arXiv:2601.03099v1 [cs.LG] 6 Jan 2026

时间感知合成控制

Saeyoung Rho哥伦比亚大学 rho@cs.columbia.edu

Cyrus Illick哥伦比亚大学 cdi2105@columbia.edu

Samhitha Narasipura哥伦比亚大学 sn3145@columbia.edu

Alberto Abadie麻省理工学院 abadie@mit.edu

Daniel Hsu哥伦比亚大学 djhsu@cs.columbia.edu

Vishal Misra 哥伦比亚大学 misra@cs.columbia.edu

摘要

合成控制（SC）框架广泛用于基于时间序列面板数据的观测性因果推断。合成控制已成功应用于各种领域，但现有方法通常认为干预前时间索引的顺序可以互换。这种不变性意味着，在存在强趋势时，这些方法可能无法充分利用时间结构。我们提出了时间感知合成控制（TASC），它采用具有恒定趋势的状态空间模型，同时保留信号的低秩结构。TASC 使用卡尔曼滤波器和 Rauch–Tung–Striebel平滑器：首先通过期望最大化拟合生成式时间序列模型，然后进行反事实推断。我们在模拟数据集和真实世界数据集上评估了 TASC，包括政策评估和体育预测。结果表明，在时间趋势显著且观测噪声水平较高的情境下，TASC 具有优势。

***K*关键词** 因果推断 ·合成控制 · 时间序列面板数据 · 状态空间模型 · 贝叶斯学习

1 引言

合成控制（SC）是观测性因果推断中一种广受欢迎的方法。合成控制（SC）常被描述为双重差分（D-in-D， [1]）的自然延伸，旨在通过构建合成反事实数据，更准确地评估干预的影响。首次应用是衡量西班牙巴斯克地区20世纪60年代恐怖主义冲突的经济影响：研究者将冲突发生前西班牙其他地区（供体单位）的国内生产总值数据结合起来，为巴斯克地区构建了在没有冲突的反事实世界中的目标单位合成国内生产总值数据 [2]。与双重差分不同，后者比较处理组与对照组结果随时间的变化，而合成控制则通过对供体进行加权组合，构建一个合成对照单位。合成控制正日益普及，应用范围也不断扩大，涵盖经济学 [2, 3, 4], 、政治学 [5, 6], 、社会科学 [7, 8], 和医疗保健 [9, 8, 10]。

合成控制方法假设时间序列面板数据源自潜变量模型，但不限制时变潜在因子之间的关系。合成控制文献中广泛采用的线性因子模型 [2, 3, 5], 就是一个例子。该模型既灵活又通用，例如，可以扩展该模型以纳入自回归分量。然而，这种灵活性也导致基于该模型构建的合成控制方法在对干预前时间索引进行置换时，仍会得出相同的估计结果。尽管这种灵活性避免了施加严格的结构假设，但也使模型无法在存在可学习趋势时捕捉预测信号。一个关键洞见是，时间序列数据通常呈现稳定趋势，而我们将这种趋势明确纳入模型。

时间序列面板数据集的另一个关键特性是，随着数据规模增大，所得数据矩阵趋于近似低秩。这一现象已由 [11], 深入分析；时间趋势越强，这种现象就越明显，因为潜在因子在各个时间点之间的变化受到限制。基于这一认识，人们提出了许多合成控制变体，以利用数据的低秩结构。这些方法通常依赖于对数据矩阵进行谱分析：例如， [12]采用主成分回归，而 [13]则将合成控制表述为一个核范数

最小化问题。然而，这些方法也不考虑时间，因为打乱时间索引不会改变矩阵的谱。

我们的贡献在于将合成控制面板数据嵌入状态空间模型中，以同时利用数据的低秩特性和时间序列特性。我们提出：1）用于面板数据集的状态空间生成模型 TASC 模型；2）用于学习 TASC 模型的 TASC 算法。本文结构如下：第2节介绍合成控制方法的必要背景知识和常见建模假设。第3节介绍 TASC 模型，这是一种基于状态空间模型的时间序列生成模型。第4节概述基于卡尔曼滤波和 Rauch–Tung–Striebel (RTS) 平滑的期望最大化（EM）式 TASC 算法。在第5节和第6节中，我们将 TASC 应用于模拟数据集和真实世界数据集，并展示我们的方法在何种情况下更具优势。

2 相关工作

2.1 合成控制方法

合成控制的时间序列面板数据集由以下部分组成。令 $Y_{i,\ t} \in \mathbb{R}$ 表示时间 t（列）时第 i 个单元（行）的观测值。第一行对应索引为 1 的受处理目标单元，而n未处理供体单元占据行 $i \in \{2, \ldots, n+1\}$。因此总共有 $N = n + 1$ 行。未受处理 观测矩阵 Y 的大小为 $N \times T$，其中目标单元在 $t > T_0$ 之后的值缺失，因为干预发生在时间 T_0之后。图 1 展示了合成控制数据集的一般结构，上标– 和 $+$ 分别表示干预前时期和干预后时期。

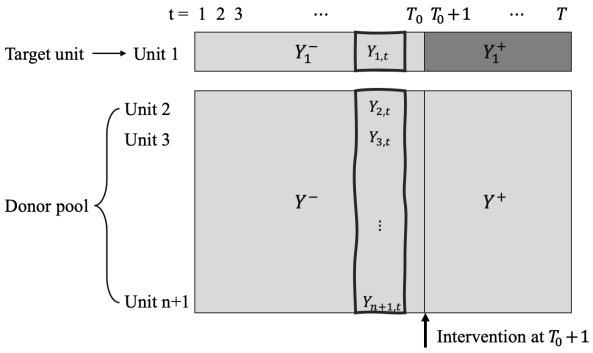

图1：合成控制的一般数据结构

基于这一数据结构，我们将合成控制方法族（算法 1）定义如下。合成控制首先利用干预前数据学习目标单元与供体之间的关系。例如， $\mathcal{M}$ 可以是一个纵向回归，其中供体的干预前列向量 $Y_{2:n+1,\ t}$ 成为标签 $Y_{1,\ t}$的输入特征，适用于所有 $t \in [T_0]$。随后，合成控制利用这些知识（f）将干预后的供体数据 Y^+ 投影并预测 $\hat{}Y_1^+$。最后，估算干预对目标单元的因果效应为 $Y_1^+ - \hat{}Y_1^+$。

算法 1： 合成控制方法族**数据：** 目标单元的干预前数据 $Y_1^- \in \mathbb{R}^{T_0}$、供体数据 $Y = [Y^-, Y^+] \in \mathbb{R}^{n \times T}$**结果：** 反事实预测 $\hat{}Y_0^+$、合成控制权重 f**1. 学习** $f = \mathcal{M}(Y_0^-,\ Y^-,\ Y^+)$ /* Y^+ 的使用是可选的 */**2. 投影** $\hat{}Y_1^+ = f(Y^+)$**3. 推断** 目标单位的干预估计因果效应为 $Y_1^+ - \hat{}Y_1^+$

$\mathcal{M}$ 指的是合成控制学习算法。大部分合成控制文献在 Y^-的凸扫描中采用了最小二乘预测器： $f = \arg\min_f ||Y_1^- - f^\top Y^-||^2$ 其中 $\sum_{i=1}^n f_i = 1,\ 0 \le f \le 1\ \forall i \in [n]$[1] [2, 3, 5]。凸扫描条件可以替换为 Lasso（$||f||_1$）或 Ridge（$||f||_2^2$）正则化 [14, 12]。一些方法在优化步骤之前使用主成分分析，仅保留数据中最前面的几个奇异值 [12, 15]。合成控制算法的其他变体则关注处理多个处理单位 [16, 17], 、处理大量供体[4, 18], 以及校正偏差 [19]等问题。有关这些技术的详细综述，请参见 [20] 。

[1]多变量时间序列和其他协变量可以采用相对重要性加权，但本文聚焦于单变量时间序列。

2.2 用于合成控制的潜变量模型

[2] 提出的第一个合成控制算法假设存在线性因子模型

$$Y_{i,t} = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \epsilon_{i,t}, \tag{1}$$

其中，δ_t 是时间趋势，$Z_i \in \mathbb{R}^p$ 和 $\mu_i \in \mathbb{R}^q$ 分别是观测预测变量和未观测预测变量的向量，其系数分别为 θ_t 和 μ_i，$\epsilon_{i,\ t}$ 是噪声。该模型意味着矩阵的信号分量的秩不超过$p + q + 1$。当这一数值显著小于矩阵的满秩时，观测矩阵就近似呈低秩。这确实是因子模型中的常见情况，许多现实世界的数据 [21, 22, 23]都验证了这一点。这启发了一系列合成控制算法利用数据的近似低秩结构。若干合成控制算法采用单纯形约束或正则化项，以最小化活跃供体单位的数量 [2, 3, 14, 24]； [18] 引入供体选择步骤，以从一开始就减少供体数量；[12, 15] 采用主成分回归；[13] 则将合成控制构建为核范数最小化问题。

合成控制（SC）中使用的面板数据还有一个特点，即其具有时间序列特性。尽管如此，许多合成控制算法的变体对于干预前数据中时间索引的排列仍保持不变。置换不变的方法虽然能够适应范围广泛的时间趋势，从而增强稳健性，但会舍弃排序信息；而当存在有意义的时间模式时，显式建模策略则可以利用额外的结构。为此，一些研究者引入了基于状态空间模型、假设存在时间趋势的算法。 [25] 设计的状态向量包含合成控制权重、局部线性趋势和季节性等元素。 [26, 27]进一步简化了这一结构，仅将合成控制权重作为潜在状态。这些方法的核心概念是允许合成控制权重随时间变化，将目标单元的时间序列定义为观测值（标量），并将合成控制权重（可能还包括其他组成部分）定义为潜在状态（至少为 n维向量，其中 n 为供体数量）。因此，这些建模方法不一定能确保观测矩阵近似为低秩。此外，由于没有显式建模供体池的随机性，这些模型可能无法充分利用供体池中的可用信息。

3 TASC 的状态空间模型

令 y_t 为未处理结果 Y的第 t 列。TASC 方法假设以下状态空间模型:

$$x_t = Ax_{t-1} + q_{t-1},\ \ q_{t-1} \sim \mathcal{N}(0, Q), \tag{2}$$

$$y_t = Hx_t + r_t,\ \ r_t \sim \mathcal{N}(0, R), \tag{3}$$

其中，我们假设初始隐藏状态为 $x_0 \sim \mathcal{N}(m_0,\ P_0)$。隐藏状态 x_t的维度为 d，而 y_t是一个 $N = n + 1$ 维向量。为保持低秩结构，我们要求 $d \ll min(n,\ T)$。

模型参数为 $\theta = \{A, H, Q, R, m_0, P_0\}$，其中 $A \in \mathbb{R}^{d\times d}$、 $H \in \mathbb{R}^{N\times d}$、 $Q \in \mathbb{R}^{d\times d}$、 $R \in \mathbb{R}^{N\times N}$、$m_0 \in \mathbb{R}^d$和 $P_0 \in \mathbb{R}^{d\times d}$。这是一个经典的线性高斯模型，我们将所有协方差矩阵 Q、 R和 P_0设为正定矩阵。如有需要，我们可以将噪声协方差矩阵 Q 和 R 限制为对角矩阵，且对角元素非零，以减少参数数量。

