

# 经济试验的协变量平衡自适应设计

陈芳露<sup>1</sup>, 李 扬<sup>1,2</sup>, 秦祎辰<sup>3</sup>, 杨昊宇<sup>1</sup>

(1. 中国人民大学统计学院, 北京 100872; 2. 中国人民大学应用统计科学研究中心, 北京 100872;  
3. 辛辛那提大学运筹商业分析与信息系统系, 辛辛那提 45221, 美国)

**摘 要** 在经济学领域的随机对照试验中, 受试者的分配通常按照完全随机的方式进行。然而, 完全随机化可能无法使基线协变量在试验组和对照组间的分布均衡可比, 导致试验的解释性与准确性降低, 甚至会得出错误的分析结果。本文在经济学随机对照试验中引入协变量平衡自适应设计, 该设计在分配过程中自适应地对协变量的平衡性进行调整, 从而能够获得协变量在组间分布相对均衡的分配方案。本文基于一项探究个性化信息是否能够影响养老金个人账户储蓄的随机对照试验案例, 分析比较了不同随机化设计对于在随机化试验中的协变量平衡以及处理效应估计等方面的影响。实证分析表明, 相比于完全随机化, 考虑协变量平衡调整的随机化设计能够降低组间的协变量的不平衡程度, 并提高后续对于平均处理效应的估计精度和检验功效。

**关键词** 随机对照试验; 自适应随机化; 马氏距离; 因果推断; 检验功效

## Covariate-adjusted Randomization Design for Economic Experiment

CHEN Fanglu<sup>1</sup>, LI Yang<sup>1,2</sup>, QIN Yichen<sup>3</sup>, YANG Haoyu<sup>1</sup>

(1. School of Statistics, Renmin University of China, Beijing 100872, China; 2. Center for Applied Statistics, Renmin University of China, Beijing 100872, China; 3. Department of Operations, Business Analytics and Information Systems, University of Cincinnati, Cincinnati 45221, USA)

**Abstract** In economic randomized controlled trials, subjects are often assigned by complete randomization. However, under complete randomization, the distribution

收稿日期: 2023-02-06

基金项目: 中国人民大学公共健康与疾病预防控制文理交叉重大创新平台 2022 年度“中央高校建设世界一流大学(学科)和特色发展引导专项资金”(2023PDPC)

Supported by Public Health & Disease Control and Prevention, Fund for Building World-Class Universities (Disciplines) of Renmin University of China (2023PDPC)

作者简介: 陈芳露, 硕士, 研究方向: 试验设计、因果效应估计, E-mail: chenfanglu1031@ruc.edu.cn; 李扬, 中国人民大学应用统计科学研究中心研究员、统计学院教授, 博士生导师, 研究方向: 相关型数据分析、模型选择与不确定性评价、潜变量建模、临床试验设计等, E-mail: yang.li@ruc.edu.cn; 秦祎辰, 美国辛辛那提大学运筹商业分析与信息系统系副教授, 博士生导师, 研究方向: 网络数据和高维数据等复杂数据分析、模型不确定性评价与可视化、临床试验设计等, E-mail: qinyin@ucmail.uc.edu; 通信作者: 杨昊宇, 博士, 研究方向: 试验设计方法、因果效应估计、网络数据分析, E-mail: haoyuyang@ruc.edu.cn.

of baseline covariates between treatment and control groups is usually incomparable, which decreases interpretability and accuracy of the experiment, or even distorts the results. In this paper, we introduce the covariate-adjusted randomization design for the economic randomized controlled trial. The covariate-adjusted randomization design adaptively adjusts the covariates balance during the allocation process so as to achieve the better covariate balance. Based on a randomized controlled trial investigating whether personalized information can affect pension savings, we compare the impact of three different randomizations on the covariates balance and estimation of the average treatment effect. Empirical analysis results show that, compared to complete randomization, covariate-adjusted randomization design can significantly reduce the covariate imbalance and thus improve the subsequent estimation precision and testing power.

**Keywords** randomized controlled trial; adaptive randomization; Mahalanobis distance; causal inference; test power

## 1 引言

伴随经济学领域日益增长的循证需求, 研究者们面临着如何将干预效应与混杂因素的影响分离开来的挑战. 随机对照试验是评价干预效果的金标准, 因为其可以通过随机化降低混杂因素对因果估计的影响. 近年来, 随机对照试验在发展经济学和劳动经济学等领域的应用愈加广泛 (Schlosser and Shanan (2022), McIntosh and Zeitlin (2022)). 2019 年, 三位经济学家因“通过随机对照试验改变了发展经济学, 并显著提高了人类与全球贫困作斗争的能力”被授予诺贝尔经济学奖. 随机对照试验提高了经济分析的可信度 (Pischke (2010)), 其在探究因果关系上有着独特的优势. 这种优势得益于随机对照试验的设计思路: 将研究对象进行随机分组, 对不同组实施不同的干预并使其他混杂因素在组间均衡可比. 因此, 随机对照可以明确地识别因果效应并通过控制干预提高内部有效性. 除此之外, 驱动经济学家大力发展试验方法的关键原因还包括: 研究人员可以在受控条件下收集试验数据, 数据的完整性和科学研究的可复制性会大大提高; 试验方法促进了试验数据和经济理论之间的良性循环, 并揭示出试验研究与经济理论相结合时最有效 (Svorenčík (2021)).

