

社会研究方法评论4期

图书在版编目(CIP)数据

社会研究方法评论·第4卷/赵联飞,赵锋主编.--
重庆:重庆大学出版社,2023.12
(万卷方法)

ISBN 978-7-5689-4322-2

I. ①社… II. ①赵… ②赵… III. ①社会学—研究
方法 IV. ①C91-03

中国国家版本馆 CIP 数据核字(2023)第 253696 号

社会研究方法评论 第4卷

SHEHUI YANJIU FANGFA PINGLUN DISIJUAN

主 编 赵联飞

策划编辑:林佳木

责任编辑:石 可 版式设计:林佳木

责任校对:刘志刚 责任印制:张 策

*

重庆大学出版社出版发行

出版人:陈晓阳

社址:重庆市沙坪坝区大学城西路21号

邮编:401331

电话:(023)88617190 88617185(中小学)

传真:(023)88617186 88617166

网址:<http://www.cqup.com.cn>

邮箱:fxk@cqup.com.cn(营销中心)

全国新华书店经销

重庆愚人科技有限公司印刷

*

开本:890mm×1240mm 1/32 印张:10 字数:243千

2023年12月第1版 2023年12月第1次印刷

ISBN 978-7-5689-4322-2 定价:42.00元

本书如有印刷、装订等质量问题,本社负责调换
版权所有,请勿擅自翻印和用本书
制作各类出版物及配套用书,违者必究

《社会研究方法评论》编委会名单（2021—2023）

（按姓氏笔画排序）

- | | |
|-----|------------------|
| 王毅杰 | 河海大学 |
| 毕向阳 | 中央民族大学 |
| 吕 鹏 | 中南大学 |
| 刘 伟 | 四川省社会科学院 |
| 刘 军 | 哈尔滨工程大学 |
| 刘成斌 | 华中科技大学 |
| 孙秀林 | 清华大学 |
| 巫锡炜 | 中国人民大学 |
| 李 煜 | 复旦大学 |
| 吴肃然 | 哈尔滨工程大学 |
| 吴愈晓 | 南京大学 |
| 邹宇春 | 中国社会科学院社会学研究所 |
| 陈华珊 | 中国社会科学院社会发展战略研究院 |
| 陈福平 | 厦门大学 |
| 周 皓 | 北京大学 |
| 赵 锋 | 中国社会科学院社会学研究所 |
| 赵联飞 | 中国社会科学院社会学研究所 |
| 洪岩璧 | 东南大学 |
| 夏传玲 | 中国社会科学院社会学研究所 |
| 高 勇 | 中国政法大学 |
| 梁玉成 | 中山大学 |
| 董海军 | 中南大学 |
| 焦开山 | 中央民族大学 |

栏 目

2022 年 2 月创刊

总第 4 卷

特 稿

- 在方法与方法论之间：围绕“田野调查实践”课程的对谈
.....周飞舟 许方毅 薛雯静 (xfy@stu.pku.edu.cn) 1

海外民族志专题

- 栏目导语龚浩群 (gonghq@aliyun.com) 36
古典学：海外民族志之羽翼杨春宇 (cyyang@cass.org.cn) 64
地方知识的生产——美国中镇田野调查过程回顾
.....梁文静 (wenjingliang07@qq.com) 100
面向发达社会的海外民族志之知识生产——基于法国田野的再思考
.....张金岭 (zhangjl@cass.org.cn) 105
走向深处：俄罗斯田野工作的经验谈
.....马 强 (maqiang@cass.org.cn) 105

论 文

- 样本结构性偏差与因果推论——基于实验数据的分析
.....周 皓 (zhouh@pku.edu.cn) 36
- 从个案到因果：过程-事件与过程-追踪——兼论以因果性为统领的
混合研究.....常富军 (jeffery_chang988@163.com) 64
- 当代中国社会学语境中对“范式”的三种理解——兼谈方法论的对话
为何难有效果？.....吴肃然 周 凡 (wusuran@hrbeu.edu.cn) 64
- 我和我的“男朋友们”：主体间性视角下的亚文化研究方法思考
.....王丽圆 (wly1213534745@163.com) 64
- 论汪晖研究在方法论上的启示：国外学术界的反响和评价
.....王晓平 (wxping75@163.com) 64

新知与书评

- 演化社会科学：历史、现状和展望...洪 泽 (zehong@um.edu.mo) 64

征稿启事

- 《社会研究方法评论》征稿启事..... 64

目 录

在方法与方法论之间：围绕“田野调查实践”课程的对谈周飞舟 许方毅 薛雯静	1
一、教学：能力而非方法	2
二、阅读：“比着做”的范本	8
三、选题：回到“内在动力”	11
四、访谈：“相人偶”的情境	16
五、汇报：“活”的材料	22
栏目导语	龚浩群 31
古典学：海外民族志之羽翼	杨春宇 32
一、传略法与巴西尔	33
二、澳大利亚的文化底色	37
三、古典学与宗教领域	42
四、在其他领域应用古典学	46
五、引用事件：典范与现实的张力	50
六、余论：以文明视角超越个案	54
地方知识的生产——美国中镇田野调查过程回顾	梁文静 59
一、选点	61
二、参与观察	62
（一）在中镇的参与观察经历	63
（二）对田野作业的反思	68
三、搜集地方资料	73

(一)做访谈	73
(二)搜集历史资料	74
(三)与当地学者互动	75
四、文化比较与翻译	76
五、总结	77
面向发达社会的海外民族志之知识生产——基于法国田野的再思考张金岭	81
一、中国社会科学的知识生产需要关注一个经验的西方世界	84
二、探究“成品社会”及其知识体系的结构化过程	87
三、发掘西方发达社会内在的多样性	89
四、理解复杂实践背后的制度脉络	92
五、剖析诸多细节背后的制度理性	95
六、发现理念与实践之间的偏离	98
走向深处:俄罗斯田野工作的经验谈	马强 107
一、走向田野的深处	109
二、走向历史的深处	115
三、走向心灵的深处	120
四、结语	123
样本结构性偏差与因果推论——基于实验数据的分析	周皓 125
一、样本结构偏离:定义与来源	129
(一)样本结构偏离的定义	129
(二)样本结构偏离的来源	130
(三)因果推断视角下的样本结构偏离	132
二、理论分析	133
(一)样本选择性偏差	133

(二)结构偏离与平均实验效应偏差	136
三、分析思路与检验方法	139
四、线性回归中的结构性偏差	142
(一)实验数据的构建	142
(二)简单线性回归的结果	144
(三)多元线性回归的结果	145
五、结构性偏差对因果推断结果的影响	148
(一)基础数据构建与分析方法	149
(二)因变量结构变动	150
(三)连续自变量结构性变化	156
(四)分类自变量结构性变化	166
(五)简单小结	172
六、对模拟结果的统计分析	172
(一)系数估计量	173
(二)估计量与标准取值和基准模型的差异	180
(三)平均实验效应的估计量及差异性	185
七、结论与讨论	189
从个案到因果:过程-事件与过程-追踪——兼论以因果性为统领的 混合研究	常富军 193
一、个案与个案研究	195
二、过程性个案研究:过程-追踪研究与过程-事件分析	197
(一)过程-追踪研究	197
(二)过程-事件分析	201
(三)过程-追踪研究与过程-事件分析的比较	207
三、以因果性为统领的混合研究	208

(一)因果与因果关系	208
(二)因果效应与因果机制	209
(三)以因果性为统领的混合研究	211
四、结论与讨论	214
当代中国社会学语境中对“范式”的三种理解——兼谈方法论的对话 为何难有效果?	吴肃然 周凡 220
一、《科学革命的结构》与“范式”	222
二、当代中国社会学家的三种范式观	226
(一)强调“整合”的范式观	226
(二)强调“规范”的范式观	228
(三)强调“演进”的范式观	229
三、三种范式观背后的学科认识与知识关怀	231
四、讨论	235
我和我的“男朋友们”:主体间性视角下的亚文化研究方法思考	王丽圆 240
一、背景与问题	242
二、亚文化研究伦理问题与方法论视角	244
(一)亚文化研究的方法伦理问题	244
(二)主体间性的方法论视角	245
三、田野场中的主体间性思考	248
(一)共享一套“话语”的亚文化世界:腐女身份入场	248
(二)谁是“他者”? ——“污名”的共鸣	253
(三)受访者日常生活的诠释与“共生文化”的建构	257
四、结论与讨论	263

论汪晖研究在方法论上的启示:国外学术界的反响和评价	王晓平 269
一、问题意识与历史化的结合	271
二、辩证法与总体性的统一	275
三、政治性与能动性的一致	280
四、结语	287
演化社会科学:历史、现状和展望	洪泽 290
一、演化理论在社会科学中应用的简史	291
二、演化社会科学的当代学术分支	298
(一)人类行为生态学	298
(二)演化心理学	299
(三)基因-文化共同演变(或文化演变)	301
(四)演化经济学	304
三、一个演化社会科学分析的具体案例:人口转型	306
(一)人类行为生态学:数量-质量权衡	307
(二)演化心理学:现代环境与心理机制的错配	308
(三)文化演变:文化信息的传播导致生育观念转变	309
(四)与传统社会科学的对比	311
四、演化社会科学——现代社会科学中的聚合力?	313
五、结语	316
《社会研究方法评论》征稿启事	324

周皓. 样本结构性偏差与因果推论——基于实验数据的分析[M]//赵联飞,赵锋. 社会研究方法评论:第4卷. 重庆:重庆大学出版社.

样本结构性偏差与因果推论——基于实验数据的分析

周皓^①

摘要: 本文利用仿真实验数据, 讨论了样本结构偏离导致对线性回归和因果推断下的平均实验效应估计量的影响。文章认为, 1) 样本结构偏离是集合性的联合分布偏离, 而非只是单变量分布的偏离; 2) 非随机的样本结构性偏差主要表现在实验效应估计量、自身系数估计量、以及与之相关(甚至无关)的变量系数估计量三个方面; 3) 因果推论视角下, 样本结构偏离会导致实验效应估计量有偏; 4) 分类变量的结构性偏差会随着与真实结构的距离增大而呈非线性加快扩大趋势, 特别是对实验效应的估计; 5) 统计方法无法从根本上解决样本结构性偏差。最后, 文章强调必须重视并从根源上解决样本结构性偏差, 以期更贴现实地从总体角度描绘社会现实, 解释社会现象, 回答社会问题。这应该是社会科学定量研究的目标。

^①作者简介: 周皓, 北京大学中国社会与发展研究中心研究员、社会学系教授。联系方式: zhouh@pku.edu.cn。

关键词：样本结构；偏差；实验效应；仿真实验

Abstract: Simulation experimental data is used to discuss the structural bias caused by the sample structure deviation and its impact on the causal inference. The analysis results indicate that 1) the structural deviation of the sample is a collective joint distribution deviation. 2) The structural bias of non-random samples mainly manifests in three estimations: the coefficients of the variable and those which are related (or even unrelated) to them, and experimental effects. 3) From the perspective of causal inference, sample structural deviation will lead to biased estimation of experimental effects. 4) The structural bias of categorical variables will increase nonlinearly with the increase of the deviation from the real structure. 5) Statistical methods cannot fundamentally address sample structural biases. Finally, the article emphasizes on paying attention to sample structural bias, data sources and sample structure, balancing the sample structure and statistical methods, and using appropriate methods to analyze (relatively) accurate data, in order to more realistically depict social reality, explain social phenomena, and answer social questions, from a population perspective.

Key words: Sample Structure; Bias; Treatment Effect; Simulation Experiment

我们能够认识什么？这是知识论中最为核心的问题。从柏拉图的可感世界和可知世界，到康德的纯粹理性，到马克思的认识与实践的统一，都是解答这一问题的尝试。大卫·休谟的问题似乎打破了经验主义一劳永逸的归纳法，把实然与应然之间的天堑毫无保留地展现了出来。

但这并不意味着经验归纳毫无意义。事实上，长期以来在统计学中实际应用的因果推断，也都建立在“经验—归纳—演绎”这个简洁的经验论逻辑关系上，统计经验的实在性有效但极其脆弱，且局限于因果律的一隅。这种脆弱性体现在经验素材的真实性和完备性中。当归纳过程中出现真实性与完备性不足的问题时，经验所得到的“知识”便会离题万里。由此得到的经验主义的因果关系无法真正地对应现实世界中的统计性因果，反过来还会导致以经验为基础的认知过程出现问题。即便采用相反步骤，从基于理性知识的先验因果出发，以实践（或经验性知识）来检验认识（或纯粹知识）时，由于经验性知识的偏差，同样也会无法真正地认识世界，无法得到理念世界中的纯粹知识。

因此，在统计学的语境下，因果推断的重要基础就是经验性知识。再进一步，就是经验性知识的完备性和真实性。抽样调查作为社会科学定量研究的重要研究方法之一，是经验性材料与知识的重要来源，调查数据的质量是社会科学定量研究的基石，事关经验性知识的准确性与可靠性。利用样本进行统计推断的目标是总体，因此，样本属性和特征决定了由样本得到的分析结果与作为总体的现

实之间的一致性或差距。^①

我国的社会科学定量研究经过不断发展,取得了长足进步,呈现出一片繁荣景象。抽样调查方法的广泛应用、抽样技术的不断改进与完善为此提供了坚实的基础,起到了重要的不可替代的作用。但不可否认的是,定量研究被误用的情况也日渐增多。单纯追求方法创新的“方法驱动型”研究而忽略分析样本的属性,尤其是忽略分析样本与总体之间的结构问题,则是潜在的、被有意无意地忽略的重要且基本的问题之一。更重要的是,在社会学研究过程中,当统计分析结果无法接受理论假设时,通常情况下,会以否定并重构理论为目标;但显然,理论假设的检验既与理论相关,也与分析结果相关。分析样本与总体之间存在的结构偏离导致统计结果出现样本结构性偏差、影响分析结果的正确性,进而会使理论检验结果出现偏差,即用“错误”的数据、得到“错误”的分析结果和研究结论,并错误地拒绝了可能正确的理论。这种错误不仅会影响到理论构建,而且更可能影响到以此为基础的政策导向的正确性。这是事关我国社会科学定量研究健康发展的一个重大基础性问题,必须予以高度重视。

为引起学界同仁甚至依赖循证决策模式开展工作的决策部门的重视,本文将着重讨论样本结构性偏差问题。当然,已有许多文献都深入讨论了样本结构对估计量的影响作用。但,虽然已有不同学

^①引自巫锡炜的评论:为了实现对总体的有效推断,样本需要满足两个基本的属性:代表性和随机性。从测量的角度看,样本就是对总体的一次测量,样本的代表性对应着测量的效度,直接决定着数据分析结果是否能够真实反映总体特征;而样本的随机性则对应着测量的信度,事关着以多个样本做重复观测推断总体特征的一致性程度。当以存在结构性偏差的样本推断总体,就好比以有问题的称称量物体,结果总是可能存在问题的。无论何种统计方法都可能无法真正地解决或消除样本结构性偏差。

科的诸多文献一直在呼吁（臧雷振、陈鹏，2015；邱忠霞、胡伟，2016；许加明、陈友华，2020），但是目前的定量研究仍然没有对样本结构偏离及其可能的危害予以必要的高度重视；且已有研究亦未从因果推论角度讨论样本结构偏离对实验效应估计量的影响问题。正因如此，本文将在简要讨论样本结构性偏差的定义与来源的基础上，从理论分析和实验性示例两方面讨论样本结构性偏差问题，以提请学界予以更多关注。