3.1 TASC 的更高级模型

我们主要关注上述 TASC 模型的构建，但我们的学习算法可以轻松修改，以适应更高级的模型。我们展示了几个有代表性的示例。

3.1.1 允许潜在状态中存在时不变部分

Wit不失一般性，我们可以添加一个常量状态 x^*，并仅让 x'_t 部分随时间变化 .

$$x'_t = Ax'_{t-1} + q_{t-1},\ \ q_{t-1} \sim \mathcal{N}(0, Q), \tag{4}$$

$$x_t = x^* + x'_t \tag{5}$$

$$y_t = Hx_t + r_t,\ \ r_t \sim \mathcal{N}(0, R). \tag{6}$$

需要额外的模型参数 $x^* \in \mathbb{R}^d$，而且只需稍作修改，就能轻松地将该模型用于我们的算法。

3.1.2 允许季节性

公式 (2) 和 (3) 中的当前模型未能捕捉季节性。幸运的是，可以轻松修改 TASC 模型以适应季节性。令 $\mathbf{1}$ 表示一个所有元素均为 1 的列向量。为纳入季节性，我们将 $s_t \in \mathbb{R}$ 定义为时间 t 的季节效应，该效应在各个单位之间保持不变，并将模型参数指定为 $\theta = \{A, H, Q, R, m_0, P_0, s_1, \ldots, s_T\}$。

$$x_t = Ax_{t-1} + q_{t-1}, \quad q_{t-1} \sim \mathcal{N}(0, Q), \tag{7}$$

$$y_t = Hx_t + s_t\mathbf{1} + r_t, \quad r_t \sim \mathcal{N}(0, R). \tag{8}$$

采用这种表述方式，s_t 可以由用户设定为超参数，也可以在 EM算法 的 M步中学习得到 .

3.1.3 允许多个时间序列

当我们观测到单位 i 在时间 t 的 m 个时间序列时，可以将它们垂直堆叠，以重新定义 y_t。设 n 个单位在时间 t 观测到的 m 个时间序列为 $y_t^{(1)}$、.....、$y_t^{(m)} \in \mathbb{R}^n$。然后，我们将其视为在时间 t 观测到 nm 个单位，并定义 $y_t = [{y_t^{(1)}}^\top, \ldots, {y_t^{(m)}}^\top]^\top \in \mathbb{R}^{nm}$。

$$x_t = Ax_{t-1} + q_{t-1}, \quad q_{t-1} \sim \mathcal{N}(0, Q), \tag{9}$$

$$y_t = \begin{bmatrix} y_t^{(1)} \\ \vdots \\ y_t^{(m)} \end{bmatrix} = Hx_t + r_t, \quad r_t \sim \mathcal{N}(0, R). \tag{10}$$

如有需要，可以通过设计另一层潜在查找表来建模来自同一单位的 m 时间序列之间的联系。

3.2 与其他模型的比较

经典合成控制 [2] 假设存在一个包含观测因子和未观测因子的线性因子模型，如公式（1）所示。这可以重新表述为状态空间模型：所有依赖时间的变量定义了潜在状态 $x_t = (\delta_t, \theta_t, \lambda_t)$，而潜在状态与各个观测值之间的映射（即 H 的第 i 行）编码了 $h_i =$（1、Z_i、μ_i）。在这种表述下，隐状态维度变为 $d = 1 + p + q$。然而，线性因子模型并未明确假设趋势矩阵 A。这种缺失可以通过允许每个时间点存在时变趋势 A_t，或通过设定 $A = 0$ 并对 $x_t = q_t$进行建模来编码；$q_t \sim \mathcal{N}$（0、Q_t）。与 TASC 的关键区别在于，TASC 强制要求 x_t 之间保持稳定关系，而线性因子模型则没有这一要求。这表明，虽然在模型设定正确时，TASC 可以实现更好的预测性能，但当时间动态较为复杂时，它也更容易受到模型设定错误的影响。

在稳健合成控制（RSC，[12]）中，假设矩阵元素遵循潜变量模型 $Y_{i,t} = g(\theta_i, \rho_t) + \epsilon_{i,t}$ 其中 θ_i 和 ρ_t 是 d 维的[2] 潜向量，用于表征第 i个单位和第 t个时间点，而 $\epsilon_{i,t}$ 是观测噪声。这是一种更一般化的表达形式，可以涵盖公式（1）中的线性因子模型。RSC 的学习算法可解释为通过将 g 视为点积并采用主成分分析来学习 θ_i 和 ρ_t：$Y = \sum_{l=1}^{min(n,T)} s_l u_l v_l^\top = \sum_{l=1}^{d} s_l u_l v_l^\top + \sum_{l=d+1}^{min(n,T)} s_l u_l v_l^\top$，其中 s_l 是按降序排列的奇异值。令 $\tilde{U}$ 的列为$l \le d$ 对应的 $s_l^{1/2} u_l$，并令 $\tilde{V}^\top$ 的行是 $l \le d$ 对应的 $s_l^{1/2} v_l$，则 $\tilde{U}$ 的行可解释为 θ_i，而V的列可解释为 ρ_t。类似地，TASC 模型提出了一种分解形式 $Y = HX + E$，其中 X 的列是隐藏状态 x_t，而 E 的列是观测噪声 r_t。在这里， H 的行与 θ_i 相对应，而X 的列与 ρ_t相对应。 $\tilde{U}\tilde{V}^\top$ 和 HX 都是精确的低秩矩阵，但它们分离噪声的方式不同。这种差异源于学习目标：RSC 的方法最小化噪声矩阵的大小（以谱范数衡量），而 TASC 则侧重于确保时间特征 x_t 随时间平缓演变，并保持恒定趋势 A。因此，经 RSC 算法滤除的噪声秩为 $\min(n, T) - d$，而 E几乎必然是满秩的（全方向）。

3.3 TASC 模型何时具有优势

TASC 方法的一个显著特点是 其对趋势 A进行显式建模。这一设计选择带来了若干 1
优势，但它 在某些情况下可能会带来局限。首先，纳入 A 可提升 TASC 模型的可解释性 y

[2] θ_i 和 ρ_t 的维度可能不同。

算法2： 期望最大化$^{\text{pre}}(Y_{\text{pre}}; N_1)$

数据： Y_{pre} 其中第（i, j）个元素为 $y_{i,t}$ $\forall$（i, t） $\in [0:\ n] \times [1:\ T_0]$（目标单位和供体的干预前数据供体）

结果： $\theta = \{A, H, Q, R, m_0, P_0\}$

初始化 $\theta^{(0)}$

对于 $i \leftarrow 1$**到** N_1 **执行**

  对于 $k \leftarrow 1$ **到** T_0 **执行**

    使用卡尔曼滤波，通过 $\theta^{(i-1)}$更新 m_k，P_k /* E步的滤波过程 */

  end

  对于 $k \leftarrow T_0 - 1$ **到** 0 **执行**

    使用RTS平滑，通过 $\theta^{(i-1)}$更新 m_k^s，P_k^s，G_k /* E步的平滑过程 */

  end

  使用 M步（算法7）更新 $\theta^{(i)}$

  其中 $T = T_0$ /* M步 */

end

返回 $\theta^{(N_1)}$

，因为它能够捕捉时间序列数据中潜在的趋势。其次，如果趋势 A 在干预点之后持续存在，TASC 便具有强大的预测能力，这在较长的时间范围内尤其有优势。第三，通过对 A进行建模，TASC 会受到时间顺序的影响，而非对其视而不见。总体而言，当存在一致的趋势时，TASC 能够利用时间结构；但当趋势较弱或不存在时（例如 $A = 0$），其优势有限。

TASC 的另一项优势是其近似低秩结构，可表示为 $Y = HX + E$，其中 HX 是精确的低秩信号，E 表示噪声。这种低秩特性常见于真实世界数据集，尤其是在数据规模增大时。在这种情况下，主成分分析（PCA）是一种广泛采用的提取低秩信号的技术。它已应用于图像处理、语音和音频分析、基因组学、金融和传感器数据分析等多个领域，也被 [12] 用于合成控制方法。然而，当存在大量观测噪声时，PCA 的性能会下降，因为它假设所选方向上学习到的主成分不含噪声。相比之下，TASC 假设噪声是全方向的（即噪声矩阵为满秩），因此在噪声条件下提取信号时可能具有更强的稳健性。

我们在附录A中使用数据处理不等式将这一直觉形式化。

4 TASC算法

本节介绍用于学习TASC 模型参数并进行反事实预测的TASC算法。首先，TASC使用干预前数据进行参数学习。我们采用期望最大化（EM）方法，M步的更新具有闭式解，可求得期望完整数据对数似然的精确最大值。随后，TASC通过额外运行卡尔曼滤波和RTS平滑过程来进行反事实估计。由于干预后目标数据被视为缺失数据，我们将目标坐标的观测噪声方差设为无穷大，从而使相应的卡尔曼增益设为零。

4.1 从干预前数据中学习

TASC使用干预前数据学习模型参数，并将EM^{pre} 作为子程序运行。对于线性高斯状态空间模型，我们可以采用经典的EM方法，其中E步由滤波过程（卡尔曼滤波）和平滑过程（RTS平滑）组成。这将得到估计值 m_k^s 和 P_k^s，用于为后验概率分布定义一个下界。算法2展示了基于干预前数据进行参数估计的主要EM算法。

对于算法2的M步，我们利用 y_t 服从多元高斯分布这一事实，计算期望完全对数似然（Q函数）的最大值。算法7展示了这种具有闭式解的方法。如有需要，也可以用对Q函数进行梯度上升来替代这一步。如果需要额外的建模参数，且无法求得闭式解，则更适合采用这种基于梯度的方法。在实现时，可以将与E步相同的步骤定义为神经网络的前向传播，并执行梯度上升。

算法 3： TASC($Y; N_1$)

数据： $y_{i,t}$ $\forall(i,t) \in [0\colon n] \times [1\colon T_0]$ 和 $\forall(i,t) \in [1\colon n] \times [T_0+1\colon T]$

结果： $\hat{\theta} = \{A, H, Q, R, m_0, P_0\}$, $\hat{y}_{0,T_0+1}, \ldots, \hat{y}_{0,T}$