随机化是随机对照试验的基本原则, 随机对照试验的内部有效性依赖于研究人员能否通过随机化得到特征可比的试验组和对照组. 传统的完全随机化 (complete randomization, CR) 能够确保每个受试者随机地获得分配, 但往往不能确保试验组和对照组在特征层面均衡可比. Fisher (1926) 指出: “许多研究者在实施完随机化后, 震惊地发现试验组之间远达不到均衡可比的程度”, 这种现象通常被称为随机对照试验中的协变量不平衡. 一般而言, 协变量是受试者的基本特征 (不包括干预), 例如年龄、性别和受教育水平等. 协变量不平衡可能会影响随机对照试验的分析结果, 例如, 研究者设计随机对照试验评估就业培训对受试者就业能力的影响, 若参加就业培训的受试者的受教育水平显著高于未参加培训的受试者的受教育水平, 研究者观察到的两组受试人群间就业能力的差异便可能来自于受教育水平的差异而不是真实的培训干预. 当随机对照试验中存在协变量不平衡时, 通过  $t$  检验或者方差分析得到的平均处理效应估计会存在偏差, 且估计量的估计精度会大打折扣 (Egbewale, Lewis and

Sim (2014)). 在数据分析阶段, 研究人员常常对组间的协变量进行均衡性检验, 并将检验结果作为是否需要协变量进行额外调整的标准. 虽然这种事后调整方法 (例如回归和倾向评分匹配等) 可以降低协变量不平衡对研究结果的影响, 但它们的估计效率往往低于在试验阶段进行协变量平衡的事前调整 (Bruhn and McKenzie (2009)). 因此, 学者建议研究者们在试验阶段尽可能地进行协变量平衡 (Imbens and Rubin (2015)).

协变量平衡调整的随机化设计被广泛应用于临床试验中 (殷赫然等 (2021), Pignata et al. (2021)). 在试验阶段, 该类设计在保证随机分配的同时对组间的协变量进行平衡调整, 以降低平均处理效应估计的估计偏差并提高试验的效率. 考虑协变量平衡调整的随机化设计包括但不限于: 分层置换区组 (stratified permuted-block, SPB) 设计 (Matts and Lachin (1988))、最小化设计 (Hu and Hu (2012))、重随机化设计 (rerandomization, RR) (Morgan and Rubin (2012)) 和协变量自适应随机化 (adaptive randomization via Mahalanobis distance, ARM) (Qin et al. (2022)) 设计. 在经济试验中引入临床随机对照试验文献中相对更复杂的协变量平衡随机化方法可以大大提高对于平均处理效应的估计精度 (Bruhn and McKenzie (2009)). 本文以一项探究个性化信息如何影响养老金个人账户储蓄的随机对照试验为例 (Fuentes et al. (2017)), 对协变量的平衡以及后续估计等方面对协变量平衡随机化设计进行比较研究. 本文旨在通过对已有设计的比较与实证分析, 突出协变量平衡随机化在试验中的重要性, 为相关研究方法的进一步拓展提供思路.

## 2 研究背景

伴随老龄化不断加剧, 许多国家采用市场化的应对策略. 譬如, 某些国家实行个人账户制的养老金模式: 缴费完全进入个人账户, 账户持有人达到退休年龄后从个人账户内提取养老金. 当实施完全积累型养老金个人账户模式时, 普通居民可能无法完全理解“今天储蓄, 明天更好退休”这一理念, 从而导致消极的储蓄行为, 进而出现养老金储蓄存在缴款率低和积累少等问题. 目前已有多项研究表明, 向养老金储蓄人推送鼓励储蓄信息这一举措在提高养老金储蓄和延迟退休年龄等方面发挥了促进作用 (Pinto (2013), Mastrobuoni (2011)). Fuentes et al. (2017) 进一步探究了向储蓄人推送个性化信息是否比通用性信息更能促进养老金储蓄行为. 在这项研究中, 研究者通过随机对照试验对干预举措的效果进行研究. 研究人员根据受试者国民身份证号最后一位数字的奇偶性进行分配, 通过“伪随机” (Peter, Richard and David (2015)) 的方式将 2604 名受试者分为试验组与对照组, 为试验组中的受试者提供个性化信息并为对照组中的受试者提供通用性信息. 接受个性化信息的个人收到了他们在维持现状、增加强制储蓄的月数、增加自愿储蓄和推迟一年退休四种情况下通过模拟得出的退休后每月可提取的养老金金额. 在养老金个性化估计时, 个人面对的是他的储蓄水平、缴款密度、收入水平和基金类型等方面的实际情况. 接受通用性信息的个人收到了关于如何增加储蓄以改善未来养老金收入的一般信息和建议, 包括增加强制储蓄月数、增加自愿储蓄和延迟退休的好处等, 但这些建议没有参考个人的情况. 然而, 该研究采用的“伪随机”分配方式可能会导致协变量在组间分布的不平衡.

