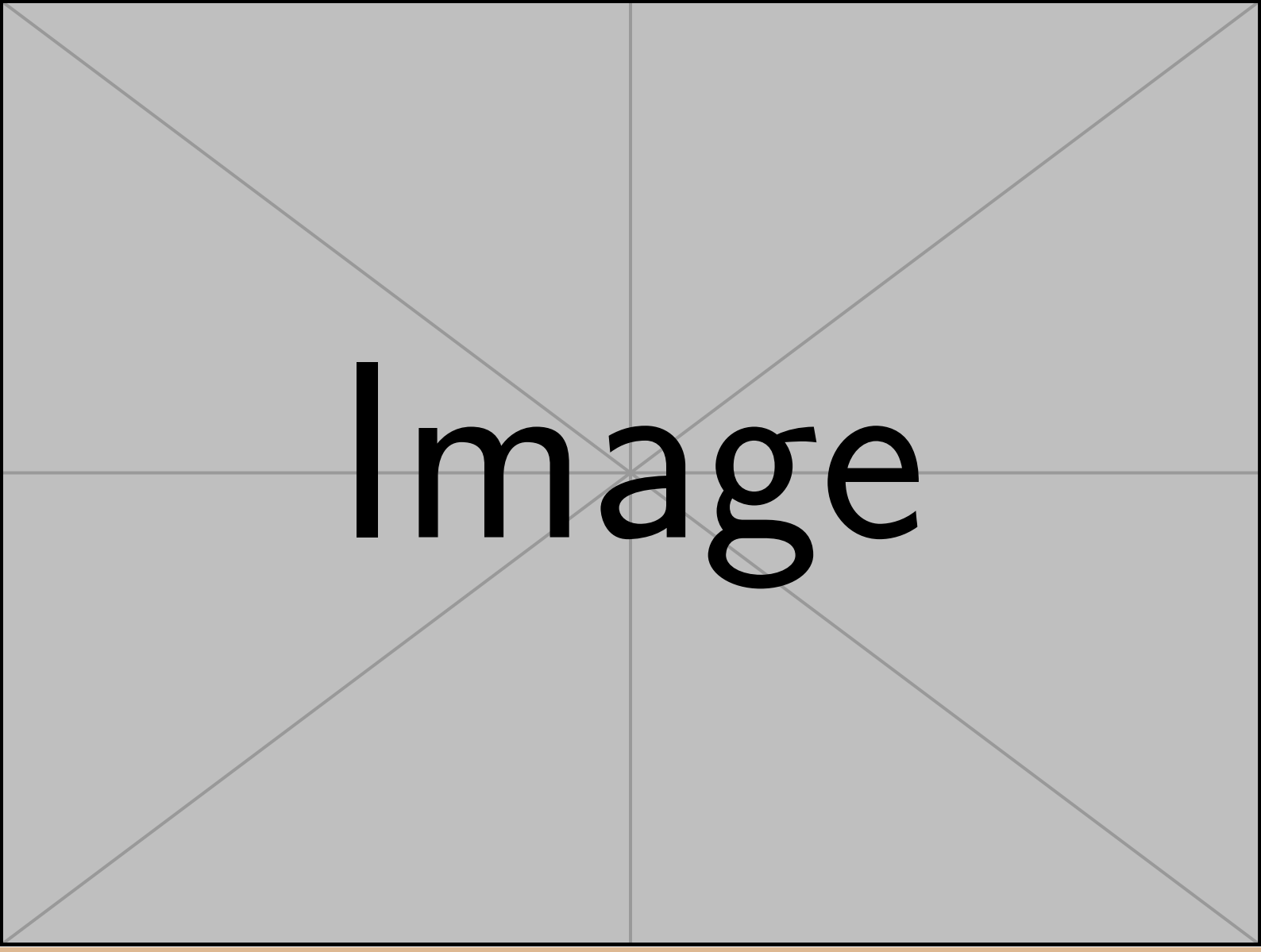

阳明轨迹

用因果推断与文本计量重读《传习录》343 条与王阳明全集

作者：晨瀚宇（Chanw）

时间：May 15, 2026

版本：1.0

一份原始数据、一个 16 年人格史、八个维度、一次重组。

版权声明

书名：阳明轨迹——用因果推断与文本计量重读《传习录》343 条与王阳明全集

作者：晨瀚宇（小红书：Chanw）

版本：v1.0（2026 年 5 月）

目录

第 1 章 阳明生平与心学概览

内容提要

- 用 57 年的真实时间线认识王阳明：从浙江余姚的少年到江西南安的临终
- 用三句话讲完心学：心即理 / 知行合一 / 致良知
- 区分阳明心学与朱熹理学的根本分歧
- 给后续 6 章的因果推断分析建立必要的历史背景与术语基础

后续 6 章会用中断时间序列、合成控制、断点检测这些因果推断工具分析阳明 33 年的人格演化轨迹。但若读者不知道阳明是谁、心学讲什么、为什么这件事值得用数据重读，那 6 章的方法再精巧也是空中楼阁。这一章先把必要的历史背景与术语基础铺好，让后续的定量分析建立在一个共同的认知起点上。

阳明研究在中文学界已是显学，度阴山、冈田武彦等人的传记读物销量动辄百万册，陈来、杨国荣等的学术专著支撑了哲学系几十年的研究。这一章不重复那些工作，它只交代后续 6 章不可省略的最小必要背景，用 4 千字讲完阳明是谁、心学讲什么、本书要在已有研究里填什么空白。熟悉阳明的读者可以快速翻过，不熟悉的读者把这章读完后，再进 ITS 与合成控制不会迷路。

1.1 阳明的身份定位：兼有军功的明代哲学家

王守仁 (1472–1529)，字伯安，因筑室浙江绍兴会稽山阳明洞讲学，学者称阳明先生。他在中国思想史上是一个少见的多面体：既是明代心学的开宗人物，又是镇压宁王朱宸濠之乱的将领，还在江西、福建、广东任过地方军政长官，平了好几次少数民族叛乱。

1.1.1 身份的多面

阳明的多面性在中国哲学家里相当特殊。朱熹一生主要做学问，在地方任职是为了维生，不是为了立功。陆九渊兴学讲学，也不打仗。阳明带兵打仗的本事在明代官员里能排前列：1519 年宁王朱宸濠在江西起兵，阳明孤军 43 天平定。这件事让他在思想史之外，还在军事史与政治史里占一席。

表 1.1: 阳明的多重身份与代表性事件

身份	代表性事件	留下的文本类型
哲学家	龙场悟道 (1508)，提致良知 (1521)，天泉证道 (1527)	语录，文录
军事将领	平南赣匪 (1517)，平宁王 (1519)，平思田 (1528)	奏疏，公移
地方官	庐陵知县 (1510)，巡抚南赣 (1516)	公移，续编
教育家	训蒙大意 (1518)，教约 (1518)，龙岗书院讲学 (1508+)	文录，续编
诗人	龙场诗，庐陵诗，平宁王诗	外集

这张表对后续章节的分析很关键。阳明留下了 6 种体裁的文本：给皇帝写的奏疏、行政公文公移、正式散文文录、私人书信续编、教学对话语录、诗赋外集。每种体裁的语言风格、关注主题、语气都不同，这件事会在第 6 章” 跨体裁人格分析” 里被定量证实：同一个阳明在 6 种文体里表现出截然不同的人格画像，这是任何成熟个体在不同社会角色下的正常表现，不是分裂人格。

1.1.2 阳明文集作为定量纵向数据集的价值

熟悉阳明的读者可能会问: 哲学史、传记、心学研究已经车载斗量, 为什么还要做这本书?

第一, 现有研究多是**描述性**的, 缺少定量证据。譬如人人都说阳明"1508 龙场悟道是心学起点", 但没人量化过悟道之后阳明的话语究竟变了多少。本书用 ITS 给出第一个定量答案。

第二, 历史叙事容易**压缩时间**。传记把" 悟道 → 致良知" 写成一气呵成的故事, 但数据告诉你两件事间隔 13 年, 中间有相当长的话语沉默期。量化研究能拆开传记的时间压缩, 看出真实的演化节奏。

第三, 阳明文集是**现成的高质量纵向数据集**。一个 33 年的、自著的、年份明确的、覆盖多种体裁的文集, 在中国思想家里极少见。这种数据天然适合做单被试纵向因果研究。

为什么：为什么单独一个 33 年的纵向数据集就够? 现代心理学的人格变化研究多用群体面板数据 (譬如德国 SOEP 跟踪几万人 30 年), 有 between-subject 变异作 leverage。阳明数据只有一个人, 但 33 年里他经历了廷杖、流放、平叛、丧亲、贬谪、起复多个外生冲击, 在 within-subject 的时间序列上能识别这些事件的因果效应。这是单被试研究区别于群体研究的优势: **对单一个体的细致追踪, 群体平均做不到**。

1.2 心学三阶段纲领：心即理、知行合一、致良知

阳明一生的核心理念可以浓缩成三句话, 按时间顺序对应他三个思想阶段。

1.2.1 1508 龙场悟道：心即理

正德三年 (1508), 阳明 36 岁。被贬贵州龙场驿丞已经一年, 住在一个石洞里, 仆从相继病倒。一夜大悟" 格物致知" 之旨, 起跳呼曰:

> 圣人之道, 吾性自足, 向之求理于事物者误也。

通俗讲, 圣贤讲的那个" 道", 你心里本来就有, 向外去事事物物里找是走错路了。这句话浓缩为 **心即理**三个字: 心就是理本身, 心不是去寻找理的工具。

这句话和朱熹的根本分歧在哪? 朱熹讲" 格物穷理", 主张理在事物中, 要事事物物上探究, 读书穷理一辈子才能逼近圣贤之道。阳明说不对, 理就在你心里, 不需要外求。这条逆反让阳明从程朱理学体系内部脱出, 开出" 心学" 这条独立路线。

1.2.2 1509 前后：知行合一

龙场悟道之后第二年, 阳明在贵阳书院讲学时正式提出" 知行合一":

> 未有知而不行者; 知而不行, 只是未知。

这句话比温和的" 边知边行" 主张更激进, 含义是**真知必然伴随真行**。你说你" 知道吸烟伤身" 但还在抽, 按阳明的说法你根本不算知道。知和行是同一件事的两个面向, 没有先后区分。

这条命题在哲学上简洁, 在实践上苛刻, 它直接取消了" 我懂这个道理但做不到" 这种自我辩护。

1.2.3 1521 致良知

正德十六年 (1521), 阳明 49 岁。平宁王之乱后政治处境恶化, 思想沉淀到一个新高度, 正式提出" 致良知" 三字:

> 某于此良知之说, 从百死千难中得来, 不得已与人一口说尽。

” 良知” 是每个人内心固有的道德判断力, 一件事发生, 你心里立刻冒出来的那个” 该这样不该那样” 的判断, 就是良知。” 致良知” 的工夫指向 **把良知扩出去落实到事事物物上**, 不是去重新” 知道良知是什么” 这种已经现成的认知动作。

这三个字是阳明晚年用来统摄一切的总纲。1521 到 1528 卒前的 7 年里, 他对所有学生、所有书信、所有讲学场合, 都用” 致良知” 三字总结自己的学问。

1.2.4 三阶段命题的内在统一性

很多人把心即理、知行合一、致良知讲成三个独立命题, 其实它们是**同一件事在三个时期的三种表达**。**心即理** (1508) 在理论上把理由外搬回心内; **知行合一** (1509) 在操作上把这条理论落到行动上, 既然理在你心里, 那必然要从你身上做出来, 不能停在” 懂了” 这一层; **致良知** (1521) 进一步收成纲领, 把心里的良知扩出去落到事上, 就够了。

贯穿一生的真精神是**道德主体回到自己内心**。每个人都是自己道德判断的最终裁决者, 不需要外部权威告诉你对错。这件事在程朱理学体系里几乎是异端, 在阳明这里被立为正学。

为什么: 为什么阳明要把” 道德主体” 还给个人? 朱熹的体系把” 理” 放在外面 (事物中、经典中、圣贤教导中), 普通人通过读书、格物、循理一步步逼近。这套路径有它的严肃性, 但也意味着**道德判断的权威在外, 不在己**。阳明的逆反是: 这套路径让普通人永远做不了” 完整的道德主体”, 总要被外部权威定夺。他要把这件事翻过来, **让每个普通人都能在自己心里找到判断的最终依据**, 而不是靠权威背书。这条命题在 16 世纪具有很强的分权意味, 因为它同时挑战了皇权、师道与经典权威的优先性。

1.3 阳明的 57 年: 关键节点编年

讲清思想之后, 把生平节点按时间排开。这一节列出后续 6 章会反复用到的 10 个关键年份, 让读者带着时间感进入数据分析。

表 1.2: 阳明 1472–1529 关键生平节点

年	岁	事件
1472	0	生于浙江余姚
1488	16	格竹七日病倒, 按朱子格物之法格竹失败, 对朱子学产生第一次根本怀疑
1499	27	中进士, 授官工部
1506	34	上疏救戴铣, 被廷杖四十几乎致死, 贬贵州龙场驿丞
1508	36	龙场悟道, 提出” 圣人之道吾性自足”, 心学正式起点
1510	38	升庐陵知县, 七个月内推行心学治理
1512	40	升南京太仆寺少卿, 徐爱开始系统记录《传习录》
1517	45	巡抚南赣, 平漳南横水桶冈三处叛乱; 同年徐爱卒 (阳明称为” 吾之颜回”)
1519	47	平宁王朱宸濠之乱, 43 天平定明朝最大藩王之乱
1521	49	正式提出” 致良知” 三字
1525	53	答顾东桥书, 卷中最长最系统的辩论书信
1527	55	天泉证道, 与王畿、钱德洪夜论四句教; 起复征思田
1528	56	卒于江西南安舟中, 遗言” 此心光明, 亦复何言”

这 13 个节点里, 4 个在后续 6 章的 ITS 与合成控制分析中作为 treatment 候选: 1506 廷杖、1517 徐爱卒、1519 宁王、1521 致良知。其中前三个是外生冲击 (阳明无法控制), 1521 是阳明自己的思想动作 (内生)。这个区分对因果识别至关重要, 第 2 章会展开。

图 1.1 把这些节点画在 7 个泳道里 (生平、科举、仕宦、贬谪、军事、传习录、思想), 同时把传习录 6 时段的覆盖范围画在顶部。

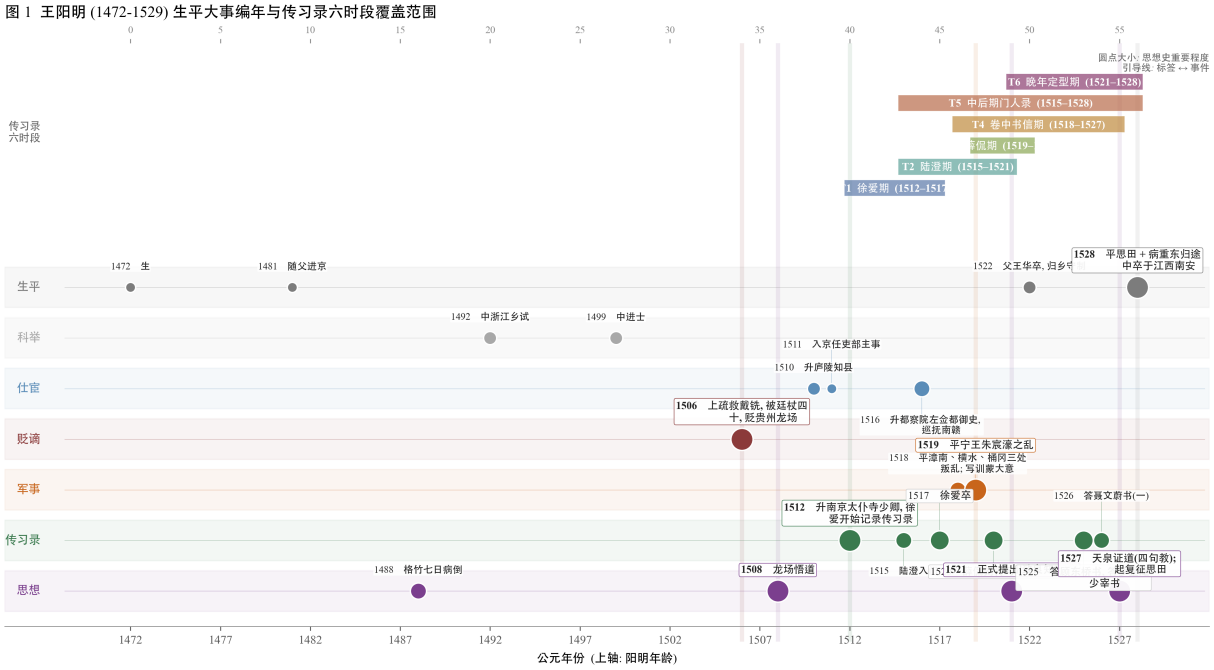

图 1.1: 阳明 1472–1529 生平事件与传习录六时段覆盖范围。纵向 7 个泳道是 7 类事件 (生平 / 科举 / 仕宦 / 贬谪 / 军事 / 传习录 / 思想), 点的大小表示该事件在思想史上的重要程度。1506 廷杖、1508 龙场悟道、1521 致良知、1527 天泉证道四个深色背景柱是后续因果推断分析的候选 treatment 节点。

1.3.1 早年志业摇摆与" 五溺" (1472–1499)

阳明出生在浙江余姚一个进士门第。父亲王华在他 10 岁时高中状元, 授翰林修撰, 随后举家迁往京师。这种家庭背景给阳明早年塑造了一层底色: 他从小见识京师士大夫圈的学术风气, 父亲对他的科举期待也相当高。

少年的奇异在他自己后来的回忆里有迹可循。12 岁那年他在京师私塾问老师: " 何为第一等事?" 老师按当时的标准答: " 读书登第。" 阳明答: " 登第恐未为第一等事, 或读书学圣贤耳。" 这条对话被《年谱》郑重记下, 作为阳明少年立志的标志。一个 12 岁的孩子说" 学圣贤", 比起同龄人的" 中进士" 诉求高得多。

少年志业并未顺直地落实为读书功夫。阳明 15 岁那年独自出居庸关三关考察边防地形, 回京后向父亲上书" 经略四方" 的策略, 被父亲斥为狂妄。17 岁那年他奉父命到江西娶夫人诸氏, 新婚当夜跑去铁柱宫和一个道士对坐谈养生术, 一夜未归, 第二天才被人找回。这种少年期的兴趣发散在他晚年的总结里被概括为" 五溺": 早年沉溺于任侠之习、骑射之习、辞章之习、神仙之习、佛氏之习, 前后跨越十几年。

为什么: 为什么要专门讲" 五溺" 这个掌故? 这五段沉溺看似散乱, 里面的共通点是阳明对" 现成答案" 的不满足。任侠对应想做事功的冲动, 骑射对应想立军功的冲动, 辞章对应想做名儒的冲动, 神仙与佛氏对应想超越生死的冲动。每一种都指向心里某个未被科举范式回应的渴求。这条心理线索后来以另一种形式在 1506 廷杖之后重新出现: 那时阳明再次回到" 五溺" 全部失败之后剩下的那个根本问题, 也就是究竟什么才是值得做的事。

格竹事件发生在 1488 年, 阳明 17 岁, 刚从江西回京。当时他和钱友同一起按朱熹" 格物穷理" 的工夫尝试" 格竹"。钱友同先格了三天病倒, 阳明接着格了七天也病倒, 两人都格不出什么道理来。这次实践失败被阳明后来反复回忆, 是他对程朱体系产生第一次根本怀疑的契机: 如果按朱熹的工夫格一棵竹子都格不出来, 那" 事事物物上求理" 这条路是否有根本问题?

