用于异质性因果效应估计的充分降维匹配

Haoran Zhao[1#], Yinghao Zhang[1#], Debo Cheng[2, 3*], Chen Li[4],Zaiwen Feng[5, 6, 7, 8, 9, 10*] [1] 华中农业大学理学院，中国武汉，430070 [2] 广西师范大学计算机科学与工程学院，中国桂林，541000 [3] 南澳大学STEM，澳大利亚 [4] 九州工业大学计算机科学与系统工程系，日本饭塚 820-8502 [5] 华中农业大学信息学院，中国武汉 430070 [6] 华中农业大学湖北省农业生物信息学重点实验室，中国武汉，430070 [7] 华中农业大学湖北洪山实验室，中国武汉 430070 [8] 华中农业大学作物遗传改良国家重点实验室，中国武汉，430070 [9] 武汉大学杂交水稻国家重点实验室，中国武汉八一路299号，430070 [10] 华中农业大学宏观农业研究院，中国武汉 430070 *通讯作者：Debo Cheng (chengdb2016@gmail.com) 和 Zaiwen Feng (Zaiwen.Feng@mail.hzau.edu.cn)

arXiv:2301.01549v1 [cs.DS] 4 Jan 2023

摘要——**因果推断在理解各领域数据生成过程的内在机制方面发挥着重要作用。由于维数灾难和数据充分性问题，利用具有高维协变量的观察性数据估计平均因果效应和个体因果效应颇具挑战。现有匹配方法无法有效估计个体因果效应，也无法解决因果推断中的维数灾难问题。为应对这一挑战，我们证明，经过充分降维（SDR）得到的约化集是用于混杂调整的平衡得分。基于该定理，我们提出采用一种充分降维方法，从原始协变量中获得约化表示集，再将该约化集用于匹配方法。具体而言，我们使用非参数模型来学习这样的约化集，从而避免模型设定错误。在真实世界数据集上的实验结果表明，所提方法优于对比的匹配方法。此外，我们进行了实验分析，结果表明，该约化表示足以平衡处理组与对照组个体之间的不平衡。**

索引词——**匹配、充分降维、因果推断、个体因果效应。**

I. 引言

近几十年来，因果推断在经济学 [1],健康 [2],统计学 [3]和计算机科学 [4]等诸多领域日益受到关注。因果效应估计在揭示因果推断中两个因素之间的因果强度方面发挥着重要作用。随机对照实验被视为识别此类因果强度的黄金标准。然而，由于成本、时间或伦理方面的限制，随机对照实验并不切实际 [1], [4]。因此，利用观察性数据估计因果效应是因果推断中的一种重要替代方法 [3], [5]–[8]。

混杂偏差是利用观察性数据估计因果效应时面临的主要挑战 [3], [9]。常见的方法是进行混杂调整，以消除利用观察性数据估计因果效应时的混杂偏差 [1]。最常见的混杂调整方法之一是匹配方法 [10]。匹配方法的核心思想是平衡处理组与对照组之间的协变量分布。具体而言，匹配的第一步是从对照组中寻找协变量与处理组个体相似的个体，以构建匹配对，然后根据匹配数据计算因果效应。匹配可通过选择协变量的不同函数，并采用不同的匹配算法来实现，例如马氏距离匹配、完全匹配、最近邻匹配和遗传匹配 [10]。

最广泛使用的函数是倾向得分，它被视为个体接受处理的概率 [11], [12]。通常，匹配方法通过倾向得分将多维协变量转换为标量，从而克服基于原始协变量进行匹配的困难 [13], [14]。事实上，在实际应用中，真实的倾向得分是未知的，因此需要根据数据估计倾向得分。然而，根据数据计算倾向得分时，模型误差或模型设定错误不可避免 [3], [15], [16]。因此，倾向得分并不是匹配方法的理想解决方案 [17]。

近年来，充分降维已成功用于从观察性数据中估计因果效应 [3], [5], [18]。例如，Luo等人提出了一种用于因果效应估计的充分维度缩减匹配方法 [19]。该匹配方法利用充分降维方法，将原始协变量缩减为降维协变量；在温和假设下，这些协变量兼具原始协变量和倾向

#本文作者对本项工作贡献相同

得分的优势。降维协变量的优势在于渐近稳定性，并且优于使用倾向得分获得的估计值。然而，该方法分别对两个子数据集（即处理组样本和对照组样本）进行充分降维，从而获得两组降维协变量，而不是在整个数据中寻找调整集。

在数据样本量不足时，所提出方法的性能会显著下降。Nabi 等人 [5]提出了一种半参数因果充分降维方法，以处理多重处理。Cheng 等人 [18] 提出了一种 CESD 匹配方法，通过核降维方法 [20] 针对处理变量对协变量进行降维，以估计平均因果效应，但该方法并不适用于异质性因果效应估计。

在本文中，我们提出了一种基于逆回归估计量（简称 MIRE 方法）的新型匹配方法，用于根据观察性数据估计平均因果效应和异质性因果效应。具体而言，我们的 MIRE方法使用一种充分降维方法，即逆回归估计量，基于全部数据学习相对于结果变量的降维协变量。随后，MIRE方法利用这些降维协变量进行匹配，以填补未观测结果（亦称反事实结果） [1], [3], [18]。我们的实验分析表明，降维协变量实现了良好的平衡，因此 MIRE 能够很好地解决混杂偏差。

