}^{Q \times N}$ 是单位特异性随机效应矩阵，$\mathcal{E} \in \mathbb{R}^{P \times N}$ 是误差矩阵。

我们始终将 A视为固定（非随机）。随机性归因于 U 和$\mathcal{E}$，因而通过（2）也归因于 Y。我们采用如下工作矩结构：

$$E(\mathrm{vec}(\mathcal{E})) = \mathbf{0}_{PN}, \qquad E(\mathrm{vec}(U)) = \mathbf{0}_{QN}, \tag{3}$$

$$\mathrm{Var}(\mathrm{vec}(\mathcal{E})) = \sigma^2 I_{PN}, \qquad \mathrm{Var}(\mathrm{vec}(U)) = \tau^2 I_{QN}, \tag{4}$$

$$\mathrm{Cov}(\mathrm{vec}(U), \mathrm{vec}(\mathcal{E})) = \mathbf{0}_{QN \times PN}. \tag{5}$$

在这些工作假设下，

$$\mathrm{Var}(\boldsymbol{y}_n \mid \boldsymbol{a}_n) = \tau^2 X X^\top + \sigma^2 I_P, \tag{6}$$

强调其与线性混合效应模型的联系，同时聚焦于重构（均值）结构。

单位层面解释。对于单元 n，

$$\boldsymbol{y}_n = X(\Theta \boldsymbol{a}_n + \boldsymbol{u}_n) + \boldsymbol{\varepsilon}_n, \tag{7}$$

其中， $\boldsymbol{a}_n$ 是固定的，而 $\boldsymbol{u}_n$ 和 $\boldsymbol{\varepsilon}_n$ 是随机的。因此， X 的列定义了 Q 共享的非负潜在剖面，术语 $\Theta \boldsymbol{a}_n$ 描述了潜在得分的系统性协变量调节，而 $\boldsymbol{u}_n$ 则捕捉特有偏离。这构成了上文介绍的均值结构视角，与基于协方差的结构方程模型表述相辅相成（Bollen，1989年；Jöreskog，1969年；Jöreskog & Goldberger，1975年）。作为特例，当 X 已知时，NMF-RE可化为随机系数增长曲线模型（Laird & Ware，1982年；Potthoff & Roy，1964年；Satoh & Yanagihara，2010年）；当 A 不存在或 U 被抑制时，则可化为标准非负矩阵分解（D. Lee & Seung，2000年；Schmidt 等人，2009年）。

关于符号和约束的备注。为了提高可解释性，我们施加 $X \geq 0$ 和 $\Theta \geq 0$，但允许 U 取实数值；岭正则化和自由度监控（第 3–3.4 节）可防止模型行为接近饱和。

2.4 约束、归一化与可识别性

非负矩阵分解类模型并非完全可识别，因为对于任意对角线元素为正的对角矩阵 D，$X(\Theta A + U)$ 在$X \leftarrow XD^{-1}$ 和 $(\Theta, U) \leftarrow (D\Theta, DU)$ 变换下保持不变；潜在成分的排列也不会改变模型。为固定尺度，我们对 X施加归一化。在我们的实现中，采用列和约束

$$X \geq 0, \quad \Theta \geq 0, \quad \mathbf{1}_P^\top X = \mathbf{1}_Q^\top, \tag{8}$$

并将由此产生的缩放吸收到 Θ中（必要时也吸收到 U 中）。这种归一化保留了可解释性：X 的每一列都可视为一个非负特征分布（例如变量上的分布），而 ΘA 的各行则表示各单位中不同成分的尺度。

除了缩放和排列不确定性之外，非负矩阵分解通常只有部分可辨识性；尽管如此，在实践中，归一化和非负性往往能产生稳定且易于解释的分解（Gillis，2014、2020；Gillis 与 Rajkó，2023）。

3 估计

3.1 通过随机效应惩罚项进行复杂度控制

我们主要将 τ^2 视为复杂度控制参数： $\lambda = \sigma^2/\tau^2$ 决定允许的个体特异性偏差。较大的 τ^2（较小的 λ）允许近饱和拟合；较小的 τ^2 则将 U 收缩至零。由此可见，有必要明确监测 U 所消耗的有效自由度（第3.4节），并设置诊断上限（第3.5节）。

3.2 工作目标函数

基于第2节中的模型，我们通过最小化惩罚最小二乘准则来估计 $(X,\ \Theta,\ U)$

$$\mathcal{L}(X, \Theta, U) = \|Y - X(\Theta A + U)\|_F^2 + \lambda \|U\|_F^2, \tag{9}$$

受限于

$$X \geq 0, \qquad \Theta \geq 0, \qquad \mathbf{1}_P^\top X = \mathbf{1}_Q^\top.$$

这里，$\|\cdot\|_F$ 是弗罗贝尼乌斯范数，λ 是第 3.1 节中引入的复杂度控制惩罚项。我们将 (9) 作为计算的工作目标函数；关于 Θ 的正式推断将在第 4 节中展开。

3.3 分块估计方案

在非负约束下联合最小化式 (9) 在计算上颇具挑战。因此，我们采用分块方案，在固定其余各块的同时，交替更新 U、 X 和Θ。这样既保留了经典非负矩阵分解算法的结构（Févotte & Idier, 2011; D. Lee & Seung, 2000），又纳入了随机效应。

3.3.1 U步：岭型更新

固定 X 和 Θ，定义 $B = \Theta A$。U 子问题为

$$\min_U \ \|Y - X(B+U)\|_F^2 + \lambda \|U\|_F^2.$$

这会使各单位之间相互解耦。对于第 n 列，

$$\boldsymbol{u}_n = (X^\top X + \lambda I_Q)^{-1} X^\top \Big(\boldsymbol{y}_n - X\boldsymbol{b}_n\Big), \tag{10}$$

其中，$\boldsymbol{y}_n$ 和 $\boldsymbol{b}_n$ 分别是 Y 和 B的第 n 列。公式（10）是一种岭型更新，类似于线性混合效应模型中的类似BLUP的随机效应估计量（Breslow & Clayton，1993年；Laird & Ware，1982年）。更新 U后，我们对每一行进行中心化，以确保各单位间的均值为零；这通过防止系统性偏移被 U吸收，提升了可识别性。

计算。对于固定的 X 和 λ，矩阵（$X^\top X + \lambda I_Q$）$^{-1}X^\top$ 对所有列都是相同的，因此可以通过构造并重复使用 $Q \times Q$求解，高效地实现 U更新。

3.3.2 X步：乘法非负更新

固定 Θ 和 U，定义

$$B_U = \Theta A + U.$$

由于 U 为实值，B_U 可能包含负元素。为保持乘法更新的非负性和数值稳定性，我们使用逐元素正部

$$B_U^+ = \max(B_U, 0)$$

作为下方更新中的稳定替代项。我们使用欧几里得乘法规则更新 X，

$$X \leftarrow X \odot \Big(Y(B_U^+)^\top\Big) \oslash \Big(X\Big(B_U^+(B_U^+)^\top\Big)\Big), \tag{11}$$

其中，$\odot$ 和 $\oslash$分别表示 Hadamard积和除法。

归一化。 更新后，我们对 X 的列重新归一化，使其满足$\mathbf{1}_P^\top X = \mathbf{1}_Q^\top$。如果 $X \leftarrow XD^{-1}$ 表示归一化，则我们应用相应的缩放（Θ，U）$\leftarrow$（$D\Theta$，DU），使拟合均值 X（$\Theta A + U$）保持不变（第 2.4 节）。

备注。 当 $U \equiv 0$ （或 $B_U \geq 0$）时，(11) 简化为标准的欧几里得乘法更新（Févotte & Idier，2011；D. Lee & Seung，2000）。引入正部稳定化仅仅是为了在 U 为实数值时确保乘法方案有定义；实际应用中，我们监控工作目标函数 (9)，以确保稳定下降。

3.3.3 Θ 步：协变量驱动更新

固定 X 和 U 后，通过最小化以下目标来更新 Θ

$$\|Y - X(\Theta A + U)\|_F^2 \quad \text{subject to } \Theta \geq 0.$$

令 $Y_U = Y - XU$。由于 Y_U 可能包含负值，我们使用正部 $Y_U^+ = \max(Y_U, 0)$ 来获得稳定的乘法更新，

$$\Theta \leftarrow \Theta \odot \left(X^\top Y_U^+ A^\top\right) \oslash \left((X^\top X\Theta)(AA^\top)\right). \tag{12}$$

这是针对协变量调整的非负三因子分解乘法更新（Ding 等人，2006），并针对实数值 U 进行了稳定化处理。

3.4 U的有效自由度

岭回归更新（10）增加了模型灵活性。对于固定的 X 和 λ，拟合得到的随机效应贡献可使用岭回归“帽子”矩阵表示

$$H_\lambda = X(X^\top X + \lambda I_Q)^{-1} X^\top.$$

我们将 U 所消耗的有效自由度定义为

$$\mathrm{df}_U = N \,\mathrm{tr}(H_\lambda). \tag{13}$$

设 $d_1, \ldots, d_Q$ 为 $X^\top X$ 的特征值。那么 $\mathrm{tr}(H_\lambda) = \sum_{q=1}^{Q} d_q/(d_q + \lambda)$，因此

$$\mathrm{df}_U = N \sum_{q=1}^{Q} \frac{d_q}{d_q + \lambda}, \tag{14}$$

这与经典的岭回归自由度表达式一致，并直接关联到惩罚估计中的复杂度度量（Efron, 2004；Tibshirani & Taylor，2012年）。这种表示形式很方便，因为对于固定的 X，df_U 随着 λ 的增大而单调递减，因此可以进行简单的一维控制。

3.5 通过 dfU 上限进行诊断控制

为避免 U 吸收大部分特异变异的近饱和状态，我们施加诊断上限

$$\mathrm{df}_U(\lambda) \le \mathrm{df}_U^{\max}.$$

由于 $\mathrm{df}_U(\lambda)$ 关于 λ 单调递减（第 3.4 节），我们定义

$$\lambda_{\mathrm{cap}}(X) = \inf\{\lambda > 0 :\ \mathrm{df}_U(\lambda) \le \mathrm{df}_U^{\max}\}.$$

每当当前的 λ 违反该上限时，我们强制执行

$$\lambda \leftarrow \max\{\lambda, \lambda_{\mathrm{cap}}(X)\},$$

在保持 X 固定的情况下，并在后续的 U 更新中使用该强制值。由于映射 $\lambda \mapsto \mathrm{df}_U(\lambda)$ 是单调的，$\lambda_{\mathrm{cap}}(X)$ 可通过一维搜索高效求得（例如，利用特征值形式 (14) 的二分法）。

解释与实际校准。dfU 上限是一种数值保护措施，而不是用于预测的调优参数。在实践中，上限比率$r^{\max} = \mathrm{df}_U^{\max}/(NQ)$通过以下方式校准：根据 $X_{\mathrm{fix}}^\top X_{\mathrm{fix}}$的特征值，将分析者对 λ 施加的下界转换为隐含饱和水平，其中 X_{fix} 通过使用 $U \equiv 0$ 的初始化拟合得到（详见附录 A）。为便于透明报告，我们会区分优化过程中上限是否曾被触发，以及最终解是否取紧约束（$\mathrm{df}_U = \mathrm{df}_U^{\max}$）或约束不取等。

3.6 收敛、初始化与实现

反复进行分块更新，直至工作目标函数 (9) 的相对变化低于容差。典型的一次迭代如下：(i) 强制执行

当前的 dfU 上限 X （第3.5节）；(ii) 按照 (10) 更新 U 并进行行中心化；(iii) 按照 (11) 更新X并重新归一化，将由此产生的缩放应用于（Θ, U）；(iv) 按照 (12) 更新 Θ。

初始化时，使用协变量驱动NMF拟合（含 $U \equiv 0$）获取（X, Θ）的初始值，随后引入 U 并通过岭回归步骤进行更新。所有更新均可通过闭式解或乘法表达式实现，并适用于中等规模和大型矩阵。

优化过程中使用的工作尺度。在某些实现中，我们会更新工作尺度以稳定有效惩罚水平，使用 $\hat{\sigma}^2_{\text{iter}} = \|R\|_F^2/(PN)$ 和$R = Y - X(\Theta A + U)$，并在需要时通过 $\hat{\sigma}^2_{\text{iter}}$ 表示 λ。这个迭代尺度量仅用于数值稳定化；对于统计推断，我们则使用自由度调整方差估计量（第4.2节）。

参考实现可在 `nmfkc` 软件包中获取（Satoh，2025），重现本文数值结果所需的所有脚本均已提供于 https://github.com/ksatohds/nmfre-paper。完整的分块估计算法（包括所有步骤的伪代码和下降保护）见附录B。

变量选择作为估计结果。 对 X 施加的非负性和低秩约束并非专门设计为显式的稀疏性惩罚，但估计得到的$\frown X$ 通常呈现稀疏的、基于部分的结构，其中每个成分仅加载于一部分观测变量——这是估计程序中自然产生的隐式测量侧变量选择。结合针对 Θ 的正式推断（第 4 节），显著性检验还能进一步识别哪些协变量会影响哪些成分，从而提供互补的协变量侧选择。