一、样本结构偏离：定义与来源

（一）样本结构偏离的定义

结构偏离本文所指的样本是**分析样本**，既非抽样样本，也非各调查机构提供的研究样本，而是研究样本经过研究人员改造并用于最终分析的样本，或者说是在研究样本的基础上剔除分析变量包含缺失值的案例后形成的用于分析的最终样本。因此，样本结构偏离是指：变量分布在最终分析样本与研究总体之间的差异超过了随机误差的范围。

样本结构偏离是**联合分布的偏离**，而非单变量分布偏离，即分析样本与总体在多个变量的联合分布（集合特征，如年龄性别结构）上存在显著的超过了抽样误差范围的偏离。一个简单的例子：假设总体中男性与女性的规模之比是1：1，但如果最终分析样本中的男女规模之比呈现出4：6或6：4的情形，且超出抽样误差的范围，则样本中性别变量的分布出现了结构偏离。实际调查会涉及很多变量，如将上例扩展成包括性别、年龄、民族三个变量，如果样本中这三个变量所形成的联合分布与总体中的联合分布存在显著差

异，则样本同样存在结构偏离。通常情况下研究者可能更关注单变量的分布，且单变量的结构偏离相对是显性的；而这种联合分布的偏离却是隐性的，很容易被忽略。当然，单变量的结构偏离必然会导致多变量的联合分布的偏离，但单变量不存在结构偏离并不意味着多变量联合分布的结构偏离就不存在。

总之，样本结构偏离是指样本与总体在多个变量的联合分布上的差异大于抽样误差范围。而样本结构性偏差是指由样本结构偏离导致的统计分析结果的偏差，意指偏离后果。

（二）样本结构偏离的来源

格罗夫斯等（Groves et al., 2009）将抽样调查过程中的各种误差归结为代表性和测量两个方面。本文认为这两个方面都会成为样本结构偏离的来源，主要包括三个大致的研究过程：1）抽样过程—抽样样本的代表性与随机性问题（如覆盖误差、无应答误差、末端抽样时的样本替代等）；2）测量过程—测量工具与访员等测量过程中的误差；3）数据处理过程—调查数据到分析样本的过程（特别是缺失数据的处理、变量的转换处理等）。

样本结构偏离表面上看是样本代表性的反面称谓，只是本文更强调分析样本与总体之间在结构上的显著差异。需要强调的是，本文中的样本结构偏离，既包括来自代表性方面的调查误差，也包括测量误差。首先，测量误差会导致变量的联合分布产生偏离，从而使样本结构与总体结构出现差异。其次，从统计结果看，“在CEV（经典变量误差[classical errors-in-variables, CEV]假定）情形下，（包含变量测量误差的）OLS回归将给出一个有偏而又不一致的估

计量”，且“衰减偏误（attenuation bias）主要取决于CEV假设”^①（伍德里奇，2010），即不论是代表性还是测量问题，都会导致估计量的无偏有效一致性假定被违反，这就是本文所强调的统计结果上的样本结构性偏差。

另外还有一点需要解释的是，下文中所用的“缺失”是为构造一个假想的分析样本而采取的数据处理方法，它对应着调查过程与分析过程中的所有可能误差（如抽样过程的非随机性、测量误差、数据处理中对缺失变量或案例的处理等）。注意：它并非完全是统计意义上的缺失案例。因此，本文中的缺失案例从本质上讲是无法通过缺失数据处理方法予以弥补的。这一点极其重要！如果再进一步，相对而言，样本代表性是相对显性化的，可以通过比较样本与相应研究总体做出评价，虽然某些情况下，要想获得研究对象在某

^①匿名审稿人认为：“随机性的抽样误差不会影响估计结果的无偏性，但会增加估计量偏差；而测量误差则不一样，不管是随机性的测量误差还是非随机性的测量误差都会导致因果推断出现偏差。举个极端的例子，如果自变量和因变量的测量都只包含随机误差，那么二者的相关性将为零，因此，即便真实变量之间存在关系，在有随机测量误差的情况下，这种相关性会出现向零的偏差。”（即“OLS 因经典变量误差而导致的衰减偏误[attenuation bias]：平均而言[或在大样本中]，所估计的 OLS 影响将会变小”[伍德里奇，2010]）但作者认为，一般情况下，（某变量的）测量误差与其他变量总会存在一定的相关性，因此，测量误差既会通过改变回归方程中各变量的方差而影响到估计量的有效性，还会通过改变变量的联合分布而破坏代表性并增加估计量的偏差。正如伍德里奇指出的那样，这种衰减偏误“有赖于 CEV 结构”（经典变量误差[classical errors-in-variables]假定）。而且，“在 x_1^* 与 x_2 和 x_3 都不相关的特殊情形中， $\hat{\beta}_2$ 和 $\hat{\beta}_3$ 都是一致的。但在实践中几乎不存在这种情况。一般来说，单个变量的测量误差会导致所有估计量都不一致。遗憾的是，偏误的大小甚至方向都不容易得到”（伍德里奇，2010）。更何况现实中更可能的是多个变量之间存在着联合性的测量误差，从而在导致估计量方差变化的条件下，还会导致联合分布的变化进而增加估计量的偏差。正如本文所强调的，样本是所有可能误差的综合体，我们根本无法区分其误差来源是隐性的测量误差，还是显性的代表性误差，更不用说现有许多研究根本不关心这些可能的误差。

个时点上对应的“研究总体”可能是比较困难，甚至是不可能的。但测量误差在许多情况下则是隐性的、无法被评价的（如个体的幸福感、收入等）。正如上文所讨论的，测量误差会改变变量自身的分布，进而改变联合分布而产生样本结构偏离。

（三）因果推断视角下的样本结构偏离

样本结构偏离并不完全等同于因果推论中的样本选择性问题。简单地讲，在因果推断视角下，分析样本中的实验组无法代表实验组对应的总体，控制组样本无法代表控制组对应的总体，从而导致因果推断中的实验效应估计量产生样本结构性偏差。以经典的教育回报研究为例，分析样本中，实验组样本与实验组所对应的总体在某些特征的联合分布上存在结构偏离，或者是控制组样本与控制组总体之间存在结构偏离。举个极端的例子，分析样本中，以毕业于某些985高校的的大学生为实验组来代表中国的大学生群体；显然，该实验组样本存在严重的结构偏离问题，进而会影响到实验效应的估计量。

某种意义上，上述的结构偏离是指分析样本因抽样、调查、数据处理等过程中的非随机因素而产生的另一种样本选择性，是更基础性的。它不仅影响样本对总体的代表性，而且还会使后续的因果推论因叠加了抽样过程、调查过程和数据处理过程的选择性（类似于样本选择性，即样本中不包含那些没有被抽中的样本）和进入实验（如是否上大学、是否迁移与流动等社会性实验）的样本选择性这两个双重选择性，从而增加了因果推论的难度与复杂性。尽管内生的选择性似乎概括了上述两种因果推断中的样本选择性问题，但却仍然无法包含由于数据处理而产生的样本结构问题。正是由于分

析样本是所有选择性的混合结果，因此，在因果推论中很难剥离各种选择性而得到更接近于现实的因果实验效应。

尽管已有众多文献讨论了因果推论中的各种问题与方法，但却始终忽略了实验组和控制组各自的结构与总体结构存在差异的情况。更重要的是，这种结构偏离对实验效应估计量的影响作用可能无法通过各种基于传统统计方法、新近发展出来的所谓高级统计方法（如IV、PSM、DID等）来解决，唯有通过改善样本结构提高其对总体的代表性来消除。所以，样本结构本身的问题及其后果，可能比内生性问题更为要紧且基本，也更容易被有意无意地忽略。只有正确认识样本结构及其代表性在整个研究中的重要性和基础性，才有可能真正体现定量研究的总体性思路。^①

二、理论分析

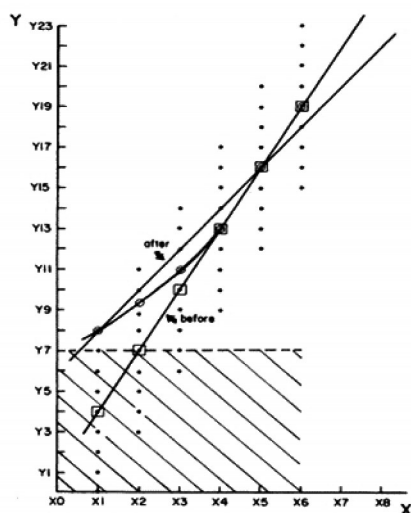
本文将首先以样本选择性偏差为例来呈现样本结构性偏差，不论是内生的还是外生的样本选择性本质上都是本文样本结构偏离的一种。在此基础上，本文将其扩展到因果推断中，以呈现样本结构偏离对于因果推断结果，特别是实验效应估计量的影响。

（一）样本选择性偏差

伯克（Berk，1983）曾在赫克曼（Heckman，1979）的基础上，讨论样本选择性所形成的偏差。温希普和梅尔（Winship &

^①被忽视的还有另外一个重要的技术性问题的：复杂抽样设计条件下的数据分析。我国目前的大型抽样调查基本上都采用多阶段分层PPS抽样的方法。而这种复杂抽样设计条件下的数据分析，应该使用复杂抽样设计的数据分析，以便尽最大可能地还原抽样过程，和更为准确地估计复杂抽样设计条件下的抽样方差及相应的统计结果。有关讨论可参见希林加等（Heeringa et al., 2010）和周皓（2012）的文章。

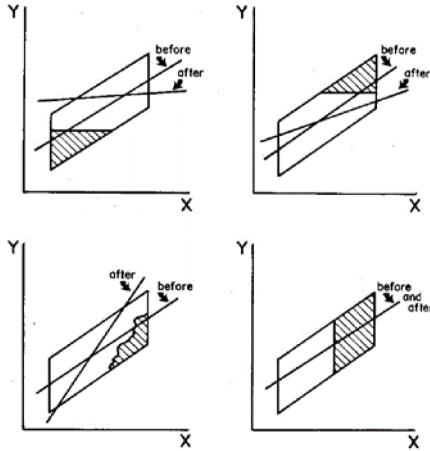
Mare, 1992) 对此做了进一步的充分讨论。如图1所示, 图中所有散点代表假设总体, 以空心方块连接的直线反映了总体中变量 X 与 Y 之间的线性关系 (以斜率表示其作用强度), 亦可将其假设为 X 与 Y 的真实关系。当存在样本选择性时, 如排除所有 Y 值小于 Y_7 的散点, 只保留大于 Y_8 及以上的散点^①, X 和 Y 间的关系则会变为图中“after”所指的这条斜率相对较小的直线, 甚至变成底部为曲线形状的非线性关系。这表明, 如果存在样本选择性, 那么根据样本得到的结果并非是 X 与 Y 之间真实的关系。图2展示了在几种不同的样本缺失情况下, X 与 Y 之间关系的变化。这些结果都表明, 样本结构偏离会导致两个变量之间关系的偏差。



来源: Berk, 1983。

图1 样本选择性偏差示意图

①这里只是以因变量为例来说明, 从自变量的角度, 相当于坐标系中的横向切割转为纵向切割, 其结果亦类似。需要再次强调的是, 在面对分析样本时, 我们根本无法真正知道样本的结构偏离到底是由因变量还是自变量的问题引起的。



来源: Berk, 1983。

图2 不同样本缺失情况下的结果差异

伯克 (Berk, 1983) 的证明过程结合了实验效应模型和选择模型两个模型。假设样本中有两个因变量 Y_{1i} 和 Y_{2i} , 且两者各有以下方程:

$$Y_{1i} = X_{1i}\beta_1 + U_{1i} \quad (1a)$$

$$Y_{2i} = X_{2i}\beta_2 + U_{2i} \quad (i = 1, \dots, I) \quad (1b)$$

其中, X 为由变量组形成的矩阵, β 为回归系数向量, U 为残差项。

假设方程 (2a) 是选择模型, 则:

$$E(Y_{1i}|X_{1i}, Y_{2i} \geq 0) = X_{1i}\beta_1 + \frac{\sigma_{12}}{(\sigma_{22})^{1/2}} \lambda_i \quad (2a)$$

$$E(Y_{2i}|X_{2i}, Y_{2i} \geq 0) = X_{2i}\beta_2 + \frac{\sigma_{22}}{(\sigma_{22})^{1/2}} \lambda_i \quad (2b)$$

其中: $\lambda_i = \frac{f(Z_i)}{1 - F(Z_i)}$; $Z_i = -\frac{X_{2i}\beta_2}{(\sigma_{22})^{1/2}}$ 。

如果存在样本选择性（如 $Y_{2i} \geq 0$ ，即只选择了概率大于0的样本），会导致方程（2a）中的估计将是有偏的。具体的相关证明过程可参见赫克曼（Heckman, 1979）、温希普和梅尔（Winship & Mare, 1992）或伍德里奇（2010）的著述。

（二）结构偏离与平均实验效应偏差

上述结果表明样本结构偏离会导致线性回归中估计量的无偏性和一致性假定被违反，那么将其扩展到因果推断中实验效应的估计时也是一样的。

首先以图示的方式讨论。假设存在实验组和控制组两个样本，当实验组与控制组没有任何结构偏离时（或在某种真实的社会现实条件下），实验组均值为 T_0 ，控制组均值为 C_0 ，平均实验效应则为 $ATE_0 = T_0 - C_0$ （图3）。若存在结构偏离，假设实验组的底部缺失（如图3中红色横线以下部分，类似于收入低于某一标准之下的案例未被包括在抽样样本中，或因调查时缺失而导致其被视为缺失样本而被剔除在分析之外），则此时实验组的均值由 T_0 变成星号所表示的 T_1 ，平均实验效应 $ATE_1 = T_1 - C_0 = (T_1 - T_0) + (T_0 - C_0) = ATE_0 + \Delta$ ；其中， $\Delta = (T_1 - T_0)$ 部分是由结构偏离引起的偏差，从而导致估计的平均实验效应 ATE_1 是有偏的，不再等于 ATE_0 。

其次从统计角度来考察。仍然假设一个样本中存在实验组和控制组，则：

$$Y = \beta_0 + \beta_1 T + \sum \beta_i X_i + e_i \quad (3)$$

其中， T 为实验（treatment），当样本为控制组时 $T = 0$ ； X_i 为一系列的控制变量； β_1 为实验效应。

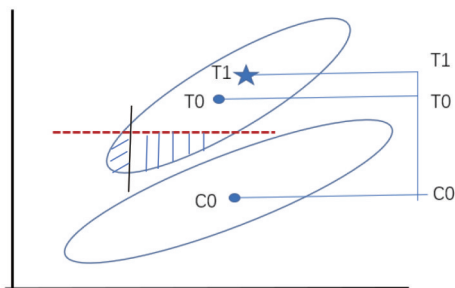


图3 结构性偏差对因果推论结果的影响示意图

如果 $cov(T, e_i) = 0$ ，则 β_1 是个无偏一致的统计量；但如果 $cov(T, e_i) \neq 0$ ，则会产生内生性问题，而 β_1 也不再是无偏一致的统计量。