针对该养老金随机对照试验案例中可能存在的协变量不平衡, 表 1 展现了对于试验组和对照组的受试者基线信息的进行描述性统计和差异性检验的结果, 其基于  $t$  检验和

表 1 协变量的描述与检验

协变量	定义	对照组均值	试验组均值	$t$ 检验 $P$ 值	K-S 检验 $P$ 值
edad	年龄	39.021	37.451	0.002	0.044
rem_actual	前六个月的平均工资	443.443	482.189	0.025	0.000
deseada_pens	希望获得的养老金	504.510	580.046	0.325	0.995
esperada_pens	预期能获得的养老金	249.946	293.929	0.309	0.601
saldo_uf	强制账户	379.505	423.898	0.126	0.878
comodidad	对养老金系统的熟悉程度	4.784	4.724	0.405	0.989
TotCotVol	过去一年自愿缴纳养老金金额	20.731	32.783	0.388	0.965
TotCot	过去一年强制缴纳养老金金额	428.165	441.202	0.524	0.562
NCotVol	过去一年自愿缴纳养老金月数	0.396	0.446	0.549	0.983
NCot	过去一年强制缴纳养老金月数	7.884	8.038	0.434	0.770
res_sim_1	模拟器计算能获得的养老金	263.551	276.069	0.322	0.034
ErrorPension	预期与模拟的差距	-17.860	13.605	0.476	0.160

Kolmogorov-Smirnov 检验判断试验组和对照组中的受试者在基线特征上是否存在显著性差异. 在十二个连续型的基线协变量中,  $t$  检验显示年龄 (edad) 和工资 (rem\_actual) 变量的均值存在显著差异, Kolmogorov-Smirnov 检验显示年龄 (edad)、工资 (rem\_actual) 和模拟器计算能够获得的养老金 (res\_sim\_1) 变量的分布存在显著性差异, 说明该随机对照试验中存在着协变量不平衡.

3 考虑协变量平衡的随机化设计

在试验中, 考虑协变量平衡的随机化设计尤为重要. 在《新英格兰医学杂志》2018 年发布的涉及随机化的临床试验中, 83.3% 的试验使用了协变量平衡的随机化设计 (刘雅琦等 (2021)). 考虑协变量平衡随机化的已有研究可被大致分为对于离散型协变量的平衡以及对于连续型协变量的平衡. 关注于对离散型协变量的平衡, 分层置换区组 (SPB) 随机化设计是目前临床试验中较为流行的随机化方法之一. 然而, SPB 设计往往受制于试验样本量和区组数, 且其能够平衡的协变量个数也有限. 当协变量个数较多时, 最小化设计的平衡能力明显提高,

其基本原理是根据已经纳入的受试者协变量的组间分布情况, 将新进入的受试者以较大概率分配到使得组间协变量分布差异最小的组别中.

平衡离散型协变量的随机化设计及理论性质在文献中已经进行了广泛的研究 (Hu and Hu (2012), Ma, Hu and Zhang (2015)). 然而, 随机对照试验中往往还包括许多连续型协变量, 而对连续型协变量的离散化处理会改变协变量的性质并导致信息损失. 因此, 直接平衡连续型协变量的随机化设计备受研究者关注. 在已有研究中, 诸多学者定义了不同的指标对协变量的不平衡性进行度量, 包括基于经验累积分布 (Lin and Su (2012))、核密度函数 (Ma and Hu (2013)) 和马氏距离 (Morgan and Rubin (2012), Li and Ding (2020), Qin et al. (2022)) 等. 已有研究结果表明, 基于马氏距离的随机化设计具备良好的理论性质 (Morgan and Rubin (2012), Qin et al. (2022)), 当被平衡的协变量与结局变量相关时, 基于马氏距离的随机化设计可以获得平均处理效应的无偏估计并提高估计精度. 结合本文实际数据中主要涉及到连续型的协变量, 因此本文将重点对两种基于马氏距离的随机化设计进行比较研究.

首先, 衡量协变量在试验组与对照组间分布差异的马氏距离定义为:

$$M \equiv (\bar{\mathbf{X}}_T - \bar{\mathbf{X}}_C)' [\text{cov}(\bar{\mathbf{X}}_T - \bar{\mathbf{X}}_C)]^{-1} (\bar{\mathbf{X}}_T - \bar{\mathbf{X}}_C) \\ = np_w(1 - p_w) (\bar{\mathbf{X}}_T - \bar{\mathbf{X}}_C)' \text{cov}(\mathbf{x})^{-1} (\bar{\mathbf{X}}_T - \bar{\mathbf{X}}_C), \quad (1)$$

其中,  $\text{cov}(\mathbf{x})$  代表协变量的协方差矩阵,  $\bar{\mathbf{X}}_T - \bar{\mathbf{X}}_C$  为试验组与对照组协变量均值差的向量,  $n, p_w$  为随机化的样本数和试验组样本数占总样本数的比例.

Morgan and Rubin (2012) 以马氏距离作为不平衡性度量提出了考虑协变量平衡的重随机化 (rerandomization, RR). RR 通过对所有受试者不断地进行完全随机化 (CR) 分配并计算每次分配下的协变量平衡程度, 直至获得满足预先设定的协变量平衡阈值  $\alpha$  的分配方案. RR 的基本流程如图 1 所示.

鉴于 RR 需基于 CR 在事先设定的平衡性阈值下对所有受试者不断地进行重复分配, 其要求受试者信息全部已知并且计算成本很大, 故研究者希望探究更为高效的随机化设计. Qin et al. (2022) 提出了基于马氏距离的自适应随机化设计 (adaptive randomization via Mahalanobis distance, ARM), ARM 遵循自适应设计的基本思路, 将分配概率设置为受试者协变量信息的函数, 不平衡性度量马氏距离根据已被分配和待被分配的受试者协变量信息计算, 随着受试者不断地被分配, 马氏距离相应随协变量信息变化自适应更新并根据马氏距离自适应地调整分配概率.