4 推断

4.1 范围与推断目标

本节针对 NMF-RE 模型中的协变量效应矩阵 Θ展开统计推断。我们的推断目标是 Θ 在给定 第 3 节中得到的正则化估计 $\widehat{\ }$ $\widehat{\ }$ 表示（ X， U ）的条件下，遵循正则化后推断的思路（Javanmard与Montanari，2014；Taylor与Tibshirani，2015；van de Geer等人，2014）。

通过三明治法和乘子自助法构造，增强方法对工作模型设定错误的稳健性。

4.2 剖面得分、信息量与方差尺度

令 X 表示第3节中估计的基矩阵，并令 λ 表示用于随机效应的最终岭惩罚。在对 Θ进行推断时，我们将（X， λ）视为固定，并对每个候选解 Θ通过岭回归剖面消去 U：

$$\widehat{U}(\Theta) = \arg\min_{U \in \mathbb{R}^{Q \times N}} \ \|Y - \widehat{X}(\Theta A + U)\|_F^2 + \lambda \|U\|_F^2.$$

由此得到逐列更新

$$\widehat{\boldsymbol{u}}_n(\Theta) = (\widehat{X}^\top \widehat{X} + \lambda I_Q)^{-1} \widehat{X}^\top \Big(\boldsymbol{y}_n - \widehat{X}\Theta \boldsymbol{a}_n\Big), \qquad n = 1, \dots, N,$$

$\widehat{\ }$ $\widehat{\ }$其在 $\Theta = \Theta$处与 U 一致。

方差尺度。我们将 σ^2 视为讨厌尺度，并使用自由度调整估计量

$$\widehat{\sigma}^2 = \frac{\|Y - \widehat{X}(\widehat{\Theta} A + \widehat{U})\|_F^2}{PN - \mathrm{df}_U - \mathrm{df}_\Theta}, \tag{15}$$

其中，df_U 是根据（X， λ）（第 3.4 节）进行岭回归随机效应更新所消耗的有效自由度。对于 Θ，一种简单的实用选择是 $\mathrm{df}_\Theta = QK$。由于 $\Theta \geq 0$可能会使某些系数落在边界上，一种更能考虑边界情况的备择方法是使用“活跃集”自由度

$$\mathrm{df}_\Theta^{\mathrm{act}} = \#\{(q,k) : \ \widehat{\Theta}_{qk} > \delta\},$$

并采用较小的容差 $\delta > 0$ （例如， $\delta = 10^{-8}$ 乘以一个尺度因子）。在我们的实验中，QK 与 $\mathrm{df}_\Theta^{\mathrm{act}}$ 之间的差异通常较小；我们会说明采用了哪种约定。在后续所有表达式中，σ^2 都将替换为其代入估计 $\widehat{\sigma}^2$。

剖面残差和得分贡献。令

$$\boldsymbol{r}_n(\Theta) = \boldsymbol{y}_n - \widehat{X}\Big(\Theta \boldsymbol{a}_n + \widehat{\boldsymbol{u}}_n(\Theta)\Big)$$

表示单元 n 的剖面残差。对剖面目标函数关于Θ 求导，并利用 $\hat{\boldsymbol{u}}_n(\Theta)$ 的一阶最优性（包络定理），可得得分贡献

$$S_n(\Theta) = -\frac{1}{\sigma^2}\Big(\widehat{X}^\top \boldsymbol{r}_n(\Theta)\Big)\boldsymbol{a}_n^\top, \qquad n = 1, \dots, N. \tag{16}$$

我们用 vec(·）表示按列堆叠向量化；推断目标是 vec(Θ) $\in \mathbb{R}^{QK}$。

岭回归帽子矩阵与收缩表示。定义岭回归帽子矩阵$\widehat{}$在 X空间中，

$$H_\lambda = \widehat{X}(\widehat{X}^\top \widehat{X} + \lambda I_Q)^{-1}\widehat{X}^\top.$$

将岭回归更新代入残差可得

$$\boldsymbol{r}_n(\Theta) = (I_P - H_\lambda)\Big(\boldsymbol{y}_n - \widehat{X}\Theta\boldsymbol{a}_n\Big),$$

$\widehat{}$因此，剖除 U 会通过 $(I_P - H_\lambda)$ 收缩 X空间中的变异。

工作信息量。 若在计算关于 Θ 的曲率时忽略剖除 U 的影响，则费舍尔型工作信息矩阵为

$$\mathcal{I}_0(\Theta) = \frac{1}{\sigma^2}\left(AA^\top \otimes \widehat{X}^\top\widehat{X}\right). \tag{17}$$

剖除所导致的信息量减少 U。剖除 U 会降低 Θ的有效信息量：

$$\mathcal{I}(\Theta) = \frac{1}{\sigma^2}\left(AA^\top \otimes \Big(\widehat{X}^\top(I_P - H_\lambda)\widehat{X}\Big)\right), \tag{18}$$

其中，内 部矩阵可简化为 $\lambda\,\widehat{X}^\top\widehat{X}(\widehat{X}^\top\widehat{X} + \lambda I_Q)^{-1}$。我们使用在 $\widehat{\Theta}$处计算的 (18) .

4.3 三明治法协方差估计

由于高斯工作模型可能存在设定错误，我们采用一种基于得分贡献的三明治型协方差估计量。令

$$\widehat{\mathcal{I}} = \mathcal{I}(\widehat{\Theta}), \qquad \widehat{\mathcal{J}} = \sum_{n=1}^{N} \mathrm{vec}\Big(S_n(\widehat{\Theta})\Big)\mathrm{vec}\Big(S_n(\widehat{\Theta})\Big)^\top.$$

$\widehat{}$vec(Θ) 的三明治协方差估计量为

$$\widehat{\mathrm{Var}}\Big(\mathrm{vec}(\widehat{\Theta})\Big) = \widehat{\mathcal{I}}^{-1}\widehat{\mathcal{J}}\widehat{\mathcal{I}}^{-1}. \tag{19}$$

当单元级得分贡献存在异方差或簇内相依时，可采用有限样本校正（例如，HC 类缩放）和聚类稳健变体（Cameron等人，2008）。

4.4 一步牛顿近似

直接进行自助法推断，需要在非负约束下反复重新优化，计算成本很高。因此，我们通过在 Θ 附近进行单步牛顿更新来近似自助复制：

$$\operatorname{vec}(\widehat{\Theta}^{*}) = \operatorname{vec}(\widehat{\Theta}) - \widehat{\mathcal{I}}^{-1} S^{*}(\widehat{\Theta}), \tag{20}$$

其中，$S^{*}(\Theta)$ 表示重抽样得分向量。这种构造利用了 Θ 的渐近线性性，是正则化后推断中的标准做法（Javanmard与Montanari，2014；van de Geer等人，2014）。

一步复制中的非负性。由于（20）是无约束的线性更新，所得 Θ^{*} 可能违反逐元素约束 $\Theta \geq 0$，尤其是对于接近边界的系数。实际实现中，我们在形成自助复制时应用逐元素投影

$$\widehat{\Theta}^{*} \leftarrow \max(\widehat{\Theta}^{*}, 0)$$

对于估计值为零（或接近零）的系数，边界效应可能导致非高斯且经截断的有限样本行为；在这种情况下，单侧检验和百分位型自助汇总通常比双侧 Wald正态近似更合适（见第 5 节）。

单步映射的渐近理论依据——即在正则性条件下，证明（20）在内部区域渐近等价于完全迭代的 Z 估计量——见附录 C（另见 Huber等人，1967；Javanmard与Montanari，2014；Taylor与Tibshirani，2015；van de Geer等人，2014；White，1980）。

4.5 得分贡献的乘子（wild）自助法

我们使用乘子自助法方案生成重抽样得分（这种方法常被称为针对估计方程贡献的野生自助法）。令 ξ_1、……、ξ_N为独立同分布的乘子，满足 $E(\xi_n) = 0$ 且 $\mathrm{Var}(\xi_n) = 1$，例如拉德马赫（± 1）、标准正态或指数中心化权重（$\xi_n = \zeta_n - 1$，其中 $\zeta_n \sim \mathrm{Exp}(1)$）（Cameron等人，2008；Mammen，1993）。在下文报告的所有实验中，我们都使用指数中心化分布。我们构造重抽样得分

$$S^*(\widehat{\Theta}) = \sum_{n=1}^{N} \xi_n \,\mathrm{vec}\Big(S_n(\widehat{\Theta})\Big),$$

$\widehat{\ }$并代入（20），以获得自助复制 Θ^*，而无需重新求解约束优化问题。

通过$\{\Theta^*\}$的经验分布，计算 Θ中各元素的自助标准误和置信区间$\widehat{\ }$。我们同时报告基于夹心估计的标准误和基于自助法的标准误，以透明地评估不确定性以及一步近似的充分性。

4.6 计算说明

克罗内克结构 $\widehat{\ }\mathcal{I} = \hat{\sigma}^{-2}(AA^\top \otimes F)$与 $F = \widehat{X}^\top(I_P - H_\lambda)\widehat{X} \in \mathbb{R}^{Q\times Q}$使得一步更新和三明治协方差可以通过对较小的 $K \times K$ 和$Q \times Q$ 求逆来计算，而无需构造完整的（QK） ×（QK）矩阵。

4.7 有效性、适用范围与实用诊断

所提出的推断目标 Θ 以所选的正则化$\widehat{\ }$ $\widehat{\ }$表示为条件（X， U）。在M/Z估计量的标准正则性条件下，三明治协方差估计量（19）即使在误设高斯工作模型下也能提供大样本稳健标准误，并且乘子自助法与一步牛顿映射相结合，可得到计算高效的重抽样近似（Cameron等人，2008；Huber等人，1967；van de Geer等人，2014；White，1980）。由于自助复制被投影到 $\Theta \geq 0$上，靠近边界的系数可能表现出

截断，因此在边界情形下建议采用方向检验和百分位自助法汇总。关于推断范围、正则性条件、边界行为以及实用诊断（包括建议报告的条目）的详细讨论见附录 D。

5 模拟研究

5.1 目标与设计

本节评估针对 NMF-RE 中协变量效应矩阵 Θ 的所提出的后正则化推断在有限样本下的表现。推断目标为 $\widehat{\Theta}$ 在拟合的低秩表示（X, U）条件下。由于（i）估计采用非负约束，并且（ii）推断在正则化拟合之后进行，因此实际中主要需要考察的问题是：一步牛顿近似（第 4.4 节）所依据的局部线性化，能否在有限样本下产生表现稳定的不确定性量化结果。为此，我们考察（a）基于夹心估计的标准误（SE）与单步乘子自助法标准误（BSE）之间的一致性，以及（b）Wald 型检验和置信区间的规模与覆盖率，并特别关注由 $\Theta \geq 0$ 引起的边界行为。

基于 Orthodont 的基线设计（$P = 4$, $Q = 1$）。模拟设计与第 6.1 节中的 Orthodont 示例相一致。我们固定 $P = 4$ 个时间点和秩 $Q = 1$。基向量设定为估计得到的 Orthodont 轮廓

$$X_{\text{true}} = (0.2308,\ 0.2409,\ 0.2566,\ 0.2717)^{\top}, \qquad \mathbf{1}_P^{\top} X_{\text{true}} = 1.$$

协变量矩阵为 $A = (\mathbf{1}, \texttt{male})^{\top} \in \mathbb{R}^{2\times N}$。对于 $N = 27$, `male` 指标遵循 Orthodont 样本。对于 $N \in \{100,200\}$，我们生成 `male` 个独立同分布的伯努利分布随机变量，其成功概率等于 Orthodont 样本中的男性比例。

数据由 NMF-RE 机制生成

$$Y = X_{\text{true}}(\Theta_{\text{true}} A + U) + \mathcal{E}, \tag{21}$$

其中， $U \in \mathbb{R}^{Q\times N}$的元素独立同分布，服从 $N(0,\tau^2)$ 分布，且 $\mathcal{E} \in \mathbb{R}^{P\times N}$的元素独立同分布，并且独立于$U$。我们设定 $\sigma^2=\tau^2=1$，并将相应的工作惩罚项 $\lambda=\sigma^2/\tau^2=1$用作基准正则化水平。

误差分布（稳健性检验）。为考察观测噪声偏离高斯分布时的敏感性，同时保持其方差与工作模型一致，我们只改变$\mathcal{E}$:

- **高斯：** $\varepsilon_{pn}\sim N(0,\ 1)$。
- **中心化指数分布：** $\varepsilon_{pn}=\xi_{pn}-1$ 其中 $\xi_{pn}\sim$ 服从 Exp(1)，因此 $E(\varepsilon_{pn})=0$ 且 $\mathrm{Var}(\varepsilon_{pn})=1$ ，但噪声呈正偏态。

除非另有说明，随机效应 U 均根据上述高斯模型生成；因此，此稳健性检验针对的是测量误差偏离高斯性的情况。

边界与内部情形。我们考虑两种情形:

- **零假设（边界原假设）：** $\Theta_{\text{true}}[\texttt{male}]=0$。由于 $\Theta\geq 0$，该零假设位于^边界上，因此 $\theta_{\texttt{male}}$ 呈现截断的有限样本分布。
- **备择假设（内部备择假设）：** $\Theta_{\text{true}}[\texttt{male}]=9.4285$，与表2中的 Orthodont 估计值一致。