再进一步，根据 T 的取值，可以将方程 (3) 改为：

$$Y_c = \beta_0 + \sum \beta_i X_i + e_i \quad (T=0) \quad (3a)$$

$$Y_T = \beta_0 + \beta_1 + \sum \beta_i X_i + e_i \quad (T=1) \quad (3b)$$

结合上例，控制组不存在样本结构偏离，则方程 (3a) 的估计量是无偏的。而实验组则由于存在样本结构偏离，方程 (3b) 将成为：

$$Y_T = \beta'_0 + \sum \beta'_i X_i + e_i \quad (T=1) \quad (3c)$$

且， $\beta'_0 = \beta_0 + \beta_1 + \Delta \neq \beta_0 + \beta_1$ ， $\Delta \neq 0$ ， $\beta'_i \neq \beta_i$ ，方程中的其他自变量的估计量会存在样本结构性偏差，由于样本结构偏离是多变量的联合分布，如果将方程 (3c) 和方程 (3a) 相结合，则：

$$\begin{aligned} Y &= \beta''_0 + \beta''_1 T + \sum \beta''_i X_i + e_i \\ &= \beta''_0 + (\beta'_1 + \Delta) T + \sum \beta''_i X_i + e_i \quad (T=1) \end{aligned} \quad (3d)$$

显然， $(\beta'_1 + \Delta) \neq \beta_1$ ，且 $\beta''_0 \neq \beta_0$ ， $\beta''_i \neq \beta_i$ 。

由此可见，当实验组存在样本结构偏离时会导致实验效应估计

量的有偏结果，甚至会由此导致其他相关变量的偏回归系数同样有偏，而这种情况对于控制组也相同。

再次，从相应的因果推断统计方法来看，以PSM为例。继续上例，当实验组不存在样本结构偏离时，用于PSM的倾向得分模型

（选择模型）为： $\text{logit}(p) = \sum_{i=0}^k b_i X_i$ ；而当实验组存在样本结构偏离

时，倾向得分模型则会变为： $\text{logit}(p) = \sum_{i=0}^k b'_i X_i$ 。样本结构偏离导致

$b'_i \neq b_i$ ，继而得到可能完全不同的倾向得分值 p_i 。不同的倾向得分既可能影响实验效应的平均值，更可能改变实验组和控制组的配对（matching）过程与结果，进而影响到实验效应的估计量。

由此可见，即便因果推断的统计方法（如PSM）可以有条件地予以部分校正，但统计方法并不能从根本上解决全部偏误问题。首先，从理论上讲，整个估计过程在上述方程组的基础上，需要再增加一个选择方程以表示是否能够被包括在分析样本中；但由于无法得到缺失案例的相关特征（因为无法真正知道总体的结构，以及结构偏离的来源与可能结果），因此这个选择方程是无解的。其次，在有结构偏离的情况下，即使是原有的选择方程，也会由于结构偏离而使选择方程及统计估计量（如PSM中的倾向得分）发生改变，进而影响到过程估计量及最终实验效应估计结果。正因如此，样本结构偏离导致了有偏的平均实验效应估计量，且可能无法被真正消除。

上述讨论已表明样本结构偏离不仅导致线性回归系数估计量不再是无偏有效的，而且因果推断的统计方法亦无法解决由结构偏离所导致的偏差问题。

三、分析思路与检验方法

为了能够更简洁地说明结构偏离的影响，本文将由易入难，利用实验数据来展示。首先，本文将分别以单变量和多变量的形式考察样本结构偏离对线性回归的影响。然后，在单个线性回归的基础上，在假设总体中加入实验效应，并使用多元线性回归和倾向得分匹配两种方法，从因果推论角度展示样本结构偏离对平均实验效应估计量的影响。最后，通过多轮（假设100次）随机化处理，改变分析样本结构偏离的情况（即以随机形式构建多个不同的分析样本），再对上述统计结果（平均实验效应和各变量系数的估计量）进行统计分析，以说明样本结构性偏差的可能条件。

本文在设计构建实验数据时使用“缺失”的方式，将“假设总体”改造为最终的“分析样本”。**需要强调的是：**本文的缺失方式看似与数据处理过程中的剔除缺失案例一样，但事实上完全不同。本文中“缺失案例”的产生，从过程上讲，包含了从抽样、测量到数据处理等所有可能导致样本结构偏离的过程，而非仅仅只是数据分析过程中剔除缺失案例这一简单过程。因此，从根本上讲，本文中的“缺失案例”无法通过缺失数据的统计方法来予以填补。只是为方便起见，下文的表述都以样本缺失来代替。

本文设计的样本结构偏离来源共包括五种情况：1) 纯随机缺失；2) 顶部缺失；3) 底部缺失；4) 局部缺失；5) 结构变化。随机缺失是指某变量值域范围内案例的缺失都是随机的，正因为是随机的，所以样本结构不一定会发生结构性变化；具体操作时，对假设总体中的每个案例赋予一个服从 $N(0, 1)$ 的随机数，将随机数

小于-1.65或大于1.65的案例标记为缺失（即缺失案例占假设总体的10%左右）；由于该随机数与任何变量均无关，因此可以达到随机缺失的效果。顶部缺失和底部缺失则是指某变量的取值超过或低于某个标准后的部分案例全部缺失；具体操作时，为使这种缺失值的标准随实验次数而发生变化，因此，在每一次实验中根据所给的随机数而决定缺失比例，如随机数为0.56，则顶部大于均值1.56（ $1+0.56$ ）个标准差，或底部小于均值1.56个标准差的案例为缺失案例。局部缺失是指变量某一取值范围内（如从均值到小于均值某个标准差的范围内）的部分特定案例被指定为非随机缺失；其中的非随机缺失通过构建模型中的残差项 $e_i < 0$ 来予以控制。结构变化是指某变量小于某个特定值（如均值）的区间内，案例按不同取值区间，赋予不同比例（概率）的非随机缺失，从而导致该变量及与其他变量的联合分布结构发生变化；其中的非随机缺失仍通过残差项控制。这五种情况中，前四种情况可能显得比较特殊，并不一定会发生在实际调查中，但第五种结构性变化在各种调查中则可能是实际存在，而且也更隐蔽，更不容易被直接观测。例如，由于抽样框等原因导致部分特定案例未被纳入（也就无法被包括在分析样本中），或案例在测量中的误报导致变量的分布变化（其实分布结构的变化是指某部分的案例在分析样本中的比例高于或低于总体中的相应比例，从而导致其分布形状与总体分布形状之间的偏离），或调查过程的无应答导致样本缺失而在分析过程中予以直接删除等。因此，以下的分析过程中需要特别注意第五种情况。

除连续变量以外，本文还将讨论分类变量的结构性变化。文中的分类变量均为二分变量，相应的五种设计分别为：0值随机缺失、0值非随机缺失、1值随机缺失、1值非随机缺失，以及用整体均值

的变化以表示结构的改变（其背后表明分析样本中两类人群的比例结构发生了变化，如性别或民族等）。其中，随机或非随机缺失的构建方法和上述连续变量相同；在0值（或1值）随机或非随机缺失的构建中，只针对取值为0（或1）这一组的案例进行缺失处理，而整体结构（均值）变化则将采用重新随机赋值方式。

本文更关注对分析样本得到的统计结果的比较与检验，比较检验的标准有两个：标准取值和基准模型。其中，“标准取值”是假设总体在因变量构建过程中各变量的实际系数，类似于代表真实的实验效应和各变量的真实作用。基准模型则是基于“假设总体”进行回归分析和倾向得分匹配后的统计结果（回归系数与平均实验效应ATT）。在构建“假设总体”时，为了体现个体差异性，每个案例在利用实际系数得到预测结果的基础上加入了服从正态分布的残差项 e 。因此，即便是基准模型，由于存在残差项，而无法得到构建标准模型时自变量对应的系数；当然，基准模型中各自变量的系数与构建模型时的标准取值基本接近，且统计检验不显著。将分析样本得到的统计结果与标准取值、基准模型这两种结果进行比较与检验才是本文的关键之所在。检验结果如果有显著的差异，则说明样本结构偏离导致了样本结构性偏差。

结构偏离效应的检验与估计量的解读一致，包括显著性、作用方向（即符号）及估计量取值大小三个方面。如果结构偏离导致某估计量（如回归系数或平均实验效应）的显著性或方向发生变化，则说明存在结构性偏差；如果显著性与方向都没有变化，但分析样本的估计量取值与两种标准存在统计显著的差异，则说明结构性偏差同样存在。同时，本文亦提请注意，在某一变量发生结构偏离时，既要注意该变量在分析中的估计结果，也要注意其他变量的估

计结果。因为结构性偏差是联合分布导致的偏差，而不只是单个自变量的问题。

方法上，本文将使用一般线性回归和简单的倾向得分匹配方法，在对分析样本的结果和标准取值、基准模型进行比较检验时则将采用不同的方法。分析样本回归分析结果与标准取值的比较检验将采用STATA中的`test...`、`mtest`的方法，即将分析样本中四个（或五个）自变量的回归系数，分别与标准取值进行联合与单独检验，以 F 统计量表示。与基准模型的比较检验将选用STATA中的`suest+test`的命令。这个命令类似于Hausman检验，但后者强调对模型系数矩阵的整体检验，而`suest`命令则是关联两个模型并对相同的每个变量的系数进行统计检验，具体请参见STATA中有关该命令的帮助文件。同时，为了检验利用倾向得分匹配方法估计得到的分析样本平均实验效应ATT与真实效应（标准取值，构建假设总体时的实验 D 所对应的系数）间的差异，本文将直接采用单变量 t 检验，即 $(ATT-0.2) / s.e. (ATT)$ 。

四、线性回归中的结构性偏差

（一）实验数据的构建

在简单线性回归中，设定两个服从标准正态分布的变量，分别命名为自变量 X 和残差项 e ，并设定案例数为10000，然后依据公式 $Y = 2X + 1 + e$ 计算得到因变量 Y ，这就构成了单变量回归所对应的“假设总体”。散点分布如图4所示。

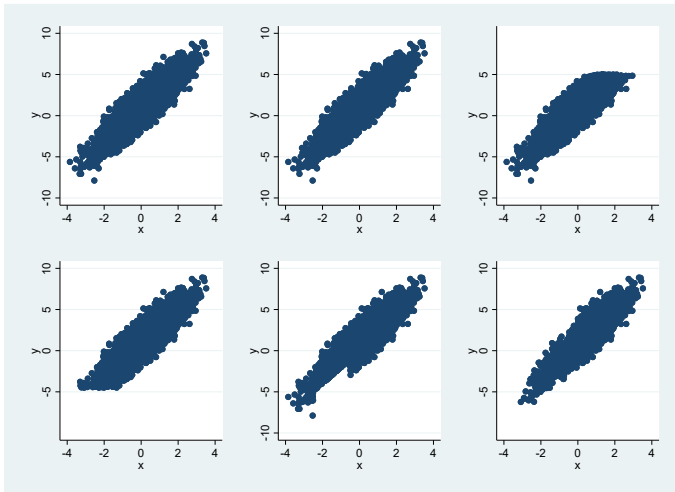


图4 基准分布及各种缺失条件下的分布情况

在多元线性回归中构建变量时，将根据自变量间的关系分为（与其他自变量）相关与不相关两种类型，并将自变量分为连续变量和二分类变量两种类型。为此，类似于上述简单线性回归，共引入四个变量： X_1 、 X_2 、 D_1 和 D_2 。其中， X_1 和 X_2 为连续变量， D_1 和 D_2 为二分变量。构建基础数据时各变量的定义如下： $X_1 \sim N(0,1)$ ， $D_1 \sim B(1,0.51)$ ， $D_2 \sim B(1,0.80)$ 。为体现变量间的相关，将连续变量 X_2 定义为与二分变量 D_1 相关，即以 D_1 分类为依据：当 $D_1=1$ 时， $X_2 \sim \chi^2(1,22)$ ；当 $D_1=0$ 时， $X_2 \sim \chi^2(1,12)$ ；将两个分布结合形成自变量 X_2 。在此基础上，因变量 Y 由方程 $Y_i = 0.5X_1 + 0.2X_2 + 0.3D_1 + 0.5D_2 + e_i$ 形成。此处， D_1 与 X_2 相关，但与其他自变量不相关。为进一步考察分类变量 D_1 和 D_2 的结构性变化可能带来的影响，实验将分别以均值0.6和0.92作为 D_1 和 D_2 的替代方案，即 $D_1 \sim B(1,0.65)$ ， $D_2 \sim B(1,0.92)$ 。

(二) 简单线性回归的结果

简单线性回归的实验以上述五种缺失情况为例。这五种缺失条件下的散点分布情况如图4所示，相应的简单线性回归结果如表1所示。

表1 简单线性回归的结果

	基准	随机	顶部	底部	局部	结构
X	1.994***	1.992***	1.925***	1.973***	1.908***	1.953***
	(0.010)	(0.011)	(0.011)	(0.010)	(0.010)	(0.010)
_cons	1.003***	1.001***	0.960***	1.011***	1.080***	1.053***
	(0.010)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)
N	10000	9006	9631	9936	9546	9765
R^2	0.801	0.799	0.773	0.793	0.798	0.808
F Test $X=2$	0.39	0.63	50.03***	7.13***	86.59***	23.97***
X 与基准模型的比较	—	—	237.07***	49.29***	379.30***	146.46***

注：(1) “ X 与基准模型的比较”一行中的数值系卡方值，自由度为1；

(2) 括号内为标准误；

(3) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ 。

由表1结果可见，尽管自变量的方向和显著性均未发生变化，

但不论与标准取值相比较，还是与基准模型相比较，四个非随机缺失的模型（顶部、底部、局部与结构非随机缺失）中的回归系数值都有显著差异。例如，结构模型中 X 的系数为 1.953，与标准取值之间也存在显著差异（ $F=23.97$ ），与基准模型（1.994）也存在显著的差异（ $\chi^2(1) = 146.46$ ）。顶部缺失、底部缺失和局部缺失的结果也相同。但随机缺失模型的结果与标准系数和基准模型都无显著差异。这说明只要变量在其值域范围内是随机缺失的情况，那么，其统计结果可能仍然是无偏的。但凡案例的缺失是非随机的，所得估计量与“可能的”真实结果之间会存在显著的差异，其关键在于非随机的缺失导致结构性偏差。

（三）多元线性回归的结果

本文的目标是检验各变量的结构性变化对分析结果的影响。此处仅呈现因变量的结构变化对统计结果的影响，不再列出由自变量结构变化而导致的估计结果偏误。因变量结构变化的分析结果如表 2 所示，该表由三部分组成：回归结果、与标准取值的比较、与基准模型的比较。

表2 不同缺失情况对多元线性回归结果的影响

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	基准	随机	顶部	底部	局部	结构	D_1	D_2
X_1	0.494***	0.492***	0.465***	0.464***	0.456***	0.469***	0.495***	0.494***
	(0.010)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
X_2	0.201***	0.202***	0.183***	0.194***	0.181***	0.191***	0.215***	0.201***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
D_1	0.334***	0.323***	0.378***	0.296***	0.166***	0.310***	-0.025	0.343***
	(0.026)	(0.028)	(0.026)	(0.025)	(0.027)	(0.027)	(0.021)	(0.027)
D_2	0.474***	0.475***	0.449***	0.453***	0.452***	0.470***	0.479***	0.045
	(0.025)	(0.027)	(0.025)	(0.025)	(0.027)	(0.026)	(0.026)	(0.036)
_cons	-0.012	-0.016	0.204***	0.202***	0.748***	0.336***	-0.059	0.324***
	(0.032)	(0.033)	(0.033)	(0.032)	(0.034)	(0.033)	(0.034)	(0.041)
N	10000	9006	9416	9643	7709	8538	10000	10000
R^2	0.754	0.754	0.690	0.743	0.745	0.756	0.750	0.746
(每个自变量) 与标准取值的比较 ($X_1=0.5$, $X_2=0.2$, $D_1=0.3$, $D_2=0.5$)								