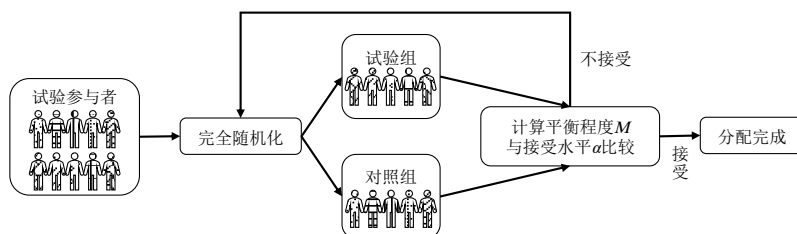


图 1 重随机化流程示意图

Qin et al. (2022) 证明了 ARM 对于协变量的平衡具有良好的收敛性质和处理效应估计最优性质. 收敛性指马氏距离以  $1/n$  的速率收敛到零, 即受试者越多 ARM 方法的平衡效果越好; 处理效应估计最优性指估计平均处理效应时, 使用 ARM 方法后的均值差估计量精度与回归调整估计量的精度相同, 即使用 ARM 方法随机化后可以不需要使用事后调整方法校正协变量不平衡. 且相较于 RR 方法, ARM 方法可以显著地降低计算成本, 并且其优势在协变量或样本数较多时更为明显. 此外, 从逻辑上讲, ARM 可以顺应在经济试验中受试者序贯进入并需要被序贯分配的实际情况. ARM<sup>1</sup> 的基本流程如下, 示意图如图 2 所示.

1) 记第  $i$  个受试者的协变量信息为  $\mathbf{x}_i$ ,  $i = 1, 2, \dots, n$ . 记第  $i$  个受试者的分配为  $T_i = 1$  (试验组) 或  $T_i = 0$  (对照组).

2) 将前两个进入试验的受试者分配分配到试验组  $T_1 = 1$  和对照组  $T_2 = 0$ .

3) 假设有  $2i$  个受试者已经被分配, 那么对于第  $(2i + 1)$  和第  $(2i + 2)$  个受试者, 有以下两种分配方式:

A 分配: 如果第  $(2i + 1)$  个受试者被分配试验组, 第  $(2i + 2)$  个受试者被分配到对照组, 计算试验组和对照组的“潜在”马氏距离  $M_A(2i + 2)$ ;

B 分配: 如果第  $(2i + 1)$  个受试者被分配对照组, 第  $(2i + 2)$  个受试者被分配到试验组, 计算试验组和对照组的“潜在”马氏距离  $M_B(2i + 2)$ .

4) 选择 A 分配, 即第  $(2i + 1)$  个受试者被分配到试验组,  $(2i + 2)$  个受试者被分配到对照组的概率为:

$$P(T_{2i+1} = 1, T_{2i+2} = 0 \mid \mathbf{x}_{2i}, \dots, \mathbf{x}_1, T_{2i}, \dots, T_1) = \begin{cases} q & \text{if } M_A(2i + 2) < M_B(2i + 2), \\ 1 - q & \text{if } M_A(2i + 2) > M_B(2i + 2), \\ 0.5 & \text{if } M_A(2i + 2) = M_B(2i + 2). \end{cases}$$

$0.5 < q < 1$ , Qin et al. (2022) 设置  $q = 0.75$ .

5) 重复进行步骤 3) 和 4) 直到所有受试者被分配完成. 如果  $n$  为奇数, 最后一个受试者被随机分配到试验组或对照组.

## 4 实证分析

由上述描述统计 (表 1) 可知, 养老金随机对照试验中的原始随机化方案并不能使试验组和对照组的协变量分布达到平衡. 因此, 本文考虑采用考虑协变量平衡的随机化设计对该试验中受试者进行重新分配, 并从对于平均处理效应估计的精度和检验功效两个方面对协变量平衡调整随机化在经济试验中的必要性进行说明. 具体而言, 本文使用完全随机化 (CR)、重随机化 (RR) 和基于马氏距离的自适应随机化 (ARM) 对实例中的受试者重新地进行 3000 次随机化分配. 基于所获得的分配方案, 本文首先比较各分配方案在协变量上的平衡性; 然后根据原始分配方案和受试者协变量数据拟合回归模型, 将重新生成的分配方案代入模型得到响应变量值, 利用不同的估计方式获得对于平均处理效应的估计并比较其估计精度和检验功效. 实证分析流程见图 3.

<sup>1</sup>本方法可使用 R 软件包 CARM 实现, 地址: <https://cran.r-project.org/web/packages/CARM>.