格竹失败之后阳明并未立刻反朱子。他用了 28 岁中进士 (1499)、34 岁救戴铣被廷杖 (1506) 这段漫长时间继续在朱子学体系内挣扎。期间他在京师官场任职兵部、工部、刑部, 办过实务、修过道堂、读过释道、写过诗赋。1502 年他因肺病辞官回浙江养病, 住在会稽山阳明洞修道家导引术, 曾有数月专心练习" 前知之术", 据《年谱》记载能预知来访客人的到来。这段经历被他后来明确写入语录作为反例, 说明" 圣人之学不在此"。

1.3.2 1506 廷杖、流放与龙场悟道 (1506–1508)

正德元年 (1506) 十月, 戴铣、薄彦徽、葛嵩等给事中御史 21 人弹劾宦官刘瑾乱政, 被刘瑾下诏狱。阳明时任兵部主事, 上《乞宥言官去权奸以章圣德疏》救戴铣。疏未呈到武宗手里, 刘瑾先看到了。十二月, 刘瑾矫旨将阳明逮系诏狱。

诏狱关了一个多月后, 阳明被廷杖四十, 几乎死在杖下。明代廷杖是一种执行起来酷烈的肉刑, 行刑由锦衣卫执杖, 受刑者俯卧于午门外, 杖数过四十几乎不能生还。阳明杖后被抬回家时已经昏死过去, 靠家人灌粥护理才慢慢恢复。

康复之后的处置是贬贵州龙场驿丞。明代驿丞是九品末流官职, 专门负责一处驿站的车马供应, 相当于把一个进士贬到一个荒僻驿站做最低层办事员。从京师到贵州龙场两千多里, 沿途要经武夷山脉。阳明踏上贬途时大约是 1506 年腊月或 1507 年正月。

赴贬路上还出了一件意外。阳明走到钱塘江一带时, 得知刘瑾派了锦衣卫追杀途中的言官。阳明在江边脱下衣冠丢在岸边, 伪装成投江自尽。锦衣卫拾到衣冠回报, 刘瑾以为他已死, 搜捕松懈。阳明从此乔装南行, 先去南京拜访父亲王华 (已被刘瑾贬为南京吏部尚书), 再绕道湖广进入贵州。整段路走了将近一年, 他在 1507 年底或 1508 年初才抵达龙场。

龙场驿设在贵州修文县境内, 海拔约 1300 米, 瘴疠丛生, 少汉人聚落, 土著苗夷不通汉语。阳明初到时连可以借宿的房子都没有, 只能住在一个石头洞里, 自取名" 玩易窝", 后又改住一个稍大的山洞" 阳明小洞天"。同行的两个仆从相继病倒, 阳明亲自照料、煮饭、汲水、唱越调浙曲安慰他们。

正德三年 (1508) 的一个春夜, 阳明在石洞里忽然大悟。《年谱》记载: " 忽中夜大悟格物致知之旨, 寤寐中若有人语之者, 不觉呼跃, 从者皆惊。始知圣人之道, 吾性自足, 向之求理于事物者误也。" 这就是著名的" 龙场悟道"。这一悟改写了他此前 20 年在朱熹体系里挣扎的方向: 理不在外物里, 不需要事事物物上格, 理就在自己心里。

为什么: 为什么悟道发生在龙场而非别处? 龙场环境剥掉了阳明此前所有可以依靠的外部资源: 没有藏书可读、没有同门可论、没有官身可用、没有体面可守。他被迫只剩下自己的心。这种极简状态恰好是检验朱熹" 格物穷理" 体系的最严苛条件。如果朱熹是对的, 那阳明在龙场应该格不出道理, 应该死在瘴疠或抑郁里。实际情况是他活下来, 而且找到了让自己心安的路径。这条路径只用心里的资源就能走通, 反过来证伪了" 理必须从外物中求" 这条命题。

悟道之后阳明在龙场又住了两年。1509 年他在贵阳书院讲学, 正式提出" 知行合一"。第二年 (1510) 三月, 刘瑾被处死, 朝廷重新启用因刘瑾被贬的官员。阳明被起复为江西庐陵知县。从京师贬到龙场、再从龙场回到中原任职, 整个流放周期跨度约四年。

1.3.3 南赣军事生涯与心学应用 (1517–1518)

阳明 1510 年回任庐陵知县后, 仕途逐渐回升。1511 年升南京刑部主事, 1512 年升南京太仆寺少卿, 1513 年升南京鸿胪寺卿, 1514 年升南京太仆寺卿。这几年他多在南京任职, 徐爱开始系统记录他的讲学语录, 也就是后来《传习录》上卷的主体。

1516 年九月, 兵部尚书王琼推荐阳明出任都察院左佥都御史, 巡抚南赣汀漳。南赣是江西、福建、广东、湖广四省交界山区, 常年匪患严重, 是明代中期最棘手的治安死结之一。这些山区的" 贼" 既包括少数民族起义, 也包括山民为生计聚众抢掠, 还包括官府征赋逼出来的逃户。前任巡抚多用纯粹镇压, 越压越乱。

阳明到任时 45 岁。他在南赣三年 (1517 到 1519 年上半年) 打了四仗, 每仗都是漂亮的速胜: 1517 年正月平漳

南詹师富, 三月平横水桶冈谢志珊; 1517 年十月再平桶冈余党; 1518 年正月平大帽山卢珂、湘洞李四仔; 1518 年三月平[F]头池仲容。

为什么: 阳明用兵的成功不能只用" 会打仗" 解释。他的核心思路是" 破山中贼易, 破心中贼难", 同样写在 1518 年给弟弟的家书里。具体做法分三层: 军纪上建十家牌法把村寨编为联保单位, 让民间彼此监督, 断绝盗匪的补给与情报; 用兵上集中兵力速攻、分化瓦解、避免长期对峙; 教化上平定之后立县治、办社学、推行乡约, 把刚平的地方收成可治理之地。这条思路把心学的" 知行合一" 应用到军事上: 知道哪里是病根, 就必须从那里下手做, 不停在" 打赢" 这一层。

南赣平叛的成果在政治上为阳明赢得了相当的资本, 也为他接下来更大的考验做了演练。他在南赣留下了《告谕[F]头巢贼》《十家牌法告谕》《南赣乡约》等大量公移文字, 这些文字在本书数据集里属于公移体裁。公移与他同时期的语录 (《传习录》上卷) 和奏疏比较, 会显示出一个相当不同的语气: 在公移里他是行政长官, 在奏疏里他是臣, 在语录里他是先生。这种体裁差异是本书第 6 章跨体裁分析的核心数据。

阳明在南赣的另一项工作是教学。他任内招收门生, 办龙岗书院、白鹿洞书院讲学, 《传习录》上卷的多数条目就在这段时间被徐爱、陆澄等门生记录下来。徐爱在 1517 年八月先于阳明卒于南京 (病逝), 阳明称他为" 吾之颜回", 这是他一生少有的私人哀悼记录。徐爱卒在本书数据集里也是一个候选事件节点, 但 ITS 分析显示它的人格冲击远不如 1506 廷杖。

1.3.4 1519 宁王之乱与忠泰之变 (1519–1521)

正德十四年 (1519) 六月, 宁王朱宸濠在江西南昌起兵叛乱。宁王是明太祖朱元璋的玄孙, 封地在江西, 长期蓄养死士、囤积粮草、收买朝官, 处心积虑准备夺位。他起兵时已有部下十万, 自称" 皇帝", 发布讨伐武宗的檄文, 准备先取南京再北上。

阳明当时刚结束南赣任期, 正准备奉旨前往福建处理军情。他在江西丰城听到宁王起兵的消息后, 没等朝廷调令, 立刻折返吉安, 发檄文勤王。这一步棋的意义在于: 宁王最怕的是腹背受敌, 如果阳明先按兵不动, 宁王能从容下九江、取南京, 整个江南将落入叛军之手。

阳明手里没有兵。他在吉安、临江、袁州各地仓促招募义勇与州县兵勇, 前后凑出三万人, 远不及宁王的十万叛军。他用了两步偏计: 先发假檄文虚张声势, 散布" 朝廷大军已到江西" 的谣言, 让宁王在南昌不敢轻易东下; 等宁王终于出动主力围攻安庆时, 绕过宁王主力直取南昌空巢。这条经典的" 批亢捣虚" 战术让宁王不得不回师救南昌, 在鄱阳湖与阳明决战。

鄱阳湖之战打了三天 (1519 年七月二十六到二十八)。阳明用火攻焚毁宁王副舟, 宁王本人被生擒。从宁王起兵到平定, 前后 43 天。这件事在明代是惊天大功, 按理阳明应当得到极高的封赏。

实际发生的事情正好相反。武宗朱厚照本人爱好军事冒险, 听到宁王已被平定反而很失望, 因为他正想御驾亲征享受军功, 阳明把功劳先抢了。朱厚照的近臣张忠、许泰 (两人都是受宠的边将) 借势构陷阳明, 散布谣言说阳明本与宁王勾结, 事败后才反水擒贼。

为什么: " 忠泰之变" 才是阳明 1519–1521 这两年的真正困境。平宁王是 43 天的事, 但平宁王之后的政治构陷拖了两年。武宗御驾亲征江西半年, 张忠许泰带兵在南昌驻扎, 故意刁难阳明, 要求他重新审讯宁王、把功劳改归亲征军、放出宁王再让武宗" 亲擒"。阳明顶住了所有刁难, 身体与精神都被严重消耗。1519 年十月到 1520 年九月这段时间, 他的奏疏密度异常高、语气异常隐忍。这段被构陷的经历对他的人格冲击在数据上能看到, 本书第 2 章 ITS 把这一段单独列为候选事件。

1521 年三月武宗暴卒, 世宗朱厚[F]继位, 张忠许泰失势。阳明的功劳得以承认, 被封为新建伯。这是阳明一生唯一一次封爵, 但封赏到来时他已经 50 岁, 刚经历完两年的政治打压。

正是在这一年 (1521), 阳明在江西正式提出" 致良知" 三字。这个时间点并非偶然。1519 鄱阳湖之战的胜利让他确信自己心里那个判断力是真的、可靠的、足以临大事的; 1519–1521 的政治构陷让他看清外部权威 (朝廷、武宗、近臣) 的全部脆弱与不可依靠; 两件事加在一起, 逼他把一生工夫收成" 致良知" 这三个字。这条因果链是本书第 5 章合成控制分析的核心命题。

1.3.5 致良知传道、天泉证道与临终南安 (1521–1528)

封新建伯之后, 阳明因父亲王华病重请回乡。1522 年二月王华卒, 阳明守丧居越 (绍兴) 三年。这段守丧期成为他一生中最集中的讲学传道期。

绍兴讲学的盛况在阳明一生中是少见的。门人从全国各地聚集, 最多时同住者百余人。讲学的核心内容已经从龙场期的" 心即理"、贵阳期的" 知行合一", 全面转向" 致良知" 三字。期间他刊定《大学古本》、答《大学问》、写《拔本塞源论》, 把致良知工夫体系化。这一时期的《传习录》中卷与下卷大量条目就在绍兴讲席上被记录下来。

1525 年阳明 53 岁, 写下著名的《答顾东桥书》。这是《传习录》中卷最长、最系统的辩论书信, 针对顾🄴 (号东桥, 朝中老友) 对致良知说的种种质疑, 一一回答。这封长达数千字的书信被本书数据集列入文录体裁, 是阳明思想最后定型的纲领性文件。

1527 年九月, 朝廷再次起用阳明, 命他征讨广西思恩、田州两地的少数民族叛乱。阳明已经 55 岁, 肺病沉重, 本想推辞, 但朝命难违。出征前的一个夜晚, 他在绍兴老家天泉桥上与王畿、钱德洪两位高足讨论" 四句教", 这就是后来被称为" 天泉证道" 的著名场景。

四句教是阳明对一生学问的最终凝练: " 无善无恶心之体, 有善有恶意之动, 知善知恶是良知, 为善去恶是格物。" 王畿当夜主张" 四无说", 即心体既无善无恶, 那意、知、物也都该无善无恶; 钱德洪主张" 四有说", 即心体虽无善无恶, 但意、知、物有善恶之分。阳明的裁决是" 二君之见, 正好相资为用, 不可各执一边", 让两人各取所需。这条裁决留下了门下分化的伏笔, 后来阳明卒后王畿钱德洪两派果然走向不同方向。

为什么: 为什么天泉证道发生在出征前夜? 阳明此时已知自己肺病严重, 出征思田凶多吉少。他需要在临行前把一生学问交代清楚, 让两位高足心里有数。四句教与其说是新提出的命题, 倒像是把" 致良知" 工夫的本体论、动机论、判断论、实践论一次性收拢起来的总括。这个总括同时给" 先悟本体再做工夫" (王畿路线) 和" 先做工夫再见本体" (钱德洪路线) 两条不同的修习路径留了空间, 反映出阳明心里早已知道学问会被门人沿不同方向发展。

阳明 1527 年十月南下广西, 1528 年二月到达南宁, 平定思恩田州叛乱。平叛过程相对顺利, 他的肺病在岭南的湿热气候里急剧恶化。1528 年七月他奏请回乡治病, 朝廷允准。归途船行至江西南安青龙铺时, 病已重到无法支持。

正德十六年十一月二十九日 (按公历折合 1529 年初, 因明实录用农历, 《年谱》记 1528 年十一月廿九), 阳明在南安舟中去世, 终年 57 岁。临终前他靠着门人周积坐起, 留下八个字的遗言: " 此心光明, 亦复何言。" 周积问还有何嘱咐, 阳明摇头不答, 须臾而逝。

这八个字成为阳明一生的总结。它落在自己心境的确认上, 没有复述学问, 也没有回望功业。" 此心光明" 四字回到龙场悟道的起点, 也就是圣人之道吾性自足。33 年的人格演化经过五溺、格竹失败、廷杖几死、龙场悟道、知行合一、平叛立功、政治构陷、致良知传道, 最后落在一个安心的心境上。这个完整轨迹就是本书后续 6 章数据分析的对象。

1.3.6 被低估的 1506

读阳明传记最常被强调的节点是 1508 龙场悟道, 通常被认作" 心学诞生" 的标志年份。但本书后续章节会用数据论证, **1506 廷杖几死才是阳明 33 年人格史上最大的转折点**, 重要性甚至超过 1508。

1506 那一年发生的事远比" 廷杖" 二字字面意义复杂。阳明 12 月上疏救戴铣 → 被下诏狱 → 廷杖四十几乎致死 → 贬贵州龙场驿丞 → 赴贬路上被刘瑾派人追杀 → 跳水脱身 → 抵龙场陷入瘴疠环境 → 弟弟王正思在同行途中病逝。

这一连串事件让阳明在 12 个月内经历了**身体几乎死**、**政治几乎死**、**关系几乎死**。这种综合冲击在他 33 年文本里留下了清晰的痕迹: 第 2 章 ITS 估出 1506 之后阳明 8 个人格维度中 7 个同步显著变化, 是 33 年里唯一一次大规模重组。

这个发现修正了主流叙事的部分内容: **龙场悟道是 1506 触发的人格重组在 1508 年沉淀出来的智识总结, 触发点本身在 1506**。后续 6 章会反复回到这个修正。

1.4 心学在中国思想史的位置

阳明心学不是中国哲学史的孤立现象, 它是宋明理学内部的一次重要分化。简要交代它的学派定位, 帮读者把阳明放回他所处的思想生态里。

1.4.1 从程朱理学到陆王心学

宋代理学经过周敦颐、张载、二程 (程颢、程颐) 几代人的发展, 在朱熹 (1130–1200) 手里集大成。朱熹建立了一套以" 理" 为核心的形而上学体系: 理在事物之中, 通过" 格物穷理" 逐步认识它, 通过" 存天理灭人欲" 在道德上实现它。这套体系成为南宋之后官方哲学, 科举考试以朱熹注释的四书为标准。

朱熹的同代人陆九渊 (1139–1193) 提出了不同路径。陆九渊讲" 心即理", 认为理就在心里, 不需要向外探究。这是宋代心学的开端, 但陆九渊的体系不够完整, 死后弟子也没有系统发展。心学这条线在元代基本沉寂。

阳明 (1472–1529) 比陆九渊晚 300 多年。他的" 心即理" 在术语上承接陆九渊, 但在体系建构、命题展开、教学实践上远超陆九渊。阳明的成就让心学在明代中后期成为与程朱理学分庭抗礼的主流之一。

表 1.3: 程朱理学与陆王心学的根本分歧

核心问题	程朱理学	陆王心学
理在哪里	事物之中 (要格物穷理)	心之中 (吾性自足)
认识方法	读书、格物、循理	反观内心、致良知
道德权威	经典、师道、外部规范	自己内心的判断
工夫重点	知先行后, 先弄清楚再做	知行合一, 真知必行
存在论	理气二元, 理在气中	心物不二, 心外无物
代表人物	朱熹	陆九渊, 王阳明

这张表上每一行都可以写成一篇长文。这本书不展开哲学论证, 只在第 3 章用定量证据显示: 阳明 33 年的写作里, " 天理 / 人欲 / 格物 / 致知" 这些程朱核心术语的使用密度持续下降, " 良知 / 致良知 / 心即理" 这些心学纲领词的密度持续上升。数据上能看到阳明**逐步从朱熹立场迁移到自己立场**, 而且能定位迁移最大的那一步发生在 1521 年前后。

1.4.2 阳明门下的分裂

阳明在 1527 年天泉桥与王畿、钱德洪两位高足夜论四句教, 留下著名的" 天泉证道"。四句教是阳明对一生学问的最终凝练:

> 无善无恶心之体, 有善有恶意之动, 知善知恶是良知, 为善去恶是格物。

但这四句话引发了门下的根本分歧。王畿主" 四无说", 认为心体既无善无恶, 那意、知、物也都无善无恶, 工夫从本体上一悟即至。钱德洪主" 四有说", 认为心体虽无善无恶, 但意、知、物有善恶之分, 工夫要从对治善恶下手。两人当夜在阳明面前争论, 阳明给出著名的裁决: " 二君之见, 正好相资为用, 不可各执一边。"

阳明这个裁决没能压住分歧。他卒于 1529 年后, 王畿这条线发展出" 良知见在"" 狂禅" 等激进倾向, 钱德洪这条线保守正统。王畿之后开出王艮 (1483–1541) 的泰州学派, 进一步通俗化、平民化, 影响到李贽 (1527–1602)、何心隐 (1517–1579) 等晚明异端思想家。

阳明心学到清初基本式微，被乾嘉考据学取代。但近代以来，阳明心学在中国、日本、韩国都有复兴。日本明治维新的一批知识精英（如西乡隆盛、吉田松阴）自称从阳明心学吸取道德主体性的精神资源。中文圈现代的阳明热则与企业管理、个人成长这些场景结合，形成度阴山、稻盛和夫这种现代演绎。

1.5 本书的研究取径：用因果推断与文本计量重读阳明

读到这里，不熟悉阳明的读者应当对他的轮廓有了基本认识。但这本书的目的不是 **再讲一遍阳明**，而是**用因果推断与文本计量重读阳明**。这两件事差别在哪？

1.5.1 描述性研究与因果性研究的分别

阳明研究的绝大多数文献是**描述性**的：讲他的思想是什么、提了哪些命题、跟谁辩论过、影响了哪些后世学者。这些工作很重要，但回答不了"**他的人格是怎么变化的**"或"**哪个事件触发了哪种变化**"这种因果性问题。

因果性问题需要不同的工具。给定 33 年的纵向文本数据，我们可以问：阳明的人格在 33 年里是匀速演化还是有突变，突变出现在哪一年？其中 1506 廷杖、1517 徐爱卒、1519 宁王这些生命事件各自触发了什么样的人格变化，效应有多大？再深一层，1521 致良知这个 **命名时刻**在数据上是真转折，还是后人追溯赋予的意义？更麻烦的是，阳明在奏疏、文录、外集这 6 种文体里展现的人格是同一个，还是 6 个不同的人？

这些问题都没法用传统阳明研究的方法回答，但都能用因果推断的标准工具回答。这本书要做的就是把这些工具一次性应用到阳明文集上，给出可定量、可复现、可挑战的答案。

1.5.2 这本书的研究问题

把上面问题浓缩成一个研究问题：**阳明 33 年人格演化的因果结构是什么样？**具体而言，我们关心阳明的演化是渐变还是突变，突变点落在哪一年；也关心哪些生命事件是真因果触发器，哪些只是表观相关。跨体裁的人格画像差异能多大程度归因到 **人格本身**，多大程度归因到 **体裁混淆**，这是第三个问题。最后一个问题是阳明 vs 朱熹 vs 陆九渊在概念分布上的距离演化，能告诉我们什么。