续表

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	基准	随机	顶部	底部	局部	结构	D_1	D_2
X_1	0.40	0.631	12.148***	13.552***	18.485***	9.207**	0.285	0.376
X_2	1.89	0.718	74.542***	13.464***	130.459***	27.025***	133.006***	0.449
D_1	0.22	0.702	8.921**	0.024	25.025***	0.146	235.700***	2.586
D_2	3.85*	0.885	4.067*	3.492	3.323	1.285	0.670	163.418***
(每个自变量)与基准模型比较								
X_1		0.381	70.095***	80.018***	33.981***	26.535***	0.224	0.011
X_2		0.603	321.595***	223.971***	446.426***	211.774***	56.731***	0.125
D_1		1.563	34.795***	46.801***	93.316***	3.597	154.264***	6.011
D_2		0.010	10.465**	5.581*	1.660	0.100	0.892	555.505***

注一：(1)表中模型(6)的结构变化是 X_1 小于0部分的案例非随机缺失(即 $X \leq 0, e < -1.288$);

(2)模型(7)中 D_1 表示 D_1 的结构性变化,即 D_1 的均值由0.51变为0.65;

(3)模型(8)中 D_2 表示 D_2 的结构性变化,即 D_2 的均值由0.80变为0.92;

(4) X_1 的取值由 D_1 决定。当 $D_1=1$ 时, $X_1 \sim \chi^2(22)$; $D_1=0$ 时, $X_1 \sim \chi^2(10)$ 。

注二：(1)表的上半部分是各种情况与构建因变量时的实际系数值的比较;表的下半部分则是与基准模型(表的上半部分的第一个模型)中各系数的比较;

(2)表的上半部分中的数值为F统计量;表的下半部分为卡方统计量;

(3)表的下半部分“与基准模型的比较”使用了suest方法;

(4)括号内为标准误;

(5)* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ 。

由结果可见,类似于上述简单线性回归的结果,多元线性回归的七个回归结果中,仅有随机缺失模型与标准取值和基准模型均不存在显著差异,其余模型都由于分析样本的结构变化而导致结构性

偏差，但各模型之间在结构性偏差上存在一定的差异。

上述两个实验数据说明：非随机的结构偏离（不论是来自样本选择性或缺失案例，还是来自其他变量的结构变化而导致的联合分布的结构性变化）都可能导致相关研究自变量的结果发生根本性变化。这就提醒我们在分析过程中必须要注意结构性偏差问题。

五、结构性偏差对因果推断结果的影响

上述结果表明，样本结构偏离会导致回归系数估计量违反无偏性假定。本部分将从线性回归扩展到因果推断中实验效应的估计问题，试图回答：分析样本如果存在结构偏离，线性回归是否能够得到无偏的平均实验效应（ATE）估计量？如果不能，是否可以利用因果推论中的相关方法改进实验效应的估计？为回答这一问题，本文将在上述模拟数据思路的基础上，再加入代表实验（treatment）的二分类变量，以讨论与检验分析样本的结构偏离对平均实验效应估计结果的影响。

本部分检验的基本逻辑如下：第一步，按照实验组与控制组分别构建基础数据（即模拟总体）。这里构建的基础数据与上述线性回归的不同之处在于其需要按照不同的组别——实验组和控制组分别构建基础数据。第二步，在控制组不变的条件下（包括变量及结构均不改变），按前述五种缺失情况改造实验组的样本结构，以检验实验组的样本结构偏离对实验效应估计量的影响。第三步，保持实验组不变、改变控制组结构，以说明控制组的结构性偏差的影响作用。第四步，将样本缺失过程随机化处理（即分析样本构建过程中的参数被随机化处理），并进行多次实验，最终将这些随机化实

验的统计结果作为分析样本再进行描述性分析，以期进一步检验样本结构偏离对估计量的影响。^①

（一）基础数据构建与分析方法

首先，定义包括 10000 个案例的数据，随机建立一个二分变量 D ，其中 $D=1$ 为实验组（设 $\bar{d} = 0.41053$ ）， $D=0$ 为控制组。然后按实验组和控制组分别构建四个变量 X_1, X_2, X_3, X_4 ，其中， X_1 和 X_2 为连续变量， X_3 和 X_4 为二分变量。在 $D=0$ 的控制组中，设 $X_1 \sim N(0.7, 0.36)$ ， $X_3 \sim B(1, 0.48)$ ， $X_4 \sim B(1, 0.79)$ ；在 $D=1$ 的实验组中，设 $X_1 \sim N(0.9, 0.36)$ ， $X_3 \sim B(1, 0.51)$ ， $X_4 \sim B(1, 0.85)$ 。这种定义的目的是使实验组和控制组在自变量上存在特征差异。同时，为体现自变量之间的相关性，将 X_2 设定为与 X_4 相关，分别定义实验组和控制组的 X_2 取值：在控制组 $D=0$ 中， $X_4=1$ 时， $X_2 \sim \chi^2(11)$ ， $X_4=0$ 时， $X_2 \sim \chi^2(21)$ ；在实验组 $D=1$ 中， $X_4=1$ 时， $X_2 \sim \chi^2(5)$ ； $X_4=0$ 时， $X_2 \sim \chi^2(35)$ 。由此构成五个自变量。

其次，构建因变量： $Y=0.2*D+0.5*X_1+0.4*X_2+0.6*X_3+0.3*X_4+e_i$ 。其中， e_i 为服从标准正态分布的随机变量； D 的系数为 0.2，即“真实”的实验效应（标准取值）。后续所有模型的结果都需要与该实验效应 0.2 进行比较。如果估计得到的统计量与标准取值 0.2 有显著差异，则说明利用带有结构偏离的分析样本无法估计得到真实结果；如果与标准取值 0.2 没有显著差异，则说明实验效应的估计不存在结构性偏差。正如多元线性回归一样，由于加入了误差项，基准回归模型的结果可能与标准取值有些许差异，因此，同样需要将

^①为节约篇幅，本文未给出所有随机过程的结果。本节中所呈现的结果仅为其中一种。感兴趣的读者可以运行本文对应的命令文件以查验。

各种样本结构的分析结果与基准模型进行整体比较并检验。

在构建的基础数据之上，本文拟采用线性回归方法和倾向值匹配方法来检验样本结构偏离对估计结果的影响。线性回归是基础，如果线性回归能够得到无偏的实验效应估计结果，则不需要再进一步利用倾向得分匹配方法来估计实验效应。但如果线性回归结果有偏时，本文将利用倾向得分匹配方法来修正，并检验其是否能够真正起到修正与调整有偏结果、得到与总体实验效应无差异的无偏作用。其中，倾向得分匹配方法将使用最近邻居法，以1:5匹配，连接函数为logit。当然，这里使用的PSM方法仅是一个粗略的估计，由于实验中涉及的变量较少，且无其他外生变量，因此，本文的倾向得分匹配可能仍然存在偏差。^①

分析样本仍然通过上述五种情况的缺失得到。从分类逻辑看，这些情境需要从实验组与控制组、因变量与自变量（共四个）两个维度分别予以检验，而自变量则又包括连续型与分类型两种不同情况，因此，总计共50种（5*2*5）不同的变化。同时，为了检验不同的结构性变化所可能带来的结构性偏差，本文将利用随机实验方法进行100次模拟实验，并记录每次实验结果（即与标准取值和基准模型的比较检验结果），最后将这5000个检验结果作为分析对象。因此，由于结果相对较多，无法逐个呈现。本文仅呈现其中一

①匿名审稿人认为：“PSM并不是一种应对样本结构性偏差的方法，且缺失数据的处理应该利用相应方法来解决。”正如前文所述，本文中的缺失并不是简单的可以通过统计方法处理的样本缺失，而是结合了所有调查误差在内的结果，且事实上本文所构建的假设总体，对于现实的调查数据而言是未知的，即总体未知。因此，本质上分析样本与总体之间的偏离是未知的。另外，此处利用PSM方法，是为了呈现在样本结构偏离的情况下是否能够得到接近现实的实验效应。对于实验组和控制组之间的实验效应的估计，仅有PSM是最合适的方法。

个结果。

(二) 因变量结构变动

首先来看因变量的结构偏离对分析结果的影响。实验按实验组和控制组分别设计。具体结果请见附表3到表7。表3下半部分的两个检验表明，如果值域范围内都是随机缺失，那么，其结果与标准取值和基准模型均不存在显著差异。但凡只要是非随机的，不论是实验组，还是控制组，其结果与两种标准的比较都会存在显著差异。例如，在顶部缺失的条件下，线性回归估计得到的结果为0.140，该估计量与标准取值和基准模型的实验效应均存在显著差异 ($F=5.88$, $\chi^2=35.96$)，即：因变量如果是非随机缺失，则所得结果必然存在结构性偏差。

表3 因变量各种缺失情况下的结果：实验组

因变量： $D=1$ 实验组

	(0)	(1)	(2)	(3)	(4)	(5)
	基准模型	随机	顶部	底部	局部	结构
D	0.214***	0.219***	0.140***	0.291***	0.117	0.382***
	(0.021)	(0.024)	(0.025)	(0.021)	(0.071)	(0.024)
X_1	0.500***	0.502***	0.506***	0.476***	0.541***	0.505***
	(0.028)	(0.029)	(0.028)	(0.027)	(0.036)	(0.030)
X_2	0.201***	0.202***	0.195***	0.198***	0.204***	0.197***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)

续表

	(0)	(1)	(2)	(3)	(4)	(5)
X_3	0.302***	0.296***	0.303***	0.285***	0.305***	0.308***
	(0.020)	(0.021)	(0.020)	(0.019)	(0.025)	(0.022)
X_4	0.493***	0.494***	0.497***	0.467***	0.497***	0.501***
	(0.038)	(0.038)	(0.039)	(0.037)	(0.041)	(0.038)
_cons	-0.014	-0.025	0.055	0.075	-0.076	0.022
	(0.053)	(0.055)	(0.058)	(0.052)	(0.064)	(0.055)
N	10000	8745	9507	9827	6083	8490
R^2	0.752	0.734	0.614	0.756	0.635	0.754
ATT	-0.1796 (0.0581)	-0.0929 (0.0664)	-0.0482 (-0.0826)	-0.02877 (0.0577)	-0.09843 (0.0594)	.2269971 (0.0606)
(每个自变量) 与标准取值的比较 ($D=0.2, X_1=0.5, X_2=0.2, X_3=0.3, X_4=0.5$)						
D	0.46	0.67	5.88*	18.76***	8.78**	123.07***
X_1	0.00	0.00	0.04	0.79	0.26	0.10
X_2	0.70	1.67	5.02*	1.87	0.28	10.53
X_3	0.02	0.04	0.02	0.61	0.26	0.02

续表

	(0)	(1)	(2)	(3)	(4)	(5)
X_4	0.04	0.02	0.01	0.81	0.02	0.19
(每个自变量)与基准模型的比较						
D	—	0.23	35.96***	161.03***	191.68***	666.11***
X_1	—	0.04	0.99	10.67**	14.93***	0.68
X_2	—	2.08	19.45***	151.67***	14.91***	307.74***
X_3	—	0.74	0.03	11.56***	6.98**	0.46
X_4	—	0.04	0.15	79.03***	14.63***	2.66

注一：(1)表中与标准取值比较时均为 F 统计量，与基准模型比较时均为卡方统计量；
 (2)括号内为标准误；
 (3) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ 。

注二：判断条件：

- (1) 随机缺失： $D=1$ 且随机数大于等于0.7；
- (2) 顶部缺失： $D=1$ 且 yd 大于等于均值加1.5个标准差；
- (3) 底部缺失： $D=1$ 且 yd 小于等于均值减去1.5个标准差；
- (4) 局部缺失： $D=1$ 且随机数大于等于0.5且误差项 e 取值小于0 ($e < 0$ 使局部缺失为非随机)；
- (5) 结构性变化： $D=1$ 且各区间内随不同的缺失比例，且误差项 e 取值小于0。

这种结构性偏差事实上并无法完全利用统计方法予以解决。以表3中的ATT一行为例。虽然在局部缺失和结构性改变条件下的实验效应估计量与标准取值和基准模型并不存在显著差异，但是，在实验组因变量顶部缺失的条件下，利用PSM估计得到的ATT为

0.0482；如果与标准取值 ($D=0.2$) 相比，则差异值的 t 统计量为：
 $t = \frac{(0.2 - 0.0482)}{0.0826} = 1.838$ ($p=0.066$)，虽然仍不显著，但与标准取值有较大差异；但如果与基准模型 (0.214) 相比，则 $t=2.007$ ($p=0.04475$)，在 0.05 水平上显著，即与基准模型中的实验效应有显著差异。底部缺失模型的结果不仅与基准模型有显著差异 ($p=0.0013$)，而且与标准取值也有显著差异 ($p=0.003$)。这表明，统计方法并不能完全解决所有形式的样本结构偏离问题。

上述结果在因变量控制组存在样本结构偏离时同样成立 (表 4)。除底部缺失模型的估计结果以外，其他三种情况与标准取值均存在显著差异，且各种条件下的估计结果与基准模型的估计结果也都存在显著差异，这说明控制组的结构性变化同样会产生结构性偏差。从 PSM 的 ATT 估计结果来看，在顶部、局部和结构模型中，PSM 的估计结果与两种标准之间同样都有着显著差异。

表 4 因变量各种缺失情况下的结果：控制组

因变量： $D=0$ 控制组

	(1)	(2)	(3)	(4)	(5)	(6)
	随机	顶部	底部	局部	结构	结构加权
D	0.210***	0.261***	0.184***	0.075***	-0.083***	0.109***
	(0.025)	(0.021)	(0.021)	(0.021)	(0.022)	(0.022)
X_1	0.497***	0.486***	0.492***	0.630***	0.463***	0.477***
	(0.033)	(0.028)	(0.027)	(0.028)	(0.029)	(0.029)
X_2	0.201***	0.198***	0.200***	0.198***	0.198***	0.199***

续表

	(1)	(2)	(3)	(4)	(5)	(6)
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
X_3	0.302***	0.289***	0.292***	0.293***	0.307***	0.314***
	(0.024)	(0.020)	(0.020)	(0.020)	(0.021)	(0.021)
X_4	0.482***	0.529***	0.481***	0.490***	0.541***	0.495***
	(0.051)	(0.038)	(0.037)	(0.037)	(0.040)	(0.041)
_cons	0.006	-0.044	0.054	0.048	0.305***	0.123*
	(0.069)	(0.054)	(0.053)	(0.052)	(0.056)	(0.056)
N	7021	9697	9922	9530	8538	8538
R^2	0.782	0.735	0.755	0.763	0.782	0.757
ATT	-0.2336 (.0654)	-0.0956 (.0524)	-0.2102 (.0568)	-0.5184 (.0588)	-0.5632 (.0621)	—
(每个自变量) 与标准取值的比较 ($D=0.2, X_1=0.5, X_2=0.2, X_3=0.3, X_4=0.5$)						
D	0.15	8.43**	0.60	33.82***	165.68***	
X_1	0.01	0.27	0.09	22.22***	1.64	
X_2	0.21	1.08	0.05	1.91	1.67	
X_3	0.01	0.28	0.15	0.14	0.10	