在两种情形下， $\Theta_{\text{true}}[\texttt{intercept}]=90.502$。

拟合与推断（基线设计）。对于每个 $N\in\{27,100,200\}$ 和每种误差分布，我们生成 $R=1000$ 个蒙特卡洛数据集。每个数据集均使用第 3 节中的分块 NMF-RE 算法进行拟合，并采用 dfU 诊断上限，其中 $\mathrm{df}_U^{\max}/(NQ)=0.21$。我们使用与第 3 节相同的收敛容差、目标函数监控和保护措施；在适用时，还会在多次初始化中保留最佳解，以缓解非凸性。

推断使用约化信息量 (18)、三明治协方差 (19) 和单步乘子自助法（第4.4–4.5节），进行 $B=1000$ 次重抽样，并投影到 $\Theta\geq 0$。对于边界零假设情形（零假设），假设检验和报告的 p 值均采用单侧计算，并与约束 $\theta_{\texttt{male}}\geq 0$ 保持一致。

方差尺度与自由度约定。在整个模拟过程中，我们使用 (15) 估计方差尺度。对于 df_Θ，我们使用 $\mathrm{df}_\Theta = QK$ （工作选择）；将其替换为活动集自由度计数（第4.2节）时，结果在定性上没有变化。

随机效应饱和的额外压力测试（dfU 上限的必要性）。为明确验证 dfU 上限作为防止随机效应饱和的安全装置所发挥的作用，我们在更灵活的设置（$P = 4$、 $Q = 3$、 $N = 100$）下开展了额外的针对性压力实验，其中有意减弱工作随机效应惩罚项，以促进近饱和。完整的设计细节和结果见附录 E。

5.2 评估指标

我们报告男性效应$\hat{}$估计量 $\theta_{\mathtt{male}}$ 的蒙特卡洛偏差、标准差（SD）和均方根误差（RMSE）。我们还报告 SE 和 BSE 的蒙特卡洛平均值。对于假设检验，我们使用基于 SE 或 BSE 的沃尔德型统计量。对于边界原假设（零假设），主要程序是单侧检验 $H_0:\ \theta_{\mathtt{male}} = 0$ 对 $H_1:\ \theta_{\mathtt{male}} > 0$。

对于区间估计，我们报告基于 SE 和 BSE 的名义95%沃尔德区间的覆盖率，以及由一步乘子自助法重复样本构造的95%百分位区间的覆盖率。我们通过 $\mathrm{df}_U/(NQ)$ 监测随机效应饱和：在基线 $Q = 1$研究中，我们汇总 $\mathrm{df}_U/(NQ)$ 的蒙特卡洛均值；而在针对性的 $Q = 3$ 压力测试中，我们还报告上尾汇总 $\mathrm{dfU}_{0.99}$，即在蒙特卡洛运行中 $\mathrm{df}_U/(NQ)$ 的第99百分位数。

5.3 结果

完整的蒙特卡洛结果报告于表 F1（附录 F）。该模拟支持本文的核心主张：条件对 Θ 在正则化 NMF-RE 拟合之后的推断，可以在受控的低秩设置中保持稳定。在备择假设（内部区域）下，标准误与 BSE 接近，且随着 N增大，覆盖率趋近名义水平，从而证实当系数远离

边界。在零假设（边界情形）下，由于向 $\Theta \geq 0$投影，BSE 系统性地小于标准误，但单侧拒绝率仍接近 5%，覆盖率略为保守。将高斯误差替换为中心化指数误差会产生定性上相似的模式，表明对中等程度的非高斯性具有稳健性。详细解释见附录 F。

压力测试结果（附录 E）表明，若没有 dfU 上限，拟合惩罚项$\widehat{}$会崩溃（$\lambda \approx 0$），导致 Θ 的推断退化；而严格上限可避免这种崩溃，并在零假设下维持接近名义水平的检验尺寸和覆盖率。（在备择假设下，刻意设置的较弱初始惩罚与更高维的$\widehat{}$随机效应空间相结合，使得所有上限设置下的 Θ 均产生显著偏差；对于旨在探查失效模式而非常规性能的压力测试场景，这一结果在预期之中。）在基准 $Q = 1$ 设计中，平均饱和比 $\mathrm{df}_U/$（NQ）保持在 0.20 左右，默认上限 $\mathrm{df}_U^{\max}/$（NQ）$= 0.21$表明，U 处于适度正则化状态。

6 实证示例

我们通过四个复杂程度逐步增加的实证示例展示 NMF-RE：正畸生长数据、Holzinger–Swineford 认知测验、起讫点流量（OD–HUB）以及一个主题模型。表1汇总了模型规模和随机效应运行状态的诊断信息。饱和比 $r = \mathrm{df}_U/$（NQ）衡量随机效应的灵活性（$0 =$ 强正则化；$1 =$ 接近饱和）；上限比率$r^{\max} = \mathrm{df}_U^{\max}/$（$NQ$）按照第3.5节进行校准（另见附录 A）。所有报告的解均为约束不取等（$\mathrm{df}_U < \mathrm{df}_U^{\max}$）。

在随后的小节中，我们报告 Θ的推断，并结合学习得到的剖面（X）与饱和度诊断（r，$r^{\max}$）来解释所估计的成分。

表1

各实证示例中模型规模与随机效应幅度的汇总。其中，N为单位数，P 为观测变量数，Q 为潜在维度，NQ 为随机效应项的总数。我们报告df_U 以及饱和比$r = \mathrm{df}_U/(NQ)$，以及用于*dfU* 上限（第*3.5*节）的上限比率 $r^{\max} = \mathrm{df}_U^{\max}/(NQ)$。“上限”列表明在优化过程中是否曾触发*dfU* 上限，λ 为所报告拟合中使用的最终（可能为强制施加的）岭惩罚。

数据集	N	P	Q	NQ	df_U	r	$r^{\max}$	λ	Cap
Orthodont	27	4	1	27	5.42	0.201	0.21	1.00	No
HS1939	300	9	3	900	861.93	0.958	0.96	0.0104	No
OD–HUB	47	43	4	188	2.38	0.013	0.02	6.75	Yes
主题模型	59	67	3	177	1.30	0.007	0.02	3.30	Yes

6.1 正畸生长数据

正畸生长数据包含 $N = 27$ 名儿童（16名男性、11名女性）在8、10、12和14岁时的垂体—翼上颌裂距离（Laird & Ware, 1982年；Pinheiro 等，2025；Potthoff & Roy, 1964年），共计 $P = 4$ 次重复测量，协变量为 $\boldsymbol{a}_n = (1,\ \text{男性}_n)^\top$。

参考基准：线性混合效应建模（线性混合效应模型）。作为参考基准，我们还拟合了一个带有随机截距和斜率的标准线性混合效应模型（详情见附录 G）。该模型证实，Orthodont 数据同时表现出显著的个体层面异质性和性别依赖的均值结构，包括11岁时显著的性别差异以及性别依赖的生长速率。

NMF-RE 规格（$Q = 1$）。对于 $Q = 1$，NMF-RE 均值结构简化为 $\boldsymbol{y}_n = \boldsymbol{x}\,(\theta_0 + \theta_{\text{male}}\text{男性}_n + u_n) + \boldsymbol{\varepsilon}_n$，其中 $\boldsymbol{x} \geq 0$、$\mathbf{1}_P^\top \boldsymbol{x} = 1$；此处的 $\boldsymbol{x}$ 是从数据中学习得到的非负轨迹形状。

估计的基函数轮廓和拟合曲线。对于 $Q = 1$，估计的基向量 X 为 8、10、12 和 14 岁的年龄分配非负权重：

$$\widehat{X} = \big(0.2308,\ 0.2409,\ 0.2566,\ 0.2717\big)^{\top}, \qquad \mathbf{1}_P^{\top}\widehat{X} = 1.$$

递增趋势表明，学习到的轨迹在较晚年龄上逐渐分配了更多权重，这与单调增长一致。图1显示了观测测量值，以及每位受试者的两条拟合曲线：固定效应分量 $X\Theta\boldsymbol{a}_n$ 和基于最佳线性无偏预测的拟合 $X(\Theta\boldsymbol{a}_n + \widehat{\boldsymbol{u}}_n)$。两条曲线之间的差异反映了随机效应所捕捉的受试者特异性偏差。

随机效应机制（自由度诊断）。在此示例中，随机效应处于常规（非饱和）机制下。随机效应的有效自由度为 $\mathrm{df}_U/(NQ) = 0.201$，而诊断上限设为 $\mathrm{df}_U^{\max}/(NQ) = 0.21$。优化过程中未触及该上限，因此它不会约束所报告的解（表1）。

协变量效应的推断。表2报告了协变量效应估计值及其标准误。 p 值通过单侧检验计算，用于检验边界原假设 $H_0:\ \theta = 0$ 与备择假设 $H_1:\ \theta > 0$，并满足非负约束 $\Theta \geq 0$。截距和男性效应均达到统计显著。此外，基于自助法的标准误与基于夹心估计的标准误接近，支持在此情形下单步乘子（wild）自助法近似的适用性。

摘要。 与需要预先指定时间基函数的线性混合效应模型不同，NMF-RE 在非负性约束下从数据中学习轨迹形状，提供简约的均值结构摘要。这种由数据驱动的 X 学习构成了测量侧变量选择：非负基会自动识别哪些时间点对生长分量有贡献。在协变量侧，对 Θ 的推断显示性别效应显著（$p = 0.001$）。随机效应部分远未达到饱和，且 SE–BSE一致性良好（与线性混合效应模型的详细比较见附录 G

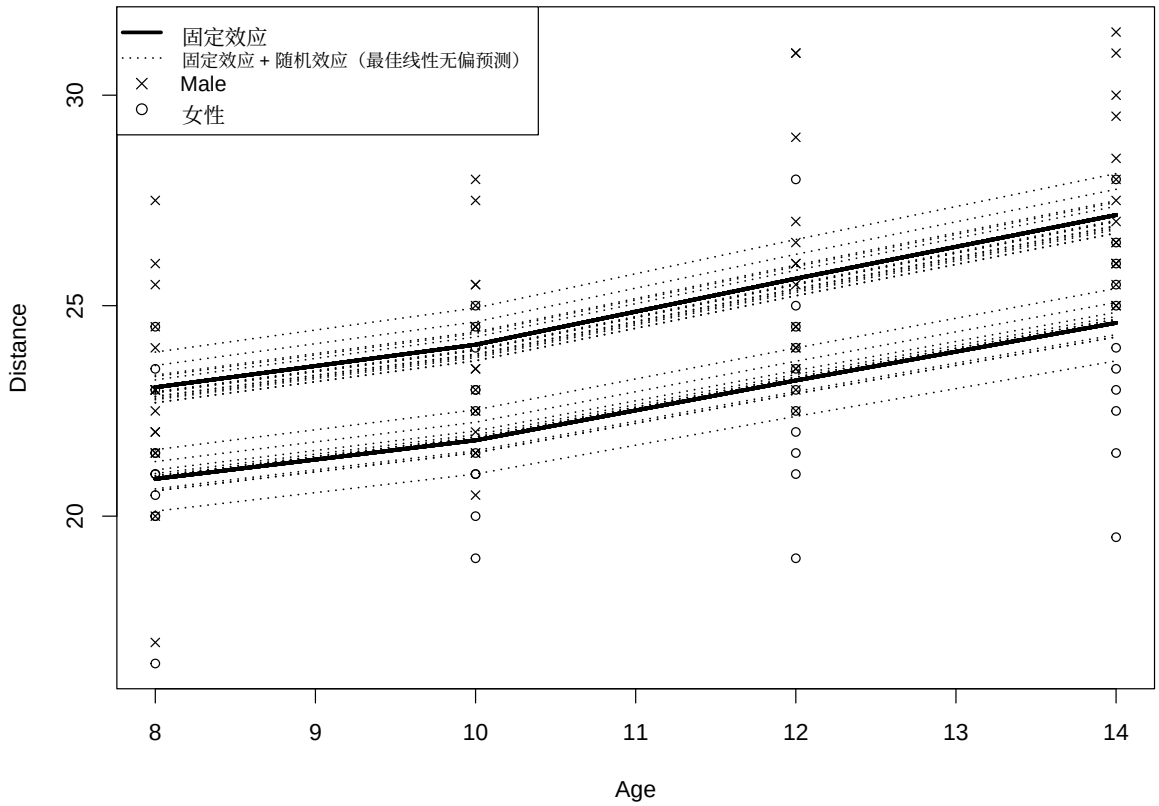

图1

正畸生长数据（$Q = 1$）：观测测量值和拟合曲线。实线表示固定效应分量的$\widehat{\ \ }$拟合值 $X\Theta\boldsymbol{a}_n$，虚线表示基于最佳线性无偏预测的$\widehat{\ \ }$拟合值 $X(\Theta\boldsymbol{a}_n + \hat{\boldsymbol{u}}_n)$。点表示观测值（圆点：女性；叉号：男性）。

基准）。

6.2 Holzinger–Swineford认知测试数据

Holzinger–Swineford（1939）数据包含针对 $N = 300$名学生的 $P = 9$ 项认知测试（Holzinger & Swineford，1939；Jöreskog，1969）。该数据集的经典验证性因子分析提出了三因子结构（视觉、文本、速度）；我们使用NMF-RE来检验在非负性约束下是否能够学习到类似的结构。作为协变量，我们使用反向年龄（`age.rev`）、性别（`sex.2`）和学校指标（`school.GW`），并将其全部重新缩放至 $[0, 1]$。