总而言之，我们的工作作出了以下贡献。

- 我们解决了如何通过充分降维，从观察性数据中估计异质性因果效应的问题。
- 我们提出了一种基于逆回归估计量 MIRE 的新型匹配方法，用于从观察性数据中进行因果效应估计。
- 大量实验表明，我们提出的匹配方法在基于观察性数据进行因果效应估计方面更加有效。

II. 背景

A. 潜在结果模型

设 T 为二元处理，包含 $T_i = 0$ （一个对照个体）和 $T_i = 1$ （一个处理个体），其中 i 表示一个个体。未被分配接受某种处理的个体构成对照组，被分配接受某种处理的个体构成处理组。集合X 表示处理前协变量的集合，即在观测处理变量和结果变量之前及之后，协变量保持不变 [1], [18]。潜在结果 Y_i定义为个体 i的结果，即 $Y_i(0)$ 和 $Y_i(1)$ 分别是个体 i 未接受处理和接受处理时的潜在结果。请注意，同一个体 i 的两个潜在结果无法同时观测到。这是因果推断中的一个基本难题 [1], [21]。在潜在结果模型下，个体因果效应和平均因果效应可定义如下。

$$ITE = Y_i(1) - Y_i(0), \tag{1}$$

$$ATE = E[Y(1) - Y(0)]. \tag{2}$$

当我们利用潜在结果模型根据观察性数据估计因果效应时，通常需要满足以下假设。

假设 *1*（稳定单位处理值假设（*SUTVA*） *[1]*）：一个个体的潜在结果不受其他个体是否接受处理的影响。这意味着，对于每个个体，其潜在结果仅取决于处理 T，且每个处理水平都不存在不同的形式或版本。

假设 *2*（无混杂性 *[1]*）：给定协变量集合 X 后，处理 T 与潜在结果 Y 相互独立，形式化表示为 $Y(1)\ Y(0) \perp\!\!\!\perp T|X$。

该假设表明，所有协变量集合为 X 的个体都是随机分配接受处理的。

假设 *3*（重叠性 *[1]*）：对于每个个体，给定协变量集合 X 后，其接受处理或进入对照组的概率均非零，即 $0 < P(T = t|X) < 1,\ t = 0,\ 1$。

该假设表明，每个个体都有一定概率被分配到处理组 t。请注意，无混杂性和重叠性假设通常统称为“可忽略性假设”。由于反事实结果不可观测，因此无法直接根据数据检验可忽略性假设 [1]。因此，就估计 T 对 Y 的因果效应而言，协变量集合 X 包含所有相关和不相关的协变量。因此，有必要从 X 中找出一个调整集 Z，以准确估计 T 对 Y 的因果效应。倾向得分在基于观察性数据进行因果效应估计时始终发挥作用，其定义如下。

定义 *1*（倾向得分 *[1]*）：倾向得分定义为：在给定协变量集合 X的条件下，个体被分配到处理组的条件概率。

$$e(x) = P(T = 1|X) \tag{3}$$

Rosenbaum 和 Rubin [12]提出了记为 $b(x)$的平衡得分，它允许使用一类函数对协变量 X建模。在实践中，平衡得分也应满足无混杂假设（见 [1]中的引理12.2）。

$$Y(1), Y(0) \perp\!\!\!\perp T|b(x) \tag{4}$$

平衡得分包括倾向得分、原始协变量等。可忽略性假设还表明，平衡得分相同或近似相等的个体具有相同的协变量分布。

B. 匹配方法

匹配方法旨在从对照组中找出协变量分布与处理组个体相似的个体，从而利用对照组个体的潜在结果，填补处理组个体缺失的潜在结果。匹配方法的核心思想是模拟随机对照实验的过程。因此，匹配后的个体可视为个体 [13], [14], [21] 的反事实结果。在匹配方法中，可根据公式 5 和 6 得到第 i 个体的潜在结果。

$$\hat{Y}_i(1) = \begin{cases} Y_i, & \text{if } t = 1 \\ \frac{1}{|\ell(i)|}\sum_{j\in\ell(i)} Y_j, & \text{if } t = 0 \end{cases} \tag{5}$$

$$\hat{Y}_i(0) = \begin{cases} \frac{1}{|\ell(i)|}\sum_{j\in\ell(i)} Y_j, & \text{if } t = 1 \\ Y_i, & \text{if } t = 0 \end{cases} \tag{6}$$

其中，ℓ 表示该样本集与对照组或处理组中的第 i个个体相匹配。

一般而言，匹配后数据中两组的协变量分布比匹配前更加相似。因此，可以利用匹配结果计算平均因果效应，从而减小混杂因素的影响。

III. 提出的 MIRE 方法

本节首先证明，通过充分降维（SDR）得到的 $\psi(X) = X^T\beta$ 是一种平衡得分，可用于解决基于观察性数据估计因果效应时的混杂偏差。随后，我们介绍所提出的 MIRE 方法（基于逆回归估计量的匹配方法），用于因果效应估计。