学习 $\theta^{N_1} \leftarrow$ 期望最大化$^{\text{pre}}(Y_{\text{pre}}; N_1)$

对于 $k \leftarrow 1$**至** T_0 **执行**

  使用算法 4 和 θ^{N_1} 更新 $m_k,\ P_k$ /* 干预前滤波 */

end

对于 $k \leftarrow T_0+1$ **至** T **执行**

  使用算法 5 和 θ^{N_1} 更新 $m_k,\ P_k$ /* 使用无穷方差进行滤波 */

end

对 $k \leftarrow T-1$ **到** 0 **执行**

  使用 $\theta^{(i-1)}$通过RTS平滑更新 $m_k^s,\ P_k^s$ /* 平滑过程 */

end

定义 $H = \begin{bmatrix} h_1^\top \\ H_2 \end{bmatrix}$

对 $k \leftarrow T_0+1$ **到** T **执行**

  $\hat{y}_{0,t} \leftarrow h_1^\top m_t^s$ /* 反事实推断 */

end

返回 θ^{N_1}, $\hat{y}_{0,T_0+1}$、……、$\hat{y}_{0,T}$

4.2 使用干预后数据进行反事实推断

利用通过期望最大化^{pre}（算法2）学到的模型参数，TASC会再次执行卡尔曼滤波器和RTS 平滑器，以进行反事实推断。然而，由于 y_k（属于目标单位）的第一个元素缺失，若不进行特殊处理，就无法完成这一过程。为此，我们将目标单位的数据视为缺失，并将数据和参数中的供体部分分离出来：$y_t = \begin{bmatrix} y_{t,1} \\ y_{t,2} \end{bmatrix}$，$r_t = \begin{bmatrix} r_{t,1} \\ r_{t,2} \end{bmatrix}, H = \begin{bmatrix} h_1^\top \\ H_2 \end{bmatrix}$ 和 $R = \begin{bmatrix} r_1 & 0 \\ 0 & R_2 \end{bmatrix}$，其中$y_{t,2} \in \mathbb{R}^n$, $H_2 \in \mathbb{R}^{n\times d}$和 $R_2 \in \mathbb{R}^{n\times n}$。随后，我们可以将供体的观测模型重写为$y_{t,2} = H_2 x_t + r_{t,2}$，其中$r_{t,2} \sim \mathcal{N}(0, R_2)$。在这个新模型中，干预后观测值不会为目标相关参数提供信息：h_1 和 r_1。这等同于在原模型中设置 $r_1 \to \infty$ 。

由于方差为无穷大，干预后目标时间序列不会影响卡尔曼滤波的结果，因此可以将其设为任意值。RTS平滑保持不变，因为它不使用 R 或 y_k 作为输入。因此，这只会将算法 4（原始卡尔曼滤波器）中干预后时间步的卡尔曼滤波部分替换为算法 5（使用 $r_1 \to \infty$的卡尔曼滤波器）。TASC的完整描述见算法 3。第一行中的期望最大化^{pre}需要 $O(N_1 T_0 N^3)$ 时间，其中 N^3和 d^3项源自使用朴素算法进行矩阵求逆。完整的TASC流程还会产生额外的 $O(TN^3)$，但假设 $T \ll N_1 T_0$，整体时间复杂度由期望最大化^{pre} 部分主导。

5 模拟数据上的实证评估

本节通过模拟数据展示 TASC 方法，并将其与三种基准算法进行比较：采用单纯形约束的合成控制（SC） [2], 采用硬奇异值阈值处理的稳健合成控制（RSC） [12], 以及采用贝叶斯建模方法的因果影响模型（CIM） [25]。

5.1 时间索引置换的影响

我们测试了时间索引置换对 TASC 性能的影响。我们从随机生成的数据集中分别打乱干预前和干预后的索引，以确保两个片段不会混杂。图 2 显示了时间索引保持原始顺序（左图）和索引经过置换（右图）时的干预后均方根误差。时间索引置换会导致 TASC 性能下降：均值增加了 48.5%，标准差增加了 25.7%。相比之下，即使时间索引经过置换，SC 和 RSC 的预测值也会按设计保持不变。

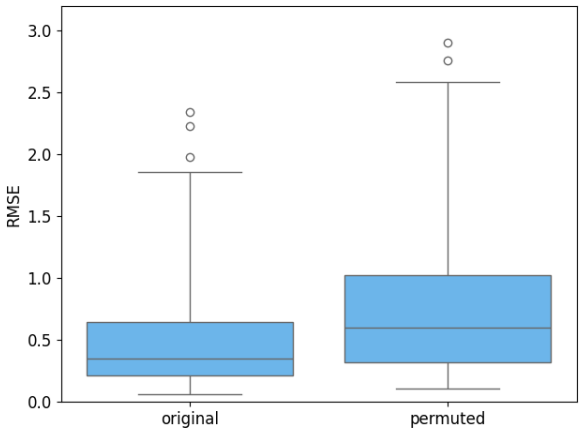

图 2：保留时间索引原始顺序（左）和打乱索引顺序（右）时 TASC 的干预后均方根误差

5.2 协方差矩阵的消融研究 ces

我们比较了在不同设置下生成隐状态扰动协方差 Q 和观测噪声协方差 R时，TASC 与三个基准的性能。对于 Q 和 R，我们分别测试了小协方差矩阵（使用 $a = 0.01$、 $b = 0.1$生成）和大协方差矩阵（$a = 0.1$、 $b = 1$）。为说明这些设置如何影响生成的时间序列，观测噪声的平均绝对值在小 R时约为 0.0839，在大 R时约为 0.8365。大 Q 会产生更高的信号绝对值均值（2.6624），而小 Q 则会产生较低的值（0.4842）。因此，信噪比（SNR）在小 R 和大Q 的设置下最高，其次是小 R 和小 Q、大 R 和大 Q，最后是大 R 和小 Q。在实践中，小Q 表示时间趋势较强（即 A的影响较强），而大 Q 会生成类似于更一般的线性因子模型的数据，其中 A 不受限于在整个时间内保持恒定。同样，小 R 对应较低的观测噪声，而大 R 意味着更高的观测噪声。

图 3 展示了在低观测噪声（较小的 R）下生成的数据集上，TASC 和基准方法的干预后均方根误差。对于较小的 Q （左图）和较大的 Q （右图），RSC 的预测准确度最高，凸显了在观测噪声较低时主成分分析的强劲表现。当 Q 较小时，SC 的表现与 RSC 相当，这表明在低观测噪声下，SC 中的单纯形约束是有效的。正如预期，在趋势 A更加明显时， Q 较小的情况下 TASC 表现相对更好，尽管在这种情况下它的表现仍逊于其他基准方法。这些结果表明，SC 和 RSC 能够捕捉潜在的

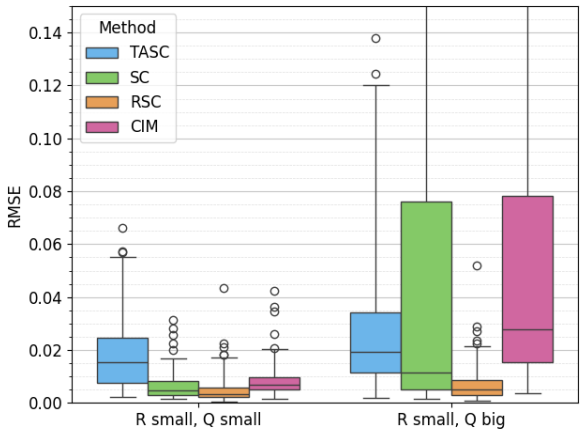

图 3：在低观测噪声（较小的 R）下生成的数据集上，TASC 与基准方法的干预后均方根误差：较小的 Q （左）和较大的 Q（右）

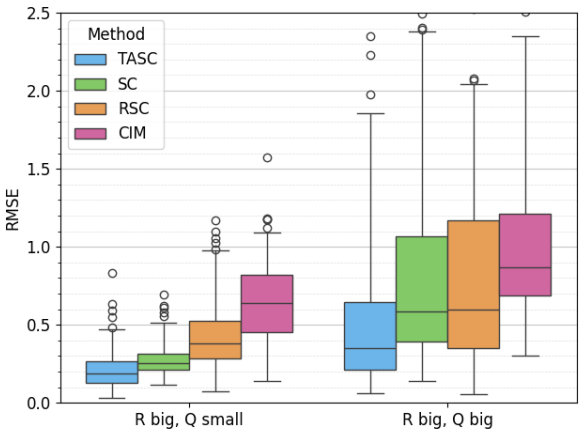

图 4：在高观测噪声（较大的 R）下生成的数据集上，TASC 与基准方法的干预后均方根误差：较小的 Q （左）和较大的 Q（右）

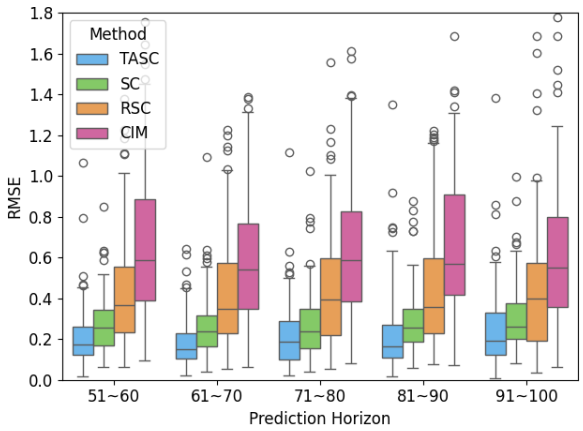

图 5：观测噪声较低且 Q 较小时，干预后的均方根误差；在五个未来预测时域（每个时域包含 10 个时间段）内进行评估

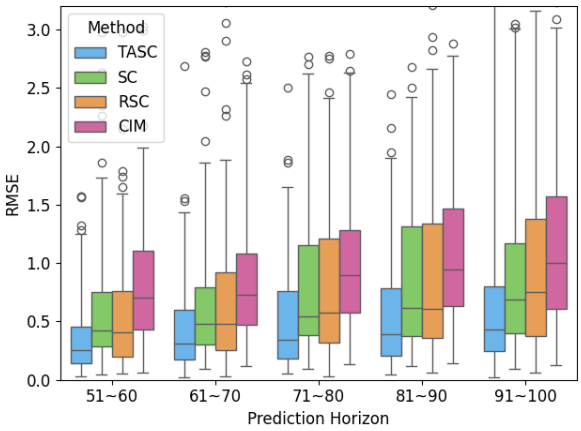

图 6：观测噪声较低且 Q 较大时，干预后的均方根误差；在五个未来预测时域（每个时域包含 10 个时间段）内进行评估

结构更加有效——即使没有显式地对时间趋势 A 建模——在 Q 较小时，时间趋势会更加明显——在低观测噪声下。

图 4 展示了一个类似的图，这次是在观测噪声较高（即 R 较大）的情况下。在这种高噪声情形下，当 Q 较小时，TASC 表现出色。随着 Q 增大，TASC 的表现有所下降，与观测噪声较低时的情况类似，但其预测准确度仍优于所有其他基准方法。有趣的是，当 Q 较小时，四种方法的表现都更好（因此趋势 A 很明显）。其中，SC 在这种情况下最为可靠，其均方根误差的方差最低。总体而言，无论时间趋势的强弱如何，TASC 在观测噪声较高时似乎都是一种稳健的选择。

最后，我们将高观测噪声下的性能按未来时间划分为五个相等的时段，并考察随着预测时域延长，性能的变化情况。图 5 展示了小型 Q 的结果，可以看出，TASC 在各个预测时段（近期或更远的未来）的表现都很稳定。图 6 展示了大型 Q 的结果，与近期未来（51~60）相比，当预测时域延伸至更远的未来（91~100）时，TASC 的性能会下降。其他基准方法也呈现出类似的趋势。

最后，我们将高观测噪声下的性能按未来时间划分为五个相等的时段，并考察随着预测时域延长，性能如何变化。图 5 展示了小型 Q 的结果，表明 TASC 在所有预测时段——无论是近期还是更远的未来——都保持相对稳定的性能。相比之下，图 6 展示了大型 Q 的结果，表明随着预测时域延长（91–100），TASC 的性能相比更早的时段（51–60）下降得更加明显。