表2

*Orthodont*生长数据（$Q = 1$）的协变量效应估计值。标准误分别由三明治估计量（*SE*）以及基于一步牛顿更新的乘子（*wild*）自助法（*BSE*）报告。*Wald*统计量为 $z =$估计值/标准误， p值针对 $H_0:\ \theta = 0$ 与 $H_1:\ \theta > 0$ 在 $\Theta \geq 0$下进行单侧检验。

协变量	估计值	SE	BSE	z	p
截距	90.502	2.471	2.450	36.62	< 0.001
Male	9.428	3.056	2.975	3.09	0.0010

NMF-RE 设定（$Q = 3$）。受经典三因子结构的启发，我们设定 $Q = 3$。随机效应处于近乎饱和但受控的状态（$\mathrm{df}_U/(NQ) = 0.958$，上限为 0.96；表1），表明除了观测到的协变量之外，还存在显著的个体层面异质性。该上限未被触发，因此报告的解并未被迫处于边界上。

学习得到的因子结构（图2）。学习得到的 X 恢复了熟悉的三领域分组：一个由视觉测验主导（x_1–x_3），一个由文本测验主导（x_4–x_6），另一个由速度测验主导（x_7–x_9），且无需预设载荷模式（图2）。由于 X 逐元素非负且列归一化，其各元素可直接按基于部分的方式解释为各项测验的相对贡献。

协变量效应的推断（表3）。年龄效应在全部三个成分上均高度显著。性别效应和学校效应因成分而异： `sex.2` 在因子2和因子3中可检测到，但在因子1中检测不到；而 `school.GW`主要影响因子2。即使在饱和比很高的情况下，SE–BSE一致性仍然成立。

补充基准：结构方程模型中的验证性多指标多原因模型／潜变量回归。作为一个基于协方差的结构方程模型基准，我们还使用 `lavaan`（Rosseel，2012），按照教科书中的三因子设定拟合了一个验证性多指标多原因模型（Bollen，1989；Jöreskog & Goldberger，1975），并计算了稳健标准误（完整内容见附录 G

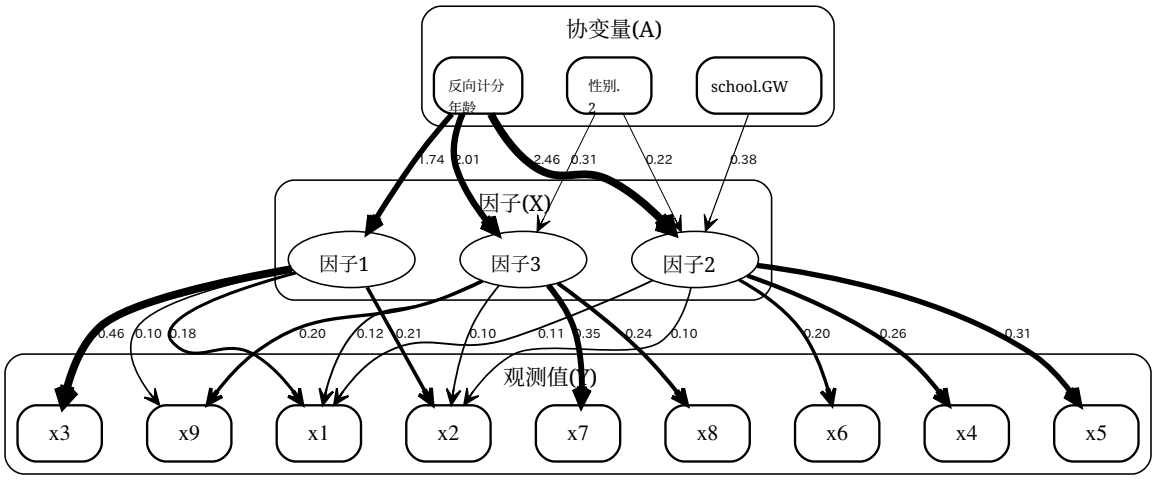

图2

*Holzinger–Swineford*认知测试数据（$Q = 3$）：*NMF-RE*模型下观测测验（Y）、潜在成分（X）与协变量（A）之间关系的估计结构。

结果）。MIMIC分析证实，在验证性设定下，三个协变量都会对潜在结构产生不可忽视的影响。一个值得注意的差异是，MIMIC模型检测到性别对视觉因子有显著的负向效应，而非负矩阵分解-RE中的非负约束 $\Theta \geq 0$ 会将其投影为零；这表明， $\Theta \geq 0$以牺牲检测不受符号限制的效应的能力为代价，换取可解释性和稀疏性。

摘要。 与验证性结构方程模型不同，NMF-RE 在非负性约束下直接从数据中学习测量结构，无需预设载荷模式即可恢复三领域分组。这展示了测量侧变量选择：仅凭 X ，九项$\widehat{}$检验便被划分为视觉、文本和速度成分。在协变量侧， Θ 的推断实现了成分特异性协变量选择——年龄始终显著，而性别和学校效应仅在特定因子中被检测到（表3）。随机效应机制接近饱和但约束不取等，且标准误与 BSE 高度一致（详细的 MIMIC 比较见附录 G）。

表3

*Holzinger–Swineford*认知测试数据（$Q = 3$）的协变量效应估计值，在 *NMF-RE*。格式同表*2*。

协变量	基	估计值	SE	BSE	z	p
反向计分年龄	因子1	1.738	0.089	0.087	19.54	< 0.001
反向计分年龄	因子2	2.456	0.100	0.100	24.46	< 0.001
反向计分年龄	因子3	2.015	0.078	0.077	25.74	< 0.001
性别.2	因子1	0.000	0.059	0.036	0.00	0.5000
性别.2	因子2	0.220	0.071	0.069	3.10	0.0010
性别.2	因子3	0.309	0.060	0.059	5.14	< 0.001
school.GW	因子1	0.000	0.058	0.032	0.00	0.5000
school.GW	因子2	0.378	0.075	0.075	5.05	< 0.001
school.GW	因子3	0.000	0.069	0.042	0.00	0.5000

6.3 基于枢纽协变量的起讫点流量数据

我们考虑来自日本全国道路交通普查（国土交通省，2021年）的起讫点（OD）流量表。令$F \in \mathbb{R}_{\geq 0}^{47\times 47}$表示该OD表；我们分析经过对数变换的数据 $\log(1 + F_{ij})$。选取四个大都市枢纽（东京、爱知、大阪、福冈）作为协变量（$K = 4$），以目的地作为单位（$N = 47$）。响应矩阵 $Y \in \mathbb{R}^{43\times 47}$ 汇集其余 $P = 43$ 个非枢纽起点的流出流量，而 $A \in \mathbb{R}^{4\times 47}$ 汇集枢纽流出流量。

NMF-RE 设定（$Q = 4$）。由于 $Q = 4$ 与枢纽数量相同，随机效应受到强烈正则化（自由度$_U/(NQ) = 0.013$；表1），这表明大部分变异由基于枢纽的均值结构解释。

类概率得分与空间可视化。由于拟合得分为$\widehat{\quad}$ $\widehat{\quad}$非负数，将 $B = \Theta A$ 归一化至列和后，可得到类单纯形表示，并据此通过 $\arg\max_q$进行硬聚类。图 3 按各目的地的

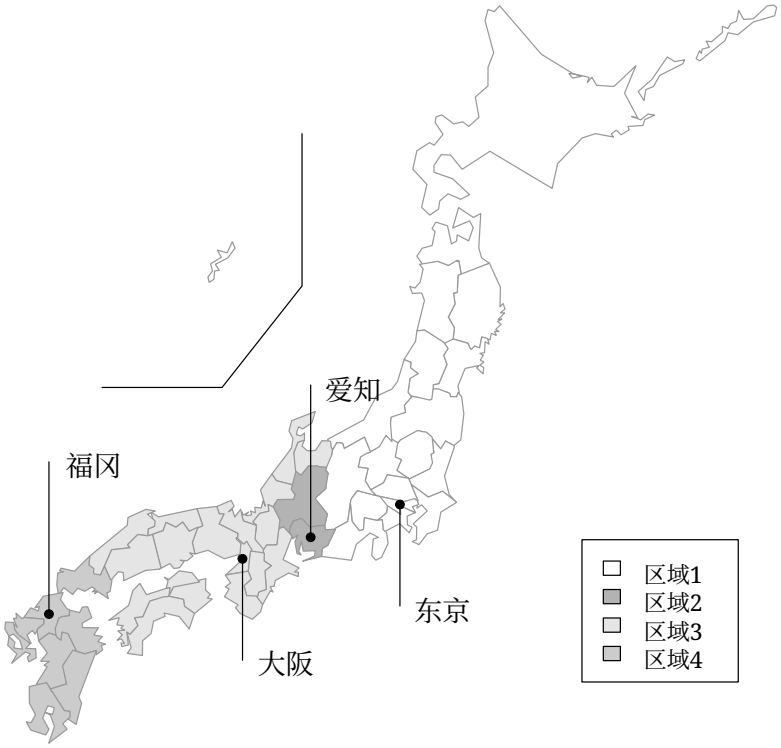

图3

*OD–HUB*数据（$Q = 4$）：*NMF-RE*模型识别出的潜在成分的空间可视化。根据拟合潜在得分，将各目的地都道府县按主导潜在成分着色。

主导成分；每个成分主要与一个枢纽对齐，从而形成清晰的地理分区。

枢纽效应的推断（表4）。 $\hat{\Theta}$呈现出近乎一一对应的枢纽–成分关联：每个成分都有一个主导枢纽系数，而在 $\Theta \geq 0$下，交叉载荷精确为零。主导系数的SE–BSE一致性良好；边界系数的BSE较小，这是由于投影到 $\Theta \geq 0$所致。

补充基准：投影寻踪回归（PPR）。作为灵活的回归基准，我们还拟合了带有 $M = 4$ 岭项的多元投影寻踪回归（Friedman, 1984; Friedman & Stuetzle, 1981）（详情参见附录 G）。估计得到的

投影方向会在每个项中混合四个枢纽，因此，可解释性取决于岭函数诊断，而非直接的枢纽-成分对应。

摘要。 NMF-RE 产生一种结构化的基于部分的分解，并得到非负得分，可用于构建类似概率的表示和清晰透明的地理聚类（图 3；详细的投影寻踪回归比较见附录 G）。测量侧$\widehat{}$选择通过 X 将每个目的地分配到其主导枢纽区域，而协变量侧选择则揭示了起点枢纽与区域成分之间清晰的一一映射（表 4）：每个枢纽协变量仅载荷于一个成分，所有非对角效应的估计值均为零。极小的饱和比表明，该划分基本由固定效应均值结构驱动；dfU 上限虽曾短暂触发，但最终解为约束不取等。

6.4 含历史协变量的主题模型

作为最后一个示例，我们将所提出的 NMF-RE 框架应用于一项基于美国总统就职演说的文本分析问题。按照使用 `quanteda`(Benoit等人，2018) 在 R语言中进行的标准预处理，我们将语料库表示为由非负词频构成的文档-词项矩阵。我们保留在语料库中至少出现 100 次的词语，得到规模适中的词汇表（$P = 67$），涵盖 $N = 59$ 篇文档。

历史协变量（时代指示变量）与编码。 为考察主题结构中的系统性历史转变，我们引入三个互斥的时代指示变量作为协变量：（i）早期（1775–1865；独立至美国内战），（ii）工业时期（1865–1945；工业化与海外扩张），以及（iii）战后（1945年–；第二次世界大战之后）。这些指示变量构成 A 不含截距的各行，从而使每个$\widehat{}$文档恰好属于一个时代。在此编码下， ΘA 的每一列等于$\widehat{}\widehat{}$特定时代的期望潜在得分向量；当文档按时间顺序排列时，固定效应重构 $X\Theta A$会在各时代之间诱导出阶跃函数（分段常数）模式。

NMF-RE设定与随机效应机制。 我们以潜在维度 $Q = 3$拟合NMF-RE，并将拟合基 X 的每一列解释为主题特定的词语画像。令 $B = \Theta A$ 表示固定效应潜在得分矩阵。对于该数据集，随机效应受到强正则化：饱和比为$\mathrm{df}_U/(NQ) = 0.007$，诊断上限为$\mathrm{df}_U^{\max}/(NQ) = 0.02$，且该上限在优化过程中至少被激活一次（表1）。因此，大部分系统性变异被归因于固定效应均值结构，而剩余的时代内特异性则通过 U被大幅收缩。

学习得到的主题。 X 的每一列都是一个主题特定的词语分布；带有$X_{pq}/\sum_{q'} X_{pq'} > 0.5$ 的词语被视为代表性词语。这三个成分分别为（仅作示例）：主题1（制度／法律）、主题2（联邦／宪法／外国）和主题3（现代公民修辞）。

固定效应主题比例（图4）。 在每个文档内对 B 进行归一化，得到主题比例 $\hat{\pi}_{qn} = B_{qn}/\sum_{q'} B_{q'n}$。图4按时间顺序展示这些比例，突显了不同时代主题流行程度系统性的阶梯式变化。