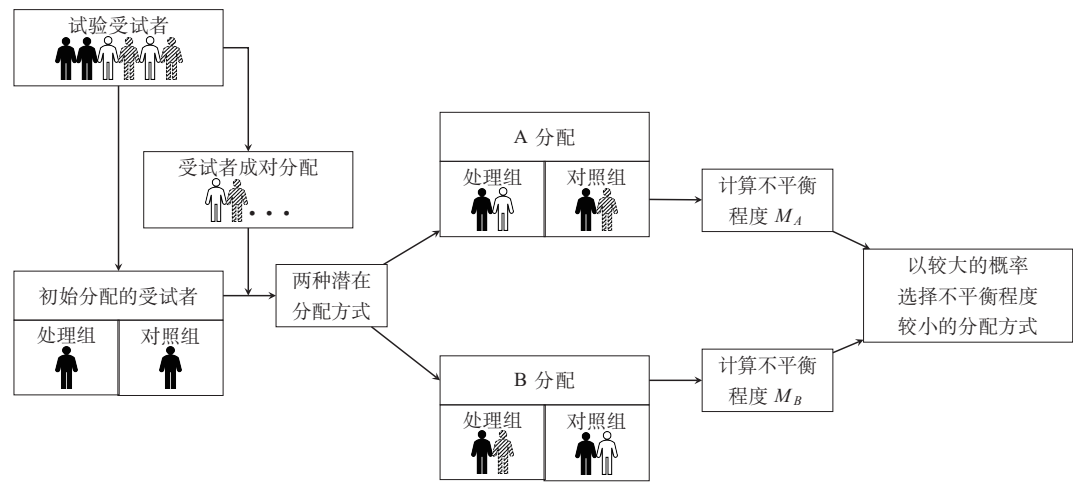


图 2 ARM 流程示意图

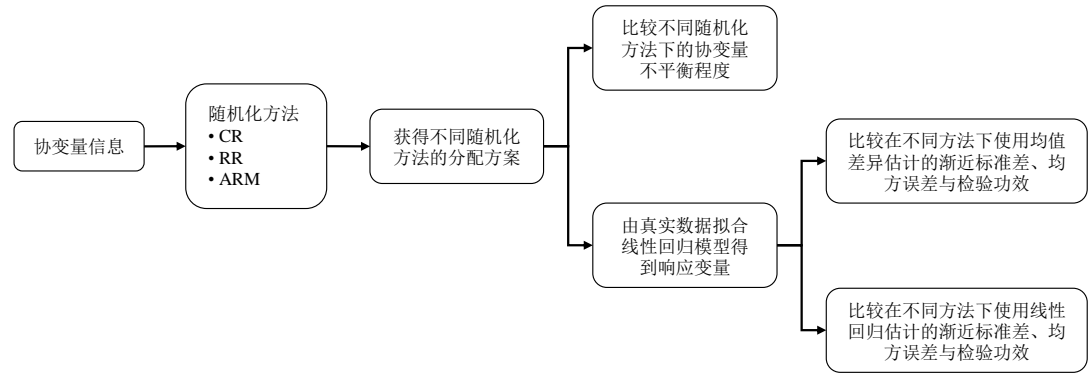


图 3 分析流程示意图

4.1 不同随机化设计在协变量平衡性方面的比较

在本文的养老金随机对照试验中，受试者按照其身份证尾号单双号被分配到试验组或对照组。在原始试验的分配方案下计算出的马氏距离为 90.988，说明该随机对照试验中存在很强的协变量不平衡。

图 4 展示了由不同随机化设计得到的分配方案所对应的马氏距离分布图。具体地，本文分别使用 CR、RR<sup>2</sup>和 ARM 方法将这些受试者分配到试验组或对照组，在多次模拟<sup>3</sup>下获得分配方案并计算相应的马氏距离。如图 4 所示，在 3 种随机化方法中基于 RR 和 ARM 的分配方案的马氏距离远小于采用 CR 的分配方案的马氏距离，且 ARM 在协变量平衡上的表现最优。

表 2 展示了在 3000 次随机化后的马氏距离均值，并展现了对年龄协变量进行差异性

<sup>2</sup> $P(M < \alpha) = 0.05$ .

<sup>3</sup>每种方法重复 3000 次.

检验的结果. 如表 2 所示, 采用协变量调整的随机化方法可以明显地减小马氏距离, 降低试验组和对照组之间协变量的不平衡. 同时, 本文对年龄协变量应用了  $t$  检验和 Kolmogorov-Smirnov 检验判断试验组和对照组间的协变量均值以及分布是否相同, 在表 2 中展示了随机化后组间仍存在显著性差异 ( $P$  值  $< 0.05$ ) 的比例. 差异性检验显示, 在完全随机化后, 在年龄协变量上仍然可能发生协变量不平衡; 经考虑协变量平衡的随机化分配后, 年龄变量在试验组和对照组间的差异均不显著, 这表明考虑协变量平衡的随机化方法能够实现对于协变量更好的平衡.

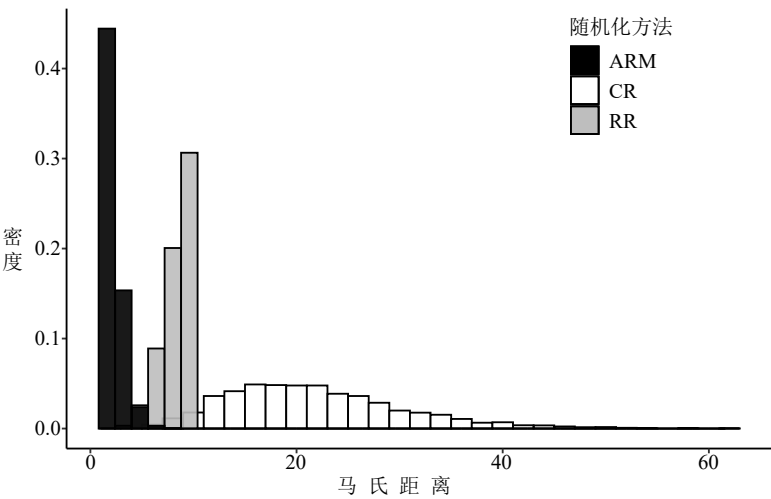


图 4 不同随机化方式对马氏距离的影响

表 2 不同随机化方法下的马氏距离均值和差异性检验结果 (以年龄为例)

随机化方法	马氏距离均值	差异显著性检验	
		$t$ 检验	K-S 检验
CR	21.84	3.63%	5.08%
RR	7.97	0.00%	0.00%
ARM	1.82	0.00%	0.00%

4.2 不同随机化设计在效应估计精度和检验功效上的比较

本文通过两种方式对该随机对照试验中干预作用的平均直接因果效应 (以下简称为平均处理效应 (average treatment effect, ATE)) 进行估计.