5.3 隐状态维度的消融研究

在我们测试的方法中，TASC 和 RSC 明确假设数据矩阵具有低秩结构，并将近似秩作为算法的超参数。我们将输入算法的超参数记为 d，将真实的数据生成参数记为 d_{true}。我们通过改变 d 和 d_{true} 进行测试，并比较 TASC 和 RSC 的性能。

首先，我们固定 $d_{\text{true}} = 5$，改变超参数 $d \in 3$，5、10、20。图7展示了 TASC 和 RSC 在 d不同取值下的干预后均方根误差。对于这两种方法，当$d = d_{\text{true}}$时性能最佳，而低估或高估都会降低准确性。低估 d 会导致明显差于高估的性能，其中 TASC 对低估的敏感性高于 RSC。在高估情况下，与 RSC 相比，TASC 对 d 的选择表现出更强的稳健性。

接下来，我们在固定 $d = d_{true}$ 的情况下改变 $d_{\text{true}} \in 3$、5、10、20。图8展示了 TASC 和 RSC 在 $d_{true} = d$ 取不同值时的干预后均方根误差。对于 d_{true} 的所有取值，TASC 始终优于 RSC。这可能归因于 TASC 的结构优势，因为真实数据生成过程与 TASC 的模型假设相一致。尽管生成的数据也与 RSC框架兼容，但 RSC 未利用时间趋势分量 A，这可能限制了其有效性。

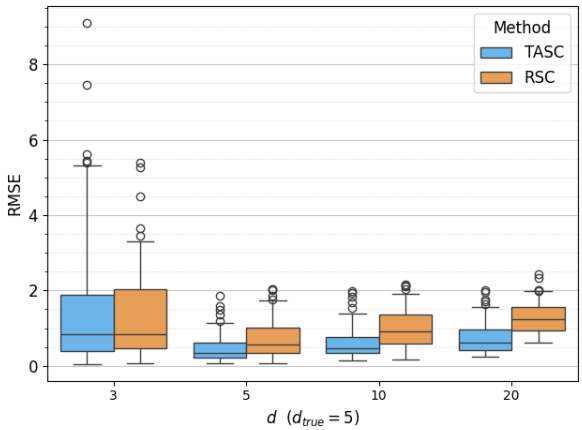

图7：不同 d 值下的干预后均方根误差（固定 $d_{true} = 5$）

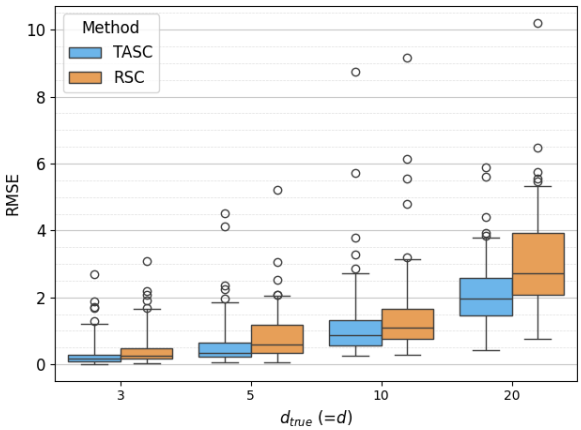

图8：不同 d_{true} 值下的干预后均方根误差（固定 $d = d_{true}$）

5.4 关于供体单元数量的消融研究

我们测试了增加数据集中供体数量的影响。我们改变了面板数据中的行数，记为 $N = n + 1$，其中 n 表示供体单位的数量，额外的一行对应目标单元。干预前时期固定为 $T_0 = 50$，预测时域从 $T_0 + 1 = 51$ 延伸至 $T = 100$。图9展示了 N 变化时的干预后均方根误差。值得注意的是，当 $N = T_0 = 50$ 时，TASC 达到最佳表现。尤其是在 N 较小时，TASC 表现出稳健性，在 $N = 10$ 和 $N = 50$ 之间的性能差异极小。其他合成控制基准方法在 N 接近 T_0 时表现也更好，而供体过多（例如，$N = 200$）和过少（例如，$N = 10$）都会对所有方法产生不利影响。

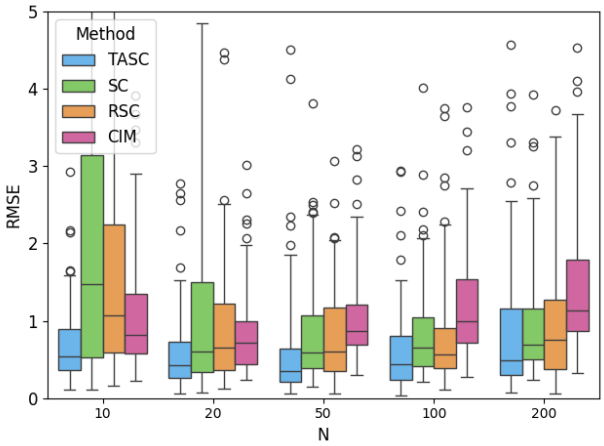

图9：随着 N 变化的干预后均方根误差（面板数据中的行数）

随着 N 从 10 增至 50，除 CIM 外的大多数方法性能都有所提升，这表明在此范围内增加训练数据是有益的。然而，尽管训练数据更多，将 N 从 50 增至 200 并未带来进一步提升。事实上，总体而言，性能反而有所下降。我们推测这一现象可能有两种解释：（1）随机数据源的信息量在达到一定程度后趋于饱和；（2）当特征数量增加而观测值数量保持不变时，模型结构难以从高维输入空间中获益（即维数灾难）。

6 真实世界数据上的实证评估

本节将展示真实世界数据集实证评估的更多结果。我们将 TASC 与上一节使用的三个基准进行比较：1）合成控制（SC， [2]）；2）稳健合成控制（RSC， [12]）；以及因果影响模型（CIM， [25]）。第 6.1 节展示了经典合成控制在评估加利福尼亚州第99号提案时的应用；第 6.2 节和第 6.3 节则分别将合成控制用于预测板球比赛和篮球比赛的得分轨迹。

6.1 评估第99号提案在加利福尼亚州的影响

在本节中，我们使用 [3]中的一个经典合成控制应用来演示我们的方法：评估第99号提案的影响。第99号提案是加利福尼亚州于1988年实施的一项政策，大幅提高了该州的香烟税。该政策实施后，香烟销量显著下降（图10中的黑线）。为了评估这一下降是否由该政策因果驱动，可以应用合成控制方法估计加利福尼亚州的反事实结果——即，如果没有实施该政策，香烟销量将会呈现何种情况。在经济分析中，通常会使用多个辅助预测变量来提高预测准确性。例如， [3]除了目标时间序列（人均香烟销量）外，还纳入了香烟平均零售价、州人均个人收入、15至24岁人口占比以及人均啤酒消费量等变量。然而，在我们的分析中，我们有意仅关注单一预测变量，即人均香烟销量，以确保不同方法之间进行公平且一致的比较。我们的目标并不是对政策影响作出最准确的估计，而是在受控环境下评估各竞争方法的表现。

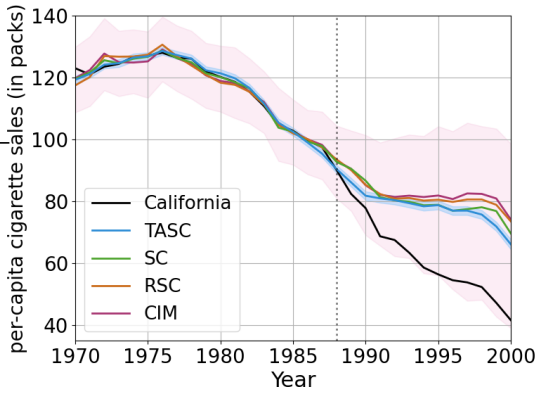

图10：加利福尼亚州人均香烟销量（以包计，黑色），以及由TASC和其他基准估计的反事实结果。蓝色和粉色阴影表示TASC和CIM的95%置信区间。

鉴于此，RSC 采用岭回归协
efficient, 与隐状态维度相同。图10以黑色显示了加利福尼亚州的观测值，所有预测值都高于观测趋势，其间的差距反映了政策效应。四种估计值总体上较为一致，仅在2000年附近略有分歧。TASC和CIM分别以蓝色和粉色阴影区域显示置信区间估计。TASC的置信区间明显窄于CIM的置信区间。值得注意的是，CIM的区间涵盖了加利福尼亚州的观测值，这表明无法从该分析中得出有力的因果结论。

为评估这些反事实估计的可信度，我们进行了安慰剂检验。由于真实的反事实不可观测，我们通过将一个供体单元视为目标单位来模拟它。在每次安慰剂检验中，我们使用其余供体预测一个供体单元的时间序列，并评估合成控制方法能否准确重建观测结果。图11展示了安慰剂检验中干预后的均方根误差（RMSE），比较对象包括

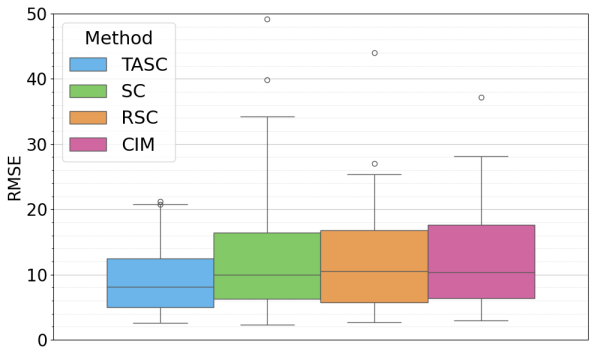

图11：安慰剂检验的干预后均方根误差

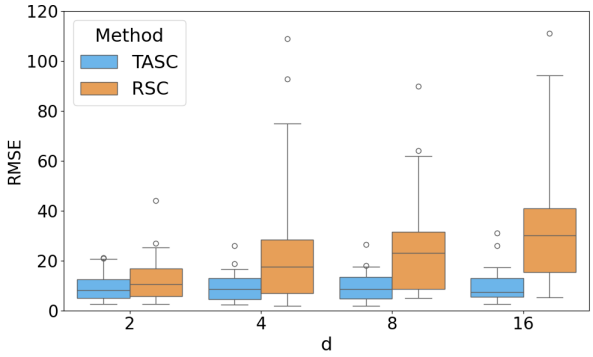

图12：不同 d下的干预后均方根误差。

TASC及其他基准。在这些方法中，TASC的RMSE中位数最低，RMSE方差也最小，这表明在本实验设置中，TASC在受测方法中提供了最可靠的估计值。

接下来，我们深入比较 TASC 和 RSC。在我们测试的方法中，TASC 和 RSC 会明确过滤数据，使其仅包含完全低秩信号。因此，这两种方法都需要超参数 d，该参数对于 TASC 表示隐维度，对于 RSC 表示近似秩。为了考察选择 d 对性能的影响，我们评估了这两种方法在不同 d 取值下的表现。图12报告了安慰剂检验中干预后均方根误差随 d 变化的情况。两种方法的误差都在 $d = 2$ 处达到最低。随着 d 增大，RSC 的性能迅速恶化，而 TASC 则相对稳定。这与模拟结果一致：模拟结果表明，TASC 对高估 d 具有较强的鲁棒性（图7）。因此，这进一步证实，在难以估计 d 真实值的情形下，TASC 可能具有优势。