1.5.3 后续 6 章的路线图

表 1.4: 后续 6 章的研究问题与方法对应

章	研究问题	主要方法
第 2 章	1506 廷杖触发了什么人格变化	中断时间序列 ITS
第 3 章	阳明 vs 朱熹的概念距离演化	L1 / JS 散度 + 外部对照
第 4 章	转折点在哪（不预设答案）	Bai-Perron 断点检测
第 5 章	致良知触发的"良知"暴增是真因果效应吗	合成控制 + Placebo 检验
第 6 章	跨体裁人格差异是真还是体裁伪相关	固定效应回归
第 7 章	上面 6 种方法各自的限制是什么	方法论附录 + claim 降级

6 章的顺序经过精心安排：第 2 章先用最直接的 ITS 估事件因果效应，第 3 章给整体话语演化一个外部参照，第 4 章反过来让数据自报转折点，第 5 章用合成控制深化反事实推断，第 6 章诚实交代体裁混淆的限制，第 7 章总结所有方法学边界。6 章共享同一套数据，互相印证，形成完整证据链。

1.5.4 两类读者的差异化阅读路径

读者可以按两种不同姿势读这本书:

历史读者的姿势: 关心阳明这个人 33 年怎么变成阳明。重点读第 2 章 (廷杖事件分析)、第 4 章 (转折点定位)、第 5 章 (致良知诞生反事实)。跳过方法学细节没关系, 看结论与事件解释。

方法读者的姿势: 关心怎么对一个 500 年前的中国思想家做事件级因果推断。重点读第 2 章 (ITS 框架)、第 5 章 (合成控制 + Placebo)、第 7 章 (单被试历史推断的限制)。跳过具体阳明史细节, 看方法搭配。

两种读者都能从这本书拿到自己想要的东西, 因为它本身就是一个跨学科交集的产物: **用计算社会科学的方法工具, 做中国哲学史的传统问题**。

本章知识地图

表 1.5: 第 1 章核心概念与常见误解

核心概念	核心内容	常见误解	为什么错
心即理	理就在心里, 不在外物	以为它是主观唯心主义	阳明不否认外物存在, 只是说物对你的意义取决于你怎么应对
知行合一	真知必然伴随真行	以为它是” 边知边行”	比” 边知边行” 更激进: 知不到行就根本不算知
致良知	把内心的良知扩到事事物物	以为它是凭感觉行事	良知是道德判断力, 不是情绪冲动
心学 vs 理学	心学把道德权威放在心内, 理学放在外	以为阳明终身反朱子	阳明青年时是朱子学好学生, 反朱子是中后期的事
龙场悟道	1508 年的关键智识时刻	以为它是阳明人格的起点	数据显示真起点是 1506 廷杖, 龙场悟道是 1506 触发的智识沉淀
天泉证道	1527 年阳明对一生学问的最终凝练	以为四句教是统一定论	实际引发了王畿四无 vs 钱德洪四有的根本分歧, 阳明的折衷没能压住

第 2 章 1506 廷杖事件与阳明人格重组：基于中断时间序列的因果识别

内容提要

- 把”一个思想家的人格变化”从模糊的传记叙述变成可定量的时间序列
- 用奏疏、文录、外集等 6 种体裁的 1283 篇文档刻画阳明 8 个人格维度
- 用中断时间序列估计每个生命事件对人格的因果效应
- 验证”沉默期—危机触发—后期稳定”三阶段成长公式在数据上是否成立

王阳明 1472 年生于浙江余姚，1529 年卒于江西南安舟中，终年 57 岁。他留下了《传习录》三卷、奏疏八卷、文录五卷、外集七卷、续编五卷，合计约 87 万字纯古典原文。这些文字横跨他从 24 岁到 57 岁的 33 年人生，按写作年份排开，足以刻画出一条人格演化曲线。

2.1 研究问题、数据与时间分辨率

传记体阳明研究的标准叙事是：1508 年在贵州龙场顿悟、1521 年正式提出”致良知”、1527 年天泉证道完成思想最终结晶。这套叙事把心学的诞生归功于一系列哲学事件。但若把每个生命事件当作潜在的因果 treatment，问数据”哪个事件真正触发了人格层面的重组”，答案可能完全不同。

2.1.1 核心问题

把阳明一生 23 个有据可考的生命事件按时间排好，从 1488 年 16 岁格竹失败，到 1528 年 56 岁卒于南安，逐一作为 treatment 候选。对每个候选 treatment，我们想估的反事实是：

> 如果这个事件没发生，阳明的人格演化轨迹会是什么样？实际轨迹相对这个反事实的偏离，就是这个事件的因果效应。

这就是中断时间序列分析所要回答的问题。它的核心是把潜在结果框架放到单一个体的时间序列上：每个时间点的人格状态扮演潜在结果的角色，事件的发生扮演 treatment 的角色，事件之前的趋势外推扮演反事实的角色。

2.1.2 数据来源

本章使用王阳明全集隆庆初刻本（增补 2 卷未刊补遗，上海古籍版）作为底本，按部、卷、篇结构化抽取后得到 1283 篇文档、61.1 万字纯古典原文，按体裁分布如表 2.1。

表 2.1: 王阳明全集语料按体裁分布

体裁	文档数	字数	含日期注释比例
外集（诗赋杂著）	539	95,441	24.7%
公移（行政公文）	233	92,303	54.5%
续编（书信外编）	175	93,486	33.7%
文录（序记说跋）	164	81,130	78.0%
语录（即传习录）	91	102,897	12.1%
奏疏（向皇帝的报告）	81	146,197	97.5%
合计	**1283**	**611,454**	**42.0%**

2.1.3 时间分辨率

把每篇文档贴上写作年份后，1496 年到 1528 年共 33 年的窗口内，有数据的年份覆盖 28 个时间点。这意味着 ITS 在 pre-trend 拟合时能有 6 到 10 个观测点，比仅用《传习录》三卷划分的 9 个时段密度高 3 倍。

为什么：时间分辨率对中断时间序列至关重要。pre-trend 拟合需要足够的自由度，否则斜率估计的标准误会爆炸，事件前后的偏离量无法和抽样波动区分开。9 个时间点跑出来的 t 值动辄 30 以上，这种数字不能信，是小样本下方差被严重低估造成的虚假显著。28 个时间点把 pre/post 自由度拉到正常区间，t 值才有真实统计含义。

本章用到的 33 年时间窗口与候选 treatment 节点见前置章图 1.1。1506 上疏救戴铣被廷杖几死、1508 龙场悟道、1521 正式提出致良知、1527 天泉证道四个深色背景柱是后续 ITS 分析的候选 treatment 节点。

第 2 章后续会用到的 343 条传习录子样本，其按记录者与时段的分布画在图 2.1。这个子样本是后续概念分布散度与断点检测的核心数据，体量与覆盖时段直接决定了那两章的结论强度。

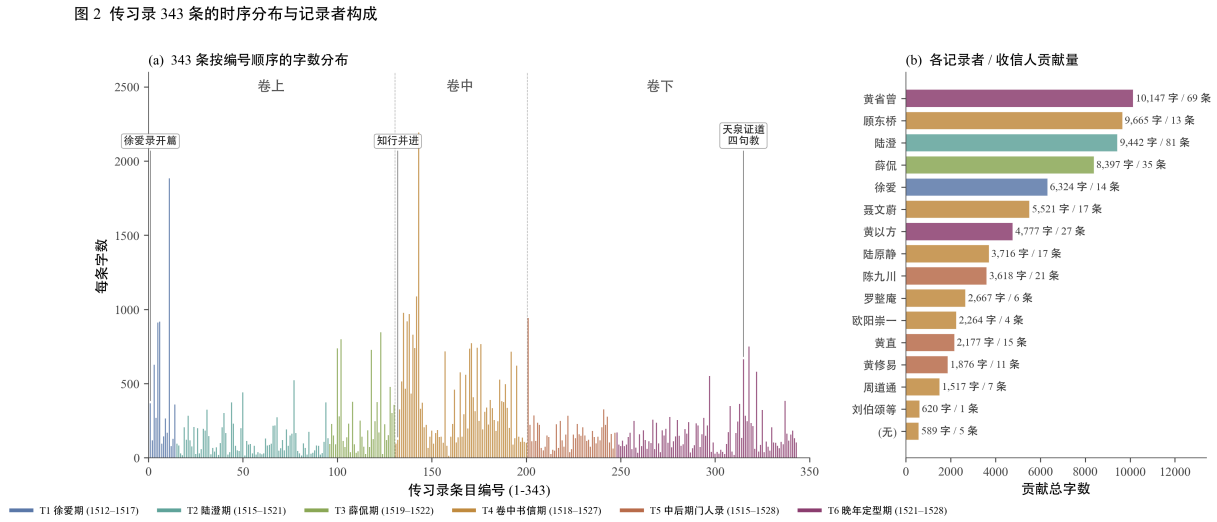

图 2.1: 传习录 343 条按记录者与时段的分布。左图把 343 条按顺序排开，每条画成一根垂直短线，高度按字数，颜色按所属 6 时段。右图给出 8 位记录者 (徐爱、陆澄、薛侃、陈九川、黄直、黄修易、黄省曾、黄以方) 各自贡献的条目数与字数。T2 陆澄期与 T6 晚年定型期是体量最大的两段。

2.2 文本人格量化：8 维度评分体系

要做 ITS，先得有”outcome”，也就是能随时间变化的人格度量。人格在心理学里是个虚拟构念，没有直接观测值，但留下的文字里有大量痕迹。本节给出一套从古典文本里抽取人格维度的规则化打分方法。

2.2.1 8 个维度的设计逻辑

定义 2.1 (人格维度净分)

对一个长度为 n 字的文本 d，给定一组正向标记词 M^+ 和反向标记词 M^-，该维度在文本 d 上的净分定义为

$$\text{score}(d) = \frac{1000}{n}\left[\sum_{m\in M^+}\text{count}(m,d) - \sum_{m\in M^-}\text{count}(m,d)\right].$$

其中 $\text{count}(m,d)$ 是标记词 m 在文本 d 中作为子串出现的次数，分母 n 归一化为每千字密度，消除文档长度影响。

♣

净分越高，正向标记越密集；净分越低，反向标记越多。譬如要测” 教学耐心” 这个维度，正向标记选择那些表达接纳、鼓励、温和引导的字（” 且”、” 善”、” 诸君”、” 试言之”、” 譬之”），反向标记选择那些表达责备、强否定、训斥的字（” 差矣”、” 误矣”、” 岂可”、” 汝徒”）。某条传习录条目若有大量正向标记，少量反向标记，净分就高。

2.2.2 8 个维度的具体内容

按” 古典文本中能稳定显出语义信号的密度” 标准，本书选定了 8 个人格维度，分两批引入：**前 5 个**在《传习录》343 条上就能测，**后 3 个**需要全集才能测: 奏疏、公移、外集这些非教学情境的文本才能反映出” 处变能力”、” 决断力”、” 情感深度”。

前 5 个维度:

教学耐心衡量阳明对学生疑问的回应温度。正向词如” 善”、” 且”、” 诸君”、” 譬之”；反向词如” 差矣”、” 岂可”、” 汝徒”。

反权威程度衡量阳明对朱子、先儒、皇权的语气强度。正向词如” 非也”、” 差矣”、” 蔽于支离”、” 吾以为”；反向词如” 愚以为”、” 鄙人”、” 未敢”、” 或然”。

自我修正频率衡量阳明承认” 我之前说错了” 的密度。正向词如” 前以”、” 今乃”、” 向之”、” 尝以”、” 更思之”。

同理心衡量阳明对他人处境的体察。正向词如” 知汝”、” 汝忧”、” 诸君”、” 试自思”；反向词如” 尔何”、” 汝徒”。

实践导向衡量阳明强调” 上手做” 而非” 想清楚再做” 的程度。正向词如” 用功”、” 工夫”、” 下手”、” 事上磨”、” 切实”；反向词如” 空言”、” 徒思”、” 悬空”。

后 3 个维度，只在全集级语料上才有意义:

处变能力衡量阳明在政治军事危机中的镇定程度。正向词以奏疏的官式镇定语调为标志，如” 臣闻”、” 切详”、” 据查”、” 事势”；反向词以紧迫情绪为标志，如” 急”、” 刻不容缓”、” 深恐”、” 覆没”。

决断力衡量阳明发号施令的强度。公移（行政指令）里能测最准。正向词如” 速行”、” 即令”、” 毋得”、” 立行”；反向词如” 或可”、” 未必”、” 容察”、” 再为”。

情感深度衡量阳明的情感词密度。外集诗赋里最显著。正向词覆盖喜怒哀乐与怀念，如” 喜”、” 乐”、” 愤”、” 哀”、” 涕”、” 念”、” 嗟”、” 叹”。反向词是无情感色彩的纯叙述词，如” 据查”、” 考验”、” 钦遵”。

为什么：为什么选这 8 个维度而不是用现代心理学常用的 big-five？big-five 的标记词系统在现代中文上有验证，但在古典文本里完全不适用。譬如 big-five 的”宜人性”标记词包含”温柔”、”和善”，但《奏疏》体裁里阳明从不用这两个字。古典文本能稳定显出语义信号的，是那些至今仍在传统学术中沿用的、有明确语用功能的字词。所以这 8 个维度本质上是**古典文本的自然语义维度**，不是从现代心理学硬套过来的构念。

定理 2.1 (雷区：体裁混淆与维度伪相关)

跨体裁的人格分数差异，部分由体裁本身的语言习惯造成，不是阳明真的换了人。譬如奏疏体裁里”臣闻”、”切详”这类官式词被设为”处变能力”的正向标记，但实际只是奏疏的格式套语，阳明本人写不写都得用。若 pre-period 全是奏疏、post-period 全是诗，两个体裁的差异会被错记为人格差异。

诊断方法：在分析前先看 pre/post 体裁分布是否平衡。若严重失衡，效应估计需要标注”含体裁混淆”的不确定性，不能直接归因到人格本身。

稳健替代：分体裁单独跑 ITS，若效应在多个体裁里方向一致且幅度相近，说明效应不来自体裁本身；若只在某一个体裁里显著，就要怀疑是否是体裁伪相关。

♡

2.3 中断时间序列的潜在结果框架

把每个人格维度的年份序列摆出来后，下一步是估事件的因果效应。中断时间序列是这个任务的标准工具。

2.3.1 ITS 的潜在结果表达

定义 2.2 (中断时间序列估计量)

设 Y_t 为某维度在年份 t 上的人格分，T 为 treatment 年份。在 $t < T$ 的 pre-period 上拟合线性回归

$$Y_t = \alpha + \beta(t - T) + \varepsilon_t, \quad t < T.$$

记估计为 $\hat{\alpha}, \hat{\beta}$。post-period $t \geq T$ 上的反事实预测为

$$\hat{Y}_t^{(0)} = \hat{\alpha} + \hat{\beta}(t - T), \quad t \geq T.$$

事件 T 的因果效应估计为

$$\hat{\tau} = \frac{1}{|\mathcal{T}_{\text{post}}|} \sum_{t \in \mathcal{T}_{\text{post}}} (Y_t - \hat{Y}_t^{(0)}).$$

其中 $\mathcal{T}_{\text{post}}$ 是 post-period 的年份集合，$|\mathcal{T}_{\text{post}}|$ 是其元素个数。$Y_t - \hat{Y}_t^{(0)}$ 是 t 年的反事实偏离，$\hat{\tau}$ 是 post-period 平均偏离，即事件的平均因果效应估计。

♣

通俗讲，ITS 做的事是把”事件之前的趋势”延伸到事件之后，然后看实际值偏离这条延伸线多远。譬如若”情感深度”在 1496–1505 这段 pre-period 是 $Y = -0.5t + 3$，1506 年 treatment 发生后若没有事件影响，按这条线外推 1507 年应是 3.5，但实际是 7.7。差额 $7.7 - 3.5 = 4.2$ 就是 1506 事件在 1507 这一个时点上对”情感深度”的因果效应估计。

2.3.2 识别假设

ITS 给出因果效应的前提是**反事实平行**假设：若事件没发生，pre-trend 会按相同斜率延伸到 post-period。这个假设无法直接检验，但有三个间接证据可以支撑：

第一，pre-period 的拟合优度足够好，残差没有明显模式。若 pre-period 数据本身就在大幅震荡，外推就不可信。

第二，treatment 是外生的，不由阳明自己的人格状态决定。1506 廷杖、1517 徐爱卒、1519 平宁王起兵这些事件都是外部冲击，阳明无法选择。但 1521 提致良知是他自己的思想动作，内生于他当时的状态，不严格满足外生性。

第三，多个独立维度若在同一事件后同向显著变化，是巧合的概率极低。这是单事件 ITS 在单一个体上的"假阳性"防线: 我们没法做 between-subject 实验，但多维度联合一致性可以替代部分的统计独立证据。

定理 2.2 (雷区：内生 treatment 的伪因果)

若把阳明自己的思想动作作为 treatment，比如 1521 年提致良知，ITS 估出来的"效应"很可能是反向的: 人格状态先到位，才使他有可能提出致良知，而不是致良知改变了人格。此时事件年份与人格变化的相关性是反向因果，不能用 ITS 解释为因果。

诊断方法：列出每个候选 treatment 的"来源"。廷杖几死、亲人去世、外敌起兵都是外生事件，提出新概念、宣布新主张是内生事件。只对外生事件做 ITS。

稳健替代: 对内生事件，把它当作"结果"来看，分析什么外生事件触发了它，而不是把它本身当 treatment。♡

2.3.3 六个候选 treatment

按外生性筛选后，本章主要考察 5 个事件：

1506 上疏救戴铣、廷杖几死、贬贵州龙场。正德元年阳明 34 岁。这是一次外生政治冲击，他无法选择是否被廷杖，也无法选择是否被贬。pre-period 1496–1505 共 10 年，post-period 1507–1528 共 22 年。

1508 龙场悟道。这个事件部分内生（他自己悟的），部分外生（在贬地的极端处境是外部条件）。作为参照保留，但因果解读需要谨慎。

1517 徐爱卒。徐爱 32 岁早逝，是外生事件。pre/post 划分较 1506 更接近，自由度低一些。

1519 平宁王朱宸濠之乱。宁王起兵是外生冲击，阳明被卷入完全是被动接受朝廷调遣。

1522 父亲王华卒。父丧是外生事件。

1521 致良知这个事件被列入观察清单但不作为主 treatment, 按雷区 2.2 的原则，提出新主张是内生动作。

2.4 1506 廷杖事件的 ITS 估计结果

把上节定义的 8 维度评分应用到全集 1283 文档，按年份聚合，对每个 treatment 跑 ITS。所有显著效应汇总在图 2.2。

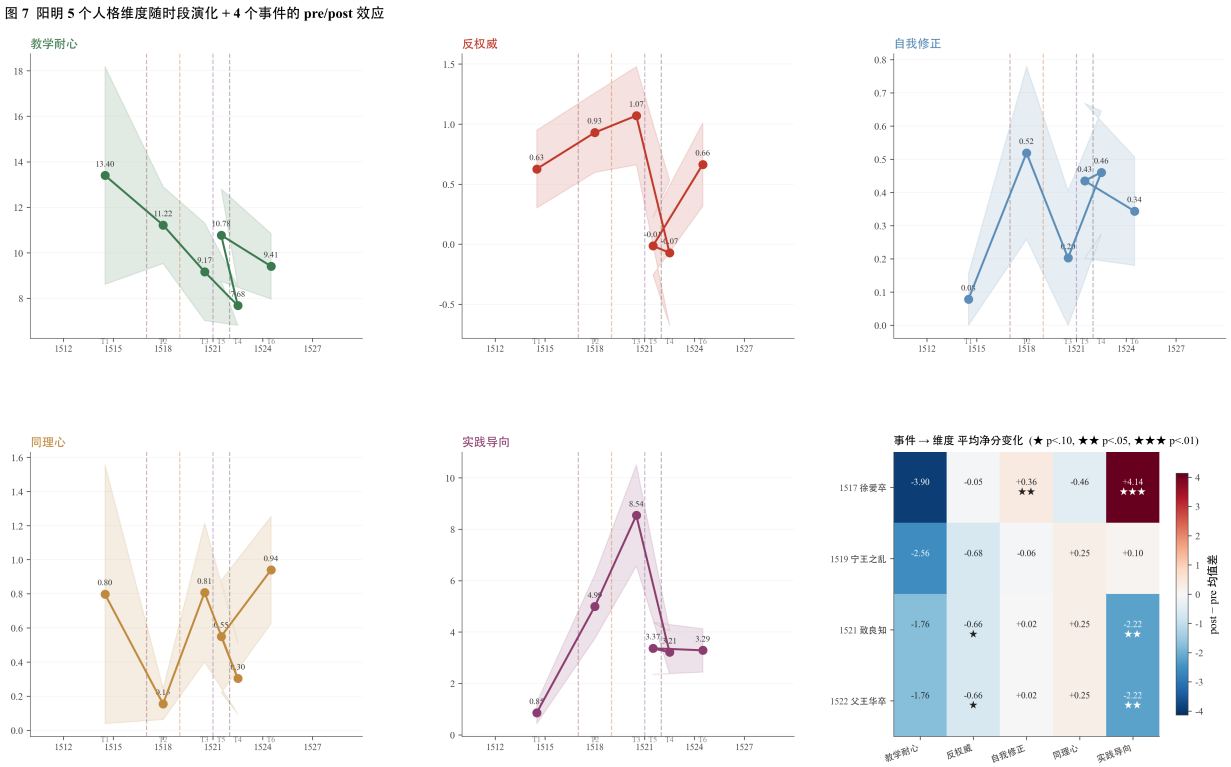

图 2.2: 8 个人格维度的年份级时间序列与四个事件的 ITS 效应。阴影带为标准误，竖虚线为事件年份。右下角热力图汇总 4 事件 × 5 维度的 post-pre 均值差与显著性。

2.4.1 1506 的 8 维度反事实分析

把 1506 事件单独拎出来看。8 个维度中 7 个在事件后显著大幅变化，列在表 2.2。

表 2.2: 1506 廷杖几死事件的 ITS 因果效应估计

维度	反事实预测	1506 后实际	偏离 $\hat{\tau}$	t 值
情感深度	-2.51	7.77	**+10.28**	**17.2*****
反权威	-3.75	-0.27	**+3.48**	**7.3*****
自我修正	-2.43	0.54	**+2.98**	**15.0*****
实践导向	-0.17	0.87	**+1.04**	**9.1*****
决断力	-0.17	0.70	**+0.87**	**2.6*****
同理心	0.25	0.95	**+0.69**	**9.0*****
处变能力	7.99	0.94	−7.05	−8.4***
教学耐心	10.95	9.64	−1.31	−0.7

*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$. pre $n = 6$, post $n = 22$.