续表

	(1)	(2)	(3)	(4)	(5)	(6)
X_4	0.12	0.55	0.27	0.08	1.06	
(每个自变量)与基准模型的比较						
D	0.10	137.4***	66.18***	370.58***	1070.94***	
X_1	0.02	4.67*	2.45	231.91***	7.80**	
X_2	0.20	31.49***	37.81***	87.33***	26.09***	
X_3	0.00	6.78**	7.17**	2.60	0.19	
X_4	0.10	5.81*	3.79	0.09	5.70*	

注一：(1)表中与标准取值比较时均为 F 统计量，与基准模型比较时均为卡方统计量；

(2)括号内为标准误；

(3) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ 。

注二：判断条件：

(1) 随机缺失： $D=0$ 且随机数大于等于0.7；

(2) 顶部缺失： $D=0$ 且 y_d 大于等于均值加1.5个标准差；

(3) 底部缺失： $D=0$ 且 y_d 小于等于均值减去1.5个标准差；

(4) 局部缺失： $D=0$ 且 $0.7 \leq X_1 \leq 1.63$ 且 $y_d \leq 2$ ；

(5) 结构性变化： $D=0$ 且各区间内缺失比例不同且误差项 e 小于0（范围内随机挑选某些案例， $e < 0$ 的案例缺失）。

逆向思考，如果样本的取值范围覆盖了总体的取值范围，即使存在样本结构偏离，PSM的估计结果与标准取值之间并不一定存在显著差异，虽然会有偏差；但如果取值范围发生变化，则必然会导致估计量存在显著的偏差。这说明PSM方法在本文中的适用性，也

说明PSM方法在因果推论中的局限性，它更多地受制于样本结构，又反过来说明样本的重要性。

(三) 连续自变量结构性变化

与其他变量无关的连续变量的结构性变化，在纯随机条件下，分析结果与基准模型和标准取值都没有显著差异，且顶部缺失和底部缺失模型也一样。但局部和结构性变化则会导致结构性偏差（表5和表6）：不仅该连续变量自身的系数（与两种标准）有明显差异，而且也会影响到实验效应的估计量。例如，在表5中，结构模型中 X_2 的系数与标准取值间差异对应的 F 统计量为24.79，与基准模型的差异的卡方统计量则达到310.18，这表明该连续变量的结果与标准取值、基准模型都存在显著差异。与此同时，局部缺失模型中ATT估计量与标准取值（0.2）的差异所对应的 t 统计量为2.954（ $p=0.003<0.01$ ）；结构模型中ATT估计量与标准取值的差异所对应的 t 统计量为2.748（ $p=0.0060<0.01$ ）。这说明，尽管在顶部或底部缺失时所对应的ATT统计量与标准取值间的差异并不显著，但PSM方法并未能纠正由于局部或结构性变化所带来的平均实验效应估计量的结构性偏差。控制组与实验组的结果基本相同。

表5 连续自变量各种缺失情况下的结果：实验组

连续自变量 X_1 ： $D=1$ 实验组

	(1)	(2)	(3)	(4)	(5)
	随机	顶部	底部	局部	结构
<i>D</i>	0.173***	0.213***	0.211***	0.329***	0.401***
	(0.025)	(0.022)	(0.023)	(0.022)	(0.023)
<i>X</i> ₁	0.500***	0.500***	0.509***	0.427***	0.359***
	(0.030)	(0.031)	(0.030)	(0.028)	(0.028)
<i>X</i> ₂	0.202***	0.201***	0.201***	0.201***	0.201***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
<i>X</i> ₃	0.301***	0.310***	0.301***	0.304***	0.300***
	(0.022)	(0.021)	(0.020)	(0.020)	(0.020)
<i>X</i> ₄	0.502***	0.495***	0.490***	0.496***	0.494***
	(0.039)	(0.038)	(0.038)	(0.038)	(0.038)
_cons	-0.024	-0.014	-0.015	0.032	0.086
	(0.056)	(0.055)	(0.054)	(0.053)	(0.054)
<i>N</i>	8353	9356	9631	9558	9337
<i>R</i> ²	0.729	0.743	0.746	0.746	0.744
ATT	-.1129 (.0711)	-.2386 (.0635)	-.1719 (.0592)	.0221 (.0602)	.0345 (.0534)

续表

	(1)	(2)	(3)	(4)	(5)
(每个自变量) 与标准取值的比较 ($D=0.2, X_1=0.5, X_2=0.2, X_3=0.3, X_4=0.5$)					
D	1.21	0.35	0.22	34.03***	78.24***
X_1	0.00	0.00	0.09	6.94**	24.79***
X_2	0.79	0.30	0.42	0.80	0.57
X_3	0.00	0.23	0.00	0.04	0.00
X_4	0.00	0.02	0.06	0.01	0.02
(每个自变量) 与基准模型比较					
D	10.42	0.10	0.26	298.83***	482.31***
X_1	0.00	0.00	0.70	197.08***	310.18***
X_2	0.10	1.16	0.91	0.12	0.04
X_3	0.02	2.13	0.27	0.12	0.22
X_4	1.03	0.21	0.54	0.52	0.10

注一：(1) 表中与标准取值比较时均为 F 统计量，与基准模型比较时均为卡方统计量；

(2) 括号内为标准误；

(3) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ 。

注二：判断条件：

(1) 随机缺失： $D=1$ 且随机数大于等于 0.6；

(2) 顶部缺失： $D=1$ 且 X_1 取值大于其均值加 1 个标准差；

(3) 底部缺失： $D=1$ 且 X_1 取值小于其均值减去 1 个标准差；

(4) 局部缺失： $D=1$ 且随机数大于等于 0.5 且误差项 e 小于 0；

(5) 结构性变化： $D=1$ 且各区间内缺失比例不同且误差项 e 小于 0。

表6 连续自变量各种缺失情况下的结果：控制组

连续自变量 X_1 ： $D = 0$ 控制组

	(1)	(2)	(3)	(4)	(5)
	随机	顶部	底部	局部	结构
D	0.217***	0.217***	0.211***	0.127***	0.108***
	(0.024)	(0.023)	(0.022)	(0.022)	(0.021)
X_1	0.515***	0.494***	0.487***	0.382***	0.311***
	(0.031)	(0.032)	(0.032)	(0.028)	(0.029)
X_2	0.202***	0.201***	0.201***	0.202***	0.201***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
X_3	0.289***	0.307***	0.300***	0.303***	0.304***
	(0.023)	(0.021)	(0.021)	(0.020)	(0.020)
X_4	0.505***	0.470***	0.493***	0.520***	0.491***
	(0.046)	(0.041)	(0.041)	(0.039)	(0.039)
_cons	-0.037	0.012	0.002	0.149**	0.263***
	(0.064)	(0.057)	(0.059)	(0.055)	(0.056)
N	7716	9047	9085	9294	9178
R^2	0.774	0.759	0.760	0.761	0.763
ATT	-.1769 (.0628)	.0776 (.0845)	-.2526 (.0627)	-.1927 (.0585)	-.2547 (.0498)

续表

	(1)	(2)	(3)	(4)	(5)
(每个自变量) 与标准取值的比较 ($D=0.2, X_1=0.5, X_2=0.2, X_3=0.3, X_4=0.5$)					
D	0.51	0.50	0.26	11.62***	18.28***
X_1	0.24	0.04	0.17	17.73***	43.59***
X_2	0.70	0.17	0.61	1.71	0.50
X_3	0.23	0.10	0.00	0.02	0.04
X_4	0.01	0.54	0.03	0.25	0.05
(每个自变量) 与基准模型的比较					
D	0.06	0.05	0.93	422.21***	492.26***
X_1	1.09	0.14	0.64	302.96***	399.57***
X_2	0.05	1.51	0.00	3.10	0.10
X_3	1.52	0.40	0.15	0.00	0.05
X_4	0.23	2.13	0.00	4.03	0.01

注一：(1) 表中与标准取值比较时均为 F 统计量，与基准模型比较时均为卡方统计量；

(2) 括号内为标准误；

(3) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ 。

注二：判断条件：

(1) 随机缺失： $D=0$ 且随机数大于等于0.6；

(2) 顶部缺失： $D=0$ 且 X_1 取值大于其均值加1个标准差；

(3) 底部缺失： $D=0$ 且 X_1 取值小于其均值减去1个标准差；

(4) 局部缺失： $D=0$ 且 X_1 取值处于均值和均值减去1.5个标准差的区间内且随机数小于等于0.5；

(5) 结构性变化： $D=0$ 且各区间内缺失比例不同且误差项 e 小于0。

表7 分类自变量（不相关）各种缺失情况下的结果：实验组

分类自变量 X_3 （与 X_2 不相关）： $D=1$ 实验组

	(1)	(2)	(3)	(4)	(5)
	0值随机	0值非随机	1值随机缺	1值非随机	X_3 结构性变化
D	0.208***	0.345***	0.218***	0.270***	0.193***
	(0.024)	(0.021)	(0.023)	(0.021)	(0.022)
X_1	0.486***	0.493***	0.496***	0.507***	0.503***
	(0.029)	(0.027)	(0.029)	(0.027)	(0.028)
X_2	0.202***	0.201***	0.202***	0.201***	0.201***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
X_3	0.313***	0.195***	0.304***	0.349***	0.184***
	(0.021)	(0.020)	(0.021)	(0.020)	(0.020)
X_4	0.493***	0.497***	0.489***	0.493***	0.493***
	(0.038)	(0.037)	(0.038)	(0.037)	(0.038)
_cons	-0.022	0.041	-0.012	-0.042	0.041
	(0.055)	(0.052)	(0.055)	(0.053)	(0.053)
N	8974	9692	8954	9851	10000
R^2	0.742	0.758	0.735	0.755	0.748
ATT	-0.1567 (.0660)	-0.0803 (.0602)	-0.1039 (.0643)	-0.0983 (.0579)	-0.2211 (.0586)

续表

	(1)	(2)	(3)	(4)	(5)
(每个自变量) 与标准取值的比较 ($D=0.2, X_1=0.5, X_2=0.2, X_3=0.3, X_4=0.5$)					
D	0.11	47.31***	0.60	10.76***	0.11
X_1	0.25	0.06	0.02	0.07	0.01
X_2	1.82	0.57	0.76	0.65	0.71
X_3	0.37	29.06***	0.04	6.24*	32.74***
X_4	0.03	0.01	0.08	0.03	0.04
(每个自变量) 与基准模型比较					
D	0.42	292.70***	0.17	132.70***	25.22
X_1	2.45	0.60	0.20	1.96	0.74
X_2	2.72	0.12	0.08	0.03	0.01
X_3	1.64	315.13***	0.06	143.54***	43.62***
X_4	0.01	0.59	0.29	0.01	0.01

注一：(1) 表中与标准取值比较时均为 F 统计量，与基准模型比较时均为卡方统计量；

(2) 括号内为标准误；

(3) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ 。

注二：判断条件：

(1) 0 值随机： $D=1$ 且 $X_3=0$ 且随机数大于等于 0.5；

(2) 0 值非随机： $D=1$ 且 $X_3=0$ 且随机误差项 $e < -1$ ；

(3) 1 值随机缺失： $D=1$ 且 $X_3=1$ 且随机数大于等于 0.5；

(4) 1 值非随机缺失： $D=1$ 且 $X_3=1$ 且随机数大于等于 0.5 且随机误差项 $e < -1$ ；

(5) X_3 结构性变化： $D=1$ 的实验组中， X_3 中取值为 1 的比例由 0.51 改为 0.65。

表8 分类自变量（不相关）各种缺失情况下的结果：控制组

分类自变量 X_3 （与 X_2 不相关）： $D=0$ 控制组

	(1)	(2)	(3)	(4)	(5)
	0值随机	0值非随机	1值随机	1值非随机	X_3 结构性变化
D	0.225***	0.081***	0.213***	0.147***	0.239***
	(0.023)	(0.021)	(0.023)	(0.021)	(0.022)
X_1	0.496***	0.497***	0.504***	0.494***	0.502***
	(0.030)	(0.027)	(0.030)	(0.027)	(0.028)
X_2	0.201***	0.201***	0.203***	0.201***	0.201***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
X_3	0.315***	0.146***	0.306***	0.381***	0.120***
	(0.022)	(0.019)	(0.022)	(0.020)	(0.021)
X_4	0.482***	0.514***	0.546***	0.482***	0.493***
	(0.043)	(0.037)	(0.042)	(0.037)	(0.038)
_cons	-0.015	0.183***	-0.077	0.031	0.053
	(0.061)	(0.052)	(0.058)	(0.053)	(0.054)
N	8482	9534	8563	9775	10000
R^2	0.767	0.773	0.768	0.764	0.747
ATT	-.2030 (.0612)	-.3231 (.0574)	-.2121 (.0624)	-.2534 (.0589)	-.1855 (.0581)

续表

	(1)	(2)	(3)	(4)	(5)
	(每个自变量)与标准取值的比较 ($D=0.2, X_1=0.5, X_2=0.2, X_3=0.3, X_4=0.5$)				
D	1.19	33.12***	0.32	6.32*	3.17
X_1	0.02	0.02	0.02	0.05	0.01
X_2	0.23	0.8	2.78	0.26	0.72
X_3	0.49	63.28***	0.07	17.06***	77.23***
X_4	0.17	0.15	1.16	0.23	0.04
	(每个自变量)与基准模型比较				
D	1.52	412.50***	0.02	190.71***	34.89***
X_1	0.11	0.12	0.13	0.63	0.42
X_2	0.53	0.02	5.82*	1.58	0.03
X_3	1.82	479.79***	0.11	217.54***	71.46***
X_4	0.25	1.86	7.12**	0.69	0.00

注一：(1)表中与标准取值比较时均为 F 统计量，与基准模型比较时均为卡方统计量；

(2)括号内为标准误；

(3) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ 。

注二：判断条件：

(1) 0值随机： $D=1$ 且 $X_3=0$ 且随机数大于等于0.5；

(2) 0值非随机： $D=1$ 且 $X_3=0$ 且随机误差项 $e < -1$ ；

(3) 1值随机缺失： $D=1$ 且 $X_3=1$ 且随机数大于等于0.5；

(4) 1值非随机缺失： $D=1$ 且 $X_3=1$ 且随机数大于等于0.5且随机误差项 $e < -1$ ；

(5) X_3 结构性变化： $D=1$ 的实验组中， X_3 中取值为1的比例由0.51改为0.65。

(四) 分类自变量结构性变化

实验之初, 本文就设定了两种分类自变量: 一种是与其他自变量相独立的 (X_3), 另一种是与其他自变量相关的分类变量 X_4 。因此, 结果也分为两类。同时, 针对分类自变量的结构性变化, 本文设计的五种缺失情况分别为: 0值随机、0值非随机、1值随机、1值非随机、以及 X_3 (或 X_4) 的结构性变化 (即使整体分布的均值发生改变, 类似于样本中两类人群的比例发生变化, 如性别或民族等)。同样地, 缺失情境分别出现在控制组和实验组上。