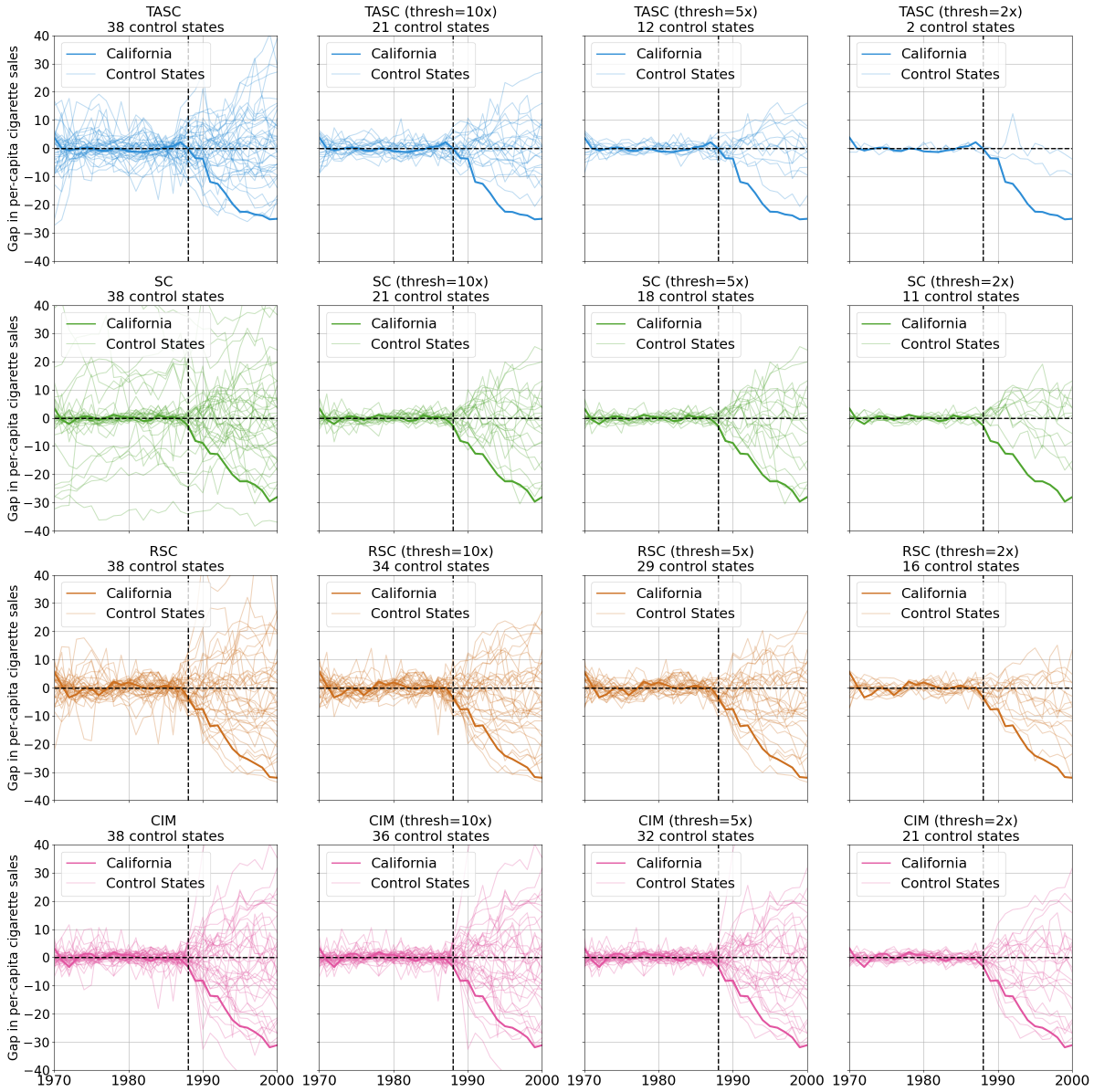

图13：观测结果与合成控制预测值之间的人均香烟销量（包数）差距（与 [3] 中的图4–7相对应）。每一行代表一种不同的算法——从上到下依次为 TASC、合成控制、RSC 和 CIM；每一列则采用不同的控制单元筛选阈值：无阈值、至多为加利福尼亚州干预前误差的 10 倍、5 倍和 2 倍（从左到右）。

最后，按照 [3]，的方法，我们绘制了加利福尼亚州的观测结果与预测反事实之间的差值，并将其与安慰剂检验中控制州的相应差值并列展示。图13第一列报告了加利福尼亚州和38个控制州的人均香烟销量（以包为单位）的预测值与观测值之间的差距。后续各列根据干预前拟合的相对质量（以均方误差（MSE）衡量）对控制州集合进行筛选。第二列仅纳入干预前 MSE 不超过加利福尼亚州10倍的控制州，而第三列和第四列则分别采用更严格的5倍和2倍阈值。

值得注意的是，随着阈值变得更加严格，最后一行（CIM）保留的控制州数量最多，这表明各州的干预前预测准确度相近。相比之下，在最严格的阈值下，与合成控制或 RSC 相比，TASC 方法（第一行）保留的控制州明显更少。这表明，使用 TASC 时，加利福尼亚州的干预前拟合比其他州更准确。请注意，加利福尼亚州是美国人口最多的州之一，因此由于平均效应，收集到的数据（人均香烟销量）可能比其他州的方差更小。在这种情况下，TASC 可能学到了更小的观测噪声方差（R），从而对加利福尼亚州得出了更准确的（干预前）拟合。事实上，加利福尼亚州对应的方差最小（2.58，中位数为 12.95，标准差为 36.17，最大值为 170.79）。在所有设定中，加利福尼亚州始终呈现出最大的差距；而在右上角的图中（阈值更严格，采用 TASC 方法），这一点更为明显。各方法对政策效应的估计相近，只有在末尾才出现分歧。接近 2000年时，TASC 和 RSC 的估计值更为平缓，而合成控制和 CIM 的估计值则持续上升。

6.2 板球比赛中的得分轨迹预测

本节将使用印度超级联赛（IPL）的板球得分轨迹数据展示我们的方法。本研究所用的数据集源自2008年4月18日至2025年3月25日期间印度超级联赛比赛逐球记录。在T20赛制中，一场标准比赛由两局组成，每局20轮，每轮投6球，每局由一支球队击球。虽然由于罚球，一局中的投球数有时会超过120次，但本分析仅关注每局前120次合法投球。我们从数据集中选取了1524条得分轨迹；这些轨迹来自2092条至少包含120次投球的轨迹。每局的得分均按每次投球进行累积汇总。

为模拟真实应用场景，我们从全部1524场比赛中随机选取一场作为目标比赛，供体池由目标比赛之前最近的 n 场比赛组成，选择依据为 $n \in \{18, 36, 72, 144\}$。每局是一段长度为 $T = 120$的时间序列，干预点固定在 $T_0 = 72$。选择 $T_0 = 72$ 是为了平衡强打阶段前36次投球与其后阶段的干预前数据。剩余的48次投球（$t = T_0 + 1, \ldots, T$）用于反事实推断和评估，预期不会产生干预效应（安慰剂检验）。

我们将 TASC 与三种基准方法进行比较：SC、RSC 和 CIM。对于这四种方法，拟合前均通过减去根据所选供体计算出的平均得分轨迹，对数据进行均值中心化。对于 TASC 和 RSC，隐状态维度和保留的奇异值数量均设为 $d = 5$。RSC 在使用 $\{10^{-1}, 10^{0}, \ldots, 10^{6}\}$ 中的取值对干预前数据进行交叉验证后，采用岭回归系数 10^3, 实现。对于每种方法和每种供体池大小，实验均重复 100 次，每次重新随机选择一个目标匹配对象。

首先，我们研究供体规模对不同方法性能的影响。图14展示了不同供体单位数量 n 下整体的干预后均方根误差。随着供体数量增加至约 $n = 72$，TASC 的性能稳步提升，之后趋于稳定。RSC 对供体规模最为敏感：当 $n = 18$ 很小时，其准确度相当，但随着 n 增大，性能显著下降，这很可能是高维过拟合所致。相比之下，受单纯形权重约束的 SC 在不同供体规模下表现稳定。类似地，CIM 在不同供体规模下也保持稳定，并在 $n = 144$ 时略有改善。当供体规模为 $n \in \{36, 72, 144\}$ 时，TASC 在各种方法中取得最低的 RMSE 中位数；而供体规模为 $n = 18$ 时，SC 的表现最低。在所有设置中，使用 $n = 72$ 个供体单位的 TASC 取得了最低的 RMSE 中位数 7.88。

接下来，我们考察预测性能在较长预测时域内的变化。图15按轮（6球区间）划分展示了均方根误差（RMSE），说明预测时域延长时预测准确度如何变化。上图对应 $n = 18$供体，在这一情形下，RSC和合成控制（SC）的表现不相上下。RSC的短期准确度较高，但随着预测时域延长，其优势逐渐消退。SC短期准确度较低，但随着预测时域延长，其表现最佳。下图对应 $n = 72$ 供体，此时TASC给出的预测值最准确，尤其擅长预测较远未来时点。两种情形下，预测误差都会随着预测时域延长而增大，反映出长期预测的难度不断增加。

