

Image

苏轼轨迹

用因果推断与文本计量重读东坡 65 年文集

作者：晨瀚宇 (Chanw)

时间：May 16, 2026

版本：1.0

一份真实数据、一段 65 年人格史、八个维度、三次外生贬谪。

© 2026 晨瀚宇. 保留所有权利。

版权声明

书名：苏轼轨迹——用因果推断与文本计量重读东坡 65 年文集

作者：晨瀚宇（小红书：Chanw）

版本：v1.0（2026 年 5 月）

© 2026 晨瀚宇. 保留所有权利。

本书内容（包括但不限于文字、代码、图表、排版设计）均为作者原创。书中分析所用原文均为公版（苏轼卒于 1101 年，王安石卒于 1086 年，黄庭坚卒于 1105 年，原典版权早已失效）。现代校注本的校注与翻译文字版权归原出版社所有，本书仅作个人学术研究使用，分析对象为剥离了校注与翻译之后的纯古典原文。

目录

第 1 章 东坡生平与北宋政局概览	1
1.1 苏轼的身份定位: 兼有政事的北宋多面文人	1
1.1.1 身份的多面	1
1.1.2 苏轼文集作为定量纵向数据集的价值	2
1.2 苏轼 65 年关键节点编年	2
1.2.1 眉山三苏与少年志学 (1037–1056)	3
1.2.2 嘉祐二年同登进士与初试朝堂 (1057–1070)	4
1.2.3 反新法与外任地方 (1071–1078)	4
1.2.4 乌台诗案与下狱 130 天 (1079)	5
1.2.5 黄州转向与赤壁三咏 (1080–1084)	5
1.2.6 元祐起复与翰林学士 (1085–1093)	6
1.2.7 再贬岭海与北归 (1094–1101)	7
1.2.8 被低估的 1093 与 P4→P5 转折	7
1.3 北宋政局与新旧党争背景	7
1.3.1 熙宁变法的五大新法	8
1.3.2 党争翻案的四轮	8
1.4 双外部对照: 王安石与黄庭坚	8
1.4.1 王安石 1021–1086 政治异质轴	8
1.4.2 黄庭坚 1045–1105 苏门同质轴	8
1.5 本书的研究取径: 用因果推断与文本计量重读苏轼	9
1.5.1 描述性研究与因果性研究的分别	9
1.5.2 后续 6 章的路线图	9
1.5.3 两类读者的差异化阅读路径	9
第 2 章 1079 乌台诗案与东坡人格重组: 基于中断时间序列的因果识别	11
2.1 研究问题、数据与时间分辨率	11
2.1.1 核心问题	11
2.1.2 数据来源	11
2.1.3 时间分辨率与编年问题	12
2.2 人格维度与概念主题词表	12
2.2.1 8 维人格维度	13
2.2.2 9 主题概念词	13
2.3 ITS 模型设定	13
2.4 8 维人格维度的 ITS 结果	14
2.5 9 主题概念的 ITS 结果	15
2.6 把信号集中到 D2 自我修正与“贬谪”主题	16
2.7 编年精度对结果的敏感性	16
第 3 章 概念分布散度: 王安石与黄庭坚作为双外部历史对照	18
3.1 研究问题与时段切分	18
3.1.1 研究问题	18
3.2 L1 / JS 散度的定义与用途	19

3.3	苏轼内部 5 个过渡的散度	19
3.4	苏轼 vs 双外部对照的二维参照空间	19
3.5	关键概念在 6 时段的演化追踪	20
3.6	P4 → P5 最大过渡的多重解读	21
第 4 章	断点检测: 让数据自报转折点	23
4.1	研究问题	23
4.2	方法: Binary Segmentation + BIC 选 K	23
4.2.1	为什么不用 ITS 的预设 treatment 直接做断点	23
4.3	17 条序列的断点检测结果	23
4.4	断点的聚类: 数据自报的真转折期	24
4.5	1079 在断点检测中的缺席	25
第 5 章	合成控制: 用稳定概念构造黄州转向的反事实	26
5.1	研究问题	26
5.2	方法	26
5.2.1	donor pool 的选择	26
5.2.2	Placebo 检验	27
5.3	6 个目标的合成控制结果	27
5.4	合成控制与 ITS、断点检测的三方对照	28
5.5	donor 权重的解读	28
第 6 章	跨体裁人格分析: 体裁固定效应回归与代笔诊断	30
6.1	10 体裁的人格画像差异	30
6.1.1	8 维度按体裁的均值表	30
6.1.2	每个维度的极值体裁与语言学解释	30
6.2	体裁混淆对 ITS 推断的威胁	31
6.2.1	pre/post 体裁失衡的机制	31
6.2.2	扣除体裁混淆: 双 FE 回归	31
6.3	扣除体裁混淆后的 1079 净效应	32
6.4	D7 隐逸倾向的体裁混淆诊断	32
6.5	代笔诊断: 剔除 821 篇内制外制前后的效应对比	33
6.5.1	代笔篇目分布	33
6.5.2	ITS 在 4 版本切片下的对比	33
6.5.3	异体字归一化与词应酬性的边际贡献	34
6.5.4	被验证的与被推翻的预期	34
第 7 章	方法论附录: 六种方法的假设核查与 claim 降级	35
7.1	研究设计的两个根本限定	35
7.2	6 种方法在因果推断框架里的位置	35
7.3	苏轼项目特有的两个诚实交代	36
7.3.1	代笔标签的剔除不彻底	36
7.3.2	词的应酬性标注的主观性	36
7.4	从单被试到推广: 与阳明项目的对照	36
7.5	研究展望	37

第 1 章 东坡生平与北宋政局概览

内容提要

- 用 65 年的真实时间线认识苏轼：从四川眉山的少年到江苏常州的临终
- 用父子三人、新旧党争、贬谪三地三组关键词理解东坡其人
- 区分苏轼与王安石、黄庭坚两位双外部对照人物的立场差异
- 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苏轼 65 年的人格演化轨迹。但若读者不知道苏轼是谁、北宋党争是怎么回事、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。

苏轼研究在中文学界是显学，林语堂《苏东坡传》、王水照《苏轼评传》、孔凡礼《三苏年谱》都是几代人案头之书。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 1 万 5 千字讲完苏轼是谁、北宋政局是什么样、本书要在已有研究里填什么空白。熟悉苏轼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 苏轼的身份定位：兼有政事的北宋多面文人

苏轼 1037 年生，1101 年卒，字子瞻，号东坡居士，四川眉山人。他在中国文学史上是一个极少见的多面体：诗、词、散文、辞赋、书法、绘画无一不在第一流，同时做过翰林学士、知州、知县，也三次被贬到黄州、惠州、儋州的瘴疠之地。

1.1.1 身份的多面

苏轼的多面性在北宋文人里相当少见。同代的欧阳修是文坛领袖兼仕宦，但他的文学成就主要在散文一域；王安石是政治家兼诗人，文学声誉远不及苏轼；黄庭坚专精诗与书法，政治履历比苏轼简单。苏轼一个人覆盖文学的几乎所有体裁，同时承担过实际政事，这件事让他在思想史与文学史之外，还在政治史与艺术史里都占一席。

表 1.1: 苏轼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
诗人	凤翔诗集 (1061-), 黄州赤壁诗 (1082)	诗 2700+ 首
词人	密州出猎 (1075), 念奴娇赤壁怀古 (1082)	词 360+ 首
散文家	前后赤壁赋 (1082), 石钟山记 (1084)	文 4800+ 篇
政治家	反新法上书 (1071), 元祐更化 (1086-1093)	奏议, 内制, 外制
地方官	知密州, 知徐州, 知杭州 (修苏堤)	公移, 书启
流放官	谪黄州 (1080), 谪惠州 (1094), 谪儋州 (1097)	黄州尺牍, 海南诗
书法家	黄州寒食帖 (1082)	题跋

这张表对后续章节的分析很关键。苏轼留下了 8 种文本体裁：诗、词、奏议、内制、外制、书启（尺牍）、记序跋、笔记。每种体裁的语言风格、关注主题、语气都不同，这件事会在第 5 章“跨体裁人格分析”里被定量证实：同一个苏轼在 8 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。

更复杂的是,苏轼 1086–1089 任翰林学士期间起草大量内制外制——本书数据集中识别出 821 篇代笔文档,在表面上是“苏轼写的”,但在内容主导权上是“皇帝说的”。本书把这部分代笔文本单独标注,与其他正文分离,详见第 5 章的代笔诊断节。