时代特定效应的推断（表5）。 采用无截距编码时，每个系数都表示一个时代特定的潜在得分期望水平。战后时期以主题3为主，而主题1和主题2则集中在较早的时代，且标准误与BSE高度一致。

补充基准：采用相同时代设计的结构主题模型（STM）。 作为稳健性检验，我们还使用相同的无截距时代编码拟合了一个结构主题模型（STM），其中包含$Q = 3$ （Roberts等，2014、2019）（完整结果见附录 G）。按照代表性词语对齐主题后，STM证实了相同的定性历史模式：公民修辞主题从早期到战后时期显著上升，而制度／法律主题在战后时期则变得不突出。

摘要。 NMF-RE 通过非负词语分布和复杂度诊断，针对经结构主题模型证实的时代信号，增加了明确的均值结构推断（详见附录 G，以了解详细比较）。 X中可以看到测量侧选择，这

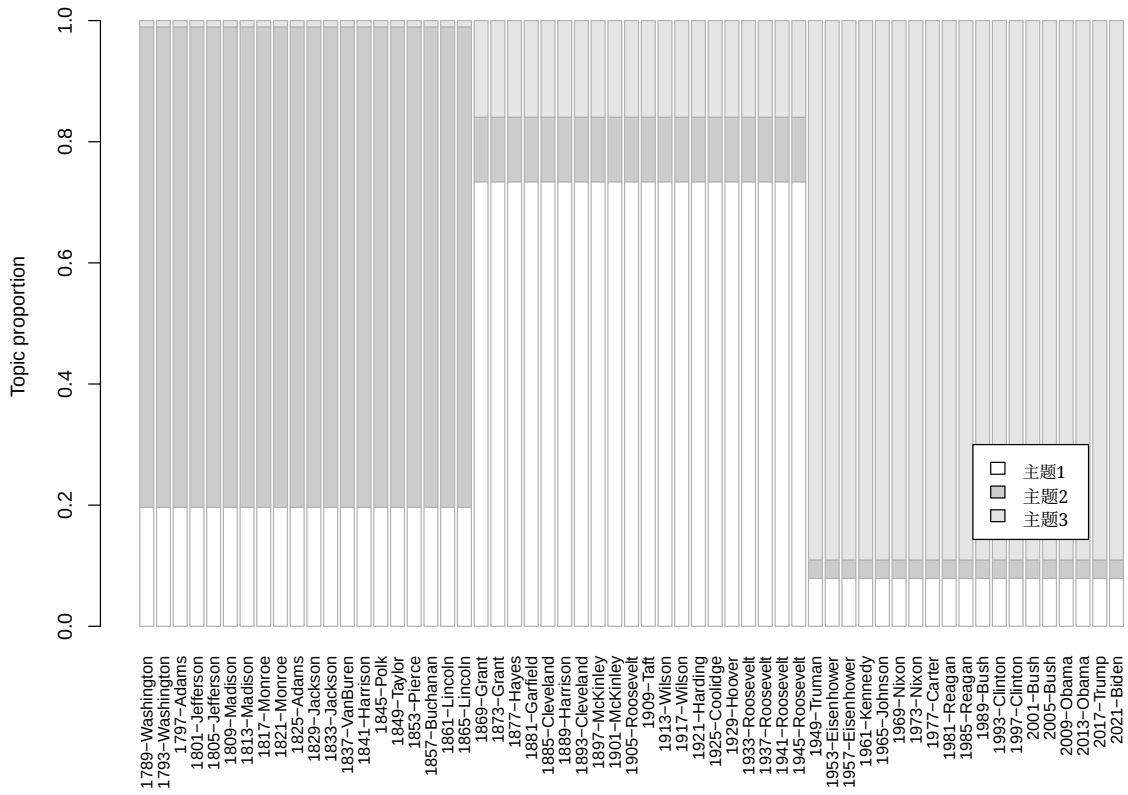

图4

带有历史协变量的主题模型（$Q = 3$）：基于$\widehat{\widehat{\ }}$固定效应分量 $X\Theta A$估计的主题比例。每个条形代表一篇就职演说（文档），展示了主题流行程度在不同历史时期之间的系统性变化。

将每个词归入其主导主题（制度／法律、公民修辞或联邦／宪法）。通过 Θ 进行协变量侧选择后发现，每个时代都有各自的主导主题——早期的主题2（联邦／宪法）、工业时代的主题1（制度／法律），以及战后时期的主题3（公民修辞）；不显著的条目则标示出缺失的时代—主题组合（表5）。随机效应部分实际上并未发挥作用（$\mathrm{df}_U/(NQ)$ 接近零），表明这一时间模式由固定效应分量驱动。

7 讨论

本文提出了 NMF-RE，这是一种协变量驱动的非负矩阵分解的均值结构扩展，将岭正则化随机效应与针对 Θ 的单步乘子自助法推断相结合，从而衔接混合效应建模与非负低秩表示。四个实证示例展示了该方法在不同情境中的应用：纵向增长（Orthodont）、认知测验组（Holzinger–Swineford）、空间起讫点流量（OD–HUB）以及总统演讲主题。这些应用的饱和比从接近零到超过.95 不等，表明 dfU 上限和复杂度诊断在各种不同的随机效应机制下仍能提供有用信息。

7.1 解释与实用诊断

NMF-RE 的一个核心实际问题是，随机效应正则化较弱时可能出现近饱和。饱和比 $\mathrm{df}_U/(NQ)$ 提供了对运行状态的无尺度摘要，而 dfU 上限可防止惩罚项崩塌和退化推断（附录 E 通过受控压力测试展示了这种失效模式）。为便于透明报告，我们建议列出 $\mathrm{df}_U/(NQ)$、上限比率、上限是否曾被触发，以及最终解是紧约束还是约束不取等。

对于推断，基于三明治法的标准误与基于自助法的标准误高度一致（SE–BSE一致性），表明单步线性化是充分的；四个实证示例均呈现了这种一致性。若差异较大，实践者应检查收敛情况、增大 B、检查 dfU 上限状态，或降低模型复杂度。对于近边界系数，应采用方向检验和百分位数汇总进行评估，而非双侧 Wald 近似（详情见附录 D）。

7.2 通过非负潜在结构进行变量选择

如引言所述，NMF-RE 在两个互补方面实现变量选择。在测量侧，非负约束发挥的作用

类似于因子分析中的旋转或稀疏性惩罚，但稀疏的基于部分的载荷是结构性结果，而非通过显式惩罚$\widehat{}$（例如 LASSO）实现。四个实证示例均证实了这一机制： X 始终能将观测变量归入可解释成分——认知测验领域、增长轨迹、地理区域或词语主题——而无需预设载荷模式。

在协变量侧</style}，对 Θ 进行正式推断可以识别哪些协变量会显著影响哪些成分，从而通过检验进行协变量选择。在各个示例中，估计的 Θ 揭示了成分特异性模式：一些协变量效应显著，而另一些的估计值恰好为零，因此无需显式的稀疏性惩罚即可形成选择性结构。综合来看，这两种机制使 NMF-RE 能够在心理测量及相关领域中作为数据驱动的变量选择工具，以一种由约束驱动且易于解释的替代方案补充传统的基于正则化的方法。

7.3 方法学局限

所提出的推断明确以估计得到的正则化$\widehat{}$ $\hat{}$解（X， U）为条件。因此，它不试图传播由估计 X 和 U所带来的不确定性。这种正则化后视角是有意为之：它在避免重复进行约束优化的同时，为低维目标 Θ 提供稳定且可扩展的不确定性量化。与此同时，条件推断可能会$\widehat{}$ $\hat{}$偏保守或偏激进，具体取决于（X， U）$\hat{}$的变异性对特定应用中 Θ分布的影响程度。与此相关的是，非负矩阵分解模型在缩放和平排列方面仅部分可识别，且非凸目标函数可能存在局部极小值；实际上，当 Q不小时，或协变量设计高度不平衡时，热启动和多次初始化可能会有所帮助。

工作高斯尺度模型用于定义得分和信息矩阵，而稳健性则通过三明治协方差估计以及对得分贡献进行乘子自助法重抽样来实现。这些选择在模型设定错误的

似然和后正则化设定中十分常见；然而，在一般非高斯数据生成过程以及对$\widehat{\ }\ \widehat{\ }(X,\ U)$进行数据自适应选择的情况下确立其形式有效性，仍是未来理论研究的重要方向。最后，非负约束 $\Theta \geq 0$ 可能导致近边界系数出现截断和偏度，因此有必要采用第 7.1 节讨论的边界感知诊断方法。

7.4 扩展与未来工作

若干扩展方向顺理成章。首先，开发选择感知推断，将（X, U）带来的不确定性传递至 Θ，有助于明确条件推断何时足够；然而，非凸性和标签切换使重抽样变得复杂。其次，以计数数据似然（如泊松似然）取代平方损失准则，可将 NMF-RE 扩展到均值依赖方差显著的情形。第三，当许多系数接近零时，若要进行有原则的总结，就需要采用超越正态近似的边界感知区间构造。第四，秩 Q 的数据驱动选择是所有低秩模型共同面临的开放问题；将信息准则或交叉验证方法纳入 NMF-RE 框架，可减少对外部秩诊断的依赖。

7.5 总结

NMF-RE 在经典混合效应思想与现代非负低秩表示之间架起了一座实用桥梁，支持以均值结构为中心的策略，用于基于约束的潜变量模型推断，并实现明确的复杂度控制和隐式变量选择。实证示例表明，基准特定方法（线性混合效应模型、因子分析、结构主题模型）可在各自领域内印证 NMF-RE 的发现；与此同时，NMF-RE 提供了一个统一框架，可在单一模型和推断体系下涵盖全部四种情形。

声明与陈述

资助。 本研究部分得到了日本学术振兴会科学研究费助成事业项目编号22K11930、25K15229、24K03007、25H00482，滋贺大学

可持续发展与环境研究中心的项目研究经费，以及藤森工艺基金会的研究资助。

利益冲突。 作者没有需要披露的相关财务或非财务利益。

伦理审批。 不适用。

数据可用性。 所有使用的实证数据集均可通过 R 软件包（`nlme`、 `lavaan`、 `quanteda`）或日本 e-Stat 门户网站 (https://www.e-stat.go.jp) 公开获取。复现本文结果所需的所有分析脚本、模拟代码和预处理数据文件均可在 https://github.com/ksatohds/nmfre-paper 获取。

代码可用性。 用于实现协变量驱动的非负矩阵分解（用于初始化）的 R 软件包 `nmfkc` 可在 https://ksatohds.github.io/nmfkc/ 公开获取。

人工智能工具的使用。 Claude Code（Anthropic，Claude Opus 4 模型，https://claude.ai/code，访问时间为2026年2月至3月）用于 LaTeX 排版，编译、交叉引用核验和稿件校对。没有任何科学内容——包括方法论、数学推导、数据分析或结果解读——由人工智能工具生成。作者对所有内容的准确性和完整性承担全部责任，对所有内容的准确性和完整性负责。

表4*OD–HUB*数据的协变量效应估计值（$Q = 4$）。格式同表*2*。

协变量	基	估计值	SE	BSE	z	p
东京	区域1	8.896	0.801	0.763	11.11	< 0.001
东京	区域2	0.000	0.302	0.165	0.00	0.5000
东京	区域3	0.000	0.292	0.163	0.00	0.5000
东京	区域4	0.000	0.250	0.153	0.00	0.5000
爱知	区域1	0.000	0.597	0.354	0.00	0.5000
爱知	区域2	4.381	0.633	0.642	6.92	< 0.001
爱知	区域3	0.000	0.524	0.309	0.00	0.5000
爱知	区域4	0.000	0.503	0.265	0.00	0.5000
大阪	区域1	0.000	0.546	0.295	0.00	0.5000
大阪	区域2	0.000	0.322	0.197	0.00	0.5000
大阪	区域3	6.895	0.637	0.656	10.82	< 0.001
大阪	区域4	0.000	0.413	0.230	0.00	0.5000
福冈	区域1	0.000	0.483	0.200	0.00	0.5000
福冈	区域2	0.000	0.190	0.095	0.00	0.5000
福冈	区域3	0.000	0.520	0.254	0.00	0.5000
福冈	区域4	5.575	0.373	0.374	14.94	< 0.001

表5主题模型（$Q = 3$）的协变量效应估计值。格式同表*2*。

时期	主题	估计值	SE	BSE	z	p
早期（1775–1865）	主题1	46.744	12.036	11.914	3.88	< 0.001
早期（1775–1865）	主题2	188.558	36.392	37.162	5.18	< 0.001
早期（1775–1865）	主题3	2.491	4.647	3.362	0.54	0.2960
工业时期（1865–1945年）	主题1	164.855	24.441	24.985	6.75	< 0.001
工业时期（1865–1945年）	主题2	23.901	13.113	12.944	1.82	0.0342
工业时期（1865–1945年）	主题3	35.862	14.349	14.026	2.50	0.0062
战后（1945–）	主题1	16.528	9.538	8.952	1.73	0.0416
战后（1945–）	主题2	6.361	7.375	5.826	0.86	0.1942
战后（1945–）	主题3	186.287	11.182	10.846	16.66	< 0.001