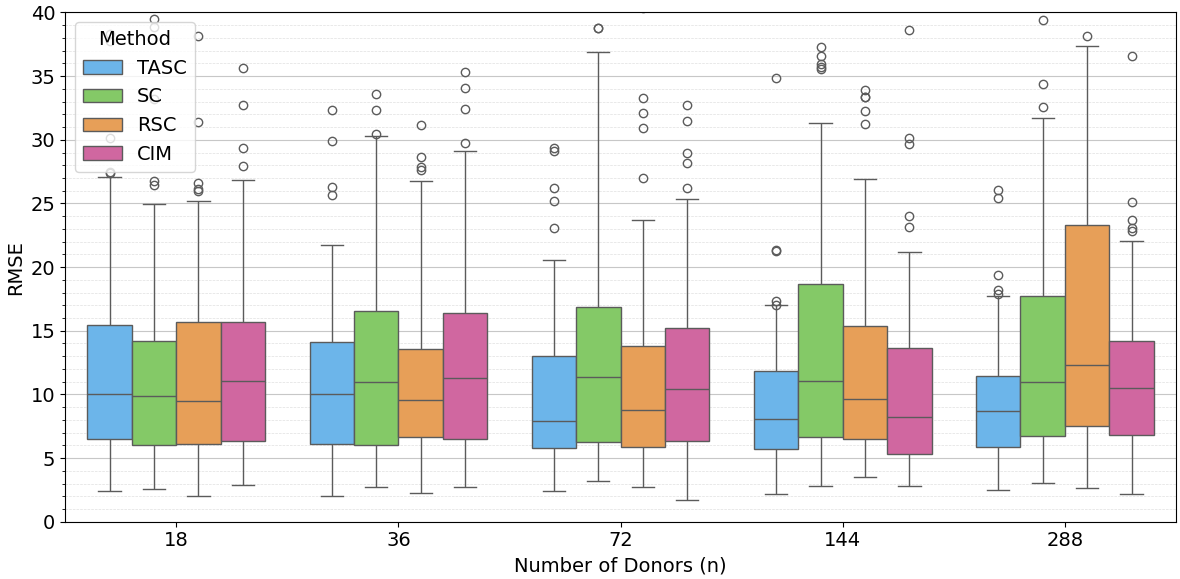

图14：不同供体规模下 TASC、SC、RSC 和 CIM 的整体干预后均方根误差。

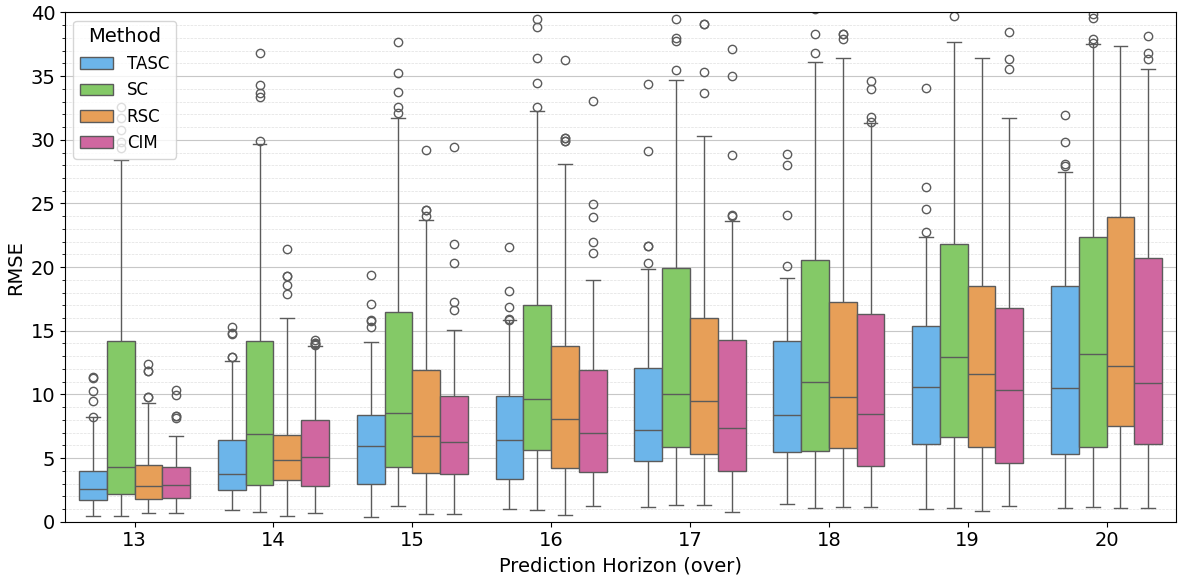

图15：按预测时域统计的每局均方根误差，适用于 $n = 144$供体。

TASC和CIM的性能下降速度比其他方法更慢，这可能是因为它们显式地编码了潜在因子的时间演化：当存在可学习趋势时，对其进行显式建模能够提高长期预测准确率。

最后，我们比较 TASC 和 CIM 的置信区间预测。在点估计（均值）方面，TASC 和 CIM 在短期范围内表现相当，而 TASC 在长期范围预测方面略胜一筹。TASC 和 CIM 不仅能生成点估计值，还能提供置信区间预测：TASC 根据潜在状态的协方差估计得出置信区间预测，而 CIM 则通过其马尔可夫链蒙特卡罗采样器获得这些预测。图 16 展示了 TASC 和 CIM 在供体数量不同的情况下，对干预后预测值的平均 95%置信区间宽度[3]。尽管 TASC 和 CIM 的均值估计表现相近，

[3]每项实验的置信区间宽度，是所有干预后时间步中置信区间上界与下界之间距离的平均值。

TASC 往往能生成更窄的置信区间，这可能有助于解读干预效应。这与第 6.1 节的结果一致：CIM 的置信区间涵盖了加利福尼亚州的观测结果，因此无法根据其估计值得出任何因果结论。

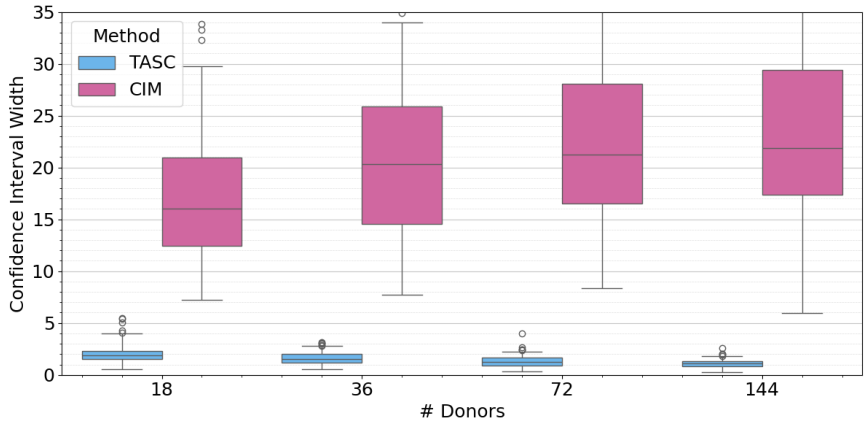

图 16：干预后　　验证集的平均置信区间宽度在线a　　供体数量的平均置信区间宽度 n

6.3 篮球比赛中的得分轨迹预测

本节将使用来自美国国家篮球协会（NBA）比赛的篮球得分轨迹数据来展示我们的方法。NBA 是北美成立于 1946 年的职业篮球联赛，目前由 30 支球队组成。一场 NBA 比赛由四节组成，每节时长 12 分钟，因此比赛总时长为 48 分钟，不包括可能的加时赛。我们使用来自 Kaggle[4] 的逐回合数据来构建得分轨迹面板数据集。原始事件数据已预处理为时间序列格式，以 15 秒的固定间隔对累计得分进行采样，从而为每场比赛生成长度为 $T = 192$ 的完整时间序列。我们仅关注比赛的前四节，并选取从 2020 年 1 月 2 日至 2023 年 6 月 9 日期间持续至少 48 分钟的比赛。

按照与板球相同的方法，从所有 7574 条可能的比赛得分轨迹中随机选取一个目标单位，供体池由目标比赛日期之前最近的 n 场比赛组成，选择范围为 $n = \{24, 48, 96, 192\}$。每条比赛得分轨迹都是长度为 $T = 192$ 的时间序列（对应 48 分钟的比赛时间，以 15 秒为间隔进行测量）。干预点为 $T_0 = 96$，对应于在中场休息时实施干预。在使用选定的供体数据拟合之前，对所有四种方法的数据都进行均值中心化。与板球分析类似，TASC 的隐维度和 RSC 的近似秩均设为 $d = 5$；使用 $\{10^{-1}, 10^{0}, ..., 10^{6}\}$ 中的值在干预前数据上进行交叉验证后，RSC 的岭回归系数设为 10^{3},。对于每种方法和每种供体池大小，实验均重复 100 次，每次使用同一组随机选取的目标单位。

我们测试了各方法在供体数据数量不断增加时的表现。图 17 展示了使用不同规模供体池 $n \in \{24, 48, 96, 192, 384\}$时整体的干预后均方根误差。对于 n，TASC 的 RMSE中位数最低，且在供体数量变化时表现稳定。CIM 对不断增加的 n表现出类似的稳健性，但 RMSE 略高于 TASC。随着 n增加，SC 和 RSC 对高维学习问题相对更敏感，其中 RSC 比 SC 更稳定。总体而言，随着 n增加，TASC 和 CIM 的 RMSE中位数逐渐降低，而 SC 和 RSC 则呈现相反趋势。

基于这些结果，我们聚焦于干预后均方根误差最低的情况（$n = 192$），并考察预测准确度如何随预测时域延长而变化。图18展示了按半刻钟区间（6分钟，包含24个时间点）划分的干预后均方根误差，其中包含 $n = 96$。在大多数情况下，TASC 的均方根误差最低，其次是 CIM、RSC 和合成控制。随着预测时域延长，所有方法的预测准确度都在下降，反映了长期预测固有的挑战。在预测近期（第三季度前半段）时，CIM 和 RSC 的表现与 TASC 相当；但随着预测时域延长，差距逐渐扩大，在第四季度后半段的预测中尤为明显。

[4]https://www.kaggle.com/datasets/wyattowalsh/篮球/data

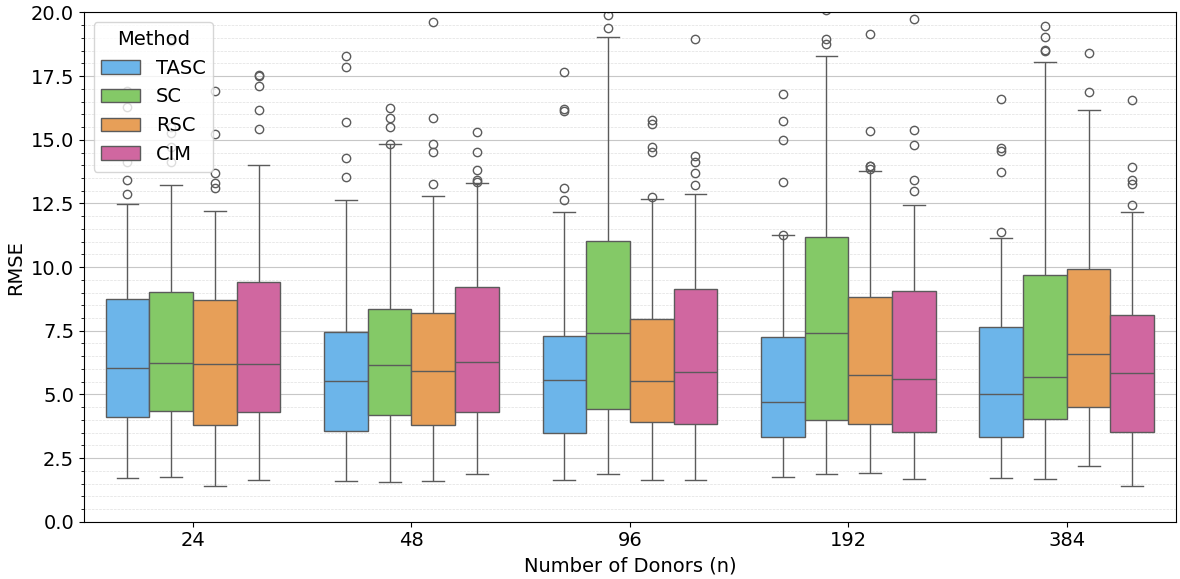

图 17：不同供体规模下 TASC、SC、RSC 和 CIM 的总体干预后均方根误差。

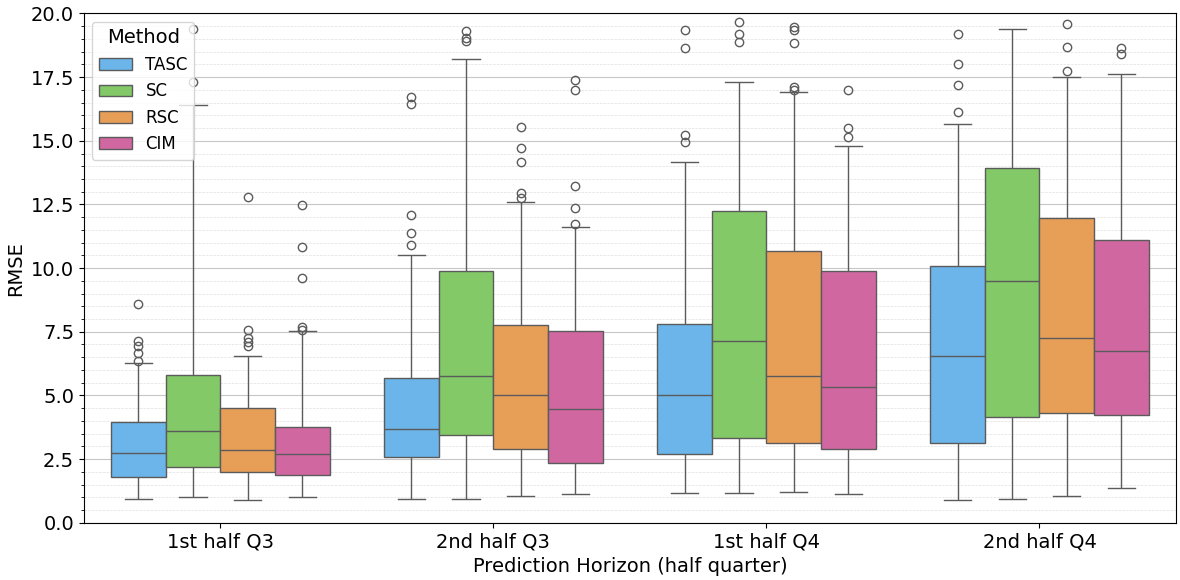

图18： 每个半刻钟（6 分钟）间隔预测时域的干预后均方根误差， $n = 19$ 2.