在潜在结果框架 (Rubin (1974)) 下, 处理效应被定义为受试者接受干预或接受对照后的两个“潜在结果”的差异, 然而受试者不能同时接受干预和接受对照, 研究者仅能观测到受试者在处理组或对照组的结局, 所以个体的处理效应是无法观测的, 一般需要通过平均处理效应等估计量对处理效应进行估计. 假设受试者接受处理  $T = 1$  的平均结果为  $E(Y_1)$ , 受试者接受对照  $T = 0$  的平均结果为  $E(Y_0)$ , 可通过  $ATE = E(Y_1) - E(Y_0)$  来表示平均处理效应.

那么估计平均处理效应最直接的方式是使用均值差异估计量  $\hat{\tau}$ , 此时需要根据受试者  $i$



接受处理  $T_i$  下得到的响应变量  $y_i$  来估计平均处理效应,  $i = 1, \cdots, n$ :

$$\hat{\tau} = \frac{\sum_{i=1}^n T_i y_i}{\sum_{i=1}^n T_i} - \frac{\sum_{i=1}^n (1 - T_i) y_i}{\sum_{i=1}^n (1 - T_i)}.$$

(2)

然而,  $\hat{\tau}$  对协变量的平衡性十分敏感 (Qin et al. (2022)). 当随机对照试验中存在协变量不平衡时, 例如在本文所分析的随机化试验中, 年龄这个变量在试验组与对照组间的分布存在显著差异. 那么, 在此情形下使用  $\hat{\tau}$  对处理效应的估计会因无法剔除协变量本身的差异而存在偏差.

通常, 研究者可通过回归对试验数据进行事后调整获得对于处理效应的无偏估计. 当假设响应变量与协变量服从线性模型时, 对于处理效应的无偏估计可基于模型的最小二乘回归获得. 即, 假设线性回归模型形式为:

$$y_i = \alpha + \tau T_i + \mathbf{X}_i^T \boldsymbol{\beta} + \epsilon_i,$$

(3)

此处, 记响应变量为  $\mathbf{Y} = (y_1, \cdots, y_n)^T$ , 受试者的协变量为  $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)^T = [x_{ij}]_{n \times p}$ , 受试者的分配方案为  $\mathbf{T} = (T_1, \cdots, T_n)^T$ ,  $\epsilon_i$  是服从正态分布的随机误差项. 那么, 记  $\tilde{\mathbf{X}} = [\mathbf{1}; \mathbf{T}; \mathbf{X}]$ , 该模型的参数为  $\boldsymbol{\beta}^* = (\alpha, \tau, \boldsymbol{\beta}^T)^T$ , 其中  $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^T$ . 此时  $\boldsymbol{\beta}^*$  的最小二乘估计的表达式为  $\hat{\boldsymbol{\beta}}^* = (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \mathbf{Y}$ , 其中对于  $\tau$  的 OLS 估计量记为  $\hat{\tau}_{\text{adj}}$ .

结合实际数据, 该模型设定结局变量  $y_i$  为接受到信息后一年的月平均强制缴纳养老金金额,  $\mathbf{X}_i^T$  为受试者的协变量信息,  $\boldsymbol{\beta}$  为各个协变量的回归系数. 表 3 给出了简单均值差估计量  $\hat{\tau}$  和线性回归估计量  $\hat{\tau}_{\text{adj}}$  对于平均处理效应的估计结果. 此外, 本文使用逐步回归选择对响应变量有重要影响的四个协变量, 表 3 给出了通过线性回归对这些协变量系数进行估计的结果.

如表 3, 由简单均值差估计量得到的估计  $\hat{\tau} = -0.367$ , 未通过显著性检验; 由线性回归估计量得到的估计  $\hat{\tau}_{\text{adj}} = -2.129$ , 其系数显著. 由于简单均值差估计量可被视为公式 (2) 中不包含任何协变量, 仅纳入干预变量的回归情形, 因此本文考虑进一步对两个回归模型的拟合优度进行比较. 分析结果显示, 仅纳入干预变量的一元线性回归模型的拟合优度为 0, 而纳入重要协变量的多元线性回归模型的拟合优度为 0.794. 该结果说明, 响应变量与协变量间的关系可以部分地被多元线性回归模型刻画; 相应地, 相较于通过简单均值差估计量获得的对于平均处理效应的估计, 研究者借助对于协变量分布不平衡的事后回归调整能够获得更准确的估计. 然而, 由于上述事后调整的回归估计十分依赖正确的模型, 而多元线性回归的  $R^2 = 0.794$  意味着该数据中仍然存在无法完全被该线性模型解释的成分. 已有诸多研究表明

表 3 平均处理效应估计与协变量调整

	均值差异估计	协变量校正的回归估计				
	处理效应	处理效应	TotCot	edad	rem_actual	saldo_uf
估计	-0.367	-2.129	0.066	-0.105	0.017	0.002
P 值	0.842	0.012	0.000	0.009	0.000	0.019
$R^2$	0.000	0.794				

(Hu, Hu and Ma et al. (2014), Imbens and Rubin (2015), Li and Ding (2020)), 在试验阶段采用考虑协变量平衡调整的随机化设计对于平均处理效应的估计而言是更优的选择.