A. 中心降维子空间是一种充分的平衡得分 *ore*

充分降维（SDR）是一种广泛用于数据处理的降维方法 [20],[22], [23]。对于响应变量 Y 和一组协变量X，充分降维旨在学习一个函数 $\psi(X)$，使得原始协变量 X 可以降至子空间 $X^T\beta \in \mathbb{R}^{p\times k}$，且满足 $k \ll p$。在本研究中，我们假设函数$\psi(X) = X^T\beta$ 存在，并可表示如下。

$$Y \perp\!\!\!\perp X|X^T\beta \tag{7}$$

其中，β 的列空间称为降维空间（DRS）。因此，在充分降维方法中，学习子空间 β 十分重要。值得注意的是，使用充分降维方法对协变量 X 进行降维所得到的 $X^T\beta$，可以视为ψ（X）的函数。当子空间 $S_{Y|Z}$ 是所有其他降维子空间的交集时，该子空间 $S_{Y|Z}$ 称为中心降维子空间（Central DRS） [20], [22]。中心降维子空间具有最小维度，且是唯一的降维子空间 [24]。因此，在我们的工作中，我们希望学习中心降维子空间 $S_{Y|Z}$。

定理*1*：给定包含处理 T、结果 Y以及处理前变量集合 X的观察性数据 O。假设中心子空间 $S_{Y|Z}$ 存在且维度为 r，其中 $r \ll p$。那么，存在一个任意的基矩阵 $\beta \in \mathbb{R}^{p\times k}$，使得 $X^T\beta$满足$Y(1)\ Y(0) \perp\!\!\!\perp T|X^T\beta$，且是一个平衡评分。

证明 *1*：根据处理前变量假设，X没有 W 或 Y的后代节点。根据可忽略性假设，基于混杂调整或平衡得分调整，可以无偏地计算 T 对 Y的因果效应。中心子空间 $S_{Y|Z}$的存在保证存在一个任意基矩阵 $\beta \in \mathbb{R}^{p\times k}$，使得根据中心子空间[5], [22] $X^T\beta \cong X$ 的不变性，$Y \perp\!\!\!\perp X|X^T\beta$成立。数学上，成立。此外，无混杂假设，即 $Y(1)\ Y(0) \perp\!\!\!\perp T|X$成立。因此，将 $X^T\beta$代入 X 在 $Y(1)\ Y(0) \perp\!\!\!\perp T|X$中，我们得到$Y(1)\ Y(0) \perp\!\!\!\perp T|X^T\beta$。因此，根据中心子空间的不变性，$X^T\beta$是一个平衡得分，并且$Y(1)\ Y(0) \perp\!\!\!\perp T|X^T\beta$。

根据定理1，$X^T\beta$ 是用于从观察性数据中进行无偏因果效应估计的平衡得分。这意味着，在匹配方法中，只需使用 $X^T\beta$ 作为平衡得分，即可实现无偏因果效应估计。值得注意的是，充分降维的另一项优势是能够在保留重要信息的同时降低原始协变量的维度。

在这项工作中，我们采用逆回归估计量（IRE）方法来学习中心降维子空间，因为IRE属于逆回归（IR）方法族，是一种渐近效率最高的最优方法 [25]。

当 Y 为连续值时，根据前述记号 [25]，将 Y 的取值范围划分为 h个切片。通过计算以下公式即可得到中心子空间 β 。

$$\beta_{\xi_y} = \sum_{y=1}^{h} Span(\xi_y) \tag{8}$$

其中 $\xi_y = \Sigma^{-1}$（E（$X|Y = y$）$- E$（X））。

此外，我们假设估计中心子空间所需的线性条件成立，即诱导出 E（$Z|P_{S_{Y|Z}}Z$）$= P_{S_{Y|Z}}Z$；基于此，中心子空间与 Z 关于 Y的逆回归相关联。

B. MIRE 的实现

在本研究中，我们使用逆回归估计量（IRE） [25]来估计我们 MIRE 方法的中心 DRS。IRE 方法通过最小化目标函数 9 来估计 DRS。首先，使用以下方程计算 IRE 方法的二次偏差。

$$\begin{aligned} F_k^{IRE}(S,C) =& (vec(\hat{\zeta}) - vec(SC))^T \hat{\Gamma}_{\hat{\zeta}}^{-1} \\ & (vec(\hat{\zeta}) - vec(SC)) \end{aligned} \tag{9}$$

其中 $\hat{\Gamma}^{-1}$ $\hat{\zeta}$ 是一个非奇异协方差矩阵。$S \in \mathbb{R}^{p\times k}$ 的列构成了 Span(ξ)的一个基，且 $C \in \mathbb{R}^{k\times(h-1)}$

$\hat{}$ 表示 ξ相对于 S的坐标。 ζ满足 $\hat{\zeta} \equiv \beta\gamma D_f A$, 其中 $\beta \in \mathbb{R}^{p\times k}$是 DRS 的一个基。 A是一个满足 $A^T A = I_{h-1}$和 $A^T 1_h = 0$的非随机矩阵。γ 是一个向量, 使得 $\xi = \beta\gamma$。 D_f 是一个对角矩阵，其对角线元素为向量 f 的各元素，其中 $\hat{}\hat{}\hat{}\mathrm{f} = (f_1, ..., f_h)$。 $vec(\cdot)$表示一种通过堆叠矩阵的列来构造向量的运算符，可形式化如下。

$$vec(C) = [(I_{h-1} \otimes S^T)V_n(I_{h-1} \otimes S)]^{-1} \times (I_{h-1} \otimes S^T)V_n vec(\hat{\xi}) \quad (10)$$