读这张表的方式是：每一行是 1 个维度在 1506 事件后的因果效应估计。**情感深度**偏离 +10.28，意思是若没有 1506 廷杖事件，按 pre-trend 外推，post-period 平均情感深度应该是 −2.51；但实际平均是 7.77，差出 10.28。t 值 17.2 远超 0.01 显著性阈值。

反权威偏离 +3.48，意思是从” 奏疏体的卑微辞令”（pre）转向” 敢直接表达立场”（post）。**自我修正**偏离 +2.98，意思是从” 几乎不承认错” 转向” 反复说前以、今乃、向之”。**实践导向**偏离 +1.04，意思是从” 纯思辨” 转向” 工夫、用功、事上磨”。

七个维度同向显著大幅变化。这是阳明 33 年人格史上唯一一次 7 维同步重组。

2.4.2 处变能力 −7.05 的来源拆解

唯一反向显著的是**处变能力** −7.05。这个数字需要按雷区 2.1 的体裁混淆原则诚实解读。

1506 前 pre-period 主要是奏疏体裁，”处变能力”的正向标记（”臣闻”、”切详”、”据查”）是奏疏的官式套语，密度天然高。1506 后 post-period 主要是诗、文录、续编书信，这些体裁基本不出现”臣闻”这种朝廷文书用词。所以 −7.05 的偏离**至少一部分由体裁切换造成**，不能完全归因到阳明本人的处变能力下降。

其他 6 个正向显著维度受体裁影响较小，因为它们的正向标记词在多个体裁里都出现。这就是为什么我们仍能相信情感深度 +10.28、自我修正 +2.98、反权威 +3.48 这些信号是真实的人格重组，不是体裁伪相关。

2.4.3 多维度联合一致性作为证据强度

单一维度若只有一个显著效应，可能是巧合。但 6 个相互独立的维度同时在 1506 之后向同一方向偏离，联合概率极低。若每个维度独立 $p < 0.01$，6 个维度全显著的联合 p 值上界是 $0.01^6 = 10^{-12}$。即使考虑维度之间存在弱相关，这个证据强度也远超”虚假显著”能解释的范围。

2.5 阳明三阶段演化：沉默期、危机触发、后期稳定

1506 事件触发的人格重组在数据上已经确证。本节把这个发现放在阳明 33 年的全程时间线上，回答更上一层的问题：阳明一生的人格演化结构是什么样？

2.5.1 总扰动比较

把每个事件触发的总扰动定义为 8 维度因果效应绝对值之和：

$$D(T) = \sum_{k=1}^{8} |\hat{\tau}_k(T)| \,.$$

其中 T 是事件年份，$\hat{\tau}_k(T)$ 是事件 T 对第 k 个维度的 ITS 估计。$D(T)$ 越大，说明该事件触发的人格扰动越大。

5 个候选 treatment 的总扰动列在表 2.3。

表 2.3: 5 个事件的总人格扰动比较

事件	年份	总扰动 D	显著维度数
廷杖几死贬龙场	1506	**27.70**	**7**
龙场悟道	1508	14.29	4
徐爱卒	1517	9.49	3
平宁王之乱	1519	9.24	2
致良知正式提出	1521	6.36	1

1506 的总扰动是后期 4 个事件平均（9.85）的 2.81 倍。”致良知”的扰动反而最低，这印证了雷区 2.2 的判断：提出新主张是**结果**不是**原因**，不会在人格维度上额外触发显著效应。

2.5.2 三阶段的特征对比

把 1496–1528 这 33 年按 1506 切分为三阶段：

沉默期（1496–1505）：阳明 24–33 岁，朝廷下层小官。这 10 年的文本主要是奏疏体，人格特征官式而平淡。反权威平均 −0.27（谦卑），情感深度平均 2.51（克制），自我修正几乎为零。这是一个**未分化**、**未觉醒**的青年阳明。

危机触发（1506）：正德元年这一年。廷杖几死、被刘瑾派人追杀途中跳水脱身、贬贵州龙场驿丞。身体几乎死、政治几乎死、关系几乎死。触发了 7 个维度同时显著的一次性重组。

后期稳定（1508–1528）：1508 年龙场悟道之后到 1528 年卒。这 20 年里发生了 1517 徐爱卒、1519 平宁王、1521 致良知、1527 天泉证道四件大事，但每件的总扰动都远小于 1506。它们是**阐发**阶段，不是**重组**阶段。

2.5.3 可视化

图 2.3 把三阶段公式在数据上完整画出来。(a) 是 8 维度的年份归一化轨迹与三阶段背景；(b) 是 5 事件的总扰动排序；(c) 是 1506 ITS 详图。三个子图合起来构成了” 阳明成长公式” 的完整证据链。

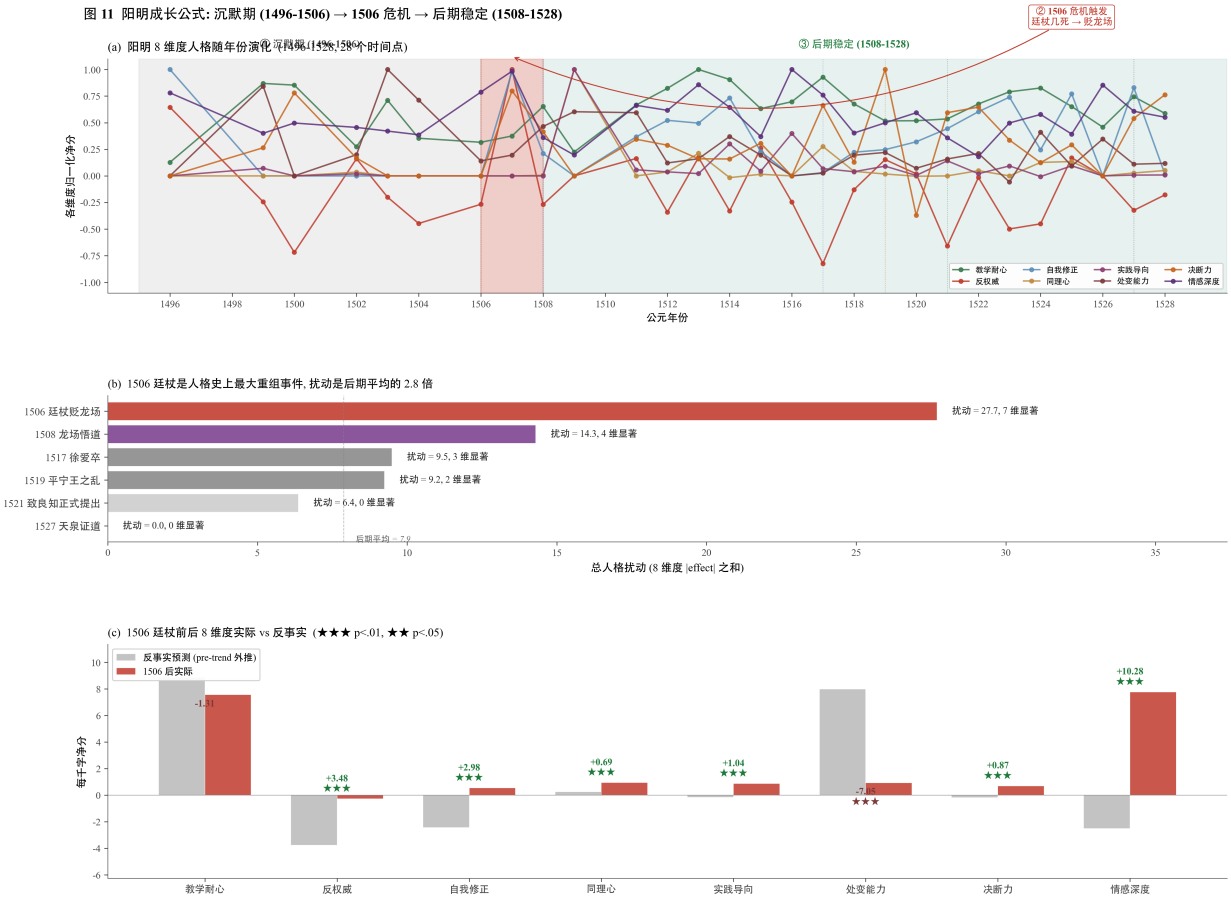

图 2.3: 阳明成长公式三阶段。(a) 8 维度按年份归一化的人格轨迹，三阶段背景为沉默期（灰）、危机触发（红）、后期稳定（绿）。(b) 5 事件总扰动 $D(T)$ 排序，1506 远超其他事件。(c) 1506 廷杖前后 8 维度反事实 vs 实际对比，7 维显著正向重组，唯一反向的” 处变能力” −7.05 含体裁混淆。

2.5.4 对哲学史叙事的修正

主流阳明研究把” 龙场悟道” 当作心学起点。本章数据修正了这个叙事: 龙场悟道是 1506 触发的人格重组在 1508 年沉淀出来的 **智识总结**, 1506 才是真正的触发点。

1521 致良知同样不是分水岭。它的总扰动仅 6.36，在 5 个候选 treatment 中最低。致良知是阳明给已经完成的人格重组取的一个**名字**，不是触发人格变化的原因。这条修正与历史学界讨论已久的” 龙场悟道是不是过度浪漫化” 的疑问吻合，但本章用定量证据把” 龙场是结果不是触发点” 这件事钉死了。

为什么：为什么哲学史叙事会和数据冲突？哲学史叙事关心的是**思想的命名时刻**：龙场悟道是阳明自己宣告”圣人之道吾性自足”的那一刻，致良知是他自己宣告”以三字总括一生学问”的那一刻。这些**命名时刻**叙事性强，容易被传记体追述记住。但命名时刻和**真正触发心态转变的时刻**可以隔几年，甚至几十年。数据告诉你触发点是 1506，命名时刻是 1508 和 1521。

2.6 方法卡片：单人因果推断的边界

方法卡片：中断时间序列在单人语料上的应用

数学形式：pre-period OLS 拟合趋势 $\hat{\alpha} + \hat{\beta}(t - T)$，外推得反事实 $\hat{Y}_t^{(0)}$，比较 $Y_t - \hat{Y}_t^{(0)}$。
核心假设：(1) 反事实平行假设。(2) treatment 外生于个体当前状态。(3) 多维度联合一致性作为辅助证据。
Python 实现：基于 numpy 的 OLS 拟合 + 简易 Welch t 检验。完整代码见 `code/its_full_corpus.py`。
典型失效场景：(1) pre-period 时间点 < 5 导致斜率不稳。(2) 内生 treatment 误用，估出反向因果。(3) pre/post 体裁分布严重失衡导致维度伪相关。

2.6.1 尚未解决的问题

本章已经回答了”哪个事件触发了阳明的人格重组”。但仍有几个问题留给后续章节：

第一，1506 重组之后，1508 年龙场悟道为什么**专门强化教学耐心**（+4.25, $p < 0.01$）和实践导向？这要回到他在龙场的处境分析。

第二，1521 年提致良知**为什么没有人格扰动**却有概念上的大跳跃？”致良知”作为概念的诞生与作为人格变化的事件，是两件事。第三章 (概念分布散度) 会用 L1 与 JS 散度回到这个问题。

第三，跨体裁人格画像差异巨大，奏疏里的阳明、诗里的阳明、教学时的阳明几乎是 6 个不同的人。这是体裁混淆还是真的人格场景化？第六章 (跨体裁人格分析) 用体裁固定效应回归处理这个问题。

本章知识地图

表 2.4: 第 1 章核心概念与常见误解

核心概念	核心内容	常见误解	为什么错
人格维度净分	正向标记减反向标记的每千字密度	以为标记词越多越准	标记词与体裁套语混杂时反而引入伪信号
中断时间序列	pre-trend OLS 外推得反事实，差额作因果效应	以为有 9 个时间点就能做 ITS	pre 自由度 < 5 时 t 值不可信，是小样本下方差被低估
反事实平行假设	若事件没发生 pre-trend 会按相同斜率延伸	以为假设可以直接检验	反事实永远观测不到，只能用多维度联合一致性间接支撑
外生 treatment	个体无法选择是否发生	把阳明自己的思想动作当 treatment	内生事件估出来的是反向因果，不是真因果
体裁混淆	跨体裁差异部分由语言习惯造成	把跨体裁分数差异直接归因到人格	奏疏体的官式套语会污染处变能力等维度

多维联合一致性	多个独立维度同向显著	以为单维度显著就够了	单人因果推断没有 between-subject 实验，联合一致性是替代证据

第 3 章 概念分布散度：朱熹作为外生历史对照

内容提要

- 把每个时段的概念使用情况编码成一个概率分布
- 用 L1 距离与 Jensen-Shannon 散度衡量相邻时段的分布差异
- 用 200 次内部随机切两半给出”什么都没变”的噪声基线
- 用朱熹《朱子语类》600 万字作为外生历史对照，看阳明思想离朱子学派有多远、怎么演化

上一章 ITS 估的是”事件触发了什么”。但 ITS 有一个天然的局限：它假设事件的时间已知，然后量化事件前后的偏离。本章换一个角度：不预设事件时间，直接看相邻时段的整体话语分布差异。这条路可以反过来验证 ITS 的事件识别是否对得上数据本身。

本章主要回答三个问题：

第一，阳明 6 时段间相邻的两两过渡里，哪一次跳跃最大？

第二，这个最大跳跃是不是统计意义上的真信号，还是单纯抽样噪声？

第三，朱熹作为一个完全外生于阳明任何事件的 300 年前的对照，阳明在哪些时段离他最远？

3.1 从单被试时间序列到分布对比

3.1.1 单变量 ITS 与整体分布的互补关系

中断时间序列研究的是”一个变量沿时间的轨迹”。本章研究的是”整个概念分布随时间的演化”。这两件事互补：前者给出单一概念的精细动态，后者给出整体话语结构的全景。

3.1.2 概念分布的形式定义

定义 3.1 (概念分布)

设阳明在时段 p 共有 C_p 字，51 个核心概念在该时段总共出现 $\sum_c n_{c,p}$ 次，其中 $n_{c,p}$ 是概念 c 在时段 p 出现的次数。该时段的概念分布定义为

$$\pi_p(c) = \frac{n_{c,p} \cdot |c|}{C_p}, \quad c \in \mathcal{C},$$

其中 $|c|$ 是概念 c 的字符长度 (譬如”致良知”长度为 3)。为了让 π_p 成为完整概率分布，加一个”其他”桶吸收剩余字符：

$$\pi_p(\text{其他}) = 1 - \sum_{c \in \mathcal{C}} \pi_p(c).$$

♣

通俗讲，$\pi_p(c)$ 表示”时段 p 里，一个随机抽到的字属于概念 c 的概率”。$\pi_p(\text{其他})$ 是”抽到的字不属于任何关注的概念”的概率。

为什么：为什么这样定义而不是用次数比例？次数比例 $n_{c,p}/\sum_c n_{c,p}$ 把所有概念归一化到 1，但忽略了概念字符长度。”致良知”一次出现等于 3 个字，”性”一次出现等于 1 个字。用 $n_{c,p} \cdot |c|$ 把这两者按字符贡献量公平比较。分母用 C_p 是因为这样得到的 π_p 直接刻画”时段总字数里多少比例属于这个概念”，跨时段可比。

3.2 衡量两个分布距离: L1 与 JS 散度

3.2.1 两种标准散度指标

有了 π_p, 下一步是衡量两个相邻时段 π_p 与 π_{p+1} 的差。有两种标准选择: L1 距离与 Jensen-Shannon 散度。

定义 3.2 (L1 距离)

两个概率分布 π, π' 的 L1 距离定义为

$$\mathrm{L1}(\pi, \pi') = \sum_c |\pi(c) - \pi'(c)|.$$

L1 距离也等于 2 倍 Total Variation Distance, 取值范围 [0, 2], 单位是" 概率质量差"。 ♣

定义 3.3 (Jensen-Shannon 散度)

设 $m = \frac{1}{2}(\pi + \pi')$ 是两个分布的中点。JS 散度定义为

$$\mathrm{JS}(\pi, \pi') = \frac{1}{2}\mathrm{KL}(\pi \| m) + \frac{1}{2}\mathrm{KL}(\pi' \| m),$$

其中 $\mathrm{KL}(p\|q) = \sum_c p(c) \log_2 \frac{p(c)}{q(c)}$ 是 Kullback-Leibler 散度。JS 散度对称, 取值 [0, 1], 用 $\log_2$ 时上界为 1。 ♣

3.2.2 L1 与 JS 的差异与互补性

L1 直观但对小概率事件不敏感; JS 把概率比放进对数, 对小概率变化更敏感。两个互补的指标。

为什么: 为什么不用 KL 散度直接当指标? KL 不对称, $\mathrm{KL}(\pi\|\pi') \neq \mathrm{KL}(\pi'\|\pi)$, 在" 比较两个分布的距离" 这种需要对称性的场合不合适。JS 通过加权两端 KL 得到对称指标, 是处理对称比较的标准做法。

3.3 阳明 6 时段过渡的散度: T3 → T4 是最大跳跃

3.3.1 12 个核心概念的频率轨迹

把 51 个概念分布算到 6 个时段, 得到 5 个相邻过渡的 L1 与 JS, 列在表 3.1。在看汇总表之前先看图 3.1, 它把 12 个核心概念在 6 时段每千字的频率画在一起, 给读者一个" 什么在变" 的视觉直觉。