本节从处理效应估计精度和检验功效两方面对不同随机化设计的表现进行比较. 在基于实际数据的模拟试验中, 本文首先基于原始数据拟合线性回归模型得到参数估计, 然后根据经 CR、RR 和 ARM 方法获得的分配方案得到基于模拟试验的响应变量  $y_i^{\text{sim}}$ .

$$y_i^{\text{sim}} = 7.479 - 2.129 \times T_i^{\text{sim}} + 0.066 \times \text{TotCot}_i - 0.105 \times \text{edad}_i + 0.017 \times \text{rem\_actual}_i + 0.002 \times \text{saldo\_uf}_i + \epsilon_i^{\text{sim}}, \quad (4)$$

其中,  $T_i^{\text{sim}}$  是通过不同的随机化方法得到的分配方式,  $\text{TotCot}_i$ ,  $\text{edad}_i$ ,  $\text{rem\_actual}_i$ ,  $\text{saldo\_uf}_i$  分别代表受试者  $i$  的协变量信息,  $\epsilon_i^{\text{sim}}$  是对拟合残差的重复抽样.

随后, 本文考虑分别通过模型 1 和 2 获得对于平均处理效应的简单均值差估计和线性回归估计.

模型 1:  $y_i = \alpha + T_i\tau + \epsilon_i$ ;

模型 2:  $y_i = \alpha + T_i\tau + \beta_1\text{TotCot}_i + \beta_2\text{edad}_i + \beta_3\text{rem\_actual}_i + \beta_4\text{saldo\_uf}_i + \epsilon_i$ .

重复 3000 次模拟试验, 表 4 给出了在不同随机化方法下, 不同估计量对于平均处理效应估计的渐近标准差和均方误差 (mean square error, MSE). 结果表明, 当使用线性回归估计量  $\hat{\tau}_{\text{adj}}$  时, 由于几乎所有的协变量不平衡都能够得到回归调整, 在同一随机化方法下的  $\hat{\tau}_{\text{adj}}$  的渐近标准差和 MSE 始终小于  $\hat{\tau}$  的渐近标准差和 MSE. 当使用简单均值差估计量  $\hat{\tau}$  时, 基于 CR 得到的估计的表现较差; 基于 RR 和 ARM 得到的估计的表现明显优于 CR, 基于 ARM 得到的估计的表现最优, 并且其表现十分接近回归估计量  $\hat{\tau}_{\text{adj}}$ . 此外, 在 CR、RR 和 ARM 方法下的均值差异估计量  $\hat{\tau}$  的 95% 置信区间分别为  $(-5.967, 1.604)$ ,  $(-4.633, 0.205)$ ,  $(-3.792, -0.396)$ . 由于基于 CR 和 RR 获得的置信区间均包含 0, 说明研究者并不能由基于 CR 和 RR 得出的均值差估计量  $\hat{\tau}$  获得对于平均处理效应的有效估计. 综上, 本节的分析表明, ARM 方法在对于协变量的平衡上有着显著的优势, 且反映出在随机化阶段平衡协变量可以提高对于平均处理效应的估计精度. 当试验组和对照组间的协变量平衡程度较高时, 在估计中研究者便不再依赖于通过事后调整的方法来估计平均处理效应.

表 4 还展示了在不纳入协变量的回归模型 (no cov) 和纳入重要协变量的回归模型 (with cov) 后对平均处理效应估计进行假设检验的结果. 本文对于检验功效计算参照 Ma et al. (2020). 结果表明, 在纳入重要协变量的回归模型下, 基于不同随机化方法的检验功效相似. 然而, 当回归模型不包括重要协变量时, 未考虑协变量平衡的随机化方法的检验功效会明显降低, 检验功效的降低幅度与随机化方法所能达到的协变量平衡程度相关. 将基于不同随机

表 4 平均处理效应的渐近标准差、MSE 与检验功效

随机化	$\text{sd}(\hat{\tau})$	$\text{sd}(\hat{\tau}_{\text{adj}})$	$\text{MSE}(\hat{\tau})$	$\text{MSE}(\hat{\tau}_{\text{adj}})$	power(with cov)	power(no cov)
CR	1.810	0.877	3.283	0.799	0.708	0.217
RR	1.301	0.875	1.734	0.772	0.701	0.324
ARM	0.861	0.828	0.758	0.691	0.720	0.681

化方法的估计精度和检验功效比较来看,分析结果表明,在随机化分配阶段对协变量的平衡性进行调整不仅能够提高后续对于因果效应的估计精度,还能使得对于效应的检验具有较高的功效.因此本文认为,考虑协变量平衡的随机化设计是研究者进行随机化试验的优良选择.

## 5 结论

基于现有的考虑协变量平衡调整的随机化方法,本文将考虑协变量调整的随机化设计引入至经济学领域随机对照试验的实证分析.本文通过对于该实证数据的模拟分析,从对于协变量的平衡程度,因果效应的估计精度,以及假设检验的功效三个方面对不同随机化设计的表现进行比较.分析结果表明,基于协变量自适应随机化设计能够获得使得协变量在组间分布更加均衡的分配方案,进而基于该随机化设计的因果效应估计以及假设检验功效具有优良表现.

本文的研究未来还可以在以下几个方面进行扩展.第一,分层置换区组设计和最小化方法在临床试验中被广泛使用.然而,这两类方法只能对离散型协变量进行平衡,因此探索能够同时平衡离散型和连续型协变量的自适应设计十分必要.第二,ARM方法在线性回归框架下的假设检验和统计推断有效性已被证明(Ma, Hu and Zhang (2015), Ma et al. (2020)),在实际问题中响应变量往往具有多种类型,在广义线性模型和比例风险模型等框架下,研究者仍需考虑如何在使用协变量自适应随机化设计后更合理地完成参数估计以及假设检验,针对这类理论问题的需进行更全面的研究.第三,在大数据时代背景下的随机对照试验中,受试者的基线协变量往往呈现高维特征.因此,为提高随机化效率,如何在进行自适应随机化的同时完成对于重要变量的识别值得进一步的探索与讨论.