附录 A

dfU 上限的实用校准

令 $r^{\max} \equiv \mathrm{df}_U^{\max}/(NQ)$ 表示上限比率，并令 X_{fix} 表示使用 $U \equiv 0$ 进行初始化后，通过协变量驱动NMF拟合得到的基。对于固定的 X，定义饱和比函数

$$r_X(\lambda) \equiv \frac{\mathrm{df}_U(\lambda)}{NQ} = \frac{1}{Q}\sum_{q=1}^{Q}\frac{d_q}{d_q+\lambda},$$

其中 $d_1, \ldots, d_Q$ 是 $X^\top X$的特征值。在对数尺度网格上计算 $r_{X_{\mathrm{fix}}}(\lambda)$，即可在运行完整的 NMF-RE 算法前，通过简单的“$\lambda \leftrightarrow r$”查找表进行查询。如果认为数值$\lambda < \lambda_{\min}$对于该应用而言不合理，则设置

$$\mathrm{df}_U^{\max} = r_{X_{\mathrm{fix}}}(\lambda_{\min})\, NQ$$

可得到一种有原则依据的上限，避免反复试错扫描。此预校准仅使用 X_{fix}来选择 $\mathrm{df}_U^{\max}$；优化过程中则使用当前的 X 通过$\lambda_{\mathrm{cap}}(X)$来实施上限。

为实现透明报告，我们区分：（i）在优化过程中上限是否曾被激活；（ii）最终收敛解是紧约束（$\mathrm{df}_U = \mathrm{df}_U^{\max}$）还是约束不取等（$\mathrm{df}_U < \mathrm{df}_U^{\max}$）。我们报告实际饱和比 $\mathrm{df}_U/(NQ)$、上限^比 $\mathrm{df}_U^{\max}/(NQ)$、最终 λ以及激活与紧约束指标。

附录B

估计算法

算法 1 展示了针对随机效应采用基于自由度的上限时，NMF-RE 的完整分块估计过程，具体如正文第 3 节所述。

备注（稳定化乘法更新与目标函数监控）。当 U 为实数时，中间矩阵 $B_U = \Theta A + U$ 和 $Y_U = Y - XU$ 可能包含负数元素，因此不能直接使用经典的欧氏乘法更新。因此，我们仅将逐元素正部 $B_U^+ = \max(B_U, 0)$ 和$Y_U^+ = \max(Y_U, 0)$ 用作稳定化替代量，以确保乘法比率定义良好，并保持 X 和 Θ的非负性。由于这些稳定化更新并非针对完整目标函数 $\mathcal{L}(X, \Theta, U)$ 推导出的精确优化最小化方案，我们会在每个分块更新后明确监控 $\mathcal{L}$ ，并在某次迭代使目标函数增大时采用简单的下降保护措施（回滚或阻尼）。在数值实验中，这种保护措施很少被触发，但它能提供透明的保障，避免出现病态步骤。

备注（热启动方差估计）。算法 1 将 $\lambda > 0$ 视为给定的初始惩罚，仅根据 dfU 上限进行上调。实际上， $\lambda = \sigma^2/\tau^2$ 取决于噪声方差 σ^2，而在拟合前该方差未知。因此，我们的实现采用热启动调度：在最初几十次迭代中， σ^2 保持为固定的初始值（例如 $\sigma^2 = 1$），以便分块更新趋于稳定（X、Θ、U），之后再通过学习率较小的指数移动平均，逐步将其更新至当前残差均方 $\|Y - X(\Theta A + U)\|_F^2/(PN)$。由于 dfU 上限为 λ设定了下限，这种自适应调度不会使 λ 降至零；它只会使工作方差尺度与数据相匹配，而不会破坏优化过程的稳定性。

算法 1 采用基于 df 的随机效应上限的 NMF-RE 分块估计**要求：** 非负数据 $Y \in \mathbb{R}_{\geq 0}^{P\times N}$、协变量 $A \in \mathbb{R}^{K\times N}$、秩 Q、初始惩罚$\lambda > 0$、上限水平 $\mathrm{df}_U^{\max}$、容差 `tol`、最大迭代次数 `maxit`。

确保： X，Θ，U以及最终的 λ。

初始化：

使用 $U \equiv 0$ 拟合协变量驱动的非负矩阵分解，得到 (X, Θ)；设置 $U \leftarrow 0$。对 X 进行归一化，使得 $\mathbf{1}_P^\top X = \mathbf{1}_Q^\top$，并将缩放吸收到 Θ中。

对于 $t = 1,2,\ldots,$ `maxit` **执行**

计算当前目标函数 $L_{\mathrm{old}} \leftarrow \|Y - X(\Theta A + U)\|_F^2 + \lambda\|U\|_F^2$。**(i) 强制执行自由度上限（诊断性保护）：** **计算** $X^\top X$**的特征值** $d_1, \ldots, d_Q$。**求取** $\lambda_{\mathrm{cap}}(X) = \inf\{\lambda' > 0 : N\sum_{q=1}^{Q} \frac{d_q}{d_q+\lambda'} \leq$ **自由度**$_U^{\max}\}$**（例如，使用二分法）。** $\lambda \leftarrow \max\{\lambda, \lambda_{\mathrm{cap}}(X)\}$。**(ii)** U**步骤（岭回归／类似最佳线性无偏预测的更新）：** $M \leftarrow (X^\top X + \lambda I_Q)^{-1} X^\top$。$U \leftarrow M\big(Y - X(\Theta A)\big)$。**对** U **进行行中心化，使每一行在各单位上的均值为零。(iii)** X**步骤（稳定化乘性更新 +再归一化）：** $B_U \leftarrow \Theta A + U$，$B_U^+ \leftarrow \max(B_U, 0)$**（逐元素）。** $X_{\mathrm{cand}} \leftarrow X \odot \Big(Y(B_U^+)^\top\Big) \oslash \Big(X(B_U^+(B_U^+)^\top)\Big)$。**归一化各列：** $X_{\mathrm{cand}} \leftarrow X_{\mathrm{cand}} D^{-1}$，**使得** $\mathbf{1}_P^\top X_{\mathrm{cand}} = \mathbf{1}_Q^\top$。**重新缩放** $(\Theta, U) \leftarrow (D\Theta, DU)$，**以保持** $X(\Theta A + U)$**不变。(iv)** Θ**步骤（稳定化乘性更新）：** $Y_U \leftarrow Y - XU$，$Y_U^+ \leftarrow \max(Y_U, 0)$**（逐元素）。** $\Theta_{\mathrm{cand}} \leftarrow \Theta \odot \Big(X^\top Y_U^+ A^\top\Big) \oslash \Big((X^\top X\Theta)(AA^\top)\Big)$。**(v) 使用真实工作目标函数进行下降保护：**

构造候选解 $(X, \Theta) \leftarrow (X_{\mathrm{cand}}, \Theta_{\mathrm{cand}})$ 并计算 $L_{\mathrm{new}} \leftarrow \|Y - X(\Theta A + U)\|_F^2 + \lambda\|U\|_F^2$。

如果 $L_{\mathrm{new}} > L_{\mathrm{old}}$ **则** （可选）应用阻尼 / 回滚，例如，设置 $X \leftarrow X_{\mathrm{old}}$，$\Theta \leftarrow \Theta_{\mathrm{old}}$ a

并跳过

或 `rep`

附录 C

单步映射的渐近论证

在给定（X，λ）且通过岭回归剖面消去 U 后，Θ 可视为一个（可能设定错误的）Z估计量，其求解方程为 $\sum_{n=1}^{N} \mathrm{vec}\{S_n(\Theta)\} = 0$。根据 Z估计量的标准正则性条件——得分映射的光滑性、目标点处雅可比矩阵的非奇异性（等价于约化信息量非奇异），以及初始估计量 Θ 是 $\sqrt{N}$-一致的，且位于目标点的一个 $O_p(N^{-1/2})$ 邻域内——单步牛顿更新

$$\mathrm{vec}(\Theta^{\mathrm{os}}) = \mathrm{vec}(\widetilde{\Theta}) - \mathcal{I}(\widetilde{\Theta})^{-1} \sum_{n=1}^{N} \mathrm{vec}\{S_n(\widetilde{\Theta})\}$$

在内部区域中，它与完全迭代的Z估计量（即估计方程的根）渐近等价，即 $\sqrt{N}\,(\Theta^{\mathrm{os}} - \widehat{\Theta}) \to 0$ 依概率收敛。这是M/Z估计量渐近线性的经典推论，也是正则化后推断中一步法/去偏构造的理论基础（例如参见Huber等人，1967；Javanmard与Montanari，2014；Taylor与Tibshirani，2015；van de Geer等人，2014；White，1980）。在我们的自助法构造中，$S^{*}(\Theta)$在一步公式中表示乘子重抽样得分（正文第4.4–4.5节）。非负约束 $\Theta \geq 0$ 通过投影一步复制来处理；在边界附近，极限分布可能是非高斯的，因此与对称的Wald 近似相比，方向检验和百分位数型汇总更为合适。

附录 D

有效性、适用范围与实用诊断

本附录对正文第4.7节概述的推断范围、正则性条件和实用诊断进行了完整讨论。

推断范围（条件目标）。本文的推断针对$\widehat{\ }$ $\widehat{\ }$协变量效应矩阵 Θ，以所选的正则化表示（X，U）及估计算法得出的最终岭水平 λ 为条件。因此，在固定协变量 A的情况下，通过逐单元（逐列）的得分贡献评估抽样变异性；而由 X 的非凸估计以及 U 的正则化拟合所带来的额外不确定性，则作为后正则化近似的一部分加以处理（Javanmard与Montanari，2014；van de Geer等人，2014；White，1980）。

$\widehat{\ }$**估计方程观点与稳健协方差**。将（X，λ）视为固定量，并通过岭回归剖面化消去 U 后，Θ 可视为一个（可能误设的） Z 估计量，其求解的估计方程为

$$\sum_{n=1}^{N} \mathrm{vec}\{S_n(\Theta)\} = \mathbf{0},$$

其中 S_n（Θ）是逐单元的得分贡献。在 M/Z估计量的标准正则性条件下——特别是，（i）各单元之间相互独立 n（或聚类稳健适配有效），（ii）得分贡献具有有限二阶矩，以及（iii）约化信息矩阵非奇异——即使高斯工作模型存在误设，三明治协方差估计量仍可提供大样本稳健标准误（Cameron等人，2008；Huber等人，1967；White，1980）。

乘子（wild）自助法：一种快速重抽样近似。乘子（wild）自助法通过随机再加权对得分贡献的经验分布进行重抽样，避免反复进行约束再优化。在温和的矩条件下，这类乘子方案能够一致地近似估计方程解的光滑泛函分布（Cameron等人，2008；Mammen，1993）。结合一步牛顿映射，可得到一种计算效率高的

后正则化设定下基于重抽样的不确定性量化近似（Javanmard与Montanari，2014；van de Geer等人，2014）。

在 $\Theta \geq 0$下的边界行为。由于一步更新不受约束，自助复制会被投影到非负正交象限。因此，边界附近的系数可能出现截断和偏度，所以与对称高斯 Wald 近似相比，方向性（单侧）检验和百分位型自助汇总通常更为适当。在这类边界情形下，基于夹心估计的标准误与基于自助法的标准误之间的差异可能是由投影机制本身造成的，不应将其解读为单步线性化失效。

实用诊断和报告。 为确保报告透明，并防止近饱和的随机效应更新，我们建议记录：(i) 饱和比 $\mathrm{df}_U/(NQ)$ 和上限水平 $\mathrm{df}_U^{\max}/(NQ)$（以及在相关情况下，记录上尾汇总，例如在迭代过程中或多次重新启动之间的 $\mathrm{dfU}_{0.99}$ ）；(ii) 在优化过程中是否曾启用 dfU 上限，以及最终解是紧约束$\widehat{}$($\mathrm{df}_U = \mathrm{df}_U^{\max}$)还是约束不取等 ($\mathrm{df}_U < \mathrm{df}_U^{\max}$)，并同时报告最终的 λ；(iii) 对于位于可行域内部较深处的系数，比较三明治法与基于自助法的标准误是否一致；以及 (iv) 在 Q 中等或较大时，对初始化的敏感性（多次初始化）和收敛行为。这些检查有助于评估所选低秩表示是否足够稳定，能够支持针对 Θ的预期条件推断。

附录 E

随机效应饱和压力测试

本附录报告了一项额外的定向压力实验，旨在验证 dfU 上限作为防止随机效应饱和的安全装置所发挥的作用。与正文中的基线模拟相比，该压力测试采用了更灵活的设置（$P = 4$、 $Q = 3$、 $N = 100$），其中工作随机效应惩罚项被有意削弱。该实验旨在演示失效模式（近饱和和 λ 崩溃），而不是作为正则化拟合条件下推断准确性的基准。具体而言，我们通过使用 $\sigma^2_{\text{fit}} = 1$ 和 $\tau^2_{\text{拟合}} = 1000$来初始化方差比，使 $\lambda = \sigma^2/\tau^2$ 极小；这会在没有保护措施的情况下促进近饱和的 U 更新，并可能使 λ趋近于零。