其中， $\otimes$ 是克罗内克积，是一种对任意尺寸的两个矩阵进行运算并生成分块矩阵的操作。 V_n 是一个正定矩阵, 等于 Γ^{-1} $\hat{\zeta}$ 的相合估计量$\hat{\Gamma}^{-1}$ $\hat{\xi}$ 。

$$\hat{S}_d = Q_{S_{(-d)}}[Q_{S_{(-d)}}(c_d^T \otimes I_p)Q_{S_{(-d)}}]^{-} \times Q_{S_{(-d)}}(c_d^T \otimes I_p)V_n\alpha_d \quad (11)$$

其中， $\alpha_d = vec(\hat{\zeta} - S_{(-d)}C_{(-d)})$， C_d 是 d的第 d行， C, $C_{(-d)}$ 由除第 d行以外的所有行组成， C，而 $Q_{S_{(-d)}}$ 在通常的内积下投影到Span($S_{(-d)}$)的正交补上。

匹配过程按照上述匹配步骤进行，距离则使用以下马氏距离进行度量。

$$D_{ij} = (\psi(X)_i - \psi(X)_j)'\Sigma^{-1}(\psi(X)_i - \psi(X)_j) \quad (12)$$

其中， Σ 是所有个体的协方差矩阵。实际应用中，使用观测数据计算协方差矩阵。

MIRE方法如算法1所示。

匹配后，协变量平衡的均值标准化差异。基于匹配数据，可以得到因果效应。平均因果效应可以使用公式9估计。第 i个个体的个体因果效应可以根据方程$\hat{Y}_i(1) - \hat{Y}_i(0)$估计。

算法1： MIRE

输入： 观察性数据 O 包括以下集合: 协变量 X ，处理 T 和结果 Y。

输出： 配对数据集

随机初始化 $S \leftarrow (s_1, ..., s_d)$
使用下式计算 B 的最小二乘系数方程 (10)，其中 S 固定
赋值 $e \leftarrow F_d(S, C)$ 和 $iter \leftarrow 0$
对于 $d \leftarrow 1$ **至** k，**执行**
  $S = (s_1, ..., s_k)$
  根据方程 11 找到新的 s_d
  $\hat{s}_d \leftarrow \frac{\hat{s}_d}{\|\hat{s}_d\|}$
  更新 $S \leftarrow (s_1, ..., \hat{s}_d, ..., s_k)$
  $C \leftarrow \arg_{c^*} \min F_k(S, C^*)$
  $e \leftarrow F_k(S, C)$; $iter \leftarrow iter + 1$
  如果 e 不再减少**则**
    $\tilde{S} \leftarrow S$
  end
  $\tilde{S} \leftarrow S$
结束
计算 Span(的有序基 $\widetilde{B}$)（即 $\hat{b}_1, \hat{b}_2$）
$X^* = X(\hat{b}_1, \hat{b}_2)$
用于 $i \leftarrow 1$ **到** n_1 **做**
  对 $j \leftarrow 1$ **到** n_2 **执行**
    $D_{ij} = (X_i^* - X_j^*)'\Sigma^{-1}(X_i^* - X_j^*)$
  end
  查找 j 中 D_{ij}最小的项
  配对个体 i, j。
结束

IV. 实验

由于我们无法获知真实世界数据集中的反事实结果 [1], [11], 因此很难使用真实世界数据集评估所提出的因果效应估计方法。为了评估所提出的 MIRE 方法的性能, 我们选取了两个半合成数据集：IHDP [26]和 TWINS [27]。这两个半合成数据集都广泛用于评估因果效应估计方法。此外，我们的实验还使用了一个真实世界数据集 Jobs [28], 因为文献中已有该数据集的实证因果效应。

为评估所提出的MIRE方法的性能，选取了七种常用的因果效应估计方法进行比较，包括：NNM（最近邻匹配 [13]）、PSM（倾向得分匹配 [12]）、贝叶斯加性回归树（BART）（ [26]）、因果森林（CF）（因果森林 [29]）、SDRM（充分降维匹配 [19]）、BCF（贝叶斯因果森林 [30]）、R-LASSO（使用 LASSO回归的 R-学习器 [31]）。这些方法因能够消除混杂偏差而被视为最高效的因果效应估计方法之一，即因果效应估计领域的最先进方法。

对于 IHDP 上的实验，我们使用异质性处理效应估计精度（PEHE）PEHE $= \frac{1}{N}\sum_{i=1}^{N} \left(y_{i1} - y_{i0} - (\hat{y}_{i1} - \hat{y}_{i0})\right)^2$ (( ) (ˆ ˆ)) 作为评估标准，以评估异质性因果效应。对于真实数据集 Jobs 上的实验，我们估计处理组平均因果效应（ATT），因为已知其经验 ATT [32]。对于 TWINS 上的实验，我们估计平均因果效应（ATE）。此外，我们~~使用均方根误差~~（RMSE）RMSE $= \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$和标准差（SD）作为评估指标，以评估所有方法的表现。