## 参 考 文 献

- 刘雅琦, 刘曼, 刘玉秀, 陈文松, 袁阳丹, (2021).《新英格兰医学杂志》临床试验随机化方法应用状况评价 [J]. 中国循证医学杂志, 21(3): 303–307.
- Liu Y Q, Liu M, Liu Y X, Chen W S, Yuan Y D, (2021). Analysis of the Application of Randomization Methods in Clinical Trials in NEJM[J]. Chinese Journal of Evidence-based Medicine, 21(3): 303–307.
- 殷赫然, 刘健, 樊小农, 沈燕, 王舒, (2021). 最小随机化在国内针灸临床研究中的应用与思考 [J]. 中华中医药杂志, 36(1): 76–79.
- Yin H R, Liu J, Fan X N, Shen Y, Wang S, (2021). Application and Consideration of Minimization in Clinical Research of Acupuncture and Moxibustion in China[J]. China Journal of Traditional Chinese Medicine and Pharmacy, 36(1): 76–79.
- Bruhn M, McKenzie D, (2009). In pursuit of Balance: Randomization in Practice in Development Field Experiments[J]. American Economic Journal: Applied Economics, 1(4): 200–232.
- Egbewale B E, Lewis M, Sim J, (2014). Bias, Precision and Statistical Power of Analysis of Covariance in the Analysis of Randomized Trials with Baseline Imbalance: A Simulation Study[J]. BMC Medical Research Methodology, 14(1): 49.
- Fisher R A, (1926). The Arrangement of Field Experiments[J]. Journal of the Ministry of Agriculture of Great Britain, 33: 503–513.
- Fuentes O, Lafortune J, Riutort J, Tessada J, Villatoro F, (2017). Personalized Information as a Tool

- to Improve Pension Savings: Results from a Randomized Control Trial in Chile[J]. Documento de Trabajo IE-PUC, 483: 1–50.
- Hu F, Hu Y, Ma Z, Rosenberger W F, (2014). Adaptive Randomization for Balancing Over Covariates[J]. Wiley Interdisciplinary Reviews: Computational Statistics, 6(4): 288–303.
- Hu Y, Hu F, (2012). Asymptotic Properties of Covariate-adaptive Randomization[J]. The Annals of Statistics, 40(3): 1794–1815.
- Imbens G W, Rubin D B, (2015). Causal Inference for Statistics, Social, and Biomedical Sciences[M]. Cambridge: Cambridge University Press.
- Li X, Ding P, (2020). Rerandomization and Regression Adjustment[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 82(1): 241–268.
- Lin Y, Su Z, (2012). Balancing Continuous and Categorical Baseline Covariates in Sequential Clinical Trials Using the Area between Empirical Cumulative Distribution Functions[J]. Statistics in Medicine, 31(18): 1961–1971.
- Mastrobuoni G, (2011). The Role of Information for Retirement Behavior: Evidence Based on the Stepwise Introduction of the Social Security Statement[J]. Journal of Public Economics, 95(7–8): 913–925.
- Matts J P, Lachin J M, (1988). Properties of Permuted-block Randomization in Clinical Trials[J]. Control Clin Trials, 9(4): 327–344.
- Ma W, Hu F, Zhang L, (2015). Testing Hypotheses of Covariate-adaptive Randomized Clinical Trials[J]. Journal of the American Statistical Association, 110(510): 669–680.
- Ma W, Qin Y, Li Y, Hu F, (2020). Statistical Inference for Covariate-adaptive Randomization Procedures[J]. Journal of the American Statistical Association, 115(531): 1488–1497.
- McIntosh C, Zeitlin A, (2022). Using Household Grants to Benchmark the Cost Effectiveness of a USAID Workforce Readiness Program[J]. Journal of Development Economics, 157: 102875.
- Morgan K L, Rubin D B, (2012). Rerandomization to Improve Covariate Balance in Experiments[J]. The Annals of Statistics, 40(2): 1263–1282.
- Peter G S, Richard H M, David A R, (2015). Field Trials of Health Interventions: A Toolbox[M]. Oxford: Oxford University Press.
- Pignata S, Lorusso D, Joly F, Gallo C, Colombo N, et al. (2021). Carboplatin-based Doublet Plus Bevacizumab Beyond Progression Versus Carboplatin — Based Doublet Alone in Patients with Platinum-sensitive Ovarian Cancer: A Randomised, Phase 3 Trial[J]. The Lancet Oncology, 22(2): 267–276.
- Pinto J M, (2012). Does Personalized Pension Projection Affect the Retirement Decision?[R]. Superintendencia de Pensiones.
- Pischke A S, (2010). The Credibility Revolution in Empirical Economics: How better Research Design is Taking the Con Out of Econometrics[J]. Journal of Economic Perspectives, 24(2): 3–30.
- Qin Y, Li Y, Ma W, Yang H, Hu F, (2022). Adaptive Randomization via Mahalanobis Distance[J/OL]. Statistica Sinica. Doi: 10.5705/ss.202020.0440.
- Rubin D B, (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies[J]. Journal of Educational Psychology, 66(5): 688–701.
- Schlosser A, Shanan Y, (2022). Fostering Soft Skills in Active Labor Market Programs: Evidence from a Large-Scale RCT[R]. CESifo Working Paper Series.
- Svorenčík A, (2021). The Driving Forces Behind the Rise of Experimental Economics[J]. Review of Political Economy, 33(2): 344–361.