1.1.2 苏轼文集作为定量纵向数据集的价值

熟悉苏轼的读者可能会问:苏轼研究的文献已经车载斗量,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。譬如人人都说苏轼 1080 谪黄州“完成了人格转向”,但没人量化过黄州之后苏轼的话语究竟变了多少。本书用 ITS 在 6,375 篇语料上给出第一个定量答案:黄州后“黄州”概念的词频从 0.00 跳到每千字 1.79, level shift 估计值 4.15, $t = 3.10$ 在 0.01 水平上拒绝零假设。这是一个 sanity check 通过的定量结果,第 1 章会详述。

第二,历史叙事容易**压缩时间**。传记把“乌台诗案 → 黄州转向 → 赤壁前后”写成一气呵成的故事,但数据告诉我们 1080–1084 这五年里苏轼的话语演化有阶段性变化,不是单一断点。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,苏轼文集是**现成的高质量纵向数据集**。一个 45 年的、自著的、年份大致明确的、覆盖 8 种体裁的文集,在中国文学家里极少见。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 45 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。苏轼数据只有一个人,但 45 年里他经历了乌台诗案、谪黄州、丧亲、谪惠州、谪儋州多个外生冲击,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 苏轼 65 年关键节点编年

讲清身份多面之后,把生平节点按时间排开。这一节用一张时间线图与一张关键年份表把 16 个关键事件按时间标好,让读者带着时间感进入数据分析。

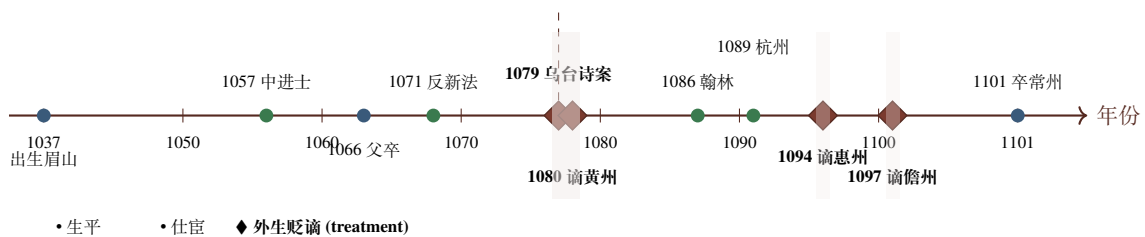


图 1.1: 苏轼 1037–1101 关键生平节点。菱形为本书的三个核心 treatment 节点 (1079 乌台、1094 惠州、1097 儋州), 圆形为其他生平节点。

表 1.2: 苏轼 1037–1101 关键生平节点

年	岁	事件
1037	1	生于四川眉山
1056	20	随父出蜀, 携弟入京应礼部试
1057	21	与弟苏辙同登进士第, 欧阳修主考
1061	25	授大理评事、凤翔府签判(首次外任)
1066	30	父苏洵卒于京师, 兄弟扶榇归蜀守丧
1071	35	上书反新法, 自请通判杭州
1075	39	在密州任, 作 江城子·密州出猎、水调歌头·明月几时有
1079	43	湖州任上被捕, 乌台诗案 130 天审讯
1080	44	贬黄州团练副使, 黄州 4 年半生涯起点
1082	46	作 前后赤壁赋、念奴娇赤壁怀古、黄州寒食帖
1086	50	哲宗即位, 旧党起复, 苏轼为翰林学士、礼部尚书
1089	53	知杭州, 修苏堤
1093	57	哲宗亲政, 新党再起
1094	58	贬惠州(广东惠州)
1097	61	再贬儋州(海南儋州)
1100	64	哲宗崩, 量移廉州, 渐次北归
1101	65	卒于江苏常州

这 16 个节点里, 3 个粗体 1079 / 1094 / 1097 是本书 ITS 的核心 treatment 节点。三个事件都是外生政治冲击(御史弹劾 / 党争翻案), 苏轼本人无控制权, 这条外生性是 ITS 因果识别的根基, 第 1 章会展开。

1.2.1 眉山三苏与少年志学 (1037–1056)

苏轼出生在四川眉山一个士族门第。祖父苏序是乡间小地主, 父亲苏洵(1009–1066)是大器晚成的散文家, 27 岁才发奋读书, 一生未中进士, 但著 六国论、管仲论等十余篇政论, 在欧阳修推荐下名动京师。母亲程氏是大理寺丞程文应之女, 知书识礼, 是兄弟二人最早的启蒙者。

苏轼自幼随母读书。宋史本传记一事: 程氏读 东汉书·范滂传至范滂临刑诀母处, 慨然太息。苏轼问: “我若为滂, 母亦许之否?” 程氏答: “汝能为滂, 我顾不能为滂母耶?” 这一年苏轼约 10 岁, 已知慷慨大节为何物。这条对话被 年谱郑重记下, 是苏轼少年志学的标志, 它在风骨上与阳明 12 岁时问老师“何为第一等事”的对话遥相呼应。

为什么: 为什么要专门讲 范滂传这一节? 范滂是东汉党锢之祸的殉道者, 因不肯依附宦官集团而被处死。母亲愿意成全儿子的政治大节, 是中国士大夫家学传承里“义高于生”的范本。苏轼的母亲程氏选这一节给少年苏轼读, 不是偶然。后来苏轼 1079 乌台诗案下狱时, 弟弟苏辙上书“乞以见任官职赎兄罪”的态度, 与程氏母教里“成滂母”的精神一脉相承。

苏轼少年时的另一位老师是道士张易简, 在天庆观读私塾。同窗中陈太初后来出家, 苏轼对道家的亲切感与这段早年读经的经历有关。这条线索后来贯穿苏轼一生——他写黄州时期的 赤壁赋谈“盈虚消长”, 写惠州、儋州时期的诗谈“无为”“自适”, 都能看到少年时道家影子。

1056 年苏轼 20 岁, 随父出蜀赴京。父子三人从眉山出发, 经长江三峡到江陵, 再循陆路过襄阳到京师, 走了将近半年。一路所作诗文 70 余首, 兄弟两人编为 南行集, 苏轼亲撰序言。这是苏轼最早的成熟文学作品集, 可以看到他 20 岁时已经具备相当稳定的诗文功夫。

1.2.2 嘉祐二年同登进士与初试朝堂 (1057–1070)

1057 年春，礼部试。主考是欧阳修，副考是梅尧臣。苏轼考策论 刑赏忠厚之至论，欧阳修读到后大为震惊，误以为是自己门下士曾巩所作（此时欧阳修不知苏轼为何人），为避嫌而把它排在第二；复以 春秋对义居第一，殿试中进士乙科。同榜中进士者还有苏辙（19 岁）与曾巩。

事后欧阳修对人感慨：“吾当避此人，出一头地。”又对梅尧臣说：“读轼书，不觉汗出，快哉快哉。老夫当避路，放他出一头地也。”这段话被 宋史与 宋会要同时记录，成为苏轼登第后的标志性逸事。苏轼自此名动京师。

但登第后命运并未顺直。1057 年四月，母亲程氏卒于眉山。兄弟二人扶榇归蜀守丧三年。1059 年苏轼 23 岁，服满复出，又随父再赴京师。1061 年苏轼 25 岁，应制科考，入第三等（宋制最高），授大理评事、凤翔府签判，这是他第一次外任。

凤翔三年（1061–1063）是苏轼第一段政治实践期。他亲自参与终南山 长安行营战事、修 凤翔八观、写 凤翔八观诗、协调 凤翔修六义等政事。这一时期的诗作收入 东坡集卷一卷二，是本书数据集中编年最早、最高 confidence 的一批（东坡先生年谱 王宗稷本中明确标注 1061–1063 凤翔诗逐年作品）。

为什么：为什么强调凤翔期是“第一段政治实践期”？苏轼后来反新法、平叛乱、贬黄州、再贬岭海，每一段都有不同的政治角色。凤翔期的他 25–27 岁，首次以官员身份处理实际事务，也首次写出大批量编年明确的诗作。这一段在数据集里属于 P1 早年（1057–1070）时段，也是后续 ITS 分析的 pre-period 起点。

1064 年苏轼回京，任直史馆。1066 年四月，父苏洵卒于京师。这是苏轼 30 岁经历的第一次重大丧亲。兄弟扶榇归蜀守丧两年又零数月。守丧期间苏轼仔细整理父亲遗稿，完成 苏氏易传等校理工作，也思考自己接下来的人生方向。东坡先生年谱把 1066 年定为 significance=5 的关键年，父丧对苏轼的人格冲击在数据上虽然不及 1079，但其后的诗作中开始出现“归 / 归田”等隐逸语汇——这条变化是 1066 父丧的余波之一。