A. 基于 *IHDP* 的异质性效应估计

因果推断中的基准数据集 IHDP 来自文献 [26] 中进行的一项模拟研究。数据来自婴儿健康与发展计划（IHDP），

该计划于1985年启动，旨在为低出生体重早产儿提供高质量家庭访视服务。计划结果显示，处理后（即接受服务后），处理组在3岁时的认知测试得分显著高于对照组。

本研究收集了多种协变量，例如儿童出生体重、头围、早产周数、出生顺序和新生儿健康指标，以及妊娠期间母亲的行为和分娩期间的一些指标。为了模拟处理组与对照组之间的不平衡，我们按照 Hill [26], 的建议，从实验数据中剔除处理组中的非随机部分，即所有非白人母亲的子女，同时保留完整的对照组。随后，通过创建响应面来模拟每个个体的潜在结果，从而可以推导出真实的个体因果效应。由于响应面是已知的，因此可以对生成响应面的协变量进行调整，以满足可忽略性假设。我们采用 Hill [26] 使用的响应面 B:

$$Y(0) \sim N(exp((X+W)\beta_B), 1)$$

$$Y(1) \sim N(X\beta_B - \omega_B, 1)$$

其中， W 是一个与 X 维度相同且每个值均为 0.5 的偏移矩阵； β_B 是一个回归系数向量，对于 6 个连续协变量，其从 (0, 0.1, 0.2, 0.3, 0.4) 中按照概率 (0.5, 0.125, 0.125, 0.125, 0.125) 随机抽样；对于 18 个二元协变量，则按照概率 (0.6, 0.1, 0.1, 0.1, 0.1) 抽样； ω_B 是为保证 ATT = 4而选择的偏移量。因此，此时可以计算真实的个体因果效应。

图 1 展示了使用不同方法对该数据进行 1,000 次模拟的结果。对于每种方法，我们取 1,000 次结果的平均值作为 PEHE 值。从图中可以清楚地看出，在所有比较的方法中，MIRE 的表现最佳，这也说明了该方法的有效性。表现第二好的方法是 R-LASSO 和 SDRM。同时，广泛使用的方法 CF（因果森林）表现也不错，但不如 MIRE、R-LASSO 和 SDRM。其余方法 NNM（最近邻匹配）、PSM（倾向得分匹配）、BART（贝叶斯加性回归树）和 BCF（贝叶斯因果森林）的表现均不如 MIRE、R-LASSO、SDRM 和 CF。

B. 因果效应估计

1) Jobs：Jobs 数据集是 LaLonde 于 1986年建立的经典数据集，用于因果推断 [28]。该数据集源自一项临时就业计划，旨在为面临经济困境或缺乏工作技能的人提供职业技能。处理变量是个人是否参与该计划，潜在结果是个人 1978 年的收入。协变量主要包括教育、年龄、种族、婚姻状况、1974 年的收入，

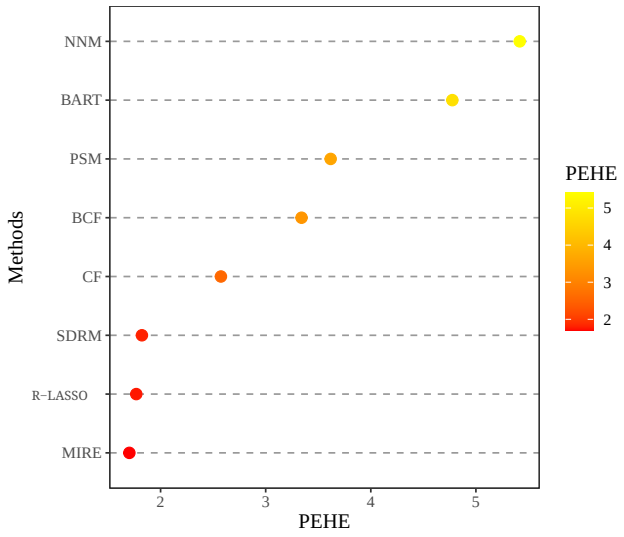


图 1. IHDP 模拟实验中使用不同方法估计的 PEHE 值。

以及 1975 年的收入。LaLonde 数据集中处理组样本的平均因果效应估计为 886 美元，标准误为 448 美元，我们使用这一估计值（886 美元）作为评估我们方法的标准 [18], [28], [32], [33]。

表 I 展示了使用不同方法估计 Jobs 数据集处理组样本平均因果效应的结果。从表 I 可以看出，MIRE 在处理组平均处理效应估计方面表现良好，估计的处理组平均处理效应（$519.09）接近基准值（$886），且标准差（$734.93）较小，也接近经验标准差（$448）。贝叶斯加性回归树和贝叶斯因果森林这两种方法在该数据集上也表现良好。而因果森林估计的处理组平均处理效应与标准值相差较大。

图 2 展示了在 Jobs 数据集上使用我们的 MIRE 降维后的散点图，显示了降维后的变量与响应变量 Y 之间的关系。从图 2 可以看出，降维后的两个协变量都与响应变量 Y 显著相关。该结果也表明，使用降维后的协变量进行匹配是合理的。

表 I Jobs 数据集处理组样本平均因果效应估计。

方法	估计的处理组平均处理效应	RMSE	SD
NNM	198.16	683.34	1,280.70
PSM	1,933.40	1,090.60	702.20
BART	931.30	2182.02	1,032.00
CF	182.24	497.65	890.61
SDRM	1,740.98	1,025.65	710.55
R-LASSO	1,271.63	385.63	825.23
BCF	697.88	544.90	511.82
MIRE	519.09	734.92	734.93