在此压力测试中，我们使用一个合成三趋势基矩阵 $X_{\text{true}} \in \mathbb{R}^{4\times 3}$，其列和归一化为 1：

$$X_{\text{true}} = \Big((0.45, 0.30, 0.15, 0.10)^\top, (0.25, 0.25, 0.25, 0.25)^\top, (0.10, 0.15, 0.30, 0.45)^\top\Big),$$

$\mathbf{1}_P^\top X_{\text{true}} = \mathbf{1}_3^\top$.截距在三个趋势之间均匀分配，男性效应仅施加于 Trend1： $\Theta_{\text{true}}[\text{Trend1}, \texttt{male}] = 9.4285$ 以及 $\Theta_{\text{true}}[\text{Trend2}, \texttt{male}] = \Theta_{\text{true}}[\text{Trend3}, \texttt{male}] = 0$， 而$\Theta_{\text{true}}[\text{Trend}q, \texttt{intercept}] = 90.502/3$ 对于 $q = 1$， 2， 3。数据由与基线模拟相同的数据生成过程生成，且包含 $U \in \mathbb{R}^{3\times N}$。

我们比较三种 dfU 上限设置：严格上限 $\text{df}_U^{\max}/(NQ) = 0.21$、宽松上限 0.99，以及无上限。目的是演示当 U 变得（近乎）饱和时会发生什么，以及 dfU 上限能否防止由此导致的 Θ 推断退化。

无 dfU 上限时的失效模式。 表 E1 是针对随机效应饱和的定向失效模式演示。在严格上限（0.21）下，$\text{dfU}_{0.99}$ 保持在接近 0.21，拟合惩罚项仍处于稳定区间（平均 Lambda 约为 1.3–1.7），从而使 U 更新远离近饱和。在该区间内，零假设下基于 BSE 的单侧检验尺寸保持接近 5%，Wald 覆盖率接近（或略高于）名义水平。

当 t 上限较宽松（0.99）或已关闭时，$\text{dfU}_{0.99}$ 趋近于 1，拟合的

惩罚项崩塌（平均 Lambda 低于 10^{-3}），对应于几乎不带惩罚的 U更新。在零假设下，这会导致退化的 推断：沃尔德检验几乎从不拒绝，覆盖率膨胀至 1.00，表明 U 吸收了几乎所有特异性变异，留给 Θ的稳定信息所剩无几。

在备择假设下，各种上限设置下的拒绝率仍然很高，但三种设置的覆盖率都很差。值得注意的是，严格上限（0.250/0.280）下的覆盖率低于宽松上限（0.650/0.640）下的覆盖率，因为严格上限对 U 的约束更强，从而^将更多偏差转移到 Θ。这正是当前压力测试设计中预期出现的情况：刻意设置的较弱初始惩罚与更高维的随机效应空间^（$Q = 3$）相结合，形成了一种使 θ 在所有上限设置下都存在显著偏差的情形，因此单凭覆盖率并不是恰当的成功判据。该表确实展示了严格上限与宽松上限／无上限设置之间的对比：没有^上限时，拟合惩罚项会崩塌（$\lambda \approx 0$），推断也会退化；而^严格上限则使 λ 保持在稳定状态，并保留了有意义（尽管并不完美）的不确定性量化。在校准良好的条件下——如基准 $Q = 1$模拟（表 F1）所示——检验水平和覆盖率都接近名义值。总的来说，压力测试^表明，dfU 上限是一项具有实际重要性的防护措施：它可防止 λ 崩塌，并阻止在其他情况下会出现近饱和 U的情形下产生病态推断。

随机效应复杂度监测。在表 F1 所报告的基准 $Q = 1$ 设计中，饱和比均值 $\mathrm{df}_U/(NQ)$ 在默认上限 $\mathrm{df}_U^{\max}/(NQ) = 0.21$ 下保持在 0.20 附近，表明 U 处于适度正则化的情形。表 E1 中的$Q = 3$压力实验说明了这项监测的重要性：如果允许 $\mathrm{df}_U/(NQ)$ 接近 1（上限宽松或关闭），拟合惩罚项可能会趋于零，而针对 Θ的推断可能会退化或严重失真。

表 E1

随机效应饱和压力测试（$P = 4$、 $Q = 3$、 $N = 100$； $R = 1000$； $B = 1000$）。拟合程序以有意设置的较弱随机效应惩罚项（$\sigma^2_{fit} = 1$、$\tau^2_{fit} = 1000$）进行初始化；这种设置（若无保护措施）会促使 $\widehat{\lambda}$ 趋近于零，并使 U 接近饱和。假设表示真实参数情形：零假设（$\theta_{male} = 0$）或备择假设（$\theta_{male} \neq 0$）。*dfU*$_{0.99}$ 是饱和比$\mathrm{df}_U/(NQ)$（在蒙特卡洛运行中的第*99*百分位数），平均 *Lambda* 是拟合惩罚项λ的蒙特卡洛均值。拒绝率是基于*BSE*的 *Wald*拒绝率（零假设下为单侧检验，备择假设下为双侧检验）；覆盖率是名义上的 *95% Wald* 覆盖率。本表旨在展示近饱和导致的退化以及 dfU 上限的保护作用，而非对不同上限下的总体性能进行比较。

假设	$r^{\max}$误差		Bias	标准差	dfU$_{0.99}$	平均 Lambda	拒绝率	覆盖率
Null	0.21	高斯	0.131	0.220	0.210	1.5988	0.050	0.970
Null	0.99	高斯	0.093	0.106	0.990	<0.001	0.000	1.000
Null	off	高斯	0.095	0.115	0.999	<0.001	0.000	1.000
Null	0.21	以实验为中心	0.105	0.209	0.210	1.6462	0.020	0.990
Null	0.99	以实验为中心	0.089	0.102	0.990	<0.001	0.000	1.000
Null	off	以实验为中心	0.090	0.104	0.999	<0.001	0.000	1.000
备择假设 0.21		高斯	-2.477	1.002	0.210	1.3792	0.960	0.250
备择假设 0.99		高斯	-1.151	0.998	0.990	<0.001	1.000	0.650
备择假设关闭		高斯	-1.710	0.785	0.999	<0.001	1.000	0.110
备择假设 0.21		以实验为中心	-2.388	1.052	0.210	1.3345	0.980	0.280
备择假设 0.99		以实验为中心	-1.118	0.797	0.990	<0.001	1.000	0.640
备择假设关闭		以实验为中心	-1.722	0.739	0.999	<0.001	1.000	0.100

附录 F

详细模拟结果讨论

本附录给出完整的蒙特卡洛结果，并提供详细解释。

SE–BSE一致性能够（以及不能）验证什么。 当系数位于可行域内部时，SE与BSE之间的一致性提供了直接证据，表明估计量能被其局部线性化很好地刻画，而这正是单步乘子自助法所使用的近似。然而，在由 $\Theta \geq 0$引起的边界情形中，SE–BSE一致性并非恰当的成功标准：将一步复制投影到 $\Theta \geq 0$ 会通过在零处增加截断而改变有限样本分布。在该情形下，相关的验证在于方向检验和区间汇总是否仍保持合理的检验水平与覆盖率。

内部区域（备择假设）：局部线性化效果良好。 在备择假设下，对于两种误差分布以及所有 N，均值(SE)与均值(BSE)都很接近，并且随着 N 增大，覆盖率趋近于名义水平。这一表现与预期的正则化后解释一致：一旦所选低秩结构稳定且系数远离边界，单步自助法便能提供准确的局部不确定性近似。

边界情形（零假设）：投影会导致 SE 与 BSE 之间出现系统性差异。 在零假设下，均值（BSE）系统性地小于均值（SE）。这是预期结果：单步自助法的重复样本被投影到 $\Theta \geq 0$ 上，使概率质量集中于零点，因此与对称高斯 Wald 近似相比，经验离散程度有所降低。尽管存在由边界引起的截断，并导致 $\wedge{\theta}_{\texttt{male}}$ 中的正偏差，单侧拒绝率仍接近名义 5% 水平，且 Wald 型覆盖率在各个 N 中略显保守。百分位数自助法区间表现出类似的覆盖率，表明在这种情形下，投影后的单步自助法仍可作为考虑边界的置信度汇总，在实践中发挥作用。

对偏斜观测噪声的稳健性。 将高斯误差替换为中心化指数误差后，偏差、标准差、均方根误差和覆盖模式在定性上相似。在零假设下，基于 BSE 的单侧检验在有限样本中略微偏于自由，但这种偏差幅度不大，也不会随着 N 增大而加剧。总体而言，这表明在这种类 Orthodont 低秩设计中，所提出的正则化后推断在测量噪声偏离高斯性的程度适中时，至少仍具有良好的稳定性。

表 F1

基于 *Orthodont* 的设计（$P = 4, Q = 1$）下，男性效应 θ_{male} 的蒙特卡洛结果，重复次数为 $R = 1000$，乘子自助法规模为 $B = 1000$。*Hypothesis* 表示真实参数情形：零假设（$\theta_{male} = 0$）或备择假设（$\theta_{male} \neq 0$）。*Reject* 表示 *Wald* 拒绝率（零假设下为单侧检验，备择假设下为双侧检验）。*Mean* 表示标准误估计（*SE* 或 *BSE*）的蒙特卡洛均值。*“a/b”* 形式的条目报告“基于 *SE* / 基于 *BSE*”的量。*PctCover* 表示 *95%* 百分位数自助法置信区间的覆盖率，*dfU* 表示 $\mathrm{df}_U/(NQ)$ 的蒙特卡洛均值。

N	假设	Bias	标准差 均方根误差	Mean	拒绝率	覆盖率 PctCover	dfU
高斯误差							
27	Null	0.373	0.526	0.645	0.853/0.572	0.064/0.082	0.956/0.951	0.956	0.203
27	备择假设	-0.013	0.919	0.919	0.842/0.826	1.000/1.000	0.912/0.902	0.910	0.201
100	零假设	0.179	0.246	0.304	0.443/0.304	0.048/0.055	0.977/0.970	0.978	0.204
100	备择假设	-0.010	0.430	0.430	0.442/0.439	1.000/1.000	0.953/0.952	0.952	0.201
200	零假设	0.144	0.185	0.235	0.324/0.226	0.057/0.062	0.970/0.966	0.970	0.206
200	备择假设	-0.011	0.325	0.325	0.324/0.323	1.000/1.000	0.949/0.948	0.946	0.201
中心化指数误差							
27	Null	0.333	0.473	0.579	0.860/0.567	0.057/0.076	0.969/0.963	0.967	0.201
27	备择假设	-0.025	0.854	0.854	0.850/0.835	1.000/1.000	0.931/0.926	0.928	0.201
100	零假设	0.186	0.267	0.326	0.444/0.304	0.063/0.074	0.967/0.963	0.966	0.202
100	备择假设	-0.011	0.457	0.457	0.443/0.441	1.000/1.000	0.939/0.937	0.938	0.201
200	零假设	0.155	0.191	0.246	0.323/0.227	0.059/0.071	0.966/0.961	0.964	0.204
200	备择假设	0.009	0.328	0.328	0.323/0.322	1.000/1.000	0.940/0.941	0.938	0.201

附录 G

补充基准比较

本附录汇集了正文中四个实证示例各自与既有方法进行的基准比较。这些比较旨在将 NMF-RE 的结果置于标准方法的背景下加以考察，并非用于进行直接的性能评估。

Orthodont：线性混合效应建模（线性混合效应模型）

针对 Orthodont 数据的标准线性混合效应模型（线性混合效应模型）会指定一个预先规定的时间基函数（通常为截距和线性年龄项），然后在该基函数上引入随机效应。对于受试者 n 在年龄 $t \in \{8、10、12、14\}$时，一种常见设定为

$$y_{nt} = (\beta_0 + \beta_{\text{sex}}\text{male}_n) + (\beta_1 + \beta_{\text{age:sex}}\text{male}_n)\, t + (b_{0n} + b_{1n}t) + \varepsilon_{nt},$$

因此，分析人员必须选择并论证时间设计事先（并在需要时加入曲率项，例如 t^2 或样条）。仅有四个观测时间时，这一设计选择可能影响重大。

为了使性别主效应在观测到的年龄范围内具有可解释性，我们还使用 $t_c = t - 11$ 拟合了年龄中心化形式：

$$y_{nt} = \beta_0 + \beta_1 t_c + \beta_2 \text{male}_n + \beta_3 t_c\, \text{male}_n + (b_{0n} + b_{1n}t_c) + \varepsilon_{nt}.$$

在这一参数化中， β_2 表示男性在 11 岁时对距离的效应。拟合模型表明，男性的平均距离高出约 2.32 毫米（$\hat{}\beta_2 = 2.321, p = 0.0054$），且通过年龄与性别交互作用，男性的生长速率也更高（$\hat{}\beta_3 = 0.305$, $p = 0.0264$），因此男性和女性的斜率分别为 0.784 和 0.480 毫米/年。随机效应估计表明，水平和斜率均存在不可忽略的受试者间异质性（标准差分别约为 1.83 和 0.18），残差标准差约为 1.31。此处的线性混合效应模型（LMM）作为参考基准，证实 Orthodont 数据同时表现出显著的个体层面异质性和性别依赖的均值结构。