先来看独立的分类自变量 X_3 的情况 (表7和表8)。首先, 不论是0值随机还是1值随机的变化, 也不论是发生在控制组还是实验组, 只要是随机缺失就不会导致结构性偏差, 这与连续变量的情况是一致的。其次, 在非随机缺失的条件下, 不论是实验组还是控制组, 0值与1值的非随机缺失都会导致该自变量本身的估计结果与两种标准间存在显著差异。如表7所示, 当实验组的0值非随机缺失时, X_3 系数与标准取值相比, F 统计量为 29.06; 与基准模型相比, 卡方统计量为 315.13; 即实验组的0值非随机缺失会导致自变量自身的估计有偏。再次, 在非随机缺失的条件下, 即便是独立的分类自变量, 其结构性变化同样可能会导致实验效应估计量的变化, 且与两种标准都存在显著差异。如表7所示, 在实验组的0值非随机缺失的条件下, 实验效应的估计量与标准取值间的差异所对应的 F 统计量为 47.31; 与基准模型的差异所对应的卡方统计量为 292.70, 均远小于 0.01 水平。可见, 即便是独立的分类自变量, 其结构性变化不仅会影响到自身系数, 而且也会影响到实验效应的准确估计, 这也就是结构性偏差的重要体现。从次, 保持其他条件不

变，只将 X_3 的比例结构替代转换（如调查样本中民族或性别等变量结构发生整体性变化），则可以发现，这种结构性变化不仅影响其自身系数，而且估计的实验效应虽然与标准取值之间的差异并不显著，但与基准模型之间存在着显著差异（表7或表8的最后一列）。最后，再来考察PSM对实验效应估计结果的校正，不论是实验组还是控制组，在0值非随机缺失的情况下，PSM估计得到的平均实验效应与两种标准都有着显著差异，而在其他情况下估计得到的平均实验效应与两种标准的差异并不显著。因此，从这个意义上讲，倾向得分匹配方法（或者是高级统计方法）尽管可以解决部分问题，却无法完全解决可能存在的某些样本结构性偏差问题。由此，统计方法并不是万能的，其基础仍然是样本结构。

再来看与其他自变量相关的分类变量（及相关的连续变量）的情况（表9和表10）。除了与 X_3 独立分类变量的情况相同以外，与其他自变量相关的分类变量的样本结构性变化所带来的最大区别在于：当其结构发生变化时，不仅会使其自身的系数和实验效应发生显著变化，而且在许多情况下也会使与其相关的连续变量的系数发生显著变化，不论是与标准取值还是与基准模型进行比较。由此可以看到，变量间联合分布的改变会导致联合性的结构性偏差。而现实情况中，分析样本中的各变量间都或大或小地存在一定的相关性，并形成联合分布；当某一自变量发生结构性变化时，必然会带来联合的结构性偏差。这一点在通常情况下却是极其容易被忽略的。

表9 分类自变量（与其他变量相关）各种缺失情况下的结果：实验组
分类自变量 X_4 （与 X_2 相关）： $D=1$ 实验组

	(1)	(2)	(3)	(4)	(5)
	0值随机	0值非随机	1值随机	1值非随机	X_4 结构性变化
D	0.221***	0.271***	0.208***	0.322***	0.170***
	(0.023)	(0.021)	(0.025)	(0.021)	(0.021)
X_1	0.495***	0.497***	0.513***	0.500***	0.503***
	(0.028)	(0.027)	(0.030)	(0.027)	(0.028)
X_2	0.202***	0.206***	0.201***	0.199***	0.189***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
X_3	0.307***	0.306***	0.289***	0.304***	0.301***
	(0.020)	(0.020)	(0.022)	(0.020)	(0.020)
X_4	0.487***	0.483***	0.490***	0.506***	0.256***
	(0.038)	(0.037)	(0.038)	(0.037)	(0.029)
_cons	-0.012	-0.066	-0.015	0.001	0.326***
	(0.055)	(0.053)	(0.055)	(0.052)	(0.040)
N	9681	9902	8307	9715	10000
R^2	0.702	0.752	0.766	0.760	0.750
ATT	-.2254 (.06998)	-.2116 (.0605)	.1176 (.0727)	-.0521 (.0587)	-.1855 (.0581)

续表

	(1)	(2)	(3)	(4)	(5)
	(每个自变量)与标准取值的比较 ($D=0.2, X_1=0.5, X_2=0.2, X_3=0.3, X_4=0.5$)				
D	0.87	11.16***	0.10	33.78***	2.01
X_1	0.04	0.01	0.18	0.00	0.01
X_2	0.83	13.75***	0.61	0.23	91.91***
X_3	0.11	0.10	0.25	0.05	0.00
X_4	0.12	0.21	0.07	0.02	72.78***
	(每个自变量)与基准模型的比较				
D	0.80	99.54***	0.28	260.30***	107.15***
X_1	1.07	0.34	1.12	0.00	1.47
X_2	0.13	71.22***	0.01	57.76***	164.24***
X_3	1.42	1.33	2.28	0.11	1.16
X_4	0.70	2.90	0.17	8.97**	79.84***

注一：(1) 表中与标准取值比较时均为 F 统计量，与基准模型比较时均为卡方统计量；

(2) 括号内为标准误；

(3) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ 。

注二：判断条件：

(1) 0值随机： $D=1$ 且 $X_4=0$ 且随机数大于等于0.5；

(2) 0值非随机： $D=1$ 且 $X_4=0$ 且随机误差项 $e < -1$ ；

(3) 1值随机缺失： $D=1$ 且 $X_4=1$ 且随机数大于等于0.5；

(4) 1值非随机缺失： $D=1$ 且 $X_4=1$ 且随机数大于等于0.5且随机误差项 $e < -1$ ；

(5) X_3 结构性变化： $D=1$ 的实验组中， X_4 中取值为1的比例由0.85改为0.90。

表 10 分类自变量（与其他变量相关）各种缺失情况下的结果：控制组
分类自变量 X_4 （与 X_2 相关）： $D=0$ 控制组

	(1)	(2)	(3)	(4)	(5)
	0 随机	0 非随机	1 随机	1 非随机	X_4 结构性变化
D	0.210***	0.161***	0.199***	0.117***	0.190***
	(0.022)	(0.021)	(0.024)	(0.021)	(0.022)
X_1	0.513***	0.500***	0.493***	0.496***	0.500***
	(0.028)	(0.027)	(0.031)	(0.027)	(0.028)
X_2	0.202***	0.197***	0.201***	0.204***	0.186***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
X_3	0.308***	0.308***	0.296***	0.300***	0.302***
	(0.020)	(0.020)	(0.023)	(0.020)	(0.020)
X_4	0.491***	0.260***	0.508***	0.614***	0.013
	(0.045)	(0.039)	(0.042)	(0.037)	(0.035)
_cons	-0.027	0.274***	-0.004	-0.036	0.555***
	(0.062)	(0.054)	(0.057)	(0.052)	(0.049)
N	9403	9832	7643	9647	10000
R^2	0.749	0.763	0.786	0.768	0.748
ATT	-.0765 (.0567)	-.2514 (.0573)	-.2715 (.0690)	-.2750 (.0574)	.0273 (.0553)

续表

	(1)	(2)	(3)	(4)	(5)
	(每个自变量) 与标准取值的比较 ($D=0.2, X_1=0.5, X_2=0.2, X_3=0.3, X_4=0.5$)				
D	0.21	3.50	0.00	15.87***	0.22
X_1	0.19	0.00	0.05	0.03	0.00
X_2	1.05	3.96*	0.70	5.43	117.13***
X_3	0.16	0.18	0.03	0.00	0.01
X_4	0.04	38.69***	0.04	9.60**	189.40***
	(每个自变量) 与基准模型的比较				
D	0.52	145.48***	1.61	308.49***	22.32***
X_1	3.38	0.00	0.20	0.26	0.00
X_2	0.47	72.83***	0.03	33.87***	127.08***
X_3	1.31	1.66	0.35	0.11	0.01
X_4	0.00	148.65***	0.69	171.73***	88.15***

注一：(1) 表中与标准取值比较时均为 F 统计量，与基准模型比较时均为卡方统计量；
 (2) 括号内为标准误；
 (3) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ 。

注二：判断条件：

- (1) 0值随机： $D=1$ 且 $X_4=0$ 且随机数大于等于 0.5；
- (2) 0值非随机： $D=1$ 且 $X_4=0$ 且随机误差项 $e < -1$ ；
- (3) 1值随机缺失： $D=1$ 且 $X_4=1$ 且随机数大于等于 0.5；
- (4) 1值非随机缺失： $D=1$ 且 $X_4=1$ 且随机数大于等于 0.5 且随机误差项 $e < -1$ ；
- (5) X_3 结构性变化： $D=1$ 的实验组中， X_4 中取值为 1 的比例由 0.85 改为 0.90。

(五) 简单小结

上述几组实验结果表明：第一，在多数随机缺失的条件下，分析结果与真实效应及各系数均不会产生显著的结构性偏差；但在非随机缺失的条件下，多数情况下，分析结果与标准取值和基准模型的结果之间都会存在显著差异。第二，独立自变量的结构性变化不仅会影响到其自身的估计结果，而且也会影响到实验效应的估计量；而不独立的自变量的结构性变化，既有独立变量的结构性偏差特征（即自身的估计量与平均实验效应都可能偏），而且还会影响到相关变量的估计结果，这也就是联合分布的结构性偏差。第三，高级统计方法在许多情况下无法在根本上解决样本的结构性偏差，PSM结果或是低估或是高估实验效应，且与标准取值和基准模型均存在显著差异。第四，在对样本进行事后加权处理时，应该是对分析中所涉及变量的联合分布的调整，而非对某一个或几个变量的加权调整，否则会由于变量联合分布的改变而导致其他变量的结构性偏差，即样本加权调整应该是对联合分布的调整，而非对单个变量的加权调整。

六、对模拟结果的统计分析

为体现样本结构性变化更多的可能性，本文在上述基础上进一步改变缺失案例的范围，即按照上述五种缺失方式，随机产生缺失值域和不同的样本。共进行100次随机化处理实验，最后将这100次实验的结果作为分析样本。分析数据包括了缺失类型、案例数、自变量的估计系数及其标准误、与标准取值比较的 F 统计量及显著性、与基准模型比较的卡方统计量及显著性，以及平均实验效应。

其中，缺失类型中又标记了缺失变量（如因变量或 X_1 到 X_4 ）、实验组或控制组的缺失等基本信息，基本上全面记录了每次实验的具体结果。以下是对这100次随机实验结果的统计分析。由于所得结果较多，本文仅呈现部分统计结果。

（一）系数估计量

在多个系数中，本文更关注实验 D 的系数在各种不同缺失条件下的估计量。因此文中仅给出变量 D 在不同情况下的分布结果（图5至图9）。

图5描绘了因变量有缺失时的情况。以因变量实验组的五种缺失情况（图5中的左图）为例，在纯随机缺失条件下（以小圆圈表示），变量 D 的系数基本上都围绕着0.2小幅波动，而因变量的顶部缺失模型中变量 D 的系数基本上围绕着0.1小幅波动，这说明顶部缺失会导致实验效应的一致性被低估。底部缺失模型中，变量 D 的系数会在0.2~0.3范围变化，即实验线的底部缺失会导致一致性地低估实验效应。由于本文中的局部缺失和结构性缺失的设计是剔除均值以下的部分案例，其中规定：局部缺失是小于均值的某个区间内误差项小于0的案例按随机数（大于0.5）的缺失；结构性缺失则将缺失范围扩大到均值以下，并将该值域分割成若干段（100段）更小的区间，规定每段小区间内误差项小于0的案例按随机数（大于0.5）表示的比例（概率）缺失。由图中的分布可以发现，结构模型中围绕着0.28左右小幅波动；但局部模型中变量 D 的系数基本上在0.3以上，且波动幅度较大。这从某种意义上说，局部缺失对实验效应估计量的影响远大于结构模型。总体来看，实验组的顶部缺失会导致实验效应的低估，而底部（包括均值以下的局部和结

构) 都会导致实验效应的高估。这一点类似于图3中的解释。

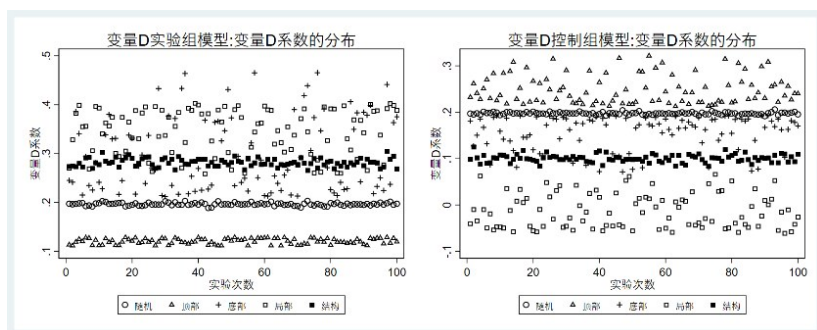


图5 因变量不同缺失条件下实验效应 (D) 线性估计结果的分布

从图5右侧的因变量控制组缺失模型来看, 随机缺失模型的结果仍然相对稳定, 但在控制组顶部缺失模型中, 变量 D 的系数都在标准取值0.2以上, 且波动范围相对较大; 底部模型中的 D 系数则都在标准取值0.2以下, 且波动范围也相对较大(0.08~0.19); 局部缺失对变量 D 系数的影响会相对更大, 变量 D 的系数不仅波动范围相对较大, 甚至会使系数发生方向性变化(有部分散点落在0以下), 即实验效应变成负向。而结构性变化使变量 D 的系数被严重低估, 但其仍然围绕着0.1小幅波动。这仍然体现了相较于结构性变化, 局部缺失对实验效应估计量的危害更大, 这一点需要引起足够的重视。这种局部缺失相当于在实际调查中缺少一部分特定特征的群体, 如收入研究缺少了部分低收入群体(或高收入群体), 从而使得估计结果既会严重偏离实际情况, 也无法判断其偏离的程度及一致性(我们原来总是认为结构偏离可能会是更严重的, 但在随机实验的参数设定中, 结构偏离的设定是将小于均值部分按非等概率缺失。这种非等概率的随机缺失, 也会呈现出一定的一致系统的偏差。这是出乎意料的, 但却极其重要!))。

上述是因变量在不同缺失条件下对实验效应估计量的影响作用，下面将讨论自变量的结构偏离对实验效应估计量的影响，具体如图6至图9所示。由于自变量被分为连续变量和分类变量两种类型，因此，以下讨论也将分成两类。