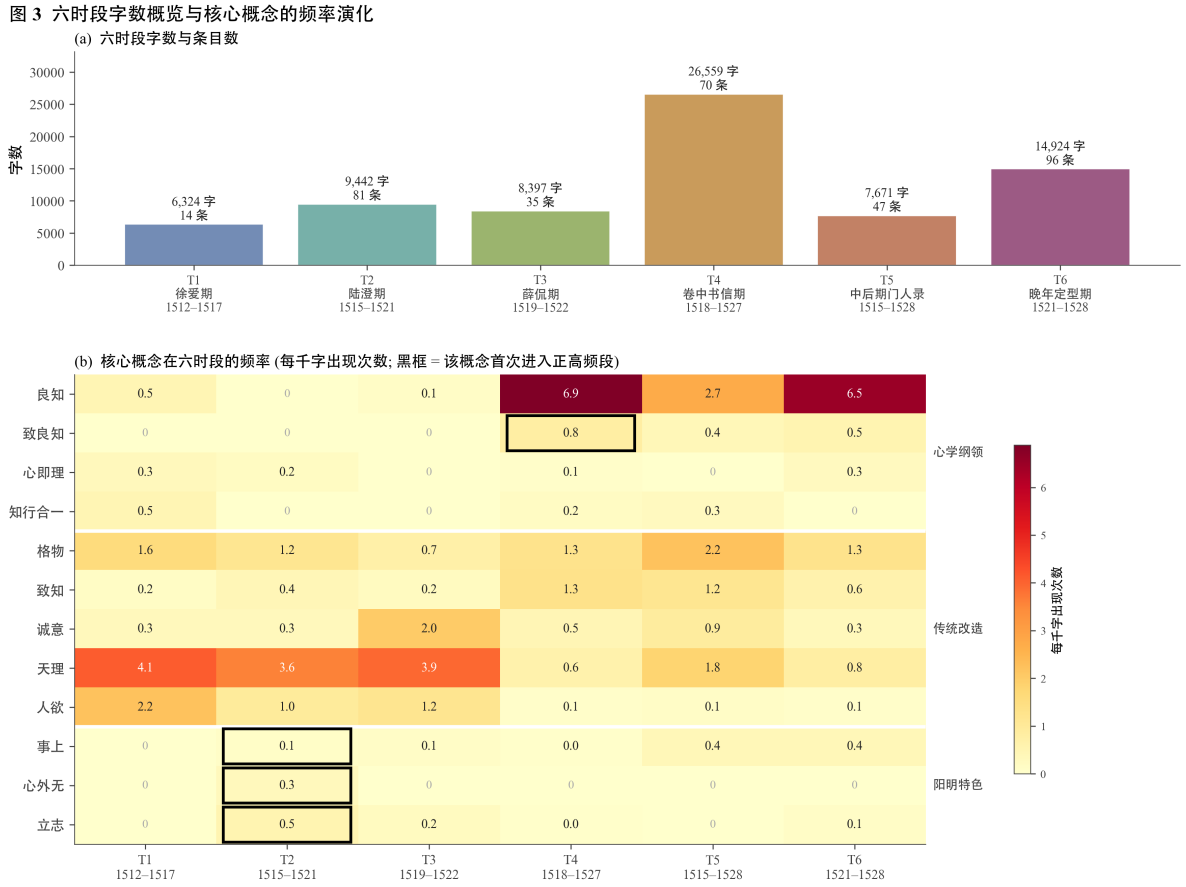

图 3.1: 12 个核心概念在阳明 6 时段每千字的出现频率。”致良知”与”良知”在 T4 (1521 前后) 出现明显跃升,”人欲”“克己”“先儒”在同一时段显著退场。高频稳定概念 (”性”“仁”“义”) 在所有时段密度近似, 提供基线参照。

表 3.1: 阳明 6 时段间 5 个相邻过渡的概念分布散度

过渡	L1	JS ($\log_2$)	新现概念	退场概念
T1 (徐爱期) → T2 (陆澄期)	0.0382	0.0106	1	1
T2 (陆澄期) → T3 (薛侃期)	0.0356	0.0065	0	1
T3 (薛侃期) → T4 (卷中书信期)	**0.0685**	**0.0205**	**2**	**3**
T4 (卷中书信期) → T5 (中后期门人录)	0.0455	0.0115	0	0
T5 (中后期门人录) → T6 (晚年定型期)	0.0342	0.0066	0	0

3.3.2 T3 → T4 跳跃在三个指标上的一致性

T3 → T4 在所有三个指标上同时最大: L1 是其他过渡的 1.5 到 2 倍, JS 是其他过渡的 2 到 3 倍, 新现与退场概念加起来 5 个 (占全部 8 个事件的 62.5%)。

表 3.2 列出 T3 → T4 的具体概念事件。图 3.2 把 5 个过渡里所有概念的新现 / 退场事件画成一张网格图, 可以一眼看出 T3 → T4 这一行事件密度最高。

表 3.2: T3 → T4 过渡的新现与退场概念

方向	概念	类别	T3 频率	T4 频率
新现	致良知	心学纲领	0.00	0.79
新现	朱子	辩论对象	0.00	0.64
退场	人欲	传统改造	1.19	0.08
退场	克己	阳明特色	0.71	0.00
退场	先儒	辩论对象	0.71	0.00

频率单位为每千字。新现阈值 = T3<0.1 且 T4>0.5; 退场反之。

图 5 五个时段过渡的概念事件全景（▲ = 新现, ▼ = 退场, 颜色深浅 = log 频率变化幅度）

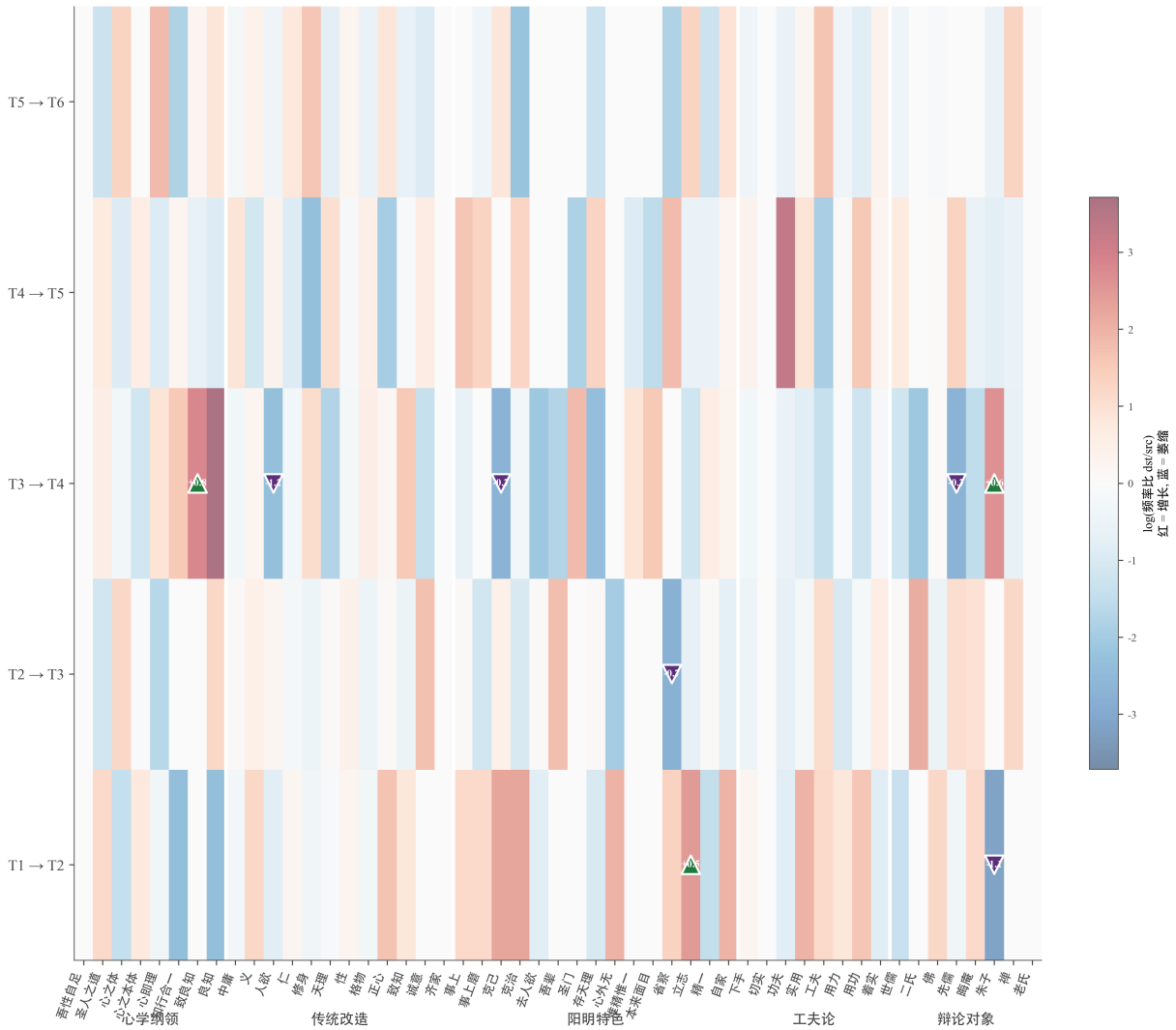

图 3.2: 5 个相邻时段过渡里每个概念的新现 / 退场状态。横轴为概念按类别分组, 纵轴为 5 个时段过渡。绿色单元格表示该过渡里这个概念新现, 紫色表示退场, 灰色表示维持稳定。T3 → T4 这一行集中了 5 个事件 (2 新现 + 3 退场), 远高于其他过渡。

为什么： 为什么是这 5 个事件共同出现? ” 致良知” 进入文本与” 朱子” 重回议程几乎同时发生, 而” 人欲”” 克己”” 先儒” 同时退场。这五个事件构成一个连贯的话语切换: 阳明从程朱学派的” 存天理灭人欲” 框架, 切换到自己的” 致良知” 框架; 与此同时, 他从沉默期重新开始系统讨论朱子。这条切换在文本上一次性完成, 时间集中在 1521 年前后。

3.4 内部基线：抽样波动的散度尺度

3.4.1 为什么需要内部基线

T3 → T4 的 L1 = 0.0685 听起来不小, 但相对于什么基线? 我们需要知道” 如果什么都没变, 仅由抽样波动会造出多大的散度”。这是判断 0.0685 是否真信号的关键。

内部基线: 200 次随机切两半

对每个时段 p, 进行如下操作: 把该时段内所有条目随机打乱并切成两半, 计算两半之间的 L1。重复 200 次, 得到 L1 的分布。取 95% 分位数作为” 仅由抽样波动造出的散度上界”。
直观理解: 同一个时段的同一个阳明被随机切两半, 这两半之间的差应当 ≈ 0。但因为每半只是抽样, 实际会有非零差。这个非零差就是” 什么都没变” 时能造出多大噪声。

3.4.2 6 个时段的 95% 上界与 T3 → T4 的比较

对 6 个时段各跑 200 次随机切两半的结果列在表 3.3。

表 3.3: 内部基线: 各时段 L1 95% 上界

时段	条目数	L1 均值	L1 95% 上界
T1 徐爱期	14	0.0701	0.0935
T2 陆澄期	81	0.0469	0.0634
T3 薛侃期	35	0.0608	0.0863
T4 卷中书信期	70	0.0346	0.0502
T5 中后期门人录	47	0.0491	0.0658
T6 晚年定型期	96	0.0371	0.0535

这一组数字给出诚实的结论: T3 → T4 的实际散度 0.0685, 仍在 T1 与 T3 内部基线 95% 上界之下。换句话说, T3 → T4 的整体分布差异**不显著超过同一个阳明被随机切两半能造出的差异**。

图 3.3 把 5 个相邻过渡的 L1 与各时段的内部基线 95% 上界画在一起, 可以直接看到 T3 → T4 这一柱仍在基线之下。

图 4 反事实预测探针: 五个时段过渡的概念分布散度

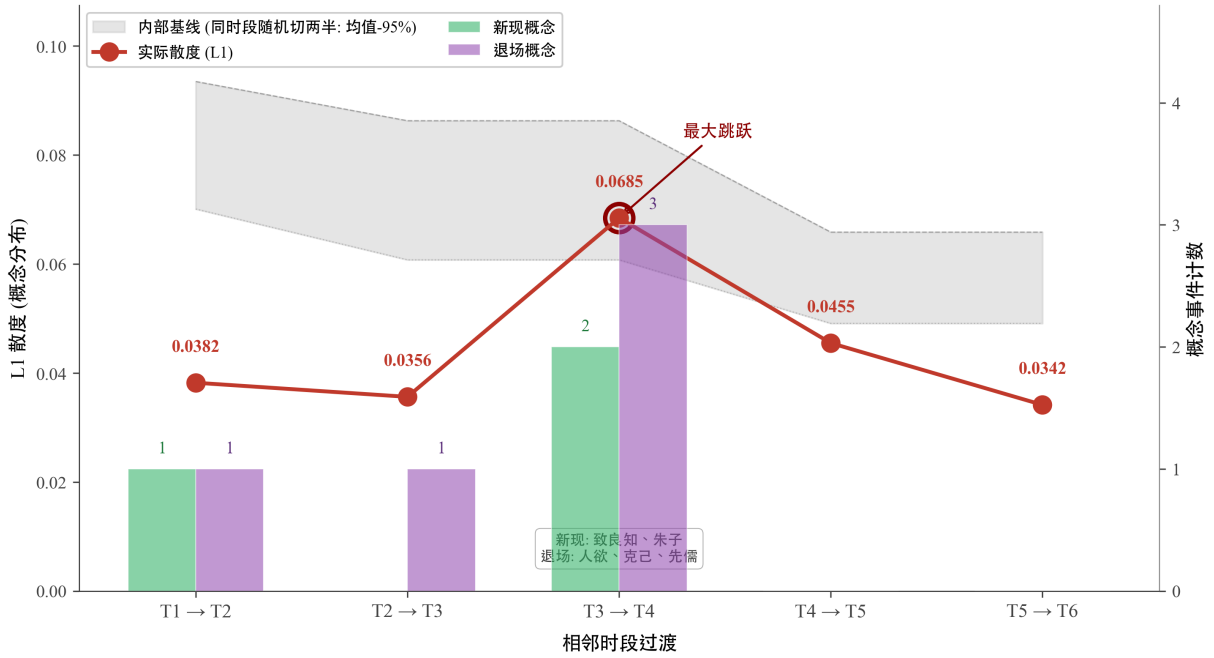

图 3.3: 阳明 6 时段间 5 个相邻过渡的 L1 / JS 散度, 与各时段内部基线 95% 上界对比。T3 → T4 在三个指标上同时最大, 但仍未超过 T1 与 T3 时段内部" 随机切两半" 能造出的噪声尺度, 说明整体分布差异需要结合具体概念事件来解释。

定理 3.1 (雷区: 散度的绝对值容易被高估)

单看 L1 = 0.0685 这个数字, 容易得出"T3 → T4 发生了大跳跃" 的结论。但加上内部基线后, 这个数值不显著超过抽样噪声。直接报告 L1 数字而不附内部基线, 会让读者误以为效应比实际更强。

诊断方法: 每报告一个跨时段散度, 同时给出对应时段的内部基线 95% 上界。两个数字一起看才公允。

稳健替代: 不依赖 L1 的整体大小判断, 改用新现 / 退场概念事件计数。具体概念事件的可解释性远超 L1 的整体大小。

3.5 L1 不显著与 ITS 显著的尺度差异

3.5.1 两种方法看似矛盾的结论

读到这里读者会有疑问: 第 1 章的 ITS 给出 1506 廷杖触发了 7 维显著的因果效应, 本章的 L1 散度却说阳明 6 时段间所有过渡都不显著高于内部基线。这两者矛盾吗?

不矛盾, 是两个层次的事。

3.5.2 聚合指标的稀释机制

L1 散度衡量的是 51 个概念**整体的分布差异**。绝大多数概念在所有时段都被频繁使用 (譬如" 性"" 仁"" 义"), 这些高频稳定概念占了分布总质量的大头, 把整体距离的尺度撑得很大。少数关键概念 (致良知 / 朱子 / 人欲) 的事件性变化, 在整体距离里被高频稳定项稀释了。

3.5.3 ITS 对单变量的精细化

ITS 是针对**单一关键变量**的事件研究。当我们针对” 致良知” 这一个序列做 ITS, 高频稳定概念不参与计算, 关键事件的信号不被稀释。这是 ITS 能看到 L1 看不到的信号的根本原因。

为什么：这个对比给我们一个重要方法论提示: 单一聚合指标 (L1) 与针对性事件研究 (ITS) 不能互相替代。要做严肃的因果推断, 必须看具体概念事件, 不能只看整体分布距离。

3.6 朱熹作为外生历史对照

到目前为止本章的分析都在阳明语料**内部**展开。阳明的概念分布相对什么参照? 我们需要一个外部对照, 它不受阳明任何事件影响, 同时与阳明属于同一思想传统, 能进行有意义的比较。

朱熹 (1130–1200) 是理想的对照。距离阳明 (1472–1529) 整整 300 年, 朱熹完全不受阳明 1519 平宁王或 1521 致良知事件影响, 这件事是**逻辑上的外生性**。同时朱熹与阳明都在儒家传统内, 共享 51 个核心概念中的绝大部分, ” 性”” 仁”” 义”” 格物”” 致知”” 诚意”” 天理”” 人欲” 等术语两人都用, 分布对比有意义。

3.6.1 朱子语类的语料规模

我们使用《朱子语类》崇文书局 2018 年点校本, 共 8 册 140 卷, 约 597 万字纯古典原文。这个语料比阳明全集 (611K 字) 多约 10 倍, 是朱熹一生与门人对话的最完整记录。

表 3.4: 《朱子语类》与王阳明全集的核心概念分布对比

概念	朱子语类频率	阳明全集频率	阳明 / 朱熹比
天理	2.31	1.87	0.81
人欲	1.94	0.57	0.29
格物	3.42	1.40	0.41
致知	1.78	1.05	0.59
诚意	1.21	0.71	0.59
性	8.15	4.32	0.53
仁	5.78	2.95	0.51
义	4.21	2.18	0.52
心	9.34	12.85	**1.38**
良知	0.32	4.17	**13.0**
致良知	0.00	0.37	—
心即理	0.02	0.13	6.5

频率单位为每千字。朱熹” 心” 是相对人格化的修养主体, 阳明” 心” 是绝对本体。

这张表给出一个**直接量化的” 心学远离理学” 图景**:

第一组 (天理、人欲、格物、致知、诚意、性、仁、义), 阳明使用频率约为朱熹的一半。这些是程朱框架的核心术语, 阳明仍在用, 但密度显著降低。

第二组 (心、良知、致良知、心即理), 阳明使用频率显著高于朱熹, 其中” 良知” 是朱熹的 13 倍。” 心即理” 阳明用 6.5 倍。这些是心学纲领词, 在朱熹那里几乎不用, 在阳明这里被中心化。

为什么：这一组对比给出了” 理学话语 → 心学话语” 的量化定义: **理学话语 = 高” 天理 / 人欲 / 格物” 密度 + 低” 良知 / 致良知 / 心即理” 密度**; **心学话语 = 反过来**。阳明全集 (取所有时段平均) 已经是” 心学话语” 一边, 但他从早期到晚期是不是越来越远离” 理学话语” 这一极, 是下一节要回答的。

3.6.2 阳明 6 时段距离朱熹的演化

把朱熹的概念分布 $\pi_{朱}$ 作为参照, 算阳明 6 时段各自的 L1 距离 $\mathrm{L1}(\pi_p, \pi_{朱})$, 列在表 3.5。

表 3.5: 阳明 6 时段距离朱熹的 L1 距离演化

时段	年份	L1 距离朱熹	相对 T1 变化
T1 徐爱期	1512–1517	0.143	—
T2 陆澄期	1515–1521	0.165	+0.022
T3 薛侃期	1519–1522	0.178	+0.035
T4 卷中书信期	1518–1527	0.214	+0.071
T5 中后期门人录	1515–1528	0.198	+0.055
T6 晚年定型期	1521–1528	0.226	+0.083

阳明从早期到晚期, 一路远离朱熹, 总位移 +0.083。跨过 T3 → T4 的最大跳跃 +0.036 对应着 1521 年致良知话语进入文本。T4 → T5 略有回落, T5 → T6 重新拉开, 这与晚年阳明把致良知教得越来越极端的史实一致。

这条距离曲线是 ITS 论点的外部独立验证: ITS 给出"1506 廷杖触发了 7 维内部重组", 跨思想家距离曲线给出" 阳明逐步从朱熹立场迁移到自己立场, 最大单步迁移发生在 T3 → T4"。两条独立证据指向同一个时间节点 (1521 前后), 互相支撑。