2）*TWINS*：TWINS基准数据集基于1983年至2000年间美国出生的双胞胎数据创建，

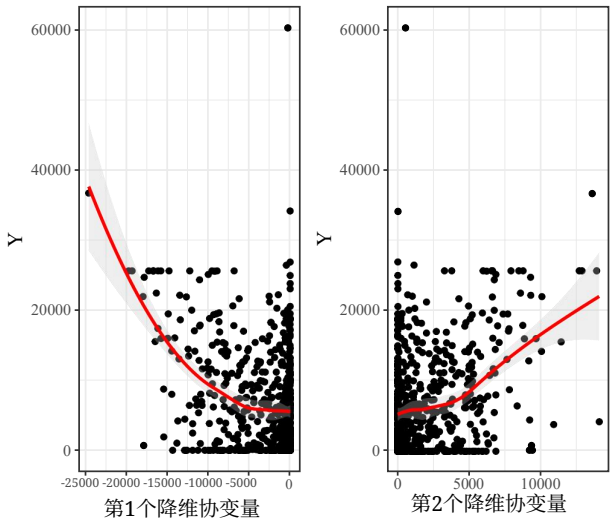


图 2. 我们的 MIRE 在 Jobs 数据集上得到的降维协变量与响应变量 Y by 之间的关系。

处理指的是出生时体重较重的双胞胎 [27]。潜在结果是双胞胎出生后第一年的死亡率。在 Louizos 等人的研究 [34], 中，他们选择性地隐藏了双胞胎中的一个，这相当于随机分配处理，因此使数据类似于随机试验数据。随后，他们选取体重低于 2 千克的双胞胎来建立数据集。该数据集包含 40 个协变量，例如父母的教育、婚姻状况、种族以及母亲分娩时的状况。我们根据以下公式模拟混杂因素的存在

[34] $W_i|X_i \sim Bern\ Sigmiod\ w'X_i\ +:$
$n \qquad W \sim U\ -0.1\ 0.1\ ^{40\times1}\ n \sim N\ 0\ 0.1$

（ () ()），其中（, ），（, ）。

我们构建的数据集的平均因果效应为−0.025。表 II 展示了使用不同方法估计 TWINS 数据集平均因果效应的结果。从表 II 所示结果可以看出，MIRE、CF、BCF 和倾向评分匹配均表现出色。这也表明，在存在混杂因素的情况下，MIRE 能够像最先进的因果效应估计量一样消除混杂偏差。

图 3 展示了对 TWINS 数据集中的协变量进行降维后的散点图，呈现了降维后的变量与响应变量 Y 之间的关系。从图 3 可以看出，降维后的协变量与响应变量显著相关。这也证实了 MIRE 能够有效地利用观察性数据估计因果效应。

V. 相关工作

潜在结果模型在因果推断中被广泛使用 [1], [13], [14], [21]。我们提出的 MIRE 方法基于潜在结果模型，并采用一些较弱的假设，例如处理前变量假设和可忽略性假设。匹配方法是因果推断中最常用的方法之一 [13],

表 II TWINS 数据集上的平均因果效应估计值。

方法	估计的平均处理效应	RMSE	SD
NNM	-0.0218	0.0029	0.0468
PSM	-0.0252	0.0043	0.0502
BART	-0.1794	0.2793	0.3188
CF	-0.0252	0.0034	0.0408
SDRM	-0.0230	0.0032	0.0480
R-LASSO	-0.0623	0.0221	0.0901
BCF	-0.0249	0.0398	0.0639
MIRE	-0.0252	0.0037	0.0002

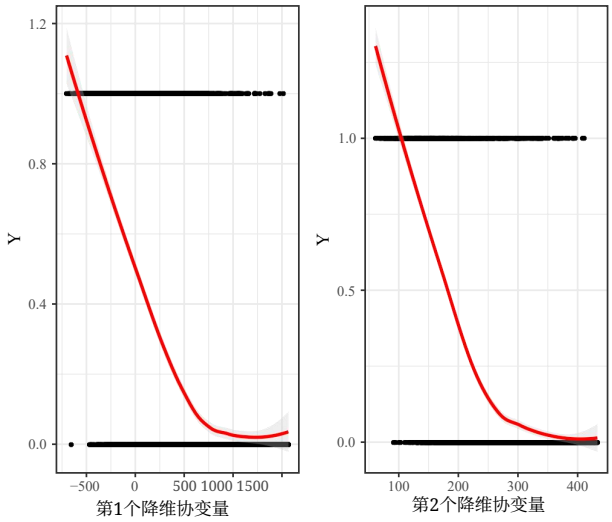


图 3. 在 TWINS 数据集上进行 MIRE 后，降维协变量与响应变量 Y 之间的关系。

[18], [19]。Stuart 总结了匹配方法的原理和步骤，以及它在许多实际问题中的应用 [10]。接下来，我们回顾一些与我们提出的 MIRE 方法密切相关的研究。

在匹配过程中，很难在两组（对照组和处理组）中找到协变量维度相同的个体。为了解决这个问题，Rubin 和 Rosenbaum [13] 引入了倾向得分，其定义为在给定协变量条件下，某个个体被分配到特定处理的条件概率。此外，倾向得分已被证明是一种平衡得分 [1]。最常用的倾向得分匹配方法是在匹配过程中用倾向得分替代原始协变量。此外，将倾向得分匹配与马氏距离匹配相结合可得到 GenMatch；GenMatch 使用遗传搜索算法获取权重，从而完成匹配 [10]。Luo 和 Zhu 使用充分降维（SDR）方法降低处理组和对照组中协变量的维度，然后基于马氏距离进行匹配 [17]。