7 结论与未来工作

在这项工作中，我们提出了 TASC，这是一种用于对合成控制类数据建模的状态空间框架，并配有模型学习和反事实推断算法。通过显式建模潜在时间因子的时间演化，TASC 实现了出色的长期预测性能。对合成数据和真实案例的实验表明，即使观测噪声较高，TASC 仍然有效；它能更好地利用更大的供体池，对超参数 d的选择具有稳健性，并且生成的置信区间比 CIM 更窄。

我们目前的工作存在一些局限。尽管许多现实场景中都有多变量时间序列可供使用，TASC 模型却只关注单个时间序列。此外，TASC 强制采用严格线性且时不变的趋势，限制了模型的灵活性。最后，当前基于 EM 的学习算法速度较慢，且对初始化敏感，往往导致拟合效果不佳。尝试多种初始化可能会改善拟合效果，但会进一步降低计算效率。

展望未来，TASC 可以从多个方向进行扩展，以推进适用于更丰富面板数据的因果推断工具。首先，可以采用更先进的模型，例如更复杂的潜变量结构和非线性状态空间模型。这些改进将使 TASC 能够纳入多个辅助时间序列并捕捉非线性趋势。其次，我们可以改进学习算法，例如采用基于梯度的期望最大化或 MCMC采样器。这也与未来进一步开发更复杂的模型架构有关。最后，设计诊断工具，以识别 TASC 最有效的适用条件，将有助于将本研究的实证发现形式化。

致谢

本研究部分得到美国海军研究办公室（N00014-24-1-2687、N00014-24-1-2700）以及哥伦比亚大学-梦之队体育人工智能创新中心博士奖学金的资助。

参考文献

[1] David Card 和 Alan B Krueger. 最低工资与就业：新泽西州和宾夕法尼亚州快餐行业的案例研究。《美国经济评论》，90(5):1397–1420，2000年。[2] Alberto Abadie 和 Javier Gardeazabal. 冲突的经济成本：巴斯克地区的案例研究。《美国经济评论》，93(1):113–132，2003年。[3] Alberto Abadie、Alexis Diamond 和 Jens Hainmueller. 比较案例研究中的合成控制方法：估计加利福尼亚州烟草控制计划的影响。美国统计学会会刊，105(490):493–505，2010年。[4] Alberto Abadie 和 Jérémy L'Hour. 针对分解数据的惩罚合成控制估计量。美国统计学会会刊，116(536):1817–1834，2021年。[5] Alberto Abadie、Alexis Diamond 和 Jens Hainmueller. 比较政治学与合成控制法。《美国政治科学杂志》，59(2):495–510，2015年。[6] Noémi Kreif、Richard Grieve、Dominik Hangartner、Alex James Turner、Silviya Nikolova 和 Matt Sutton. 使用合成控制法评估涉及多个处理单位的卫生政策。《卫生经济学》，25(12):1514–1528，2016年。[7] Michael W Robbins、Jessica Saunders 和 Beau Kilmer. 高维微观数据合成控制方法框架：评估社区特定犯罪干预。美国统计学会会刊，112(517):109–126，2017年。[8] Giacomo Vagni 和 Richard Breen. 母职导致的收入与薪酬惩罚：使用个体合成控制方法对英国女性的估计。《欧洲社会学评论》，37(5):834–848，2021年。[9] Kristian Thorlund、Louis Dron、Jay JH Park 和 Edward J Mills. 临床试验中的合成对照与外部对照——研究人员入门指南。《临床流行病学》，457–467页，2020年。[10] Dennis Shen、Anish Agarwal、Vishal Misra、Bjoern Schelter、Devavrat Shah、Helen Shiells 和 Claude Wischik. 从阿尔茨海默病随机对照试验中获得个性化预测。《科学报告》，15(1):1671，2025年。[11] Madeleine Udell 和 Alex Townsend. 为什么大数据矩阵近似为低秩？《*SIAM*数据科学数学杂志》，1(1):144–160，2019年。[12] Muhammad Amjad、Devavrat Shah 和 Dennis Shen. 稳健合成控制。《机器学习研究杂志》，19(22):1–51，2018年。[13] Susan Athey、Mohsen Bayati、Nikolay Doudchenko、Guido Imbens 和 Khashayar Khosravi. 因果面板数据模型的矩阵补全方法。美国统计学会会刊，116:1716–1730，2021年。[14] Nikolay Doudchenko 和 Guido W Imbens. 平衡、回归、双重差分与合成控制方法：综述，2016年。NBER工作论文22791。[15] Muhammad Amjad、Vishal Misra、Devavrat Shah 和 Dennis Shen. mRSC：多维稳健合成控制。《*ACM*测量与分析计算系统会议录》，3(2):1–27，2019年。[16] Arindrajit Dube 和 Ben Zipperer. 使用合成控制汇集多个案例研究：最低工资政策的应用，2015年。IZA讨论论文第8944号。[17] Yiqing Xu. 广义合成控制方法：使用交互固定效应模型进行因果推断。《政治分析》，25(1):57–76，2017年。

[18] Saeyoung Rho、Andrew Tang、Noah Bergam、Rachel Cummings 和 Vishal Misra. Clustersc：通过供体选择推进合成控制。载于人工智能与统计国际会议，109–117页。PMLR，2025年。[19] Eli Ben-Michael、Avi Feller 和 Jesse Rothstein. 增强合成控制法。美国统计学会会刊，116(536):1789–1803，2021年。[20] Alberto Abadie. 使用合成控制：可行性、数据要求与方法论方面。《经济文献杂志》，59(2):391–425，2021年。[21] James H Stock 和 Mark W Watson. 通货膨胀预测。《货币经济学杂志》，44(2):293–335，1999年。[22] Allan W Gregory 和 Allen C Head. 生产率、投资和经常账户的共同波动与国家特定波动。《货币经济学杂志》，44(3):423–451，1999年。[23] Mario Forni、Marc Hallin、Marco Lippi 和 Lucrezia Reichlin. 广义动态因子模型：识别与估计。《经济学与统计学评论》，82(4):540–554，2000年。[24] Victor Chernozhukov、Kaspar Wüthrich 和 Yinchu Zhu. 针对反事实与合成控制的精确稳健保形推断方法。美国统计学会会刊，116(536):1849–1864，2021年。[25] Kay H Brodersen、Fabian Gallusser、Jim Koehler、Nicolas Remy 和 Steven L Scott. 使用贝叶斯结构时间序列模型推断因果影响。《应用统计年鉴》，9(1):247–274，2015年。[26] Danny Klinenberg. 系数随时间变化的合成控制：采用贝叶斯收缩的状态空间方法。《商业与经济统计杂志》，41(4):1065–1076，2023年。[27] Junzhe Shao、Mingzhang Yin、Xiaoxuan Cai 和 Linda Valeri. 基于状态空间模型的广义合成控制方法。载于*NeurIPS 2022*现实世界影响因果关系研讨会，2022年。[28] James Durbin 和 Siem Jan Koopman.《状态空间方法的时间序列分析》，牛津统计科学丛书第38卷。牛津大学出版社，牛津，第2版，2012年。

TASC 的理论依据

在本节中，我们将进一步阐述 TASC 背后的理论依据。我们通过证明经典合成控制及其许多变体具有置换不变性，可能会忽略 TASC 旨在恢复的预测信息，从而将第3.3节中概述的直觉严格化。这种联系可以通过数据处理不等式加以形式化。

我们框架的一个核心主张是，经典合成控制法及其许多变体在时间上具有置换不变性，因此无法利用嵌入在时间顺序中的预测信息。相比之下，我们的 TASC 方法利用状态空间模型和卡尔曼滤波器，提取用于预测未来结果的最小充分统计量。现在，我们使用数据处理不等式将这一主张形式化。令 $\mathcal{F}_{T_0}^{\text{seq}}$ 表示σ

-由有序处理前 -处理轨迹 $\{y_t\}_{t=1}^{T_0}$，和 $\mathcal{F}_{T_0}^{\text{bag}}$ σ的 -由相同对象生成的代数

^数据被视为由列组成的无序多重集。令 $\hat{x}_{T_0|T_0}$ 和 $P_{T_0|T_0}$ 分别表示基于截至 T_0 的观测，在 T_0 处的滤波隐状态和协方差。

命题 A.1. 考虑满足方程(2) 和 (3)的面板数据。则以下结论成立。

^*1. (*卡尔曼充分性*)* 从卡尔曼滤波器获得的滤波状态$\hat{x}_{T_0|T_0}$ 和协方差 $P_{T_0|T_0}$ 是最小充分统计量，可用于预测 $\{y_t\}_{t>T_0}$。也就是说，

$$y_t \perp\!\!\!\perp \mathcal{F}_{T_0}^{\text{seq}} \,\big|\, (\hat{x}_{T_0|T_0}, \hat{P}_{T_0|T_0}), \quad t > T_0.$$

2.（置换不变性导致的信息损失）由于 $\mathcal{F}_{T_0}^{\text{bag}} \subset \mathcal{F}_{T_0}^{\text{seq}}$ 对任意 $T_0 > 1$均成立，数据处理不等式（即布莱克威尔信息结构排序）意味着，

$$I\Big(\{y_t\}_{t>T_0}; \mathcal{F}_{T_0}^{\text{bag}}\Big) \;\le\; I\big(\{y_t\}_{t>T_0}; \mathcal{F}_{T_0}^{\text{seq}}\big)\,,$$

严格不等 当且仅当 $A \neq 0$和 Q有限时成立（即时间动态携带预测信号） .