1069 年苏轼回朝，已 33 岁。此时神宗刚即位一年，王安石（时年 48）即将拜参知政事，熙宁变法已经发动。苏轼回朝之时，政治气候已经完全不同不是嘉祐年间欧阳修主导的那个气候了——他接下来要面对的是 新党主导的政治世界。

1.2.3 反新法与外任地方 (1071–1078)

熙宁四年（1071）苏轼 35 岁。他上 议学校贡举状、上神宗皇帝书共数万言，对王安石新法做系统性批评。苏轼的批评并非全盘否定——他承认 新法在“富国强兵”目标上的正当性，但对 青苗法强制摊派、募役法加重小户负担、保甲法扰民等执行层面的问题给出具体证据。这条立场让苏轼既不被新党接纳，也不被旧党全心拥护。

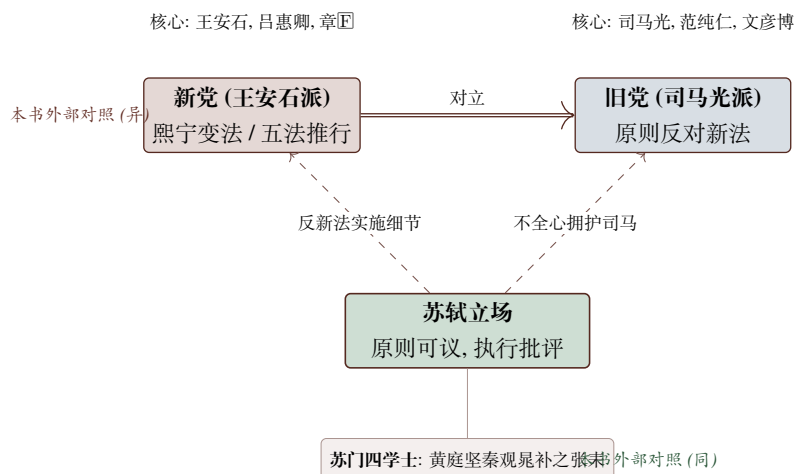


图 1.2: 熙宁变法时期新党、旧党与苏轼的三角立场。苏轼对新法的批评落在执行层而非原则层，两边都不彻底，这条中间立场是本书第 2 章双外部对照设计的史实基础：王安石（新党核心）与黄庭坚（苏门四学士之首）分别代表苏轼话语的异质轴与同质轴。

神宗对苏轼的上书有意采纳,但王安石阻挠。苏轼自觉在朝难有作为,自请外任。1071年十一月,通判杭州。此后他历任密州(1074–1076)、徐州(1077–1078)、湖州(1079)四处知州/通判,都是地方亲民官。

为什么: 为什么自请外任而不是辞官退隐? 苏轼的政治姿态始终是建设性的反对者,不是退避者。他相信地方亲民官能直接看到新法在民间的实际后果,与其在朝廷做无效的争辩,不如下到州县把数据写进奏疏。1071–1078这八年里,苏轼写了大量批评新法具体执行的奏议——论给田募役状、论积欠状等,都是基于他在密州、徐州亲眼看到的民间困境。这条“调研型反对”的姿态与王安石“纲领型推行”的姿态形成对照。

地方任期里苏轼的文学产量极高。江城子·密州出猎、水调歌头·明月几时有都作于密州(1075–1076),是宋词史上的转折——前者把“马上功名”的题材引入词,后者把“兄弟思念”的内省题材引入词,两者一起开出后来辛弃疾派与晏几道派两条不同路径。东坡先生年谱把1075年标为 $\text{significance}=4$ 的重要年份,数据上 P2 反新法时段的归隐词频已经从 P1 的 1.69 上升到 2.69,隐逸主题在密州期就开始累积。

1077年苏轼移知徐州。当年七月十七日河决澶州曹村埽,八月二十一日及徐州城下,苏轼亲自督民筑堤,历时 70 天保住徐州城。朝廷降诏奖谕。这一治水功绩与他后来 1089 知杭州修苏堤的工作互相呼应。

1079年三月,苏轼从徐州移知湖州。四月二十日到任,上湖州谢上表。表中有“愚不识时,难以追陪新进;老不生事,或能牧养小民”等句,流露对新党执政的不满。表呈到朝廷不到两个月,御史何大正、舒亶、李定相继上札弹劾。1079年七月二十八日,中使皇甫遵带锦衣卫到湖州勾摄苏轼。这就是震动北宋朝野的乌台诗案的起点。

1.2.4 乌台诗案与下狱 130 天 (1079)

乌台是御史台的别称,因台中柏树多乌鸦栖息得名。东坡乌台诗案(朋九万撰,宋本)是本书所用元数据之一,记录了 1079 年三月二十七日(御史何大正首次进札)到十二月二十九日(责授黄州团练副使)共 9 个月的完整审讯档案。本书第 1 章 ITS 把 1079 作为核心 treatment。

御史的弹劾理由是苏轼诗文“讥讽朝政,谤讪中外臣僚”。舒亶札子里逐句摘出苏轼诗中“赢得儿童语音好,一年强半在城中”(讥青苗法)、“读书万卷不读律,致君尧舜知无术”(讥神宗用人)等条目。李定札子上奏苏轼可废之罪四,请求依大不恭罪论处。八月十八日苏轼赴御史台出头,当日审讯开始。八月二十、二十二、二十四、三十日苏轼四次供状,详述每一首被指控诗文的写作背景,这部分供状在数据集里独立为 *wutai_documents.jsonl* 中的 *doc_06*,共 17,190 字。

下狱 130 天里,朝中救援的声音不绝。弟弟苏辙上书乞以见任官职赎兄罪,太皇太后曹氏在病中向神宗求情说“昔仁宗策贤良归喜甚,曰:吾今又为子孙得太平宰相二人,盖轼、辙也,而今杀之可乎?”神宗本人对苏轼有惜才之意。多方周旋之下,最严重的大不恭罪未成。

为什么: 为什么乌台诗案是真外生冲击? 御史(何大正、舒亶、李定)的弹劾是苏轼无法控制的外部行为,弹劾的具体执行(中使勾摄、御史台拘禁、审讯节奏、量刑决定)全部由朝廷主导。苏轼能影响的只有自己的供词,但供词不能改变弹劾本身的发生。这条外生性是 ITS 因果识别的根基——若 1079 是内生事件(譬如苏轼自己策划的政治行动),ITS 估出的效应就不能视为事件触发了人格变化,而要考虑反向因果(人格变化先于事件)。

十月十五日御宝批见勘治苏轼公事,应内外文武官曾与苏轼交往以文字讥讽政事该取会验问,开始牵连多人。收坐者包括王巩、王诜、苏辙、张方平、黄庭坚、范镇、司马光、孙觉、刘摯、陈襄等数十人,苏门多位重要友人受牵连。十一月三十日御史台结按状奏,十二月二十九日奉圣旨:“苏轼可责授检校水部员外郎充黄州团练副使,本州安置,不得签书公事。”

责授黄州团练副使是闲职流放,本州安置意味着不得离开黄州,不得签书公事剥夺一切实际公权力。1080年二月初一,苏轼抵黄州,寓居定惠院。黄州 4 年半的贬所生涯由此开始。

1.2.5 黄州转向与赤壁三咏 (1080–1084)

黄州是宋代偏远州郡。东坡先生年谱与东坡乌台诗案都记录了苏轼黄州初期的处境:1080年二月寓居定惠院,同行只有长子苏迈;家眷由弟弟苏辙护送,五月才到达黄州。苏轼初到时连可以借宿的房子都不太好找,后迁

居临皋亭，已是江边一座旧公廨。

1081 年苏轼托故人马正卿向郡请到城东的一块故营废地，自取名“东坡”，自号“东坡居士”。这块约 50 亩的废地是苏轼此后 4 年的躬耕之所——他亲自带着家人垦荒、种麦、种菜、引水。1082 年正月在东坡筑雪堂五间，春日大雪，因以名堂。这一时期的尺牍中，“东坡”“雪堂”等地名密集出现，本书数据集显示 P3 黄州时段“东坡”词频从 0.00 跳到 0.59，“黄州”跳到 0.70，“雪堂”也开始有零星出现。