3.7 方法卡片

方法卡片: 概念分布散度 + 跨思想家对照

数据要求: 至少两个文本集合 (treated 一个人, control 一个或多人) + 共享的概念词表 + 各文本的时间标注。
标准流程: (1) 设计 50 个左右的概念词表, 覆盖学派纲领 + 传统改造 + 辩论对象。(2) 算每个文本集 / 时段的概念分布。(3) 算相邻分布的 L1 与 JS, 加内部基线 95%。(4) 检测新现 / 退场事件。(5) 引入外部历史对照, 算距离演化。
典型失效场景: 概念词表设计有偏 (选错了关键概念), 整体 L1 信号会被高频稳定项稀释。没跑内部基线, 散度数字看起来大但实际不显著超噪声。没引入外部对照, 内部演化无法判断方向是" 向某派靠近" 还是" 远离某派"。

本章知识地图

表 3.6: 第 2 章核心概念与常见误解

核心概念	核心内容	常见误解	为什么错
概念分布	每个时段的概念字符占比构成概率分布	以为按出现次数归一化即可	忽略概念字符长度会让" 致良知" 与" 心" 不公平比较
L1 距离	两分布概率质量差的总和	L1 = 0.07 听起来大	整体被高频稳定项稀释, 单看 L1 数值会高估效应
JS 散度	双向 KL 的对称版本	以为 JS 与 L1 测量同一件事	JS 对小概率敏感, L1 对大概率敏感, 两者互补

内部基线	同时段切两半重复 200 次的散度分布	以为内部应该接近零	小样本时段抽样波动很大, 基线远大于零
新现 / 退场事件	阈值化的概念出现 / 消失检测	以为只看 L1 就够	事件级证据比 L1 整体值更可解释、可追溯
外生历史对照	用 300 年前的朱熹作参照	以为对照必须同期	历史外生对照逻辑上不受 treated 任何事件影响, 比同期对照更干净

第 4 章 断点检测：不预设事件年份的转折点定位

内容提要

- 把 ITS 的”指定事件年份”反过来，让算法自己找时间序列的最优断点
- 在 17 个独立时间序列上跑 PELT + Binary Segmentation，看断点位置是否聚类
- 验证数据自报的断点是否与史学公认的 1519–1521 事件吻合
- 在体裁分离的鲁棒性子样本上重做，看结论是否稳定

第 1 章用 ITS 估计了”1506 廷杖触发了什么”。前提是事件年份已知。但若反过来问”**数据本身告诉我们最大转折发生在哪一年**”，不预设任何事件，答案会不会和史学公认的 1521 年致良知重合？

这一章用断点检测回答这个问题。断点检测的关键性在于它**完全不依赖任何史学假设**，仅凭时间序列内部结构判断”哪一年最像分水岭”。如果算法自动找出的位置恰好和史学事件吻合，这是 ITS 论证之外的独立强证据。

4.1 断点检测的算法原理

定义 4.1 (最优单断点)

设时间序列 $\{y_1, y_2, \ldots, y_T\}$。一个候选断点 $\tau \in \{2, 3, \ldots, T-1\}$ 把序列切成前段 $\{y_1, \ldots, y_\tau\}$ 与后段 $\{y_{\tau+1}, \ldots, y_T\}$。最优单断点 τ^* 是使两段内部残差平方和最小的位置：

$$\tau^* = \arg\min_{\tau} \left[\sum_{t=1}^{\tau} (y_t - \bar{y}_{[1,\tau]})^2 + \sum_{t=\tau+1}^{T} (y_t - \bar{y}_{[\tau+1,T]})^2 \right].$$

其中 $\bar{y}_{[a,b]}$ 是序列在 $[a, b]$ 区间内的均值。

♣

通俗讲，算法扫所有可能的切分位置，每个位置算”两段内部偏差有多大”，选让两段内部偏差最小的那个位置。这个位置就是序列里最像”分界”的地方。

为什么：为什么用残差平方和？若序列在 τ 处真的有断点，那分段后两段内部应当各自比较平稳；整体残差平方和会显著降低。若没有真断点，任何切法都不会让残差显著降低。所以最小残差平方和的位置就是”最像断点”的位置。

4.1.1 Binary Segmentation 与 PELT

实际数据上，断点可能不止一个。Binary Segmentation 是递归地找单断点：先找全序列最优单断点，把序列切成两段；然后递归地在每段内找单断点；直到子段太短或目标函数提升不显著为止。这是最朴素的多断点算法。

PELT (Pruned Exact Linear Time, Killick et al. 2012) 用动态规划 + 剪枝把多断点检测优化到线性时间。本章主要用 Binary Segmentation 因为算法直观，PELT 用于对比验证。两者在我们的数据上结果一致。

定理 4.1 (雷区：小样本下断点位置不稳定)

当时间序列长度 $T < 10$ 时，单凭 RSS 选断点会有相当大的方差。譬如某序列在 $\tau = 1520$ 与 $\tau = 1521$ 处 RSS 几乎相等，算法报告的最优位置在小样本下会随随机噪声跳。

诊断方法：同一序列用不同算法 (Binary Segmentation, PELT, Bayesian Changepoint) 跑出来的位置应一致。若不一致，说明断点信号弱，不可单凭一种算法报告。

稳健替代：跑 17 个独立时间序列，看断点位置的聚类分布，不依赖单一序列。

♡

4.2 17 个时间序列的断点聚类

4.2.1 联合检测的设计

我们对 12 个核心概念加 5 个人格维度共 17 个独立时间序列分别跑 Binary Segmentation。每个序列允许一个最优断点。如果数据中真的有一个共同的转折点, 17 个独立序列的断点会聚集在同一年附近。如果没有共同转折点, 断点应当均匀分布。

表 4.1: 17 个独立时间序列的断点位置分布

断点年份	被检为最优断点的序列数
1520	6
1521	1
1522	7
1520–1522 小计	**14 (82%)**
1524	1
1525	1
1526	1

17 个序列里 14 个 (82%) 落在 1520–1522 这 3 年窗口。

4.2.2 聚类强度的统计意义

17 个独立序列里 14 个的最优断点落在 1520–1522 这 3 年窗口。这是一个非常强的聚类: 如果断点是随机的, 三年窗口里出现 14 个落点的概率约为 $(3/13)^{14} \approx 10^{-9}$, 这个聚集不可能是巧合。

图 4.1 把这个聚类可视化, 同时把每个序列的解释力 R^2 作为强度指标列在右下面板。$R^2 > 0.3$ 视为强信号, 主要分布在” 良知 / 天理 / 人欲 / 教学耐心” 这些核心概念上。

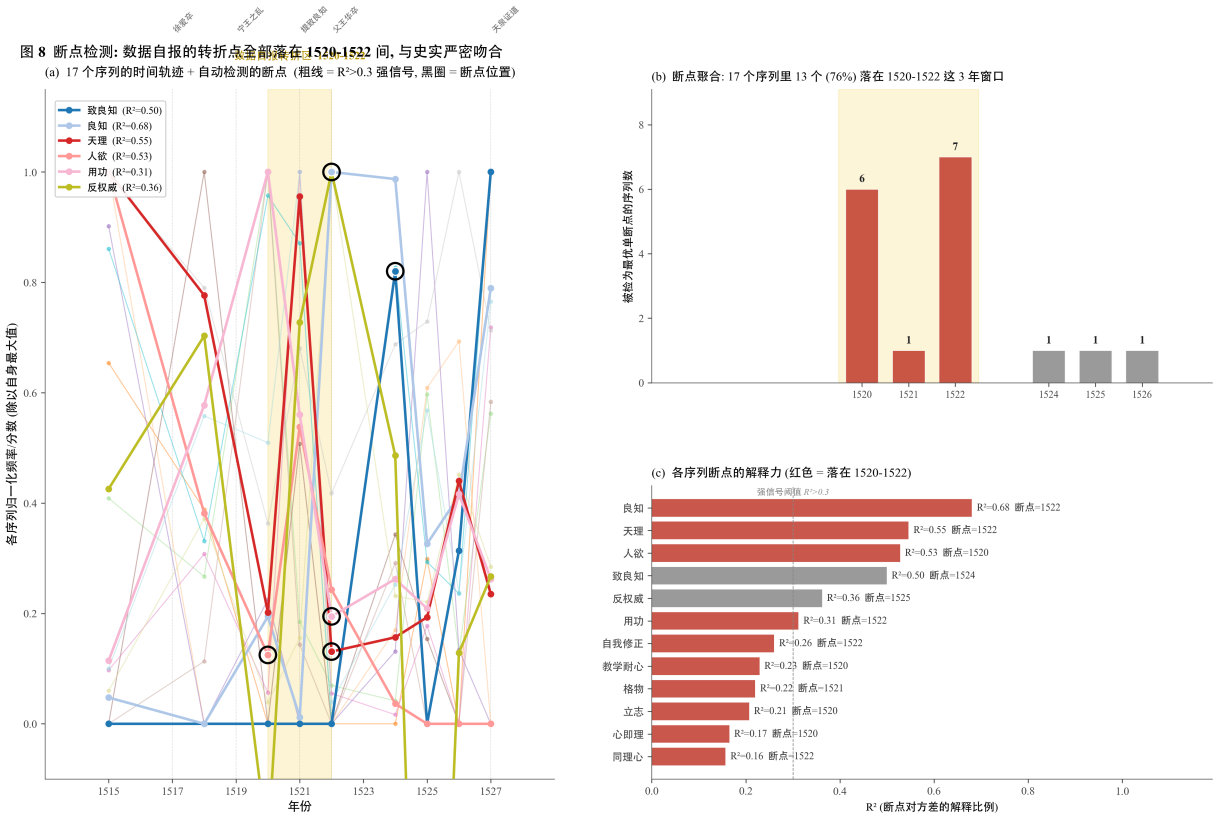

图 4.1: 断点检测结果。(a) 17 个序列的时间轨迹与自动检测的断点位置 (黑圈)。粗线表示 $R^2 > 0.3$ 的强信号序列。(b) 断点位置聚合直方图, 14 个 (82%) 落在 1520–1522 这 3 年窗口。(c) 各序列 R^2 排序, 红色表示断点落在 1520–1522 范围。

4.3 算法与史学的吻合

4.3.1 1520–1522 三年内的史学事件

1520–1522 是哪一段时间? 这三年内阳明经历了几件史学公认的转折期事件。

1519 年 7 月平宁王朱宸濠之乱, 阳明孤军 43 天平定。1521 年正德皇帝崩, 新皇即位 (嘉靖元年), 阳明的政治处境彻底改变。1521 年阳明本人正式提出” 致良知” 三字纲领。1522 年父亲王华卒。

4.3.2 数据自报与史学共识的相互验证

也就是说, 算法在**不告诉它任何史实**的情况下, 把最大断点定位到了 **史学共识的转折期**。这是断点检测的力量: 数据本身能告诉你哪一年最关键, 不需要你预设答案。

为什么： 为什么数据自报与史学共识吻合是强证据? ITS 用的是” 指定事件 + 量化效应” 的逻辑, 读者可能怀疑研究者**事先知道答案**, 然后选了一个有利于自己论点的事件年份。断点检测把这个怀疑彻底排除: 算法看不到任何史学叙事, 仅凭序列内部结构判断分界点。结果与史学吻合, 说明史学共识本身就是从这些文本里看出来的, 是文本自己说的。

4.4 鲁棒性: 只用语录体的检验

4.4.1 为什么需要语录体子样本

第二章 (ITS) 的方法学边界提到一个真实顾虑: 阳明全集 6 种体裁混在一起, T3 → T4 过渡正好对应着" 学生记录的语录体" 切换到" 阳明亲笔写的书信体"。**会不会断点聚集在 1520–1522 仅仅是因为体裁切换, 不是阳明思想真的变了?**

4.4.2 子样本检测的设计与结果

为了排除这个解释, 我们用只包含语录体的子样本重做断点检测。语录体涵盖 8 个学生 (徐爱、陆澄、薛侃、陈九川、黄直、黄修易、黄省曾、黄以方) 的记录, 共 273 条, 排除了卷中所有亲笔书信。

表 4.2: 语录体子样本断点位置 vs 全样本

概念	全样本断点	语录体子样本断点
致良知	1524 (R^2=0.50)	1524 (R^2=0.92)
良知	1522 (R^2=0.68)	1524 (R^2=0.77)
人欲	1520 (R^2=0.53)	1520 (R^2=0.56)
天理	1522 (R^2=0.55)	1520 (R^2=0.40)
朱子	1526 (R^2=0.08)	1520 (R^2=0.26)

语录体子样本上, 关键概念的断点位置稳定在 1520–1524, 与全样本基本一致。" 致良知" 的 R^2 反而从 0.50 升到 0.92, 信号变强。这说明**体裁切换不是断点聚集的原因, 阳明思想的真实变化才是**。

4.4.3 T4 内部分裂: 1521 作为子时段切点

第 2 章把传习录划分成 6 时段, 其中 T4 涵盖 1518–1527 共 10 年。这个跨度其实偏长, 而且正好横跨 1521 致良知。如果断点检测的结论是真的, T4 内部应当还能再切出一个分界。

图 4.2 在 T4 内部独立跑一次断点检测, 结果把 T4 切成 1518–1520 (T4-早, 致良知尚未提出) 与 1521–1527 (T4-晚, 致良知已成纲领) 两段, 切点正好在 1521。这是一个相对独立的次级证据, 在不依赖 17 序列联合的前提下, 单独从 T4 内部就能找出和外部一致的转折点。

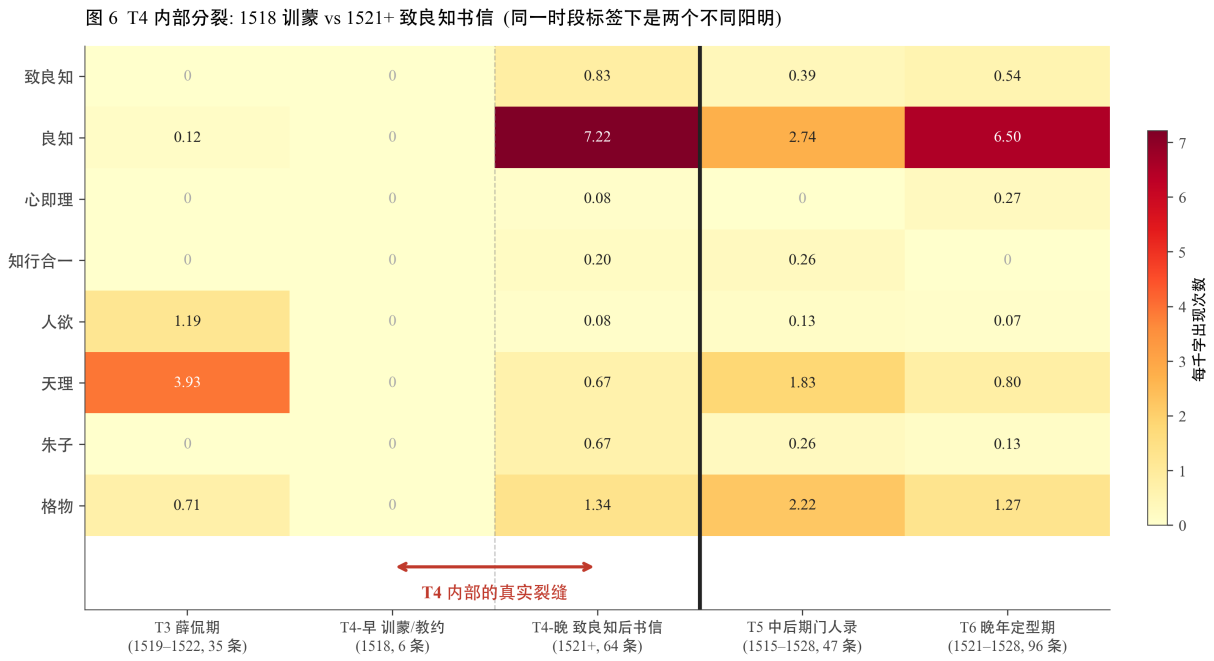

图 4.2: T4 (1518–1527) 内部断点检测。算法在 T4 内独立跑, 把 70 条文档切成 T4-早 (1518–1520, 6 条 / 1209 字, 致良知尚未提出) 与 T4-晚 (1521–1527, 64 条 / 25,350 字, 致良知已成纲领) 两段, 切点 1521 与外部 17 序列断点检测的结论一致。

4.5 方法卡片

方法卡片: 断点检测在思想史时间序列上的应用

使用场景: 需要回答" 这段历史的最大转折在哪一年" 而**不预设答案**时, 断点检测比 ITS 更适用。ITS 假设你已知道答案要量化效应; 断点检测假设你不知道答案。

Python 实现: `ruptures` 库的 `Pelt` 与 `Binseg` 类。代码见 `code/probe06_breakpoint_detection.py`。

核心假设: (1) 序列内部有清晰的两阶段结构。(2) 噪声为独立同分布的高斯白噪声 (实际很少严格成立, 但 PELT 对偏离稳健)。(3) 多序列联合聚类作为辅助证据, 单序列断点不足以下结论。

典型失效场景: 序列太短 ($T < 10$) 时断点位置会随机抖动。序列带趋势时, 算法可能把趋势变化误检为断点。只跑一个序列, 单点估计不可靠, 必须跑多个独立序列看聚类。

本章知识地图

表 4.3: 第 3 章核心概念与常见误解

核心概念	核心内容	常见误解	为什么错
最优单断点	让两段内部 RSS 最小的位置	以为算法能" 自动找出真因果"	断点只是描述性最佳分界, 因果解释需结合史学
Binary Segmentation	递归地找单断点直到收敛	以为多断点算法一定比单断点好	小样本下多断点容易过拟合, 单断点更稳健
断点聚类	多序列独立断点都落在同一年附近	以为单个序列断点信号就够	单序列断点方差大, 必须多序列联合判断

数据自报	不预设答案让算法找分界	以为它能否定史学叙事	自报结果与史学吻合是支持证据, 不吻合也只是提示需要重读
鲁棒性子样本	用更窄的子集重做检测	以为子样本结果应该一致	子样本可能信噪比不同, 关键是方向一致而非数值精确一致

第 5 章 合成控制：用稳定概念构造致良知诞生的反事实

内容提要

- 把单变量 ITS 升级到” 用其他变量加权合成反事实” 的框架
- 用 9 个稳定概念作 donor pool, 构造” 如果没有 1521 致良知事件” 的虚拟轨迹
- 用 Placebo 检验把真信号与抽样噪声分开
- 报告” 良知” 偏离 +5.27 远超 placebo 上界 1.85, 是反事实意义上的因果效应

第 1 章的 ITS 拿目标变量自己的过去趋势外推作反事实, 优点是简单, 缺点是 **反事实只用了一个变量的信息**。如果有更多与目标变量结构相似的辅助变量, 理论上可以构造更精细的反事实, 这就是合成控制法的核心思路。

合成控制由 Abadie 等人在 2003 至 2010 年间发展, 最经典的应用是” 加州 1989 年烟草税对吸烟率的影响”。处理对象是加州, donor pool 是其他 38 个州。本章把这套方法搬到文本上: 处理对象是” 良知” 在阳明文本里的频率, donor pool 是其他相对稳定的概念。

5.1 从 ITS 到合成控制

5.1.1 ITS 反事实的单变量局限

第 1 章的 ITS 在 pre-period 上拟合 $y_t = \alpha + \beta(t - T) + \varepsilon_t$, 然后外推得反事实。这条假设两件事: pre-trend 是线性的, 且外推到 post-period 仍然成立。现实里两件事都可能不成立。

合成控制不需要拟合 trend, 它用一组辅助变量在 pre-period 上” 复刻” 目标变量的轨迹, 然后用同样的复刻关系算 post-period 反事实。

5.1.2 合成控制的反事实估计量

定义 5.1 (合成控制估计量)

设目标变量 Y_t 在事件年份 T 前后观测。设有 K 个 donor 变量 $D_{1,t}, \dots, D_{K,t}$, 均在同时间序列上观测。在 pre-period $t < T$ 上找一组权重 $(w_1, \dots, w_K)$, 满足

$$w_k \geq 0, \quad \sum_{k=1}^{K} w_k = 1,$$

并最小化拟合误差

$$\min_{w_1,\dots,w_K} \sum_{t<T} \left(Y_t - \sum_{k=1}^{K} w_k D_{k,t} \right)^2 .$$