与 MIRE 最相关的工作是文献 [18], [19] 提出的 SDR 匹配。Luo 和 Zhu 的工作 [19]

考虑对处理组单位和对照组单位的子数据集进行降维，得到两个降维协变量作为匹配的平衡得分。由于将全部样本划分为两个子数据集会导致数据不足，该方法可能存在偏差。Cheng 等人 [18] 旨在根据观察性数据估计平均因果效应，但并非用于异质性因果效应估计。相比之下，我们的理论发现支持一种用于异质性因果效应估计的数据驱动方法。

近年来，大量基于深度学习的方法被提出，用于从观察性数据中估计因果效应 [35]–[39]。基于深度学习的方法的主要优势在于，神经网络能够学习变量之间复杂的非线性关系，并且能够很好地处理高维数据集。然而，参数调优效率较低，而且可解释性也不佳。

另一类研究关注从含有潜在混杂因素的数据中进行因果效应估计 [7], [8], [39]。当给定工具变量（IV）时，即使数据中存在潜在变量，也可以从数据中无偏地计算 T 对 Y 的因果效应 [29], [40]–[43]。由于基于工具变量的估计量不依赖于可忽略性假设，因此它们与我们的MIRE方法没有直接关系。

VI. 结论

在本研究中，我们证明，通过充分降维方法得到的中心降维子空间是一种平衡得分，足以控制基于观察性数据进行因果效应估计时的混杂偏差。我们的研究结果为使用降维后的协变量进行匹配提供了理论支持。根据所提出的定理，我们提出了一种数据驱动方法，即 MIRE，用于在温和假设下估计观察性数据中的因果效应。首先，MIRE 利用逆回归估计量对原始协变量进行降维，然后使用降维后的协变量进行匹配。实验验证了所提出的 MIRE 的优势。首先，基于 Jobs 数据集的平均因果效应估计结果表明，与其他匹配方法相比，我们的方法所得估计结果更接近先前研究推荐的基准值（$886）。其次，个体因果效应估计结果表明，基于降维协变量进行匹配，更容易在匹配过程中为个体找到另一组中的配对对象。与其他匹配方法相比，我们的方法在个体因果效应估计方面表现出显著优势。基于 IHDP 数据集的估计结果表明，我们的方法在匹配过程中能够匹配更多个体。与其他因果推断方法相比，我们的方法在异质性效应估计准确度（以 PEHE 表示）和混杂因素控制方面也表现出一定优势。

致谢

本研究项目部分由湖北洪山实验室重大项目（项目编号：2022HSZD031）资助，部分由中国海洋国防技术创新中心创新基金（项目编号：JJ-2021-722-04）资助，部分由国家自然科学基金（项目编号：62076041 和 61806027）资助，部分由中央高校基本科研业务费（项目编号：2662020XXQD01、2662022JC004）资助；此外，还获得武汉大学杂交水稻国家重点实验室开放基金以及 Huzhong Agricultural University 作物遗传改良国家重点实验室开放基金（项目编号：ZK202203）的资助。

参考文献

[1] G. W. Imbens 和 D. B. Rubin，《统计学、社会科学与生物医学科学中的因果推断》。剑桥大学出版社，2015年。[2] A. F. Connors、N. V. Dawson 等，“重度慢性阻塞性肺病急性加重后的结局：SUPPORT 研究者（理解预后以及对结局和治疗风险的偏好研究）”，《美国呼吸与重症医学杂志》，第154卷，第4期，第959–967页，1996年。[3] T. Ghosh、Y. Ma 等，“用于可行且稳健地估计平均因果效应的充分降维”，《辛尼卡统计学》，第31卷，第2期，第821页，2021年。[4] J. Pearl，《因果关系》。剑桥大学出版社，2009年。[5] R. Nabi、T. McNutt 等，“高维处理的半参数因果充分降维”，arXiv 预印本 arXiv:1710.06727，2017年。[6] C. Li、A. Annisa、A. Zaman、M. Qaosar、S. Ahmed 和 Y. Morimoto，“使用区域天际线查询的基于 MapReduce 算法的位置推荐”，《算法》，第11卷，第12期，第191页，2018年。[7] D. Cheng、J. Li 等，“含隐藏变量的观察性数据中的因果查询”，载于《欧洲人工智能会议论文集》，2020年，第2551–2558页。[8] D. Cheng、J. Li、L. Liu 等，“迈向从含隐藏变量的数据中进行唯一且无偏的因果效应估计”，《IEEE 神经网络与学习系统汇刊》，第1–13页，2022年。[9] M. D. Cattaneo，“可忽略性假设下多值处理效应的高效半参数估计”，《计量经济学杂志》，第155卷，第2期，第138–154页，2010年。[10] E. A. Stuart，“因果推断的匹配方法：综述与展望”，《统计科学：数理统计学会评论杂志》，第25卷，第1期，第1–21页，2010年。[11] D. B. Rubin，“使用多变量匹配抽样和回归调整控制观察性研究中的偏差”，《美国统计学会杂志》，第74卷，第366a期，第318–328页，1979年。[12] P. R. Rosenbaum 和 D. B. Rubin，“倾向得分在因果效应观察性研究中的核心作用”，《生物统计学》，第70卷，第1期，第41–55页，1983年。[13] D. B. Rubin，“通过匹配消除观察性研究中的偏差”，《生物统计学》，第29卷，第159–183页，1973年。[14] ——，“因果效应观察性研究的设计与分析：与随机试验设计的相似之处”，《医学统计学》，第26卷，第1期，第20–36页，2007年。[15] J. Hahn，“倾向得分在平均处理效应高效半参数估计中的作用”，《计量经济学》，第66卷，第2期，第315–331页，1998年。[16] K. Greenewald、K. Shanmugam 等，“用于样本高效处理效应估计的高维特征选择”，载于《国际人工智能与统计会议论文集》，2021年，第2224–2232页。[17] W. Luo、Y. Zhu 等，“使用充分降维估计基于回归的因果效应”，《生物统计学》，第104卷，第1期，第51–65页，2017年。