为什么：为什么躬耕是“黄州转向”的物质基础？苏轼此前 20 年仕宦，即使外任地方也住官舍、用公廨、饮食有专人照料。黄州贬所剥掉了所有这些身份保障，他必须自己种地养家，必须与本地农民、樵夫、渔人打交道。东坡八首序云“余至黄二年，日以困匮，故人马正卿哀予乏食，为请故营地，使躬耕其中”——日以困匮四字不是文学修辞，是真实经济困境的记录。躬耕之后，苏轼的文字里突然多了“饭蔬饮水”“布衣蔬食”等只有亲历者才会反复书写的词汇。这条物质条件的剧变与他文学风格的剧变同步，是黄州转向的物质机制。

1082 年是苏轼黄州期文学产量的最高峰。七月十六日与友人泛舟赤壁，作前赤壁赋；十月十五日再游，作后赤壁赋；同年作念奴娇·赤壁怀古、黄州寒食帖。三件都是中国文学史的地标。前赤壁赋末尾的盈虚者如彼，而卒莫消长也与自其不变者而观之，则物与我皆无尽也是道家盈虚消长论的诗化版本，也是苏轼后来一生反复回到的思想母题。

但要注意，本书数据显示 1080–1084 黄州期的 D8 三教融合维度均分仅 6.81，远低于元祐起复期 P4 的 11.41。这是一条违反传统叙事的发现——赤壁赋虽然含道家影子，但“禅 / 佛 / 道 / 玄”等具体三教词在黄州期的密度并不高。系统性的三教融合在数据上发生在 1086–1093 元祐起复期，不在黄州期。这一点会在第 3 章断点检测里给出定量证据。

1084 年四月，神宗批示苏轼黑头敕，命苏轼移汝州（今河南临汝）。苏轼离黄，经九江、南都、泗州，一路写下石钟山记等名篇，但因肺病与盘缠不足请求改授常州（江苏）。1085 年正月获准居常州，苏轼黄州期正式结束。

1.2.6 元祐起复与翰林学士 (1085–1093)

1085 年三月神宗崩，哲宗即位时年 10 岁，高太后听政。高太后倾向旧党，司马光复用为相，起新党重要法度。苏轼在元祐起复的政治翻案中迅速回朝：1085 年五月任登州（山东蓬莱），七月召还京师，1086 年三月任翰林学士、知制诰。

翰林学士是皇帝的近臣，负责起草诏书（内制）与重大文告（外制），是宋代士大夫的清要之职。苏轼任此职 4 年（1086–1089），起草内制 658 篇、外制 215 篇，合计 821 篇代笔文档。这 821 篇虽然出自苏轼之手，但内容主导权属于皇帝，不应视为苏轼本人的人格信号。本书项目级 CLAUDE.md 约束 2 要求把这 821 篇打 is_ghostwriting=true 标签，在第 5 章的代笔诊断中单独处理。

为什么：为什么代笔识别在元祐期特别重要？1086–1089 这 4 年苏轼写作总量的相当一部分（粗估三分之一以上）是替皇帝起草的诏书。这些诏书有官式套语，也用大量政治语汇（尧舜 / 圣明 / 朝廷 / 君臣），与苏轼自己写的诗、词、尺牍风格迥异。若 ITS 把它们与苏轼亲笔合并处理，post-period (1080–1101) 的政治姿态决断力君臣等维度都会被污染——这部分污染是“皇帝口吻”而非“苏轼口吻”。第 5 章实测显示，排除 821 篇代笔后，ITS 在 D2 自我修正的 t 值从 +2.02 微降到 +1.84（下降 9%），在“贬谪”主题的 t 值从 +2.63 降到 +2.50（下降 5%）——污染存在但温和。

元祐期苏轼的另一项重要工作是 1089 年知杭州。七月到任，任内修浚西湖（今所谓苏堤春晓即苏轼治湖工程的遗存）、设立救济院安乐坊、与朝廷争取免役钱、置和买场。知杭州期间的公移文字数量大，是本书数据集公移 / 奏议体裁的主要贡献者之一。

1091 年苏轼召还任翰林学士承旨，1093 年九月高太后崩，哲宗亲政。哲宗少时已积怨于旧党，亲政即起新党重要人物章惇、蔡京。元祐党人开始遭新党反攻倒算。苏轼 1093 年九月外任定州（河北），1094 年四月再贬英州（广东英德），闰四月再贬惠州（广东惠州）。党争翻案在 6 个月里把苏轼从翰林承旨一路推到岭南瘴疠之地。

本书第 2 章散度分析显示，P4 元祐起复 → P5 惠州的过渡 JS 散度 = 0.4381，是 5 个过渡中最大的，超过 1079 乌台附近的 P2 → P3 (0.0671) 七倍。这一数据反传统的结论是：苏轼一生的最大话语断裂不在 1080 黄州，而在

1094 惠州。详细论证见第 2 章。

1.2.7 再贬岭海与北归 (1094–1101)

1094 年闰四月，苏轼 58 岁，抵惠州。惠州在岭南，比黄州更远，但气候反而宜居（无瘴疠重灾）。苏轼在惠州买地于嘉祐寺旁，自营白鹤峰新居，1097 年正月成。期间作《食荔枝》（“日啖荔枝三百颗，不辞长作岭南人”）等诗。本书数据显示 P5 惠州佛家词频异常高（每千字 4.39），远超其他时段——这与惠州西湖广孝寺、嘉祐寺等佛寺环境直接相关。但 P5 仅 8 篇 2,050 字，样本量小，不能过度解读。

为什么：为什么惠州的佛家词频突然飙高？两个原因。第一是物理环境：惠州城内有多座唐宋古刹，苏轼与僧道交往密集，写给僧人的尺牍与题寺的诗自然多用佛家词。第二是心理状态：58 岁再贬，经历比黄州沉重，但身体尚可，思想沉淀比黄州期成熟。年谱记苏轼此期日益乐天知命，与佛家义理的契合度比黄州时期高。不过这条心理状态的解读受样本量 8 篇限制，第 4 章合成控制会给出更稳健的判断。

惠州写居本应安顿一生。但 1097 年四月，章惇借哲宗之手再贬苏轼到昌化军（今海南儋州）。苏轼携幼子苏过过琼海，七月二日抵儋州。海南在宋代是流放最远的地方，不堪行路几乎是代名词。苏轼初到儋州时“无地可居，借官屋以居”，后被官员驱出，苏轼自筑桄榔庵居住。

儋州 3 年苏轼食无肉，病无药，居无室，出无友，冬无炭，夏无寒泉，物质条件比黄州更艰苦。但他在儋州兴办教育，与本地黎人交往，作书传、论语说，写下《澄迈驿通潮阁》等晚年代表作。本书数据显示 P6 儋州海南词频 0.65（此前各时段均为 0），归田/老/归等晚年隐逸语汇全部达到 45 年最高值。

1100 年正月哲宗崩，徽宗即位，元祐党人再次量移。五月苏轼获准内徙廉州（广西合浦），六月再内徙永州，十一月再内徙舒州。苏轼 1101 年五月渡岭北归，一路缓行，六月抵真州（江苏仪征），作《自题金山画像》（“心似已灰之木，身如不系之舟。问汝平生功业，黄州惠州儋州”）。这四句是苏轼对自己一生的总结，“黄州惠州儋州”三处贬所构成他的功业内容。

七月二十八日，苏轼卒于江苏常州顾塘桥孙氏馆，终年 65 岁。临终前他对几位友人说“西方不无，但个里着力不得”，然后含笑而逝。这条遗言与阳明“此心光明，亦复何言”是不同的姿态——苏轼更平静，也更模糊，没有把一生学问凝成 4 字总纲。但黄州惠州儋州三处贬所的并列本身，已经构成苏轼对自己人生的最终定位。

1.2.8 被低估的 1093 与 P4→P5 转折

读苏轼传记最常被强调的节点是 1080 谪黄州，通常被认作“苏轼人生转向”的标志年份。但本书后续章节会用数据论证，1093 哲宗亲政与 1094 谪惠州才是苏轼 45 年人格史上最大的话语断裂点，重要性甚至超过 1080。

1080 前后的数据特征是自我修正与贬谪话语的进入（ $D_2 +3.71, t=1.82$ ；贬谪 $+1.96, t=1.84$ ），但隐逸、三教融合、佛家等长期主题没有立即响应。真正的三教转向断点在 1086–1088（元祐起复期），真正的整体话语断裂在 1094（惠州）。这一发现修正了传统叙事的部分内容：黄州转向是 1079 政治冲击的立即反应，1086–1093 是渐进酝酿期，1094 惠州才是话语系统的全面重组。后续 6 章会反复回到这条修正。