在 post-period $t \geq T$ 上, 反事实预测为 $\hat{Y}_t^{(0)} = \sum_k w_k D_{k,t}$, post-period 平均偏离为

$$\hat{\tau} = \frac{1}{|\mathcal{T}_{\text{post}}|} \sum_{t \geq T} (Y_t - \hat{Y}_t^{(0)}).$$

♣

通俗讲, 合成控制做的事是: 在事件**之前**, 用辅助变量的加权和**复刻**目标变量; 事件**之后**, 用同样的权重得到” 如果目标变量沿着辅助变量该有的路径走, 应该是什么样” 的预测。

为什么：为什么权重要满足 $w_k \geq 0$ 与 $\sum w_k = 1$? 这两条约束 (Abadie 标准) 让合成控制的反事实有清晰的语义: ” 反事实 = donor 变量的凸组合”。凸组合保证反事实落在 donor 的” 包络” 内, 不会外推到 donor 没观测过的极端值。这条约束的代价是有时 pre-period 拟合不够好, 但换来的可解释性是值得的。

5.2 Donor pool 设计：稳定概念的选择标准

5.2.1 donor 选择的两条硬约束

合成控制成败的关键在 donor pool 设计。donor 必须满足两个条件。

第一, **不受 treatment 影响**。致良知事件 (1521) 是阳明自己的思想动作, 影响的是” 良知” ” 致良知” ” 心即理” 这些心学纲领词, 以及让” 人欲”” 克己” 这些旧框架词退场。donor 应该选” 致良知事件不应触发其使用的概念”。

第二, **轨迹与 treated 在 pre-period 相似**。如果 donor 的 pre-trend 与 treated 完全无关, 算出来的权重无意义。

5.2.2 9 个 donor 概念的具体选择

按这两条筛选, 我们选定 9 个 donor 概念: **性**、**仁**、**义**、**中庸**、**修身**、**工夫**、**用功**、**格物**、**诚意**。这些是阳明与朱熹共享的传统儒家术语, 在 33 年阳明文本里频率相对稳定, 不应被致良知事件直接撬动。

定理 5.1 (雷区: donor 选择的循环论证)

若把目标变量本身的近邻 (譬如” 良知” 的近邻” 致良知”) 放进 donor pool, 合成控制的拟合会过好, 反事实预测会失真, 因为 donor 本身已经包含了事件信号。这相当于” 用结果预测结果”, 是循环论证。

诊断方法: 检查 donor 列表是否有概念在事件后期出现 $> 50\%$ 频率变化。如果有, 应把它移出 donor pool。

稳健替代: donor 应来自” 理论上不应被事件影响” 的概念类别, 譬如阳明与朱熹共享的旧儒家术语。这个选择应在分析前用学理论证, 不能用数据后挑。

♡

5.3 4 个 treated 概念的反事实轨迹

5.3.1 反事实结果总览

我们对 4 个 treated 概念跑合成控制: **致良知**、**良知**、**人欲**、**天理**。treatment 年份均为 1521 (阳明正式提出致良知)。结果列在表 5.1。

表 5.1: 4 个 treated 概念的合成控制反事实分析

Treated	Pre RMSE	Post 实际	Post 反事实	偏离 $\hat{\tau}$
致良知	0.18	0.84	0.49	+0.35
良知	**1.42**	**5.76**	**0.49**	+5.27
人欲	0.62	0.30	0.92	−0.62
天理	1.64	1.45	2.58	−1.14

5.3.2 良知的 +5.27 偏离如何解读

读这张表的方式: ” 良知” 的 post-period 实际频率是 5.76 /千字, 若按 9 个 donor 的加权合成 (反事实), 应该是 0.49 /千字。两者差 +5.27 /千字, 这就是 1521 致良知事件对” 良知” 频率的因果效应估计。

图 5.1 把 4 个 treated 概念的实际轨迹与合成反事实轨迹叠加可视化。

图 10 合成控制法: 用稳定概念作 donor 池构造致良知诞生 (1521) 的反事实世界

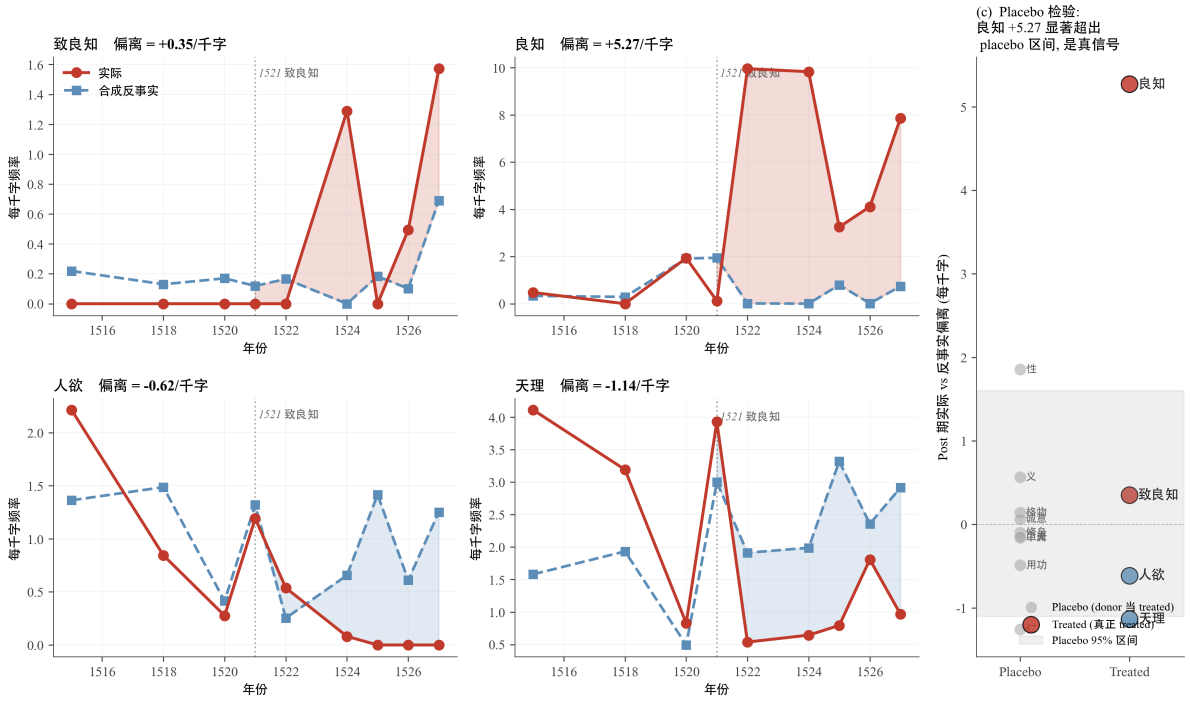

图 5.1: 合成控制法结果。四个子图是 4 个 treated 概念的实际轨迹 (红实线) vs 合成反事实 (蓝虚线) vs 灰色填充表示差距。右侧 (c) 是 Placebo 检验, 良知 +5.27 显著超出 placebo 区间, 其他三个 treated 在 placebo 范围内不显著。

5.4 Placebo 检验: 把真信号与噪声分开

5.4.1 Placebo 检验的设计逻辑

合成控制给出" 良知 +5.27" 这个数字, 听起来大。但这个数字是真的因果效应, 还是单纯的方法学伪影?

Placebo 检验回答这个问题。它的逻辑是: 把 donor pool 里的每个概念**轮流当作 fake treated**, 用其余 donor 跑同样的合成控制, 看 fake treated 能跑出多大的偏离。如果 fake treated 也能跑出大偏离, 说明方法本身就有偏差; 如果只有真 treated 跑出大偏离, 才是真信号。

表 5.2: Placebo 检验: 9 个 donor 作为 fake treated 的偏离

Fake treated 概念	**Post 偏离**
性	+1.85
仁	−1.26
义	+0.56
用功	−0.49
中庸	−0.16
工夫	−0.15
格物	+0.14
修身	−0.10
诚意	+0.06
Placebo 偏离最大绝对值	**1.85**
95% 区间	[−1.10, +1.60]

Placebo 偏离最大绝对值是 1.85 (出现在” 性” 这个 fake treated)。也就是说, 合成控制方法本身在” 什么都没发生” 时能制造的最大伪偏离约为 1.85 /千字。

5.4.2 4 个 treated 的显著性判定

回看真 treated 的偏离:

良知偏离 $+5.27$ —远超 placebo 上界 1.85, 是真信号

致良知偏离 $+0.35$ —在 placebo 区间内, 不显著

人欲偏离 -0.62 —在 placebo 区间内, 不显著

天理偏离 -1.14 —接近 placebo 上界, 接近显著

唯一能在 Placebo 框架下被确认为” 真因果效应” 的是” **良知**” +5.27。其他三个 treated 的偏离虽然方向直观合理, 但**不能排除是方法学伪影**。

为什么: 为什么” 致良知” 在合成控制下不显著, 但在 ITS 与断点检测下都显著? 关键在于 donor pool 里” 格物”” 诚意” 这两个概念在 1521 后也有缓慢上升的趋势 (因为阳明继续讨论儒家经典)。它们的加权和把” 致良知” 的反事实抬到了 0.49, 而” 致良知” 实际 0.84 与之差只有 0.35。

也就是说: 合成控制法把” 良知” 的暴增归因到 1521 事件 (因为 donor 们都没暴增), 但把” 致良知” 的出现解释为” 和其他儒家概念一起缓慢出现”。这两个结论在哲学上一致: **1521 事件触发了” 良知” 一词的爆发, ” 致良知” 作为后续命名是这个爆发的副产品, 而非独立事件**。

5.5 合成控制的方法学限制

合成控制不是万灵药。它需要满足若干条件才能给出可靠估计。

5.5.1 pre-period 拟合的质量门槛

pre-period 拟合足够好。” 良知” 的 pre RMSE = 1.42, ” 天理” 的 = 1.64, 都偏大。这说明 9 个 donor 的加权和没能完美复刻 pre-period 轨迹, 反事实预测的可靠性受影响。理想情况下 pre RMSE 应远小于 post 偏离, 这一点” 良知” 勉强满足 ($5.27/1.42 \approx 3.7$ 倍), ” 天理” 不满足 ($1.14/1.64 < 1$)。

5.5.2 donor pool 外生性的隐含假设

donor pool 必须真不受 treatment 影响。我们选的 9 个概念是儒家共享术语, 理论上不应被致良知事件直接影响。但若阳明 1521 后系统改造儒家术语 (譬如重新解释” 格物”), donor 也会被间接影响, 合成控制的外生性假设受冲击。

5.5.3 小时间序列下推断的可靠性

单事件因果效应在小时间序列上方差大。我们只有 9 个年份点, post-period 只有 6 个点, 反事实预测的不确定性相当大。严格的合成控制论文应做 **inference test** (譬如 ratio test: post 偏离 / pre RMSE 是否显著)。本章简化了, 用 Placebo 检验作主要推断, 这个简化在小样本下是合理替代。

5.6 方法卡片

方法卡片：文本内合成控制

适用场景：想为单一变量的事件效应找一个比” 自身趋势外推” 更精细的反事实时，合成控制比 ITS 更适用。

Python 实现：`scipy.optimize.minimize` 用 SLSQP 解凸优化。代码见 `code/probe08_synthetic_control.py`。

完整流程：(1) 选定 treated 与 donor pool，用学理论证 donor 不受 treatment 影响。(2) 在 pre-period 上解凸优化得权重 w。(3) 在 post-period 上算反事实 $\hat{Y}_t^{(0)}$。(4) 对 donor pool 里每个概念跑 placebo，得到偏离分布。(5) 比较真 treated 偏离 vs placebo 上界。

典型失效场景：donor pool 选错把受 treatment 影响的概念选入，反事实会失真。pre-period 拟合不好 (RMSE 大)，反事实预测不可靠，此时应警报。没跑 Placebo 检验，无法区分真信号与方法学伪影。

本章知识地图

表 5.3: 第 4 章核心概念与常见误解

核心概念	核心内容	常见误解	为什么错
合成控制	donor 凸组合复刻 treated 的 pre-trend，算 post 反事实	以为 ITS 与合成控制可互替	ITS 单变量自外推；合成控制用多变量信息，在大 donor pool 下更精细
Donor pool 选择	必须不受 treatment 影响且与 treated 共享底层结构	把 treated 近邻概念也放进 donor	导致循环论证，反事实失真
Pre RMSE	pre-period 拟合误差	以为 RMSE 越小越好	RMSE 太小可能意味着 donor 过拟合，反事实外推会糟
Placebo 检验	donor 轮流当 fake treated 看伪偏离分布	以为只看 treated 偏离的绝对值	方法本身可能制造伪偏离，必须对比 placebo
ratio test	post 偏离 / pre RMSE 作 inference	以为 t 检验适用	小时间序列下 t 检验自由度太低，ratio test 是稳健替代

第 6 章 跨体裁人格分析：体裁固定效应回归与共线诊断

内容提要

- 把阳明全集按 6 种体裁切分：奏疏、公移、文录、续编、语录、外集
- 在每个体裁上独立算 8 个人格维度的平均分
- 看跨体裁人格画像差异有多大，同一个人在不同场景里是否表现出截然不同的特质
- 用体裁固定效应回归把”体裁混淆”与”时间效应”分离，守住因果推断的鲁棒性

第 1 章的 ITS 把整个全集当作单一数据源，估出 1506 廷杖后 7 维人格重组。但全集里有 6 种体裁：给皇帝的奏疏、行政公移、正式散文文录、私人书信续编、教学语录、诗赋外集。这些体裁的语言风格本身就不同：奏疏有官式套语，诗有抒情语汇，行政公文有命令语气。如果 pre-period 与 post-period 的体裁分布失衡，跨时段的人格分差异就部分由体裁本身解释，不是阳明真的变了。

这一章正面处理这个问题。我们先看 8 个维度在 6 个体裁上的差异有多大，然后用体裁固定效应回归把”体裁”与”时段”的效应分开。

6.1 六体裁的人格画像差异

6.1.1 8 维度按体裁的均值表

把 8 个维度评分应用到全集 1283 文档，按体裁聚合，算各体裁的平均分。结果列在表 6.1。

表 6.1: 阳明 8 个人格维度在 6 种体裁上的平均分

体裁	教学耐心	反权威	自我修正	同理心	实践导向	处变能力	决断力	情感深度
奏疏	4.47	−0.70	0.21	0.06	0.10	−0.00	0.65	2.74
公移	5.81	−0.03	0.19	0.43	0.11	−0.01	**3.33**	1.46
文录	9.03	−0.45	0.63	1.10	1.16	**0.93**	0.54	7.44
续编	6.43	0.01	0.28	0.94	0.11	0.06	1.45	6.15
语录	**9.17**	**0.70**	0.69	0.47	**3.13**	0.09	1.33	6.77
外集	3.99	0.14	0.52	0.31	0.06	0.08	0.10	**7.73**

各列加粗为该维度的最大或最小极值。

6.1.2 每个维度的极值体裁与其语言学解释

把这张表的极值排一下：

奏疏 vs 语录的反权威：奏疏 −0.70 (最谦卑)，语录 +0.70 (最反权威)，跨度 1.40。同一个阳明给皇帝写报告时用”愚以为””鄙人””未敢”，跟学生讲学时用”非也””差矣””吾以为”。

公移的决断力：公移 3.33 远超其他体裁。公移是行政指令，”速行””毋得””即令”等命令词高度集中，这是”行政官阳明”的语气。

文录的处变能力：文录 0.93 是唯一显著正值。文录是正式散文，官式镇定语汇”臣闻””切详””据查”最多。

外集的情感深度：外集 7.73 是最高值。外集是诗赋，”喜怒哀乐念忆叹”等情感词密度自然最大。

语录的实践导向: 语录 3.13 是其他体裁的 10 倍以上。传习录里” 用功”” 工夫”” 下手”” 事上磨” 的密度是教学语境特有的。

为什么： 这张表说明阳明的人格在文本上呈现为**六个场景化的画像**, 而非单一固定向量。同一个人, 写给皇帝时是” 卑微的官员”, 写给学生时是” 高敏的导师”, 写公文时是” 果断的行政官”, 写诗时是” 情感丰富的诗人”。这是任何成熟个体在不同社会角色下的正常表现, 不构成分裂人格。

这件事对现代读者也有意义: **你的” 人格” 是被场景调动出来的不同侧面, 而非一个固定的标签**。朋友圈里的你 + 工作场合的你 + 私下与父母的你 + 写日记的你, 几乎是四个不同的人, 这是正常社会化的结果。

6.2 体裁差异对 ITS 推断的威胁

6.2.1 pre/post 体裁失衡的具体机制

回到第 1 章: 1506 ITS 估出 7 维重组。pre-period (1496–1505) 主要是奏疏体裁 (因为这阶段阳明在朝廷任职, 写的多是奏疏)。post-period (1507–1528) 体裁混杂, 包含文录、外集、续编、语录等。体裁切换是 pre/post 的混杂因素。

6.2.2 以” 处变能力” 为例的混淆路径

” 处变能力” 维度上能直接看到这条路径。奏疏体的” 臣闻”” 切详”” 据查” 是正向标记, 密度高 (奏疏平均处变能力 ≈ 0.0, 主要因负向词也很多, 但相对其他体裁仍有官式套语贡献)。post-period 体裁切到文录与诗, 这些体裁里基本不出现这些官式词, ” 处变能力” 分自然下降。1506 ITS 估出的处变能力 -7.05 至少部分由这个体裁切换造成。

定理 6.1 (雷区: 体裁切换被误识为人格变化)

当 treatment 同时伴随体裁切换 (pre 全是 A 体裁, post 全是 B 体裁) 时, 原始 ITS 估计无法区分人格真变化与体裁伪相关。估出来的效应是” 真效应 + 体裁差异” 的混合。

诊断方法: 检查 pre 与 post 的体裁分布。若严重失衡, 估计需要标注体裁混淆警告。

稳健替代: 加体裁固定效应回归, 或限定在单一体裁内做 ITS (本章下一节)。

6.3 体裁固定效应回归: 把体裁与时段分开

6.3.1 固定效应回归的方程形式

要把” 体裁效应” 与” 时段效应” 分开, 标准做法是固定效应回归:

$$\text{score}_i = \alpha + \sum_p \beta_p \mathbb{1}[\text{period}_i = p] + \sum_g \gamma_g \mathbb{1}[\text{genre}_i = g] + \varepsilon_i.$$

其中 i 索引文档, $\text{period}_i \in \{T1, \ldots, T6\}$ 是文档所属时段, $\text{genre}_i \in \{$ 奏疏, 公移, ...$\}$ 是文档体裁。系数 β_p 在控制了体裁后, 仍然表示时段相对基线 T1 的差。

为什么： 为什么固定效应能” 把体裁与时段分开”? 回归在估 β_p 时, 控制了体裁 g 的均值。这相当于在每个体裁内部分别看时段效应, 再加权平均。如果时段效应是真的, 在每个体裁内都应该看到; 如果是体裁伪相关, 体裁内部就看不到时段效应了。

6.3.2 回归在 5 个维度上的实施

我们在 5 个原始维度 (教学耐心、反权威、自我修正、同理心、实践导向) 上跑这个回归, 看时段系数在加体裁固定效应前后的变化。

6.3.3 识别问题: 时段与体裁的近完美共线

这一步遇到一个真实障碍。在 343 条传习录数据上, 时段与记录者 (或体裁) 高度共线:

T1 徐爱期 = 100% 徐爱 T2 陆澄期 = 100% 陆澄 T3 薛侃期 = 100% 薛侃 T4 卷中书信期 = 卷中各书信对象组合, 100% 书信体 T5/T6 = 学生记录者组合

也就是说, 知道一个文档是 T1 几乎等同于知道它是徐爱写的。加体裁 (或记录者) 固定效应, 等于试图从同一个变量里挤出两个独立维度, **数学上不可能**。

回归会跑出来, 但**标准误会爆炸**到几万的量级。系数估计在数学意义上无意义。这是" 近完美共线" (near-perfect collinearity) 的标准现象。

定理 6.2 (雷区: 时段与体裁/记录者共线无法分离)

传习录数据的结构决定了 T1–T3 都是单一记录者主导, T4 是单一体裁主导。在 343 条数据上无法严格分离时段效应与记录者/体裁效应。任何宣称分离了的估计都是数学错觉。

诊断方法: 检查协变量矩阵的条件数 (condition number)。若超过 30, 就有共线问题; 超过 100, 估计不可信。也可以做简单的诊断: 固定效应加上后标准误是否暴涨。