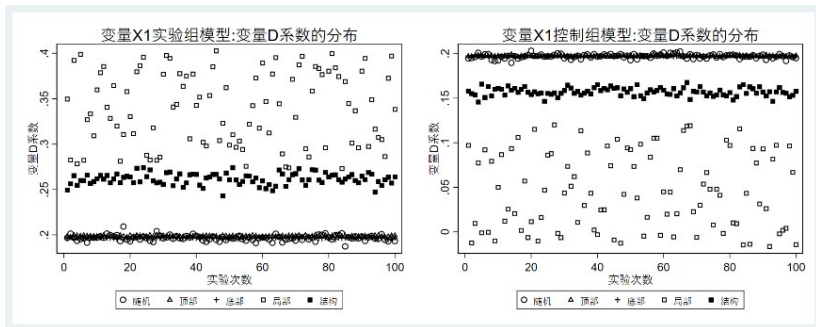


图6 X_1 不同缺失条件下实验效应 (D) 线性估计结果的分布

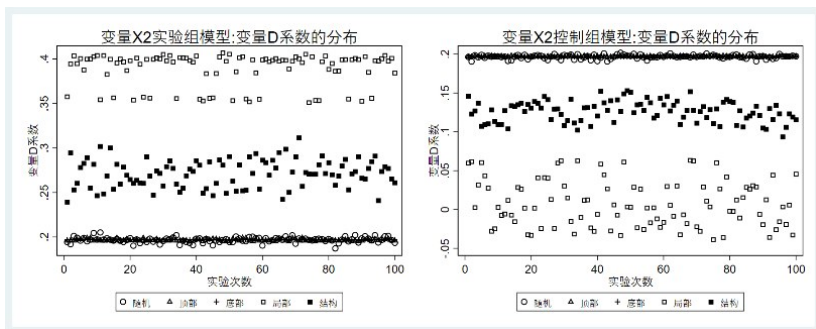


图7 X_2 不同缺失条件下实验效应 (D) 线性估计结果的分布

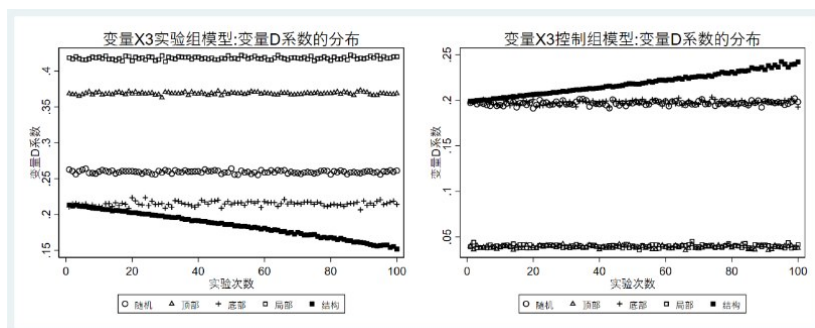


图8 X_3 不同缺失条件下实验效应 (D) 线性估计结果的分布

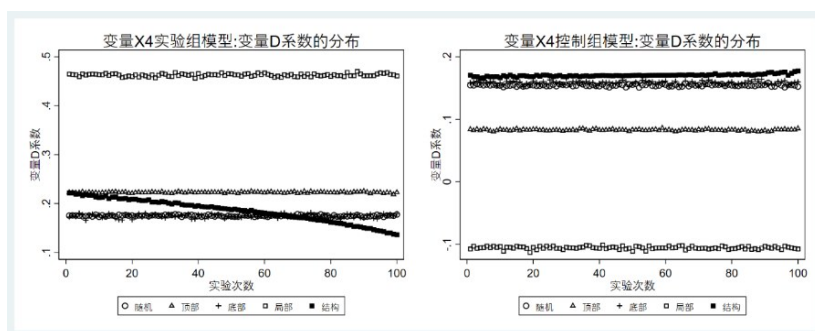


图9 X_4 不同缺失条件下实验效应 (D) 线性估计结果的分布

从连续变量的角度看, 以 X_1 为例, X_1 的随机、顶部和底部模型对实验效应 D 的估计并没有太大的影响, 但是在结构模型中, 实验效应 D 的估计却会发生系统性的偏差; 局部模型中, 实验效应 D 的估计量偏差相对更大, 且波动幅度也更大。同样地, 当实验组中 X_1 小于均值部分呈局部缺失或结构性缺失都会使实验效应 D 被高估; 而控制组的情况正好相反。在控制组的 X_1 呈局部缺失时, 由于波动范围过大而使实验效应 D 的方向都有可能发生变化。连续变量 X_2 的作用与 X_1 基本相同。在此不再赘述。

分类自变量的缺失有0值随机缺失、0值非随机缺失、1值随机缺失、1值非随机缺失和结构性变化（即总体均值改变）五种情况，具体结果如图8和图9所示。从 X_3 来看，实验组0值随机缺失时，会导致实验变量 D 的系数被一致性地高估；而1值随机缺失时，实验变量 D 的系数却基本围绕着标准取值小幅波动。但 X_3 的0值与1值均为非随机缺失时，实验变量 D 的系数被较大幅度地高估，0值非随机时，该系数基本上在3.6左右波动，而1值缺失时，该系数则在4.1左右波动。由此可见，非随机缺失情况会导致系数的大幅高估，而随机缺失情况还需要结合 X_3 的分布情况来确定这种高估的幅度。

但在控制组中， X_3 的缺失却呈现出另一种情况。当 X_3 是随机缺失时（不论 X_3 的取值是0还是1），实验变量 D 的系数仍然是围绕着标准取值小幅波动；当 X_3 是非随机缺失时，实验变量 D 的系数将被严重低估，结果具有一致性，都在0.3附近小幅波动。由此可以看到， X_3 的结构性变化对实验效应的影响受到具体实验组或控制组的影响，而不仅仅与是否随机缺失有关。

最后来看 X_3 的结构性变化。实验中设定 X_3 的总体比例由0.45逐步提高到0.75。这时，在实验组中，实验变量 D 的系数随比例的提高而逐步下降；相反，在控制组中，实验变量 D 的系数随比例的提高而逐步提高。即，前者被低估，而后者却被高估。这说明，实验效应 D 会随着二分变量的结构性变化而出现样本结构性偏差。样本偏离总体越严重，实验效应估计量的偏离也会越严重。而实验组与控制组的偏离方向相反则说明两组样本各自的结构变化对实验效应估计量的影响作用正好是相反的。

分类变量 X_4 的变化与 X_3 的表现有些不同。首先，不论是实验组

还是控制组，也不论 X_4 是0值还是1值的缺失，只要是随机的，都会导致实验效应 D 被小幅低估。其次，在实验组中， X_4 的0值非随机缺失对实验效应 D 会产生较为一致的小幅高估；但 X_4 的1值的非随机缺失会导致实验效应 D 的严重高估。在控制组缺失条件下，与实验组的情况正好相反，0值或1值的缺失将会低估实验效应。再次，实验组中 X_4 的结构性变化，会使实验效应 D 的系数的被略微高估，以非线性方式下降逐步变成被低估；系数与标准取值的偏差幅度随结构偏离程度的增加而提升。但在控制组中，这种结构性的偏离尽管会略微低估实验效应 D 的系数，但其结果的波动幅度相对较小。

上述结果是对实验效应变量系数分布的讨论，可以看到，如果缺失是随机的，那么，对于实验效应的估计不会有太大的偏差；但如果是非随机的，则实验效应估计量必然存在或多或少的偏差。从现实角度来看，我们很难判断这种结构偏离的存在及其来源，也更难判断这种结构性偏差到底有多大。但，现实中分析样本与总体之间的差异还是有可能反映样本结构偏离情况的。例如，2020年“七普”中汉族人口占全国总人口的91.11%，少数民族人口占全国的8.89%。如果某个调查的民族构成与全国总人口中的比例存在显著差异（如汉族人口仅为80%），那么在加入民族变量以后就有可能导致比较严重的样本结构性偏差。按照上述结果，偏离总体越远，则结构性偏差就可能越大。另外，从流动人口构成来看，2020年“七普”中的流动人口构成为26.62%。那么，在相同定义之下，调查得到的分析样本中流动人口比例与总体比例之间的差异如果太大，那么，分析结果不仅会影响到流动人口这一变量自身的系数，而且也会影响到实验效应的估计。而这只是某变量的独立分布情况，还未涉及联合分布的情况。

除了实验效应估计量以外，本文还涉及与其他自变量相关的自变量系数估计的问题。在本例中， X_2 是依据 X_4 而构建的，因此， X_2 与 X_4 相关。 X_4 的结构性变动对 X_2 系数的影响作用如图10所示。首先，在实验组中， X_4 变量的各种不同缺失都会在一定程度上低估 X_2 的作用（ X_2 的标准取值也是0.2），其中，低估量最小的是0值非随机缺失，其次是结构性改变；偏差相对较大的是0值或1值的随机缺失，且这两种随机缺失得到的结果基本重合；偏差最大的是1值的随机缺失。其次，在控制组存在缺失的条件下， X_4 的随机缺失结果与其标准取值也基本相同，但0值非随机缺失会低估 X_2 的系数，而1值的非随机缺失会使 X_2 的系数被高估。这也正是实验组与控制组相反的地方。再次，实验组的结构性变化随着实验次数的增加（即实验组所对应的 X_4 的均值比例[1值所占比例]的提高）而呈现出逐步向标准取值靠拢的趋势；控制组的结构性变化在各控制组 X_4 均值变化过程中基本不变，但所引起的 X_2 系数的偏差却是最大的，且严重低估了 X_2 的作用。由此，在强调对实验效应无偏估计的同时，还应考察与缺失变量相关的其他自变量的系数估计量问题。这也就是联合分布偏差效应。

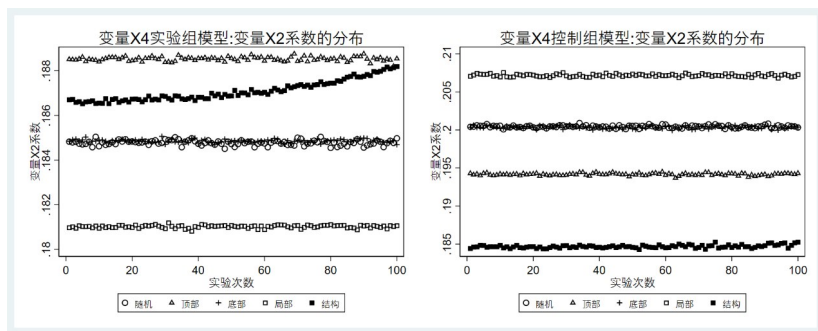


图10 X_4 缺失条件下，相关自变量的系数变化情况

(二) 估计量与标准取值和基准模型的差异

除了上述估计量与标准取值之间的数值比较以外，还需要统计检验系数估计量与标准取值和基准模型间的差异。由于所得结果相对较多且繁杂，本文将使用在 100 次随机实验中两种统计检验显著的次数来表示，具体结果如表 11 所示。

表 11 各种缺失条件下系数与标准取值、基准模型间差异显著的次数

		与标准取值的比较					与基准模型的比较				
		随机	顶部	底部	局部	结构	随机	顶部	底部	局部	结构
	YD	sigF	sigF	sigF	sigF	sigF	sigC	sigC	sigC	sigC	sigC
控制组	D	0	98	81	100	100	8	100	100	100	100
	X_1	1	22	18	15	0	6	100	100	17	70
	X_2	0	100	54	2	35	4	100	100	91	100
	X_3	3	0	4	10	0	4	100	99	11	35
	X_4	0	43	0	15	0	8	100	100	66	43
实验组	D	0	100	98	100	100	2	100	100	100	100
	X_1	2	0	40	90	0	2	1	100	91	6
	X_2	0	100	83	71	100	5	100	100	95	100
	X_3	2	0	39	66	0	4	0	100	66	15
	X_4	0	0	53	19	0	5	61	100	80	30
	X_1	sigF	sigF	sigF	sigF	sigF	sigC	sigC	sigC	sigC	sigC

续表

		与标准取值的比较					与基准模型的比较				
控制组	D	0	0	0	100	100	2	0	0	100	100
	X_1	1	0	0	100	100	1	0	0	100	100
	X_2	0	0	0	0	0	5	0	73	6	7
	X_3	3	0	0	0	1	2	0	0	0	6
	X_4	0	0	0	0	0	6	0	36	7	3
实验组	D	0	0	0	100	100	6	3	0	100	100
	X_1	4	79	0	100	100	6	31	0	100	100
	X_2	0	0	0	0	0	2	0	0	4	6
	X_3	5	0	0	0	0	8	0	13	0	1
	X_4	0	0	0	0	0	2	0	0	0	3
控制组	X_2	sigF	sigF	sigF	sigF	sigF	sigC	sigC	sigC	sigC	sigC
	D	0	0	0	100	100	3	0	0	100	100
	X_1	3	0	0	3	0	2	0	0	2	1
	X_2	0	0	0	99	48	6	0	0	100	100
	X_3	2	0	0	7	1	3	26	0	4	7
	X_4	0	0	0	100	38	4	0	0	100	100

续表

		与标准取值的比较					与基准模型的比较				
实验组	D	0	0	0	100	100	5	0	0	100	100
	X ₁	4	0	0	3	0	7	0	0	3	0
	X ₂	0	0	0	100	95	3	0	0	100	100
	X ₃	4	0	0	0	0	3	14	0	6	4
	X ₄	0	0	0	0	1	7	0	0	46	89
		0值 随机	0值 非 随机	1值 随 机	1值 非 随 机	结 构 变 化	0值 随 机	0值 非 随 机	1值 随 机	1值 非 随 机	结 构 变 化
控制组	X ₃	sigF	sigF	sigF	sigF	sigF	sigC	sigC	sigC	sigC	sigC
	D	0	100	0	100	61	5	100	3	100	100
	X ₁	5	5	0	0	0	5	2	9	2	5
	X ₂	0	0	0	0	0	7	1	4	3	46
	X ₃	1	100	2	100	100	5	100	4	100	100
	X ₄	0	0	0	0	0	2	2	3	2	89
实验组	D	100	100	43	100	53	100	100	100	100	97
	X ₁	0	0	4	4	0	9	13	7	7	48
	X ₂	0	0	0	0	0	0	0	16	22	14
	X ₃	100	100	100	93	100	100	100	100	100	100
	X ₄	0	0	0	0	0	31	10	15	13	0

续表

	X_4	与标准取值的比较					与基准模型的比较				
		0值 随机	0值 非 随机	1值 随 机	1值 非 随 机	结 构 变 化	0值 随 机	0值 非 随 机	1值 随 机	1值 非 随 机	结 构 变 化
控制 组	D	100	100	100	100	100	100	100	100	100	100
	X_1	0	0	0	1	0	5	0	0	0	0
	X_2	0	100	0	100	100	9	100	0	100	100
	X_3	100	100	100	100	0	100	100	100	100	0
	X_4	0	100	0	100	100	4	100	0	100	100
实 验 组	D	100	100	100	100	57	100	100	100	100	92
	X_1	0	0	3	4	0	0	1	1	0	0
	X_2	100	100	100	100	100	100	100	100	100	100
	X_3	6	8	7	2	0	46	66	5	0	0
	X_4	100	100	100	100	100	100	100	100	100	100

1. 因变量控制组缺失、与标准取值的比较。除随机缺失以外，其他四种缺失模型中的实验D系数与标准取值的差异基本上都统计显著。例如，顶部缺失时有98次实验显著；底部模型中有81次显著；而在局部和结构模型中，所有实验都呈显著。同时还应该注意到，因变量控制组的顶部缺失在导致实验效应显著以外，还会导致 X_2 和 X_4 的系数发生变化：所有实验中 X_2 的系数与标准取值都有显

著差异，且43%的实验中 X_4 的系数与标准取值会有显著差异，但其对 X_1 和 X_3 的影响相对较小，特别是在结构模型中。这表明因变量的缺失不仅会导致实验效应估计量的变化，而且也可能导致其他变量的系数估计量的显著变化。

因变量实验组缺失的情况，与上述结果基本相同，即，既会影响实验效应的估计，也会影响 X_2 的估计结果。而在底部和局部模型中，甚至 X_1 和 X_3 、 X_4 这三个变量都可能与标准取值有显著差异。

2. 因变量控制组缺失、与基准模型的比较。和标准模型的结果十分类似，除随机模型以外，其他四种缺失既会使实验 D 系数全部都显著，而且 X_2 系数估计量在多数情况下也都显著。甚至于在顶部和底部缺失模型中， X_1 和 X_3 、 X_4 这三个变量的系数估计量与基准模型都有显著差异。即使在局部和结构模型中，仍然会分别有66%和43%的 X_4 结果呈现为显著差异。而 X_1 的系数与基准模型有显著差异的比例，在结构模型中达到了70%。