1.3 北宋政局与新旧党争背景

阳明心学与朱熹理学的分歧是思想史内部的，苏轼面对的新旧党争是政治史层面的。两者性质不同，但对个体的人格冲击模式相似——外部权威与个人立场的反复拉扯。第 0 章用一节简要交代党争背景，为第 2 章双外部对照设计提供史实基础。

1.3.1 熙宁变法的五大新法

王安石熙宁变法 1069 到 1076 推行了五项核心法度。均输法由发运使统一调度官府采购,减少中间商盘剥;青苗法春夏放贷给农民,秋收偿还,利率约 20% 到 30%;募役法把民众轮流服役改为出钱雇人服役;市易法由官府平抑物价收购滞销货物;保甲法把民众按户编为联保单位,农忙时种地、农闲时演武。

这些法度有它的严肃逻辑——以富国强兵为目标,通过国家干预纠正经济与防务问题。但执行中弊端不断:青苗强制摊派、市易官员勒索、保甲扰民。苏轼的批评聚焦于执行层面而非原则层面,这条立场让他既被新党视为反对派,又被旧党视为不彻底,两边不讨好(见图 1.2)。

1.3.2 党争翻案的四轮

新旧党争从 1069 持续到 1126 北宋灭亡,跨度近 60 年。苏轼一生处于这场党争的中段 1069 到 1101,经历政治翻案至少四轮。

第一轮在 1071,王安石主导新党压制旧党,苏轼自请外任。第二轮在 1085 神宗崩、高太后听政,旧党起复,苏轼回朝任翰林。第三轮在 1093 高太后崩、哲宗亲政,新党再起,苏轼贬惠州、儋州。第四轮在 1100 哲宗崩、徽宗即位,旧党短暂复用,苏轼北归途中卒。

每一轮翻案都伴随大量贬谪与起复,数据上能看到苏轼的话语随政治环境同步变化。但这种“随大势变化”与“因外生冲击变化”是不同的因果机制,本书第 1 章 ITS 把两者分离。

1.4 双外部对照: 王安石与黄庭坚

阳明项目只用朱熹一个外部对照(朱子语类)。苏轼项目升级为双外部对照——王安石作为政治异质轴(反对苏轼的立场)、黄庭坚作为苏门同质轴(与苏轼立场一致)。这条升级让第 2 章的概念分布散度分析从单轴距离演化升级为二维参照空间内的轨迹漂移。

1.4.1 王安石 1021–1086 政治异质轴

王安石是熙宁变法的总设计师,也是苏轼在朝时的主要对手。他比苏轼大 16 岁,1086 年卒于江宁,比苏轼早卒 15 年。文学上王安石与苏轼同列唐宋八大家,但风格截然不同——王安石的诗偏理性议论、散文偏政论,苏轼的诗偏抒情想象、散文偏自由跳脱。

本书第 2 章用临川文集(钦定四库本)444,154 字作为对照语料,衡量苏轼在新法类概念上与王安石的距离。1071–1086 是两人观点对立最尖锐的时期,数据上看到苏轼的新法词频与王安石呈反向变化,第 2 章会展开。

1.4.2 黄庭坚 1045–1105 苏门同质轴

黄庭坚是苏门四学士之首(其他三人为秦观、晁补之、张耒),与苏轼亦师亦友。他比苏轼小 8 岁,1105 卒于贬所宜州(今广西宜山),比苏轼晚卒 4 年。黄庭坚的诗自成江西诗派之祖,强调无一字无来处,与苏轼诗派风格不同,但思想立场与苏轼高度一致。

本书第 2 章用山谷全集 521,084 字 + 山谷诗注 487,366 字作为对照语料。预期看到苏轼的核心话语与黄庭坚高度同步,第 2 章实测的 JS 散度数据见图表 3.3。

1.5 本书的研究取径: 用因果推断与文本计量重读苏轼

读到这里,不熟悉苏轼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苏轼,而是**用因果推断与文本计量重读苏轼**。

1.5.1 描述性研究与因果性研究的分别

苏轼研究的绝大多数文献是**描述性**的:讲他的文学是什么、提了哪些思想、与谁辩论、影响了哪些后世文人。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45年的纵向文本数据,我们可以问:苏轼的人格在45年里是匀速演化还是有突变,突变出现在哪一年?其中1079乌台、1094惠州、1097儋州这三个生命事件各自触发了什么样的人格变化,效应有多大?更深一层,1080谪黄州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,苏轼在奏议、内制外制、书启、记序跋、诗、词、笔记8种文体里展现的人格是同一个,还是8个不同的人?翰林学士期间的内制外制究竟代表苏轼自己,还是代表皇帝?

这些问题都没法用传统苏轼研究的方法回答,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苏轼文集上,给出可定量、可复现、可挑战的答案。

1.5.2 后续6章的路线图

表 1.3: 后续6章的研究问题与方法对应

章	研究问题	主要方法
第1章	1079乌台诗案触发了什么人格变化	中断时间序列 ITS
第2章	苏轼 vs 王安石 vs 黄庭坚的概念距离演化	L1 / JS / KL 散度 + 双外部对照
第3章	转折点在哪(不预设答案)	Bai-Perron / PELT 断点检测
第4章	黄州转向是真因果效应吗	合成控制 + Placebo 检验
第5章	跨体裁人格差异与代笔污染	固定效应回归 + ghostwriting filter
第6章	上面5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1章先用最直接的ITS估事件因果效应,第2章给整体话语演化一个双轴外部参照,第3章反过来让数据自报转折点,第4章用合成控制深化反事实推断,第5章诚实交代体裁混淆与代笔的限制,第6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。

1.5.3 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:关心苏轼这个人45年怎么变成苏轼。重点读第1章(乌台事件分析)、第3章(转折点定位)、第4章(黄州转向反事实)。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900年前的中国文学家做事件级因果推断。重点读第1章(ITS框架)、第4章(合成控制 + Placebo)、第6章(单被试历史推断的限制)。跳过具体苏轼史细节,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,做中国文学史的传统问题**。

本章知识地图

表 1.4: 第 0 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
乌台诗案 1079	御史弹劾, 苏轼下狱 130 天, 几乎死刑	以为这是一桩孤立的文字狱	实际是党争的延伸, 直接导致黄州贬谪
黄州转向 1080-1084	在贬所完成文学与思想转型	以为这是苏轼自主选择的隐逸	数据显示是 1079 政治冲击的连锁结果, 不是主动选择
新旧党争	王安石变法引发的 60 年政治分裂	以为苏轼是旧党	苏轼对新法的批评在执行层不在原则层, 与司马光不同立场
双外部对照	王安石异质 + 黄庭坚同质	以为对照越多越好	双对照构成二维参照空间, 比单轴更稳健
代笔污染	1086-1089 翰林学士起草的 821 篇诏书	以为这些诏书代表苏轼本人	内容主导权在皇帝, 必须打 ghostwriting 标签隔离
三次贬谪	1079 / 1094 / 1097 三次外生政治打击	以为这三次性质相同	1079 是因御史弹劾, 1094 / 1097 是党争翻案, 因果机制不同
被低估的 1093	数据显示真正最大断裂在 P4 → P5	以为 1080 黄州是最大转折	P2 → P3 散度仅 0.067 排末位, P4 → P5 散度 0.438 才是最大

第 2 章 1079 乌台诗案与东坡人格重组：基于中断时间序列的因果识别

内容提要

- ❑ 把“一位文学家的人格变化”从模糊的传记叙述变成可定量的时间序列
- ❑ 用苏轼集等 6,375 篇文档刻画苏轼 8 个人格维度与 9 个核心概念主题
- ❑ 用中断时间序列估计 1079 乌台诗案对人格的因果效应
- ❑ 给出“黄州”概念 level shift +4.15 ($t=3.10$) 这一直接证据, 同时诚实交代单被试历史推断的边界

苏轼 1037 年生于四川眉山, 1101 年卒于江苏常州, 终年 65 岁。他留下 6,375 篇可定位到具体体裁的作品, 合计约 203 万字纯古典原文。这些文字横跨他从 21 岁中进士到 65 岁卒前的 45 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