Holzinger–Swineford：结构方程模型中的验证性多指标多原因模型 / 潜变量回归

为了将 NMF-RE 的均值结构方法与经典的基于协方差的结构方程模型进行比较，我们使用 `lavaan`package 拟合了一个验证性多指标多原因模型（Bollen, 1989年；Jöreskog & Goldberger, 1975年）（Rosseel, 2012年）。按照教科书中的三因子设定，将九项检验分配给 `visual` (x_1–x_3)、 `textual` (x_4–x_6) 和 `speed`(x_7–x_9)，并对这三个因子进行协变量回归。我们采用稳健（三明治法）标准误（MLR），并将因子方差固定为 1（`std.lv=TRUE`），以便更稳定地解释回归系数。

表 G1 报告了潜变量回归系数。在该验证性分析中，`sex.2` 与 `visual` 因子显著相关，而 `school.GW` 与 `textual`因子高度相关； `age.rev`表现出因子特异性效应，包括与 `speed` 的明显关联。将每个协变量的三个回归系数同时约束为零的联合沃尔德检验，对 `age.rev`、`sex.2`和 `school.GW`均予以拒绝，表明在验证性设定下，三个协变量都对潜在结构产生了不可忽略的影响。作为总体参考，拟合的 MIMIC模型的 CFI=0为 0.893，TLI=0为 0.840，RMSEA=0为 0.092，SRMR=0为 0.061。

OD–HUB：投影寻踪回归（PPR）

作为一种不施加低秩分解约束的灵活回归基准模型，我们还拟合了多元投影寻踪回归（PPR）（Friedman，1984；Friedman 与 Stuetzle，1981）。将目的地视为观测值，我们使用 $M = 4$ 个岭项，将 43 维响应 $\boldsymbol{y}_n \in \mathbb{R}^{43}$ 回归到 4 维枢纽流量预测变量 $\boldsymbol{a}_n \in \mathbb{R}^4$ 上：

$$\boldsymbol{y}_n = \sum_{m=1}^{M} \boldsymbol{\beta}_m \, g_m(\boldsymbol{\alpha}_m^\top \boldsymbol{a}_n) + \boldsymbol{\varepsilon}_n,$$

其中， $\boldsymbol{\alpha}_m \in \mathbb{R}^4$ 是各响应变量共享的投影方向， $g_m(\cdot)$ 是平滑的单变量脊函数，$\boldsymbol{\beta}_m \in \mathbb{R}^{43}$ 是响应特异性权重。在 R语言中使用 `ppr` 并结合$M = 4$ 项（内部允许最多 10 个候选解），拟合模型达到了

表 G1

Holzinger–Swineford 数据的验证性多指标多原因模型/潜变量回归结果（结构方程模型）：以协变量回归三个教科书因子。标准误在稳健最大似然估计下采用三明治法计算，*p* 值为双侧检验。

协变量	因子	估计值	SE	z	p
反向计分年龄	视觉	0.075	0.075	0.998	0.318
反向计分年龄	文本	0.183	0.059	3.131	0.002
反向计分年龄	速度	-0.264	0.073	-3.605	< 0.001
性别.2	视觉	-0.427	0.162	-2.642	0.008
性别.2	文本	0.119	0.128	0.934	0.350
性别.2	速度	0.166	0.146	1.138	0.255
school.GW	视觉	-0.259	0.184	-1.404	0.160
school.GW	文本	0.521	0.129	4.053	< 0.001
school.GW	速度	-0.139	0.151	-0.922	0.357

在训练数据上呈中等程度的多变量拟合（在汇总所有起点和目的地的平方误差时，总体 $R^2 = 0.632$ ）。估计的投影方向通常在每一项内混合四个枢纽（例如，一项主要加载于大阪和福冈，而另一项则混合东京和爱知），因此可解释性取决于岭函数诊断，而非直接的枢纽-成分对应。

主题模型：结构主题模型（STM）

作为使用现成概率主题模型的稳健性检验，我们还使用 `stm` 包拟合了一个包含三个主题的结构主题模型（STM）（Roberts等，2014、2019）。为了与NMF-RE协变量编码相对应，我们使用了流行度公式

$$\texttt{prevalence} \;=\; \sim\ 0 \;+\; \texttt{era},$$

因此，结构主题模型会为每个时代（早期、工业时期、战后）估计一个各时期特定的预期主题流行度，且不含截距。表 G2 报告了这些时期层面的流行度

估计值，并给出全局不确定性下基于模拟的标准误；p 值为双侧检验。在这种参数化下，较小（接近零）的系数表明某个主题在某个时代几乎不存在；从定性角度看，这与 NMF-RE 均值结构编码中“接近零”的解释直接可比。

由于主题标签只能在排列意义下被识别，我们通过检查代表性词语（例如通过 `labelTopics` and Pr(主题 |词语)）将 STM 主题与 NMF-RE 主题标签进行对齐。在此语料库中，STM 主题2对应我们的主题1（制度／法律；宪法／法律／国会），STM 主题3与我们的主题2（工会／宪法／外国；战争／工会／外国）对齐，而 STM 主题1对应我们的主题3（现代公民修辞；今天／美国／自由／世界）。因此，表 G2 报告的是重新标记主题后的 STM 结果，使其与 NMF-RE 主题1–3的顺序一致。

经过这种对齐，STM 印证了相同的定性历史模式。公民修辞主题（主题3）从早期到战后时期显著增加，而制度／法律主题（主题1）在战后时期变得较小。联盟／外国主题（主题2）在早期之后有所下降，但仍不可忽略。

表 G2

就职演说语料库的 *STM* 时代层级主题流行度估计（$K = 3$），使用 `prevalence = ~ 0 + era`（无截距）拟合。重新标记主题，使其顺序与 *NMF-RE* 主题*1–3* 一致（*STM* *2*→主题*1*，*STM* *3*→主题*2*，*STM* *1*→主题*3*）。标准误基于全局不确定性的模拟；*p* 值为双侧检验。

协变量	主题	估计值	SE	t	p
早期（1775–1865）	主题1	0.51913	0.05369	9.669	< 0.001
早期（1775–1865）	主题2	0.40326	0.04381	9.205	< 0.001
早期（1775–1865）	主题3	0.07741	0.03688	2.099	0.040
工业时期（1865–1945年）	主题1	0.47425	0.05336	8.887	< 0.001
工业时期（1865–1945年）	主题2	0.19953	0.04014	4.971	< 0.001
工业时期（1865–1945年）	主题3	0.32591	0.04181	7.796	< 0.001
战后（1945–）	主题1	0.05729	0.04878	1.175	0.245
战后（1945–）	主题2	0.15039	0.04002	3.758	< 0.001
战后（1945–）	主题3	0.79233	0.04005	19.782	< 0.001

参考文献

Benoit, K., Watanabe, K., Wang, H.,Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). Quanteda：用于文本数据定量分析的 R 软件包。开源软件期刊，*3*(30), 774. https://doi.org/10.21105/joss.00774 Bollen, K. A. (1989年)。含潜变量的结构方程(第 25 卷)。威立。Breslow, N. E., & Clayton, D. G. (1993年)。广义线性混合模型中的近似推断。《美国统计学会杂志》，*88*(421), 9–25。

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008)。基于自助法的聚类误差推断改进。经济学与统计学评论，*90*(3), 414–427。

Cemgil, A. T.（2009）。非负矩阵分解模型的贝叶斯推断。计算智能与神经科学，*2009*(1)，785152。

Ding, C., Li, T., Peng, W., & Park, H.（2006）。用于聚类的正交非负矩阵t因子分解。第*12*届 *ACM SIGKDD* 知识发现与数据挖掘国际会议论文集，126–135。

Efron, B.（2004）。预测误差的估计：协方差惩罚与交叉验证。《美国统计学会杂志》，*99*(467)，619–632。

Févotte, C., & Idier, J.（2011）。基于β-散度的非负矩阵分解算法。神经计算，*23*(9), 2421–2456。

Friedman, J. H.（1984）。智能用户指南 （技术报告第1号）。计算统计实验室，斯坦福大学。https://purl.stanford.edu/pd898dq6156

Friedman, J. H., & Stuetzle, W. (1981). 投影寻踪回归。《美国统计学会杂志》，*76*(376)，817–823。

Gillis, N. (2014). 非负矩阵分解的原理与方法. 正则化、优化、核与支持向量机，*12*(257), 257–291. Gillis, N. (2020).非负矩阵分解. 美国工业与应用数学学会.

Gillis, N., & Rajkó, R. (2023). 非负矩阵分解的部分可辨识性 .

*SIAM*矩阵分析与应用杂志，*44*(1)，27–52。Grabski, I. N., Trippa, L., & Parmigiani, G.（2025）。用于突变特征的贝叶斯多研究非负矩阵分解。《基因组生物学》，*26*(1)，98。Holzinger, K. J., & Swineford, F.（1939）。因子分析研究：双因子解的稳定性。教育补充专著。Huber, P. J., 等（1967）。非标准条件下最大似然估计的行为。第五届伯克利数学统计与概率研讨会论文集，*1*(1)，221–233。Javanmard, A., & Montanari, A.（2014）。高维回归的置信区间与假设检验。《机器学习研究杂志》，*15*(1)，2869–2909。Jöreskog, K. G.（1969）。验证性最大似然因子分析的一般方法。《心理计量学》，*34*(2)，183–202。Jöreskog, K. G., & Goldberger, A. S.（1975）。对具有多个指标和多个原因的单一潜变量模型进行估计。《美国杂志》

n

统计学会，*70*(351a), 631–639。Kenward, M. G., & Roger, J. H. (1997)。基于限制性最大似然的固定效应小样本推断。《生物统计学》,983–997。Laird, N. M., & Ware, J. H. (1982)。纵向数据的随机效应模型。《生物统计学》, 963–974。Lee, D., & Seung, H. S. (2000)。非负矩阵分解算法。载于 T. Leen, T. Dietterich, & V. Tresp（编），神经信息处理系统进展 （第 13 卷）。MITPress。https://proceedings.neurips.cc/paper_files/paper/2000/file/f9d1152547c0bde01830b7e8bd60024c-Paper.pdf Lee, D. D., & Seung, H. S. (1999)。通过非负矩阵分解学习物体的组成部分。《自然》，*401*(6755), 788–791。

Mammen, E. (1993)。高维线性模型的自助法与野生自助法。《统计年刊》，*21*(1), 255–285。国土交通省。（2021）。全国道路交通调查：都道府县间起讫点表 [访问日期: 2026-01-10]。https://www.e-stat.go.jp/stat-search/files?stat_infid=000040170612 Phelan, G., & Campbell, D. A. (2025)。检验话语主题的协变量效应假设。*arXiv*预印本 *arXiv:2506.05570*。Pinheiro, J., Bates, D., & R 核心团队。（2025）。*Nlme*：线性与非线性混合效应模型 [R 软件包版本 3.1-168]。
https://doi.org/10.32614/CRAN.package.nlme Potthoff, R. F., & Roy, S. N. (1964)。一种尤其适用于生长曲线问题的广义多元方差分析模型。*Biometrika*，*51*，313–326。

Roberts, M. E., Stewart,B. M., & Tingley, D. (2019). 结构主题模型：用于结构主题模型的 R 软件包。统计软件杂志，*91*，1–40。

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand,D. G. (2014). 用于开放式调查回答的结构主题模型。美国政治科学杂志，*58*(4)，1064–1082。

Robinson, W., Sharan, R., & Leiserson, M. D.(2019). 癌症中突变过程活性的临床与分子协变量建模。生物信息学，*35*(14), i492–i500. Rosales, R. A., Drummond, R. D., Valieris, R.,Dias-Neto, E., & Da Silva, I. T. (2017). Signer：一种用于突变特征发现的经验贝叶斯方法。生物信息学，*33*(1), 8–16. Rosseel, Y. (2012). Lavaan：用于结构方程模型的 R 软件包。统计软件杂志，*48*(1), 1–36. Satoh, K. (2023). 关于使用高斯核作为协变量的非负矩阵分解[日文]。日本应用统计学杂志，*52*(2), 59–74.

Satoh,K. (2025). *nmfkc*：带协变量和结构扩展的非负矩阵分解 [R 软件包版本 0.5。8]. https://ksatohds.github.io/nmfkc/

Satoh, K., & Yanagihara, H.（2010年）。生长曲线模型变系数的估计。《美国数学与管理科学杂志》，*30*(3-4), 243–256。

Schmidt, M. N., Winther, O., & Hansen, L.K.（2009年）。贝叶斯非负矩阵分解。独立成分分析与信号分离国际会议，540–547。

Taylor, J., & Tibshirani, R. J.（2015）。统计学习与选择性推断。《美国国家科学院院刊》，*112*(25), 7629–7634。

Tibshirani, R. J., & Taylor, J.（2012年）。Lasso问题中的自由度。《统计年刊》，*40*(2), 1198–1232。https://doi.org/10.1214/12-AOS1003

van de Geer, S., Bühlmann, P., Ritov, Y., &Dezeure, R.（2014）。关于高维模型的渐近最优置信域与检验。《统计年刊》，*42*(3), 1166–1202。https://doi.org/10.1214/14-AOS1221

Vonesh, E. F., & Carter, R. L. (1987年). 使用非平衡数据对随机系数生长曲线模型进行有效推断。《生物统计学》, 617–628。

White, H. (1980年). 一种异方差一致协方差矩阵估计量及异方差直接检验。《计量经济学》：计量经济学会期刊, 817–838。