3.（支配性）因此，基于卡尔曼充分统计量的贝叶斯最优预测器 $\hat{y}_{0,\,t}^{\text{KF}} = \mathbb{E}[y_{0,t} \mid \hat{x}_{T_0|T_0}, P_{T_0|T_0}]$ 所达到的均方误差，理论上不大于在这些假设下关于 $\mathcal{F}_{T_0}^{\text{bag}}$ 可测的置换不变预测器 $\hat{y}_{0\,t}^{\text{bag}}$ 的均方误差。

$$\mathbb{E}\big[(y_{0,t} - \hat{y}_{0,t}^{\text{KF}})^2\big] \;\le\; \mathbb{E}\Big[(y_{0,t} - \hat{y}_{0,t}^{\text{bag}})^2\Big]\,.$$

证明概要。 设 $Y^- \in \mathbb{R}^{N\times T_0}$ 为处理前供体矩阵（行对应 = 个供体，列对应 = 个时间索引），$y_0^- \in \mathbb{R}^{T_0}$ 为目标单元的处理前向量。经典合成控制求解

$$f^\star = \arg\min_{f\in\Delta^{n-1}} \|y_0^- - f^\top Y^-\|^2,$$

其中 Δ^{n-1} 是 n 维单位单纯形。对于作用于时间索引（列）的任意置换矩阵 $\Pi \in \mathbb{R}^{T_0\times T_0}$，我们有

$$\|y_0^-\Pi - f^\top Y^-\Pi\|^2 = \|y_0^- - f^\top Y^-\|^2,$$

因为欧几里得范数在重排下保持不变。因此，SC 和 RSC 目标函数不受时间置换的影响，这证实了置换不变性。

第 1 部分。由线性高斯模型中卡尔曼滤波器的标准充分性结论可得：($\hat{x}_{T_0|T_0}$，$P_{T_0|T_0}$) 对任意 $t > T_0$ 而言都是 p ($y_t \mid y_{1:T_0}$) 的充分统计量（例如参见 [28]）。

第 2 部分。由 $\mathcal{F}_{T_0}^{\text{bag}} \subset \mathcal{F}_{T_0}^{\text{seq}}$ 可得；根据数据处理不等式和 Blackwell 排序，受限于 $\mathcal{F}_{T_0}^{\text{bag}}$ 的预测器无法胜过使用 $\mathcal{F}_{T_0}^{\text{seq}}$ 的预测器，退化情形除外（$A = 0$）。

第 3 部分。由此可直接得出：基于卡尔曼的预测器使用来自 $\mathcal{F}_{T_0}^{\text{seq}}$ 的充分统计量，而具有置换不变性的合成控制受限于 $\mathcal{F}_{T_0}^{\text{bag}}$，这意味着其均方误差表现不劣；存在时间结构时则严格占优。

□

注释 A.2。 这一结果启发了一种自然的两阶段方法：*(i)* 从时间结构中提取卡尔曼充分统计量，然后 *(ii)* 对这些统计量应用 *SC*，而不是对原始观测值应用 *SC*。我们的 *TASC* 公式可以被视为将这一两阶段的卡尔曼滤波器与合成控制过程统一到生成模型中的方法，并且

在模型设定错误和含噪参数估计下具有更强的稳健性。从经验上看，可以通过置换压力测试对此进行验证：当处理前时间索引被打乱时，置换不变方法应不会出现性能下降，而 *TASC* 的性能则应下降，这证实了它利用了经典方法所缺失的时间信息。

B 基本算法

本节给出了构成 EM 方法的基本算法的完整伪代码：卡尔曼滤波器（算法 4）、无限方差卡尔曼滤波器（算法 5）、RTS 平滑器（算法 6），以及采用最大似然估计方法的 M 步（算法 7）。

算法 4： 卡尔曼滤波器

输入： $y_k \in \mathbb{R}^{n+1}$，前一时刻的估计值 m_{k-1}，P_{k-1}，当前参数 $\theta = \{A, H, Q, R, m_0, P_0\}$
输出： m_k，P_k
$m_{k|k-1} \leftarrow Am_{k-1}$ /* 上一时间步的预测 $k-1$ */
$P_{k|k-1} \leftarrow AP_{k-1}A^\top + Q$ /* 上一时间步的预测 $k-1$ */
$v_k \leftarrow y_k - Hm_{k|k-1}$
$S_k \leftarrow HP_{k|k-1}H^\top + R$
$K_k \leftarrow P_{k|k-1}H^\top S_k^{-1}$ /* 卡尔曼增益 */
$m_k \leftarrow m_{k|k-1} + K_k v_k$ /* 观测后更新 y_k */
$P_k \leftarrow P_{k|k-1} - K_k S_k K_k^\top$ /* 观测后更新 y_k */

Alg法 5: 无限方差卡尔曼滤波器

输入： $y_k \in \mathbb{R}^{n+1}$ 目标单位（首个）元素缺失时，先前的估计值 m_{k-1}，P_{k-1}，当前参数 $\theta' = \{A, H, Q, R, m_0, P_0\}$，其中 $R'_{1,\ 1} = \infty$ **输出：** m_k，P_k**定义** h_1，H_2，R_2 根据 $H = \begin{bmatrix} h_1^\top \\ H_2 \end{bmatrix}$，$R' = \begin{bmatrix} \infty & 0 \\ 0 & R_2 \end{bmatrix} y_k \leftarrow [h_1^\top m_{k|k-1}, y_{1,k}, \ldots, y_{n,k}]^\top$ /* 扩充目标值 */$m_{k|k-1} \leftarrow Am_{k-1}$ /* 根据上一个时间步进行预测 $k-1$ */$P_{k|k-1} \leftarrow AP_{k-1}A^\top + Q$ /* 根据上一个时间步进行预测 $k-1$ */ $v_k \leftarrow y_k - Hm_{k|k-1}$ /* 第一个元素为零 */$S_k \leftarrow HP_{k|k-1}H^\top + R' S_k^{-1} \leftarrow \begin{bmatrix} 0 & 0 \\ 0 & (\end{bmatrix}$ $H_2 P_{k|k-1} H_2^\top + R_2)^{-1}$] /* 根据舒尔补 */$K_k \leftarrow P_{k|k-1}H^\top S_k^{-1}$ $m_k \leftarrow m_{k|k-1} + K_k v_k$ /* 观测到 y_k 后更新 */$P_k \leftarrow P_{k|k-1} - K_k S_k K_k^\top$ /* 观测到 y_k 后更新 */

算法 6： Rauch–Tung–Striebel（RTS）平滑器

输入： 卡尔曼滤波器估计 m_k，P_k、平滑估计 m^s_{k+1}，P^s_{k+1}、当前参数 $\theta = \{A, H, Q, R, m_0, P_0\}$
输出： m^s_k，P^s_k
$m_{k+1|k} \leftarrow Am_k$ /* 根据卡尔曼滤波器估计 m_k 进行预测 */
$P_{k+1|k} \leftarrow AP_kA^\top + Q$ /* 根据卡尔曼滤波器估计 P_k 进行预测 */
$G_k \leftarrow P_kA^\top P^{-1}_{k+1|k}$
$m^s_k \leftarrow m_k + G_k[m^s_{k+1} - m_{k+1|k}]$ /* $m^s_t = m_t$ 对于最后一个时间步 $t = T_0$ 或 T */
$P^s_k \leftarrow P_k + G_k[P^s_{k+1} - P_{k+1|k}]G^\top_k$ /* $P^s_t = P_t$ 最后一个时间步 $t = T_0$ 或 T*/
返回 $m^s{}_k, P^s_k, G_k$

算法7： 基于最大似然估计方法的参数更新（M步）

输入： 当前参数 $\theta = \{A, H, Q, R, m_0, P_0\}$，序列长度 T，RTS参数 m_k^s，P_k^s，G_k对所有 $k \in \{0, \ldots, T\}$，观测值 y_k对所有 $k \in \{1, \ldots, T\}$**输出：** θ'**定义**$\Sigma = \frac{1}{T}\sum_{k=1}^{T} P_k^s + m_k^s m_k^{s\top}$
$\Phi = \frac{1}{T}\sum_{k=1}^{T} P_{k-1}^s + m_{k-1}^s m_{k-1}^{s\top} B = \frac{1}{T}\sum_{k=1}^{T} y_k m_k^{s\top} C = \frac{1}{T}\sum_{k=1}^{T} P_k^s G_{k-1}^\top + m_k^s m_{k-1}^{s\top}$
$D = \frac{1}{T}\sum_{k=1}^{T} y_k y_k^\top$**更新**$A' \leftarrow C\Phi^{-1}H' \leftarrow B\Sigma^{-1}Q' \leftarrow$ 对角化$(\Sigma - 2CA^\top + A\Phi A^\top)$ /* 对角化$(\cdot)$仅保留输入的对角元素 */$R' \leftarrow$ 对角化$(D - 2BH^\top + H\Sigma H^\top) m_0' \leftarrow m_0^s P_0' \leftarrow P_0^s + (m_0^s - m_0)(m_0^s - m_0)^\top$**返回**
$\theta' = \{A', H', Q', R', m_0', P_0'\}$

C 基准合成控制算法

本节将完整介绍第5节和第6节中使用的基准合成控制算法。

经典合成控制（SC）采用带有单纯形约束的垂直回归 [3]。算法8展示了我们研究中实现的经典SC，其中重要性矩阵 V被设为单位矩阵。请注意，我们在第99号提案研究（第 6.1 节）中的实现与原始分析（ [3]中）略有不同，因为我们仅使用感兴趣的目标时间序列（人均烟草销量，单位为包），不包含任何额外的协变量。

算法8： 合成控制 [3]

数据： 目标单元的干预前数据 $Y_1^- \in \mathbb{R}^{T_0}$，供体数据 $Y = [Y^-, Y^+] \in \mathbb{R}^{n \times T}$
结果： 反事实预测 $\hat{Y}^{1+}$，合成控制权重 f
1. 学习
$f = \arg\min_f ||Y_1^- - f^\top Y^-||^2$ 其中 $0 \le f \le 1$，$\sum_{i=1}^{n} f_i = 1$ /* 单纯形约束 */
2. 投影 $\hat{Y}_1^+ = f(Y^+)$
3. 推断 目标单位的干预估计因果效应为 $Y_1^+ - \hat{Y}_1^+$

稳健合成控制（RSC，[12]）将硬奇异值阈值处理（HSVT）作为预处理步骤，以去除观测数据中的噪声。然后，它使用干预前部分的数据学习纵向回归模型，并利用干预后的数据进行投影，以进行反事实推断。算法 9 描述了我们采用原始算法并使用观测概率 $p = 1$（无缺失数据）的做法。

算法 9： 稳健合成控制 [12]**数据：** 目标单元的干预前数据 $Y_1^- \in \mathbb{R}^{T_0}$、供体数据 $Y = [Y^-, Y^+] \in \mathbb{R}^{n \times T}$、保留的奇异值数量 d**结果：** 反事实预测 $\hat{}Y_1^+$、合成控制权重 f**1-1. 去噪**$Y = \sum_{i=1}^{min(n,T)} s_i u_i v_i^\top$**/*** 奇异值分解 (SVD) ***/**$\tilde{Y} = \sum_{i=1}^{d} s_i u_i v_i^\top$**/*** 硬奇异值阈值化 (HSVT) ***/****1-2. 学习** $f = \arg\min_f ||Y_1^- - f^\top \tilde{Y}^-||^2 + \lambda||f||^2$**2. 投影** $\hat{}Y_1^+ = f^\top \tilde{Y}$**+3. 推断** 目标单位所受干预的估计因果效应为 $Y_1^+ - \hat{}Y_1^+$