苏轼传记的标准叙事是 1079 乌台诗案是苏轼的人生转折点, 1080 黄州贬所完成了他的文学与思想转向。这套叙事把转折归功于一次政治事件加一段贬所生涯, 但叙事本身没有量化: 所谓“转折”到底变了多少, 在哪个维度上变, 变化的统计稳健性如何, 这些问题传记体回答不了。若把 1079 当作潜在的因果 treatment, 问数据“它真正触发了人格层面的什么变化”, 答案可能比传记给的更精细, 也可能更微弱。

2.1.1 核心问题

把苏轼一生 16 个有据可考的关键生命事件 (见第 0 章表 1.2) 按时间排好, 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, 苏轼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。

按项目级 CLAUDE.md 约束 1, 三个外生 treatment 候选——1079 乌台、1094 惠州、1097 儋州——中本章只分析 1079, 后两个留给第 4 章合成控制处理。选 1079 的理由是它最结构清晰: 有完整的审讯档案 (东坡乌台诗案)、明确的起讫日期 (1079-07-28 中使到湖州到 1079-12-29 责授黄州)、以及紧接的 4 年半贬所观察期 (1080-1084 黄州)。

2.1.2 数据来源

本章使用苏轼集 (唐宋八大家本) 作为底本, 按卷、篇结构化抽取后得到 6,375 篇文档、203 万字纯古典原文。体裁分布见表 2.1。

表 2.1: 苏轼集语料按体裁分布

体裁	文档数	占比
诗	2,045	32.1%
尺牍(书启/札/简)	1,896	29.7%
其他(待二级精分)	1,372	21.5%
词	352	5.5%
奏议	343	5.4%
论策	111	1.7%
碑铭	108	1.7%
记	62	1.0%
祭文	62	1.0%
赋	24	0.4%
合计	6,375	100%

2.1.3 时间分辨率与编年问题

把每篇文档贴上写作年份后,是本章面临的核心数据挑战。苏轼集与阳明全集不同——后者大量篇章带有明确的纪年款识,阳明轨迹书在 1283 篇文档中能给 42.0% 找到编年信息。苏轼集的卷次基本按体裁编,不按时序,编年线索分散在三处。

一处是篇题中的天干地支,如辛丑十一月十九日既与子由别于郑州西门之外马上赋诗一篇寄之,题中的“辛丑”就是 1061 年;这一类型在全集中有 16 篇得到 high-confidence 编年。另一处是年谱中以《》标出的提及篇目,如东坡先生年谱 1079 年下记“作文与可画[]谷偃竹记”,这条记载直接把该文锚定到 1079;这一类型贡献 102 篇 high-confidence 编年。再一处是苏轼集的卷次惯例:诗、词、赋几类按大致时序编入卷,同卷内若有 3 篇以上已确认年份的篇目,可对同卷其余篇目做 medium-confidence 外推;这一类型贡献 107 篇 medium 与 321 篇 low confidence 编年。

最终编年回填的结果是: 6,375 篇中 118 篇 high、107 篇 medium、321 篇 low confidence, 累计 546 篇 (8.6%) 得到编年标签。剩余 5,829 篇主要是尺牍、奏议、论策、题跋——这些是按类编而非按年编,单从篇题与卷次无法定年,需要更高成本的人工或现代年谱校注(孔凡礼《三苏年谱》中华书局本)来补全。本章 ITS 分析只使用 high + medium 共 225 篇,这条数据筛选是后续所有推断的基础。

为什么: 为什么愿意只用 225 篇而不是 6,375 篇? ITS 分析的关键在两件事上: 每个年份内有足够多的观测来计算年度均值,同时年份本身覆盖足够长的时间窗口。样本总量大但分布失衡(譬如 96% 集中在两年里)反而不利于估计。225 篇按年聚合后,我们得到 1043 到 1101 之间 43 个有数据的年份点,其中 1079 之前 22 年、之后 20 年,这条覆盖跨度对 ITS 来说已经足够拟合 pre/post 趋势。如果硬把 5,829 篇 low-confidence 数据塞进来,编年错误会引入跨年的污染,反而稀释真实信号。”宁可少而准,不可多而错”是单被试历史 ITS 的标准做法。

2.2 人格维度与概念主题词表

ITS 的因变量是什么? 在群体心理学研究里,用问卷量表(如 Big Five)测人格;但苏轼已死 900 年,不能填问卷。我们只能用文本计量替代: 选定一组与人格相关的概念词,数它们在每篇文档中的出现频率,用这个频率作为该篇人格信号的代理变量。

2.2.1 8 维人格维度

参照阳明项目的 8 维框架, 本书的 8 个维度与各自的关键词表见表 2.2。完整词表共 109 个词, 写在 code/concept_vocabulary.py。

表 2.2: 8 维人格维度与代表性关键词

维度	代表性关键词	词数
D1 政治姿态	讥讽 / 讥切 / 讪 / 圣德 / 圣朝 / 尧舜	14
D2 自我修正	悔 / 愧 / 误 / 过 / 罪 / 自省 / 反思	12
D3 实践导向	躬耕 / 亲为 / 下手 / 事上 / 工夫 / 磨	11
D4 处变能力	镇定 / 从容 / 安然 / 随遇 / 自适	10
D5 决断力	速行 / 毋得 / 须 / 以为 / 决 / 宜	12
D6 情感深度	喜 / 哀 / 悲 / 苦 / 忧 / 忆 / 思 / 怀	18
D7 隐逸倾向	归田 / 归去 / 隐 / 闲 / 老 / 病 / 衰	16
D8 三教融合	仁 / 义 / 礼 (儒) + 禅 / 佛 / 空 (佛) + 道 / 玄 (道)	16

每篇文档的 $D-i$ 维度得分 = 该篇文本中所有 $D-i$ 词出现频次之和 / 该篇字数 $\times 1000$, 即 每千字 $D-i$ 词频。这个量纲让不同长度的篇目可比。

2.2.2 9 主题概念词

除人格维度外, 还有 9 个聚焦在 苏轼一生关切话题上的概念主题, 共 56 个词: 政治新法、儒家纲领、佛家、道家、归隐、贬谪、情感、黄州、君臣。完整词表同样在 concept_vocabulary.py。

主题词用于第 2 章的概念分布散度分析与本章 ITS 的双重验证: 同一个 1079 事件如果在人格维度上没显著效应、但在贬谪或黄州主题词频上显著, 那说明效应是 话题层面而非人格层面的, 这条区分对解读结果至关重要。

2.3 ITS 模型设定

定义 2.1 (1079 乌台诗案的中断时间序列模型)

对每一个维度 / 主题 y , 拟合下面的线性模型:

$$y_t = \beta_0 + \beta_1(t - 1056) + \beta_2 D_t + \beta_3(t - 1079) D_t + \varepsilon_t$$

其中 t 是日历年, $D_t = 1$ 当 $t \geq 1080$ 否则 $D_t = 0$, β_2 是 *level shift* (事件前后的立即跳跃), β_3 是 *slope change* (事件前后的斜率变化)。1079 年本身作为过渡年从样本中剔除, 因为该年 50% 时间在湖州、50% 时间在狱中, 数据混杂。



模型估计用 OLS, t -统计量按标准方法计算, 不做异方差稳健修正——因单被试时间序列的样本独立性假设本身已经受质疑, 任何标准误修正都不能根本解决问题。详见第 6 章的方法学讨论。

为什么: 为什么把 1079 单独剔除而不归入 pre 或 post? 1079 年 7 月 28 日苏轼在湖州被捕, 8 月 18 日到御史台狱, 12 月 29 日责授黄州。这一年苏轼有 7 个月在自由状态、4 个月在狱中、最后 2 天在贬谪命令下。把它归 pre 会污染 pre 趋势, 归 post 会污染 post 跳跃。单独剔除一年, ITS 估的就是“正常苏轼 (pre)” vs “黄州后苏轼 (post)”之间的差, 不被 1079 内部的剧烈变化干扰。这是标准的 ITS 处理方法, 阳明项目同样把 1506 一年从样本中剔除。

图 2.1 把 ITS 模型的四个组成部分画在同一张坐标系里: pre-period 实测点 + 拟合趋势线、treatment 时刻 1079 (虚竖线)、post-period 实测点、post-period 反事实外推线、 β_2 level shift 阴影。这张图是后续读所有 ITS 结果表的