因变量实验组缺失情况下的结果与控制组的基本相同，在此不再赘述。

3. 自变量各种缺失条件下的结果与因变量的缺失略有不同。以连续变量 X_1 为例，不论是实验组还是控制组的缺失， X_1 的缺失主要导致局部和结构性变化模型中实验效应和其自身系数显著不同于两种标准，但对其他变量的影响相对较小。但从连续变量 X_2 来看，由于它和 X_4 相关，不仅会导致实验效应和 X_2 的系数显著不同于两种标准，而且也会导致 X_4 的系数与基准模型有显著差异，特别是在 X_2 的控制组局部模型和结构模型中。

4. 在分类变量的模型中， X_3 与其他变量不相关，因此它的缺失主要导致实验 D 系数和自身系数的显著，对其他变量的影响相对较

小。但由于 X_4 和连续变量 X_2 相关，因此， X_4 的结构性变化不仅会导致实验 D 系数和 X_4 系数的显著差异，而且同样也会导致 X_2 系数的估计量与标准取值和基准模型之间存在显著差异。

这些结果表明，样本结构偏离不仅会影响到实验效应的估计，而且也会影响其他自变量的估计，分析中应更多关注由于联合分布的结构偏离而导致的对其他变量系数估计量的影响问题。

（三）平均实验效应的估计量及差异性

图 11 到图 15 分别呈现了不同情况下利用倾向得分匹配方法计算得到的平均实验效应，但很奇怪的是，所有的结果基本上均呈现出负向的结果。导致这种情况的可能原因是：PSM 的基本要求是第一阶段的倾向值估计模型与第二阶段的匹配模型间至少要有一个重要变量是有区别的；但本实验中仅涉及四个简单变量，且无其他变量可以借鉴以计算倾向得分值。当然，由此也可以看到倾向得分匹配方法自身的局限性，在模型设定有问题的情况下，估计结果可能与真实效应之间存在明显的差异。但前面与基准模型比较则剔除了模型设定问题，可以将两种估计结果看成是控制了模型设定以后的差异。因此，本文将参照标准设定为基准模型的平均实验效应估计量，而不再是标准取值（即真实的实验效应）。

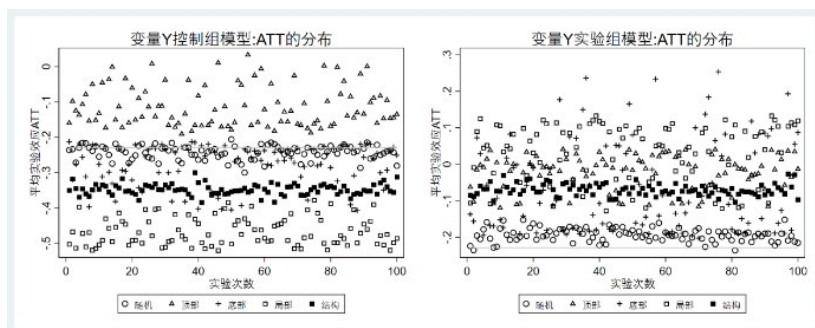


图 11 因变量缺失条件下 ATT 估计量的分布

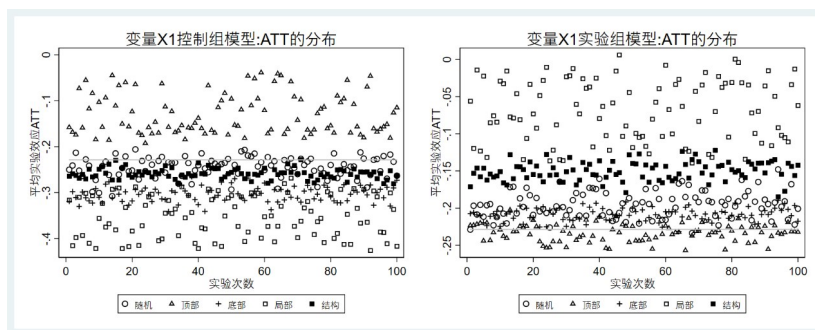


图 12 X_1 缺失条件下 ATT 估计量的分布

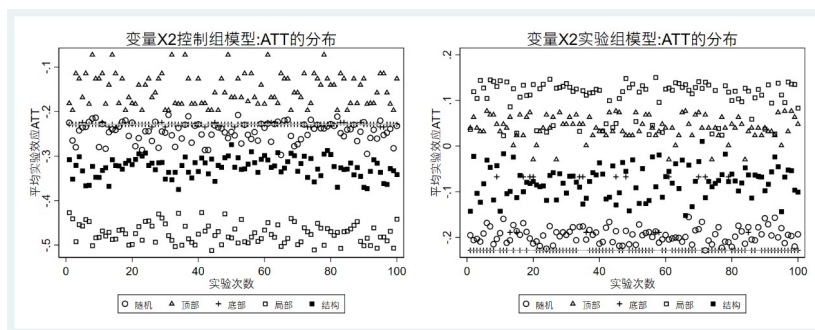


图 13 X_2 缺失条件下 ATT 估计量的分布

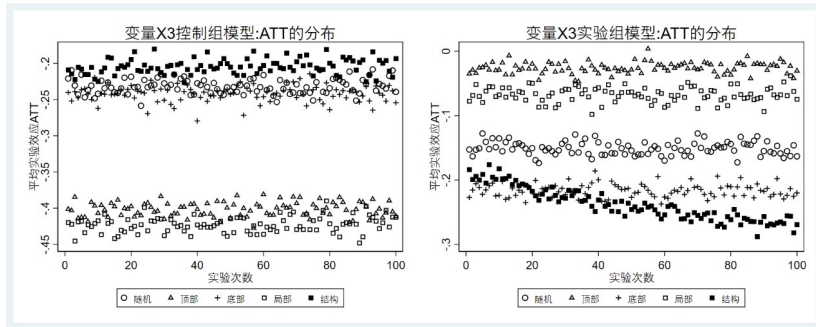


图 14 X_3 缺失条件下 ATT 估计量的分布

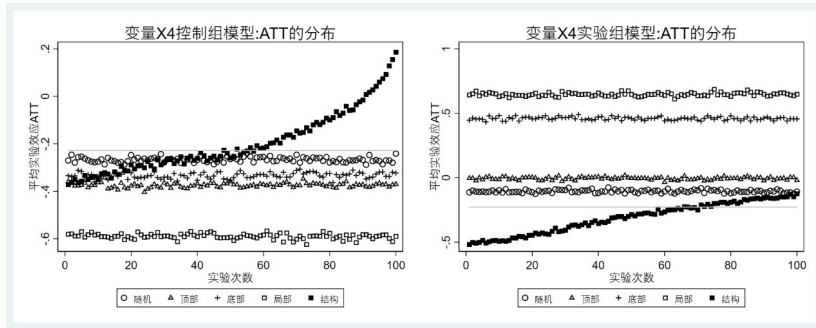


图 15 X_4 缺失条件下 ATT 估计量的分布

由于图示较多，本文仅以因变量缺失条件下的平均实验效应 ATT 估计结果为例（图 11）。在控制组模型中，根据五种因变量的缺失对应的 ATT 估计结果，1) 随机缺失模型仍然表现出相对较好的一致性，其估计的平均实验效应基本上都在基准模型 ATT 的周围小幅波动；2) 控制组顶部缺失时，相当于控制组本身的均值被低估，在实验组不变的情况下，两组之间的均值差有所减小（以绝对值看），因此，控制组的顶部模型低估了平均实验效应；且低估的幅度受顶部缺失值域范围的影响；3) 控制组底部缺失的结果正好

与顶部缺失的结果相反，会显著高估平均实验效应，且波动幅度也相对较大；4) 结构性缺失模型中所得的平均实验效应的波动相对最小，但其与基准模型估计量的差异不仅大于随机模型，也大于底部模型；5) 相对较为严重地高估平均实验效应模型的是局部模型，不仅与基准模型有较大的差异，而且其变动幅度也相对较大，这种局部模型的波动性主要还是源于局部缺失的值域区间。

在实验组中，ATT的估计量几乎都大于基准模型的估计结果。随机缺失模型估计的ATT与基准模型最为接近，但基本上都偏高；而结构模型的结果虽然也偏高，但波动幅度相对较小；在顶部缺失、底部缺失和局部缺失这三种模型中，估计量的变动范围相对较大，而且更重要的是，在某些情况下，平均实验效应估计量的方向与基准模型都是相反的，这说明这三种非随机缺失无法真正准确地估计平均实验效应。

在分类变量 X_4 的不同缺失情况下，平均实验效应的估计结果与基准模型间的差异，与连续变量的表现并不一致。控制组模型中， X_4 的随机缺失会一致性地略微低估实验效应。顶部和底部模型中的ATT估计量偏差相对大于随机模型，而局部模型的低估相对比较严重。当 X_4 发生整体结构性变化时（即 X_4 的均值从最初的0.55，随实验次数逐步上升到0.95，步长为0.004），平均实验效应估计量呈非线性上升（图15）；当 X_4 取最小（0.55）时，平均实验效应被低估的幅度相对更大，且离基准模型的结果也相对更远；随着 X_4 均值的提高，平均实验效应估计量也逐步提高，并逐步趋近于基准模型；但随着 X_4 取值逐步增大，平均实验效应估计量离基本模型的结果也逐渐变远，且呈非线性上升趋势；当 X_4 取最大（0.95）时，两者之间的差异更大，即离假设总体均值越远，估计

量偏差越大。而且应该注意到，平均实验效应估计量最后甚至出现了方向的改变。 X_4 在实验组中的变化与控制组中的变化基本相同，只是结构模型估计得到的平均实验效应并没有发生方向性变化，但估计量与结构之间的关系同样是非线性的。这说明，二分变量的结构性变化，不仅会导致无法准确估计平均实验效应，而且还可能导致估计量的方向性变化。

七、结论与讨论

本文利用仿真实验数据，讨论了结构性偏差及其对因果推断中平均实验效应估计量的影响。分析结果表明：1) 非随机的样本结构性偏差主要表现在三个方面——实验效应的估计量、自身系数的估计量，以及相关（甚至无关）自变量系数的估计量。在因果推论中，不论是模型中哪个变量的样本结构偏离，既会带来有偏的实验效应估计量，也会导致该变量的系数估计量的偏差。2) 样本结构性偏差源自变量的联合分布，而不只是单变量分布的变化，因此，某个变量的结构偏离会导致其他相关变量，甚至不相关变量的估计量有偏差。3) 分类变量的结构性偏差会随着与真实结构的距离增大而呈非线性加速扩大趋势，特别是对实验效应的估计。4) 在某种意义上，统计方法无法从根本上解决这种样本结构性偏差（尽管在某些情况下有助于减少实验效应估计量的偏差）。因此在实证分析过程中，并不能完全迷信于统计方法，而更应该严格检验样本与总体之间的差异量，以使统计结果能够真正反映总体的情况。

本文将样本结构性偏差的影响作用从线性回归扩展到因果推论中平均实验效应的估计，这是本文的主要目标。由上述结论可以看

到，在非随机缺失的情况下，样本结构偏离不仅使线性回归的系数估计量受到影响，而且平均实验效应估计量也会发生一定的偏差。特别是，如果二分变量存在结构偏离，那么由此引起的结构性偏差随着与“真值”的距离越远而变得越大，这一点很有现实意义。例如，某调查数据的年龄性别等基础结构可能在加权调整后与总体并不存在太大的差异，但民族结构（假设以汉族和少数民族为分类）却存在较大的差异。如果分析不包括民族结构，则民族变量的结构偏离不一定会对估计量产生太大的影响；但如果分析纳入了民族变量，则不仅民族变量的系数不再是无偏一致的，而且它还会使其他相关变量的系数存在偏差，更会影响到包含民族变量的因果推论的结果，且偏离越远，偏差越大。而这种偏差在目前的各种抽样调查样本中却是普遍存在的，但却又是被数据使用者和分析者们或有意或无意地忽略的。许多研究更关心分析结果与分析方法，而并不关心分析样本的数据来源及其结构问题。根本上讲，已有的许多定量研究的结果是值得商榷、有待检验的。因此，提请定量研究人员能够更加关注数据来源与样本结构，以期能够更贴现实地从总体角度描绘社会现实，解释社会现象，回答社会问题，这既是本文的目的之一，更应是研究人员思考与追求的目标之一。

同时，我们更不能完全迷信于高级统计方法。统计方法只是分析工具，包括回归分析和PSM在内的各种统计方法都应该服务于理论（检验）。之所以使用各种统计方法，不是说某些所谓的高级统计方法所获得的结果必然会优于线性回归（有时有可能正好相反），而只是为解决一般线性回归中可能存在的问题（如无偏有效一致性或分析单位或时间单位等问题）。但不论哪种统计方法都脱离不了工具属性，都会有其适用性与局限性。目前的定量研究普遍存在过

度关注统计方法，却相对忽略了工具的对象与基础的问题。样本代表性（结构性偏差）是统计分析中重要的基础性问题，是定量研究结果可信性与可靠性的基础，是后续理论检验与政策建议的重要基础。注重分析样本的代表性问题，既可以解决或缓解结构性偏差，也可以为更加正确地利用各种统计方法讨论真实的因果关系提供坚实的基础。更重要的是，统计方法并不是万能的，其并不一定能完全解决样本结构性偏差问题。因此，将样本与统计方法并举，争取做到“使用适合的方法分析（相对）准确的数据，以得到更加接近现实的结果”才应该是社会科学定量研究的目标。

回到本文开篇，我们使用统计学所用的因果推断，仅仅是出于因果推断的工具化和简单化，这也是最狭隘意义上的“因果”关系。这种工具性的因果关系可以直接观察，或者用思维工具推导出。当然，这一过程有赖于经验材料的真实性和完备性，在思考过程中也暂时性地忽略了实然与应然之间的关系。而经验论和理性论所讨论的因果关系，事实上也是大的因果律，这其中自然包含了统计学中的局部性使用的因果推断。在认知过程中，经验素材存在的偏差，会在技术上造成归因问题。经验论的脆弱和理性论的霸道也影响了知识的形成过程，关系到我们能够认识什么，以及如何认识的问题。

参考文献

杰弗里·M. 伍德里奇，2010，《计量经济学导论（第四版）》，北京：中国人民大学出版社。

邱忠霞、胡伟，2016，《我国社会科学定量研究方法问题的反思》，《学术论坛》

第11期。

许加明、陈友华，2020，《数据质量、前提假设与因果模型——社会科学定量研究之反思》，《社会科学研究》第02期。

臧雷振、陈鹏，2015，《选择性偏差问题及其识别》，《世界经济与政治》第04期。

周皓，2012，《流动儿童社会融合的代际传承》，《中国人口科学》第1期。

Berk, Richard A. 1983, "An Introduction to Sample Selection Bias in Sociological Data." *American Sociology Review* 48(3).

Groves, Robert M., Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer & Roger Tourangeau 2009, *Survey Methodology*, Second Edition Hoboken. Hoboken, New Jersey: Wiley.

Heckman, James J. 1979, "Sample Selection Bias as a Specification Error." *Econometrica* 47(1).

Heeringa, Steven G., Brady T. West & Patricia A. Berglund 2010, *Applied Survey Data Analysis*. Boca Raton: Chapman and Hall/CRC.

Winship, Christopher & Robert D. Mare 1992, "Models for Sample Selection Bias." *Annual review of sociology* 18(1).