[18] D. Cheng、J. Li、L. Liu、T. D. Le、J. Liu 和 K. Yu，“用于平均因果效应估计的充分降维”，《数据挖掘与知识发现》，第36卷，第3期，第1174–1196页，2022年。[19] W. Luo 和 Y. Zhu，“使用充分降维进行因果推断的匹配”，《商业与经济统计杂志》，第38卷，第4期，第888–900页，2020年。[20] K. Fukumizu、F. R. Bach 等，“基于再生核希尔伯特空间的监督学习降维”，《机器学习研究杂志》，第5卷，1月号，第73–99页，2004年。[21] D. B. Rubin，“估计随机与非随机研究中的处理因果效应”，《教育心理学杂志》，第66卷，第5期，第688页，1974年。[22] R. D. Cook，“二元响应回归的图形表示”，《美国统计学会杂志》，第91卷，992，1996年，435，983号，第–页。[23] ——，《回归图形：通过图形研究回归的思路》。约翰·威立父子公司，2009年，第482卷。[24] A. F. Connors、T. Speroff 等，“右心导管插入术在危重患者初始护理中的有效性”，《美国医学会杂志》，第276卷，第11期，第889–897页，1996年。[25] R. D. Cook 和 L. Ni，“通过逆回归实现充分降维：一种最小差异方法”，《美国统计学会杂志》，第100卷，第470期，第410–428页，2005年。[26] J. L. Hill，“用于因果推断的贝叶斯非参数建模”，《计算与图形统计学杂志》，第20卷，第1期，第217–240页，2011年。[27] D. Almond、K. Y. Chay 等，“低出生体重的成本”，《经济学季刊》，第120卷，第3期，第1031–1083页，2005年。[28] R. J. LaLonde，“使用实验数据评估培训项目的计量经济学评估”，《美国经济评论》，第76卷，第4期，第604–620页，1986年。[29] S. Athey、J. Tibshirani 等，“广义随机森林”，《统计年鉴》，第47卷，第2期，第1148–1178页，2019年。[30] P. R. Hahn、J. S. Murray 和 C. M. Carvalho，“用于因果推断的贝叶斯回归树模型：正则化、混杂与异质性效应（含讨论）”，《贝叶斯分析》，第15卷，第3期，第965–1056页，2020年。[31] X. Nie 和 S. Wager，“异质性处理效应的准预言机估计”，《生物统计学》，第108卷，第2期，第299–319页，2021年。[32] K. Imai 和 M. Ratkovic，“协变量平衡倾向得分”，《英国皇家统计学会杂志B辑（统计方法论）》，第76卷，第1期，第243–263页，2014年。[33] A. Diamond 和 J. S. Sekhon，“用于估计因果效应的遗传匹配：一种实现观察性研究平衡的通用多变量匹配方法”，《经济学与统计学评论》，第95卷，第3期，第932–945页，2013年。[34] C. Louizos、U. Shalit 等，“使用深度潜变量模型进行因果效应推断”，载于《神经信息处理系统会议论文集》，2017年，第6446–6456页。[35] U. Shalit、F. D. Johansson 等，“个体处理效应估计：泛化界与算法”，载于《国际机器学习会议论文集》，2017年，第3076–3085页。[36] L. Yao、S. Li 等，“从观察性数据中估计处理效应的表征学习”，载于《神经信息处理系统会议论文集》，2018年，第2638–2648页。[37] J. Yoon、J. Jordon 等，“Ganite：使用生成对抗网络估计个体化处理效应”，载于《国际学习表征会议论文集》，2018年，第1–22页。[38] S. M. Shortreed 和 A. Ertefaie，“结果自适应套索：用于因果推断的变量选择”，《生物统计学》，第73卷，第4期，第1111–1122页，2017年。[39] N. Kallus、X. Mao 等，“未观测混杂下个体层面因果效应的区间估计”，载于《国际人工智能与统计会议论文集》，2019年，第2281–2290页。[40] D. Cheng、J. Li、L. Liu、J. Zhang、T. duy Le 和 J. Liu，“无需完整知识的因果推断祖先工具变量方法”，载于《国际人工智能联合会议》，2022年，第4843–4849页。[41] D. Cheng、J. Li、L. Liu、K. Yu、T. D. Lee 和 J. Liu，“从观察性数据中发现用于因果推断的祖先工具变量”，arXiv 预印本 arXiv:2206.01931，2022年。

[42] M. A. Hern´an 和 J. M. Robins，“用于因果推断的工具：流行病学家的梦想？”《流行病学》，第 17 卷，372，2006 4，360 期，第 – 页。[43] E. P. Martens、W. R. Pestman等，“工具变量：应用与局限性”，《流行病学》，第 17 卷，第 3 期，第 260–267 页，2006 年。