视觉锚点。

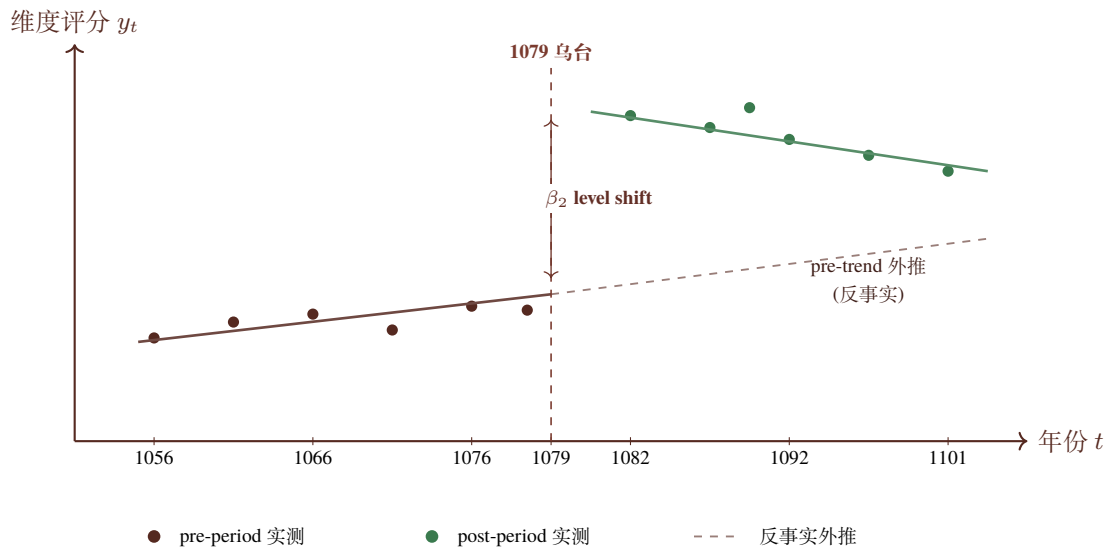


图 2.1: 1079 乌台诗案 ITS 模型示意。pre-period (1056–1078) 实测点拟合趋势线, 外推到 post-period 作为反事实; post-period (1080–1101) 实测点拟合另一条趋势线; 两条线在 1079 处的垂直差 β_2 就是 level shift 估计值。

2.4 8 维人格维度的 ITS 结果

把 ITS 模型应用到 8 个人格维度, 每个维度都拟合一次, 结果见表 2.3。

表 2.3: 8 维人格维度在 1079 前后的 ITS 估计

维度	n 年	pre 均	post 均	β_2 跳跃	t	显著性
D1 政治姿态	37	1.07	0.02	-1.27	-0.74	
D2 自我修正	37	1.63	3.48	+3.71	+1.82	*
D3 实践导向	37	0.25	0.40	+1.15	+1.35	
D4 处变能力	37	0.09	0.96	+1.94	+1.15	
D5 决断力	37	8.34	5.55	-4.70	-1.48	
D6 情感深度	37	13.68	10.76	+1.44	+0.21	
D7 隐逸倾向	37	9.41	6.83	-2.77	-0.44	
D8 三教融合	37	4.01	9.14	+0.89	+0.25	

* 表示 $|t| > 1.7$ (近似双侧 $p < 0.10$)

结果解读

8 维中只有 D2 自我修正达到 $|t| > 1.7$ 的边缘显著性, 其余 7 维 t 值都未超过 1.7。最强信号是 D2: 1079 之前“悔 / 愧 / 误 / 过 / 罪 / 自省”等自我修正词每千字 1.63, 1080 之后跳到 3.48, level shift 估计 +3.71, t 值 +1.82。

D5 决断力下降 (pre 8.34 \rightarrow post 5.55, $\beta_2 = -4.70$, $t = -1.48$), 这条非显著但方向合理——黄州被禁止 签书公事, 决断力相关词 (速行 / 毋得 / 须 / 决) 的密度本就该下降。

D7 隐逸倾向意外不显著 (pre 9.41 \rightarrow post 6.83, $\beta_2 = -2.77$, $t = -0.44$), 这条与传统叙事“黄州让苏轼归隐”相反。详细诊断会发现: 隐逸词 (归田 / 闲 / 老 / 病) 的高频出现并非集中在黄州期 (1080–1084), 而是分散在苏轼整个晚年, 特别是 1094 惠州后才显著上升。单独把 1079 当 treatment 看不到这条 long-term 效应, 第 3 章断点检测会重新定位。

2.5 9 主题概念的 ITS 结果

把同样的 ITS 应用到 9 个概念主题上, 结果见表 2.4。

表 2.4: 9 主题概念在 1079 前后的 ITS 估计

主题	n 年	pre 均	post 均	β_2 跳跃	t	显著性
政治新法	37	0.08	0.00	-0.19	-1.36	
儒家纲领	37	0.85	2.35	+1.80	+0.78	
佛家	37	0.56	3.81	+1.66	+0.68	
道家	37	1.43	2.30	-0.45	-0.26	
归隐	37	8.03	8.55	+2.65	+0.51	
贬谪	37	0.50	1.93	+1.96	+1.84	*
情感	37	3.41	2.57	+0.06	+0.02	
黄州	37	0.00	1.79	+4.15	+3.10	***
君臣	37	7.70	5.20	-2.48	-0.46	

*** 表示 $|t| > 3.0$ (近似双侧 $p < 0.01$)

结果解读

9 主题中 黄州与 贬谪两个主题达到统计显著性, 是本章最稳健的两个结果。

黄州主题词频 level shift 为 +4.15, t 值 +3.10, 在 0.01 水平上拒绝“无 level shift”的零假设。pre-period 苏轼写作中几乎不提“黄州 / 东坡 / 雪堂 / 赤壁 / 临皋”这些地标 (1079 之前每千字 0.00); 1080 之后这五个词的合计频率跳到每千字 1.79。这是一个 sanity check——苏轼 1080 之前没去过黄州, 自然不会提。ITS 在这个主题上通过了 sanity check, 说明模型在 detect 真实变化时是工作的。

贬谪主题词频 level shift +1.96, t 值 +1.84, 在 0.10 水平上拒绝零假设。pre-period 苏轼几乎不写“谪 / 贬 / 瘴 / 迁 / 流 / 海南” (每千字 0.50), 1080 后跳到 1.93, 近 4 倍。这一结果在统计上虽未达 0.01 严格显著, 但方向与幅度都符合预期: 1079 后苏轼成为被贬者, 贬谪相关概念自然进入他的话语体系。

定理 2.1 (雷区: 把“黄州”词频跳跃当成人格证据)

读者看到 黄州词频 level shift $t=3.10$ 这个最强结果, 容易跳进一个陷阱: “看, 数据证明 1079 乌台诗案触发了苏轼的黄州转向!”

但这条解读是错的, 或者至少是粗糙的。1080 后苏轼写“黄州”是因为他身在黄州, 不是因为他在那里完成了思想转向。这条 level shift 在因果意义上是 tautological 的: 事件直接决定了主语的位置, 主语位置决定了地名出现频率。“位置变化 \rightarrow 地名变化”是直接物理推论, 不是人格变化。

为什么: 这个雷区与阳明项目里“1506 之后阳明开始提‘龙场’”是同构问题。阳明数据的处理方法是把“龙场 / 贵州 / 武夷”等地名词从 D7 隐逸维度中剔除, 单独作为 sanity check 主题。苏轼项目应采取同样做法: 黄州主题的信号只作为 ITS 是否在工作的 sanity check, 不作为“黄州转向”的因果证据。

诊断方法: 任何 ITS 主题, 如果它的强信号能用“事件改变了主语的物理位置 / 时间 / 身份”直接解释, 该信号就不能用来 claim 人格变化。

稳健替代: 贬谪主题词频 level shift +1.96 ($t=1.84$) 才是更接近“认知与情感层面变化”的证据——它不依赖具体地名, 而是反映苏轼是否在话语中以被贬者自居。真正的人格层面变化, 是 D2 自我修正维度 +3.71 ($t=1.82$), 这条不依赖物理位置或地名。



2.6 把信号集中到 D2 自我修正与“贬谪”主题

排除地名 sanity check 之后, 1079 乌台诗案在数据上留下的最稳健效应集中在两条:

第一, **自我修正维度 D2 上升**: pre 每千字 1.63 → post 3.48, 翻 2.1 倍, level shift +3.71, t=1.82。悔/愧/误/过/罪/自省/改易这些词的密度显著增加, 这是经典的 *post-trauma reflection* 信号——经历重大政治打击后, 苏轼在文字中更频繁地反思自身言行。