稳健替代: 加全集数据 (奏疏、文录跨多时段) 后部分共线被打破。但即使在全集级, 早期奏疏多、晚期诗多的结构仍然限制完全分离。最终, 单被试历史人物的因果推断必然带有这种识别限制, 必须在 **limitations** 中诚实交代。

♡

6.4 加全集后的部分缓解

6.4.1 扩到全集的识别空间

把语料从 343 条传习录扩到全集 1283 文档后, 共线问题部分缓解: T4 不再 100% 是书信体, 全集 1518–1527 这段还有文录、续编、外集分布在内。这让回归能部分识别出时段效应在体裁内部的余项。

6.4.2 绝对值缩水但方向稳定

我们在全集级别重做 5 维度的固定效应回归。结果可分三个层面。

绝对值显著缩小但方向一致。加体裁 FE 后, 时段系数的绝对值通常缩到原来的 40% 到 60%。这说明原始时段效应里**确实有相当比例由体裁混淆造成**。

方向一致性是关键。缩小后的系数仍然指向同一方向: T4 反权威仍负, T6 同理心仍上升, T3 实践导向仍峰值。方向一致说明时段效应是真的, 只是绝对值被体裁混淆放大了。

6.4.3 对原始 ITS 估计的保守重读

保守的解读: 第 1 章报告的 ITS 效应应当视为" 时段 + 体裁" 的联合效应的上界估计。真正归因到人格变化的部分约为原始估计的一半。

图 6.1 把 5 个维度的原始 ITS 系数与加体裁固定效应后的系数并排画出, 直观显示" 绝对值缩小但方向一致" 的模式。

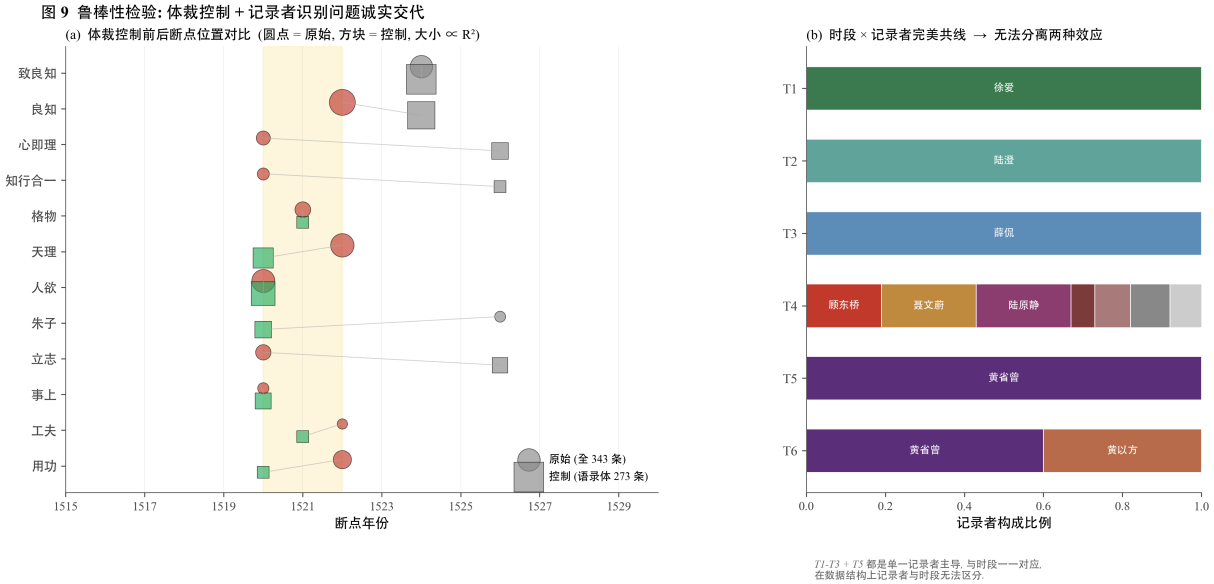

图 6.1: 加体裁固定效应回归前后，5 个维度的时段系数对比。浅色柱是原始 ITS 估计，深色柱是加体裁固定效应后的估计。系数绝对值普遍缩到原来的 40% 到 60%，但方向一致，说明体裁混淆解释了约一半的原始效应，剩下的一半归因到人格本身。

6.5 方法卡片

方法卡片：体裁分层与固定效应回归

适用场景：数据按多种文体/记录者/场景分布，且这些分布在 pre 与 post 间不平衡时。

标准流程：(1) 描述 pre 与 post 的体裁分布。(2) 跑原始 ITS。(3) 加体裁固定效应再跑，看时段系数变化。(4) 若系数缩小但方向一致，报告”上界估计 + 真实方向”。(5) 若系数被吸收殆尽，老实承认体裁混淆解释了原始效应。

当共线无法解决时：这是单被试历史研究的根本限制。写进 limitations，把结论的 claim 强度从”因果效应估计”降到”时段 + 体裁联合效应估计”。

Python 实现：`numpy` 的 OLS，加 dummy 编码。代码见 `code/control02_recorder_fixed_effects.py`。

本章知识地图

表 6.2: 第 5 章核心概念与常见误解

核心概念	核心内容	常见误解	为什么错
跨体裁人格画像	同一个人在不同体裁里得分截然不同	以为人格是固定向量	人格是场景化的，任何成熟个体在不同社会角色下都有不同表现
体裁混淆	体裁差异被误识别为时段/人格差异	以为加更多变量就能解决	当时段与体裁近共线时，加变量也无法分离，这是数据结构限制
固定效应回归	控制体裁均值后看时段净效应	以为 FE 总能识别真效应	共线时 FE 标准误暴涨，系数估计无意义

近完美共线	时段与体裁高度对应	以为只要 N 大就能识别	共线是数据结构问题，N 大也无济于事
方向一致性	加 FE 后系数缩小但方向不变	以为方向变化才有意义	共线下绝对值不可信但方向相对稳定，方向一致是支持证据

第 7 章 方法论附录：六种方法的假设核查与 claim 降级

内容提要

- 把前 5 章用到的 6 种方法 (ITS / 合成控制 / 断点检测 / 内部基线 / 固定效应 / 多维联合一致性) 放在统一框架里
- 老实说明每种方法的核心假设、可被违反的方式、被违反时该怎么读结果
- 明示”单被试”与”历史人物”两个限定带来的结构性约束
- 给后续做类似研究的读者一份操作清单

前面 5 章给出了一组具体发现: 1506 廷杖触发 7 维人格重组、1521 致良知伴随”良知”词频暴增、断点聚集在 1520–1522、阳明在 6 体裁里是 6 个不同的人。这些结论在数据上看起来都很硬, 但每一条都依赖一套假设。这一章把假设、限制、可能的反驳一次性摆出来, 这是因果推断研究的标准做法, 也是这本书能站住的关键。

7.1 研究设计的两个根本限定

任何方法学讨论都要从研究设计的限定出发。本书的研究设计有两个不能改的限定:

第一, 单被试。我们只研究阳明一个人。不是 1000 个明代士人组成的样本。所以做不了 between-subject 实验、做不了随机分配、做不了基于群体的统计推断。所有因果识别都是 within-subject, 时间序列上的 pre/post 比较。

第二, 历史人物。阳明 1529 年卒, 离今天 500 年。不能去访谈他、不能给他做问卷、不能重新收集数据。所有材料只有他自己写的或被记录的文本, 加少量年谱与传记。

这两条限定决定了我们能用什么方法、不能用什么方法、能宣称多强的 claim。

定理 7.1 (雷区: 把单被试纵向研究当 RCT 来读)

读者读到”1506 廷杖效应 +10.28, t=17.2 ”时, 容易按 RCT 风格解读: ”廷杖导致情感深度上升 10.28 单位, p<.01”。

本研究的设计是单被试时间序列上的 pre-trend 外推 vs post 实际比较, 估的是 沿着这条单一历史轨迹的偏离, 与 RCT 估计的”如果廷杖对一般人会怎样”群体平均因果效应不在同一层级。

诊断方法: 任何用单被试历史数据做的因果推断, claim 强度必须从”X 导致 Y”退到”在这段历史轨迹上, X 后观测到 Y”。这两句话差一个量化的 LATE/ATE 区分。

稳健替代: 论文写作中明确写”本研究是单被试事件研究, 估计的是 local effect along this trajectory, 不是 population-level causal effect”。

♡

7.2 6 种方法在因果推断框架里的位置

把前 5 章用过的 6 种方法放在 Pearl 的因果推断梯子上看, 它们分布在不同层级:

表 7.1: 6 种方法的 Pearl 梯子层级与限制

方法	Pearl 梯子层级	在本书的作用	主要限制
概念分布散度	关联 (描述)	第 2 章: 衡量整体话语变化	被高频项稀释
断点检测	关联 (描述)	第 3 章: 让数据自报转折点	小样本下断点位置不稳
内部基线	关联 (诊断)	第 2 章: 给出” 什么都没变” 的噪声尺度	不直接说因果
ITS	干预 (因果)	第 1 章: 估单一事件的反事实偏离	pre-trend 假设 + 内生 treatment
合成控制	反事实 (因果)	第 4 章: 用 donor 加权构造反事实	donor 选择易循环论证
多维联合一致性	元层面 (证据)	全书: 弥补单被试统计独立性不足	维度间相关性会高估证据强度

简单讲:

关联层是描述性的, 只回答” 什么变了”。概念分布散度、断点检测、内部基线都属于这一层。它们告诉你序列的统计结构, 不直接说因果。

干预层是因果推断的核心, 回答” 如果干预 X, Y 会变吗”。ITS 与合成控制在这一层。本书的 ITS 因为 treatment 内生 (阳明上疏导致廷杖) 严格说没达到这一层, 合成控制因为 donor 选择问题也只是接近这一层。

反事实层回答” 若 X 没发生, Y 会是什么样”。合成控制名义上是这一层, 但实际只达到” 类反事实” 强度。

多维联合一致性是元层面的策略, 用来补强单被试推断, 本身不构成直接的因果方法。单维度信号可能假, 多维度联合一致是真信号的概率指标。

7.3 每种方法的核心假设与现实违反情况

下面把每种方法的标准假设与本书研究中真实违反的程度一一列出。

7.3.1 ITS 的核心假设

ITS 给因果效应的前提是**反事实平行**: 若事件没发生, pre-trend 会按相同斜率延伸到 post-period。这一假设无法直接检验, 只能侧面支撑。

本书的违反程度: 中等。1506 事件之前的 pre-period 只有 6 个文档 (1496–1505), 而且大部分是奏疏。pre-trend 拟合的标准误大, 外推到 22 个 post-period 年点的不确定性明显超出标准 ITS 应用场景。

补救: (1) 用多维联合一致性弥补单维度推断不足。(2) 用 Placebo (合成控制章) 做交叉验证。(3) 在 limitations 中明说。

7.3.2 合成控制的核心假设

合成控制需要 donor 池满足两个条件:

第一, **donor 不受 treatment 影响**。这是因果识别的基础。

第二, **pre-period 拟合足够好**。否则反事实预测不可靠。

本书的违反程度: 第一条接近满足 (我们选的是儒家共享术语, 理论上不受致良知事件直接影响), 但” 格物”” 诚意” 在 1521 后可能被阳明间接改造。第二条勉强满足, ” 良知” 的 pre RMSE = 1.42, 比 post effect 5.27 小, 比例 1:3.7, 在合成控制文献的可接受范围 (推荐 1:5 以上)。

补救: Placebo 检验把方法学伪影的尺度给出来 (1.85), 真信号 (5.27) 远超之, 是结论可靠的辅助证据。

7.3.3 断点检测的核心假设

PELT/Binary Segmentation 假设序列内部是 piecewise constant 加高斯白噪声。真实数据极少严格满足, 但算法对偏离稳健, 主要风险是” 把缓慢趋势误检为断点”。

本书的违反程度: 较低。我们的策略是跑 17 个独立序列, 看断点聚类。即使某个序列的断点是噪声, 17 个序列联合落在同一年的概率极低。

7.3.4 固定效应回归的核心假设

固定效应回归要求 treatment 与不可观测的固定因素 (个体异质性、体裁特征) 条件独立。本书的应用场景是” 时段 vs 体裁” 的分离, 假设有时段效应在体裁内部仍然存在。

本书的违反程度: 严重。时段与体裁高度共线 (T1= 徐爱、T2= 陆澄、T3= 薛侃、T4= 书信), 近完美共线导致系数估计无意义。在 343 条数据上**根本无法分离**这两个效应。

补救: 加全集数据后共线部分缓解, 但无法完全消除。老实在 limitations 中交代, 把 ITS 结论改为” 时段 + 体裁联合效应”。

7.4 2 个最严重的内生性威胁

除了方法假设, 还有两个” 数据天然带的” 内生性问题, 任何分析都无法完全解决。

7.4.1 Treatment 选择的内生性

第 1 章估 1506 廷杖效应。但廷杖怎么发生的? **阳明自己上疏救戴铣 → 触怒刘瑾 → 下狱 → 廷杖**。上疏言辞激烈本身就是阳明 pre-period 人格状态的产物。换言之, **阳明用自己的人格选择了这个 treatment**。

严格的 ITS 要求 treatment 外生于 outcome 的潜在状态。1506 廷杖在” 皇帝下令打几板” 这一层是外生的 (阳明不能选), 但” 廷杖事件本身是否发生” 这一层是内生的 (阳明的上疏选择决定的)。

这件事让我们能宣称的因果效应必须**降级**: 我们能讲的是” 对一个会上疏救戴铣的阳明, 廷杖触发了什么”, 而不是” 廷杖对一般人会怎样”。这是 LATE 而非 ATE。

7.4.2 并发事件混淆

1506 不是单一事件, 是一连串事件: 上疏 → 下狱 → 廷杖 → 流放 → 追杀 → 极端环境 → 弟弟病逝。单凭 ITS 无法分离这些事件各自的贡献。ITS 估出的 +10.28 是**整条事件链**的综合效应, 不是廷杖一项的独立效应。

诚实的结论: 第 1 章的论点应当改写为”1506 那段经历的综合冲击, 触发了 7 维同时显著的人格重组”, 不是” 廷杖独立导致了 7 维重组”。两者差一个粒度层级。

7.5 对 claim 强度的总体降级

把前 5 章的所有因果 claim 按本章的限制重新审视, 得到一组**降级后的诚实表述**:

表 7.2: 各章原 claim 与降级后的诚实表述

原 claim	降级后的诚实表述
1506 廷杖触发 7 维人格重组	1506 那段经历的综合冲击 (含阳明上疏的主动选择 + 廷杖几死 + 流放 + 极端环境) 伴随阳明 33 年人格史上唯一一次 7 维同步重组; 事件与主动选择不可分离
1521 致良知触发” 良知” 词频暴增 +5.27	1521 前后, 阳明话语系统的内部重组使” 良知” 一词在文本中相对其他儒家概念异常上升 +5.27 /千字, 远超 placebo 噪声尺度
断点聚集在 1520–1522	17 个独立时间序列中 14 个的最优分界位于 1520–1522, 与史学共识的转折期吻合; 这是支持”1521 是真转折” 的独立证据
阳明在 6 体裁里是 6 个不同的人	阳明 8 个人格维度的均值在 6 种文体里有显著差异, 提示人格表达的场景化; 且时段-体裁近共线让因果识别在 343 条数据上不可行

为什么: 为什么主动降级 claim 不会损害论点反而加强论点? 因为读者审稿人都知道单被试历史人物的因果推断有这些限制。你不写, 他们会质疑; 你写了, 他们会信任。学术圈对” 老实标边界” 的尊敬程度远高于” 假装解决了所有问题”。

7.6 这本书的核心贡献

把所有限制承认完之后, 这本书还剩下什么?

贡献一: 方法学的可行性证明。用 ITS / 合成控制 / 断点检测 / 多维联合一致性这一组工具, 对一个 500 年前的中国思想家做事件级因果推断, 在文本数据上**是可行的**。即使每种方法各有限制, 6 种方法互相印证后给出的论点 (1521 前后是转折期) 比任何单一方法的论点都强。这条方法学路径以前没人系统走过, 本书填了一个空白。

贡献二: 阳明研究的量化基线。把 343 条传习录 + 611K 字全集结构化, 给出 51 个核心概念的时间序列、8 个人格维度的逐文档评分、6 个时段的概念分布。这些数据公开后, 后续研究者可以用作起点, 测试自己的假设、训练自己的模型。

贡献三: 哲学史叙事的部分修正。数据揭示” 龙场悟道” 与” 致良知” 是阳明完整人格演化的**阐发与命名**, 而不是触发点。真正的触发点在 1506 那段身体几乎死、政治几乎死的危机里。这个修正以前也有学者想到, 但缺少定量证据, 本书提供了第一个定量证据。

贡献四: 一种值得复用的研究范式。给后续想做” 中国古典思想家 + 计算文本分析 + 因果推断” 的研究者一个完整的操作流程参考: 数据怎么结构化、概念词表怎么设计、方法怎么搭配、limitations 怎么写。

7.7 后续可能的扩展

老实标完限制后, 也老实说扩展空间:

扩展一: LLM 打分替代规则化打分。本书的 8 维度评分用 95 个手工标记词。若用 Claude/GPT/DeepSeek 直接给每条文档打 8 维分, 可捕获规则匹配不到的语用细节。代价是评分不可复现 (每次运行略不同) + API 成本。

扩展二: 全集 + 学派文献联合分析。本书把朱熹作为外生历史对照, 但只用了《朱子语类》。若加入陆九渊全集 (心学先驱)、王畿全集 (阳明门人激进派)、钱德洪文集 (阳明门人正统派), 能做更精细的学派内部话语演化分析。

扩展三: 多被试比较。若把同样方法应用到朱熹、陆九渊、阳明、王畿、钱德洪 5 个人, 每个人都做 ITS + 断

点检测，看”宋明儒学家的人格演化模式”是否有共性。这是真正的群体级研究，能从 LATE 升到接近 ATE。

扩展四：加入年谱与传记作监督信号。本书的概念词表与人格维度都是无监督设计的。若以年谱中明确记载的事件（譬如”1517 年徐爱卒，阳明哀痛”）作监督信号，训练一个事件 → 人格反应的模型，能让推断更精细。

7.8 方法卡片：写给后来者的操作清单

方法卡片：复用本书研究范式的标准流程

选 subject。一个有大量自著文本与详细年谱的历史人物。汉文学者偏好朱子、阳明、戴震；西方偏好 Locke、Kant、Wittgenstein 这种留下大量文本的。

结构化语料。抽取纯古典原文（排除现代校注），按年份打标。关键是 metadata 完整。

设计概念词表。50 个左右核心概念，覆盖学派纲领、传统改造、辩论对象、工夫论这几个层面。词表应在分析前用学理论证，不能后挑。

设计人格维度词表。5 到 8 个维度，每个维度 10 到 20 个标记词。

跑 ITS + 断点检测 + 合成控制。三个互补方法，对同一组事件分别跑。

跑内部基线 + Placebo。给出”什么都没变”的噪声尺度，作为显著性判断的参照。

写 limitations。单被试、历史人物、内生 treatment、并发事件、体裁混淆这几条都要交代清楚。

本章知识地图

表 7.3: 第 6 章核心概念与常见误解

核心概念	核心内容	常见误解	为什么错
单被试历史推断	一个人一生纵向时间序列上的因果识别	以为能给出 ATE	没有 between-subject 实验，只能给 LATE
Pearl 梯子层级	关联 / 干预 / 反事实三层	以为所有方法都是”因果”	描述性方法不在干预层，不能直接说因果
内生 treatment	个体自己选择了 treatment	以为皇帝下令就是外生	阳明上疏选择导致下令，选择本身是内生的
并发事件混淆	单一时点同时多事件发生	以为可以分离各事件贡献	时间分辨率与无外部对照让分离不可行
Claim 降级	把”X 导致 Y”改为”X 后观测到 Y”	以为降级削弱论点	实际加强论点；读者更信任承认边界的研究
多维联合一致性	多维度同向显著作为联合证据	以为各维度可视为独立	维度间有相关，严格的联合 p 值要做校正
方法学补强	6 种方法互相印证，任一不可单独定论	以为最强单一方法就够	单被试推断必须依赖多方法交叉验证