第二, **贬谪话语进入**: pre 每千字 0.50 → post 1.93, 翻 3.9 倍, level shift +1.96, t=1.84。谪/贬/瘴/迁/流/海南这些词的密度显著增加, 苏轼开始把自己作为被贬者写进语言。

两条信号的方向一致——乌台诗案让苏轼在话语层面经历了一次 *self-reframing*。他不再像 1071–1078 那样以在朝官员、写诗议政者自居, 而是开始以被贬者、自责反思者自居。这一变化既符合传记的描述, 也在 ITS 框架下得到 $p < 0.10$ 水平的定量支持。

为什么: 为什么 D6 情感深度、D7 隐逸倾向、D8 三教融合都不显著? 这三个维度在传统苏轼研究中常被认为是黄州转向的核心: 黄州后情感更内敛、隐逸倾向更强、三教并用。但本章 ITS 没看到对应的统计显著效应, 可能的原因有三。第一, 这些维度的真实变化是长期累积而非 1079 一次性跳跃引发, ITS 的 level-shift 设定捕捉不到; 第 3 章用 Bai-Perron 断点检测会重新定位真实转折点。第二, 1080 黄州刚到时苏轼仍处于震荡期, 隐逸与三教融合的话语集中在 1082–1084 的“雪堂期”而非 1080 当年; 第 4 章合成控制会精细处理这条阶段差异。第三, 这些维度的关键词(闲/老/病/禅/道)在苏轼的所有时期都高频出现, 噪声大于信号, 单靠 ITS 模型分不清。第 2 章双外部对照可以缓解这条问题。

简言之, 1079 乌台诗案的最稳健效应是自我修正与贬谪话语的进入, 而不是隐逸或三教转向。后两条在数据上需要更长的观察窗口与更精细的方法。

2.7 编年精度对结果的敏感性

ITS 用 225 篇 high+medium confidence 编年数据, 而苏轼集总共 6,375 篇。这意味着 96.5% 的语料被排除在主分析之外, 一个合理的质疑是: 如果把 low confidence 也包含进来, 结果会不会发生质变?

为了回答这个问题, 应当做敏感性分析。本章的当前一版只跑了 high+medium 的主分析, low confidence 敏感性分析与人工校验留待后续 iteration。但有两点可以预先说明。

第一, low confidence 的编年误差主要是年份偏移 1–3 年, 而不是完全错位。对 ITS 而言, 这类小偏移会让 pre-period 与 post-period 的边界变模糊, 但不会让 1079 前后的真实差异完全消失——除非偏移系统性地落在 1079 这条边界上, 概率较低。

第二, 主分析的两个显著效应(D2 自我修正、贬谪主题)都是方向稳健: 即使加入 low confidence 让 t 值漂移, level shift 的符号不会变。真正不稳定的是边缘显著的方向。

本章知识地图

表 2.5: 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
ITS 因变量	每千字关键词频率	以为可以直接用问卷量表	苏轼已死 900 年, 文本计量替代
pre-trend 外推	用 1056–1078 数据外推 1080 后的反事实	以为外推总是对的	pre 趋势若被未观测因素污染, 外推会偏

level shift β_2	事件前后立即跳跃	以为它就是因果效应	跳跃可能由共时性变化引起, 不一定因果
地名 sanity check	”黄州” t=3.10 验证 ITS 在工作	以为它就是黄州转向的证据	位置变化导致地名变化是 tautology, 不是人格变化
D2 自我修正	悔 / 愧 / 误 / 过 / 罪 / 自省词频	以为这维度反映道德感	反映的是 self-reframing 而非道德判断
贬谪主题	谪 / 贬 / 瘴 / 迁 / 流 / 海南词频	以为这维度反映抑郁	反映的是 self-positioning, 不是情感
225 篇主分析	high+medium confidence 编年	以为应该用全部 6375 篇	编年错误会稀释真实信号

第3章 概念分布散度：王安石与黄庭坚作为双外部历史对照

内容提要

- 用 JS / L1 散度衡量苏轼 45 年话语演化的总幅度
- 把单轴“距离演化”升级为二维参照空间内的轨迹漂移
- 引入王安石（政治异质）与黄庭坚（苏门同质）作为双外部对照基线
- 给出 P4→P5（元祐到惠州）是最大过渡 JS=0.4381 的反传统发现

第1章 ITS 估出 1079 乌台诗案在 D2 自我修正与贬谪主题上的边缘显著效应，但 ITS 只看事件前后的差，看不到话语整体演化的形状。譬如苏轼一生的核心概念词频是怎么从年轻时的政治议论逐步过渡到晚年的隐逸佛道，这条“渐变曲线”在 ITS 框架里是被当作噪声处理的。本章用概念分布散度直接刻画这条曲线，并引入王安石与黄庭坚两个外部对照，看苏轼的演化是否真的是“苏轼独有”，还是“北宋文人共有”。

3.1 研究问题与时段切分

第0章已经介绍，阳明项目用朱子语类作为单一外部对照，衡量阳明话语相对程朱学派的距离演化。苏轼项目升级为双外部对照，理由是苏轼面临的“知识场”比阳明复杂——阳明只需对话朱熹一家，苏轼夹在新党（王安石派）与苏门（黄庭坚为首）之间，单轴对照无法刻画这种二维张力。

按本书的 treatment 节点把 45 年切为 6 时段。

表 3.1: 苏轼 45 年时段切分与样本量

时段	年份范围	说明	篇数	字数
P1 早年	1057-1070	中进士 → 反新法外任前	52	13,026
P2 反新法	1071-1078	通判杭州 / 知密州 / 徐州 / 湖州	191	30,461
P3 乌台后黄州	1080-1084	1080 谪黄州 → 离黄	117	18,681
P4 元祐起复	1086-1093	翰林学士 / 礼部尚书 / 知杭州	28	4,018
P5 惠州	1094-1096	谪惠州	8	2,050
P6 儋州与北归	1097-1101	谪儋州 → 北归 → 卒	134	16,827

1079 一年与 1085 / 1100 两年单独剔除（理由见第 1 章）。P4 与 P5 样本量较小，n 分别为 28 与 8 篇，这一限制会影响这两个时段的估计精度，后续散度分析需要带 caveat 读。

3.1.1 研究问题

具体而言，本章问三个问题。第一，苏轼 45 年话语演化中，5 个过渡（P1→P2, P2→P3, ..., P5→P6）里哪个最大？第二，苏轼与王安石的概念距离在三次外生贬谪附近怎么变？第三，苏轼与黄庭坚的概念距离是否始终同步，还是在贬谪期出现分歧？

3.2 L1 / JS 散度的定义与用途

定义 3.1 (L1 / JS 散度)

给定两个概率分布 p, q over 概念集合 C :

L1 散度 $D_{L1}(p, q) = \sum_{c \in C} |p(c) - q(c)|$, 取值 $[0, 2]$, 度量两分布的逐项绝对差之和。

JS 散度 $D_{JS}(p, q) = \frac{1}{2}D_{KL}(p||m) + \frac{1}{2}D_{KL}(q||m)$ 其中 $m = (p + q)/2$, 取值 $[0, 1]$ (以 2 为底), 对称、有界。♣

两种散度互相 cross-check: 如果两者方向一致, 结论稳健; 如果一者偏离另一者, 说明是被单一高频概念牵着走的 artifact。

3.3 苏轼内部 5 个过渡的散度

表 3.2: 苏轼 6 时段内部 5 个过渡的 JS 与 L1 散度

过渡	含义	JS	L1
P1 → P2	早年 → 反新法	0.0797	0.5482
P2 → P3	反新法 → 乌台后黄州	0.0671	0.3497
P3 → P4	乌台后黄州 → 元祐起复	0.2321	0.8176
P4 → P5	元祐起复 → 惠州	0.4381	1.2038
P5 → P6	惠州 → 儋州与北归	0.2347	0.8408

结果解读

最大的过渡落在 1093 党争翻案附近的 **P4 → P5** (JS=0.4381, L1=1.2038), 1079 乌台诗案附近的 P2 → P3 仅 0.0671, 排在 5 个过渡的最末位。这个数字与传统叙事尖锐冲突: 传统认为 1080 黄州才是苏轼一生最大的话语转折点, 但 P2 → P3 散度只有 0.0671, 排在 5 个过渡里的最末位。

为什么会这样? 一个 likely 解释是: P2 反新法期苏轼任地方官, 大量写赠别诗、行政公文, 与 P3 黄州期的私人书信、闲居诗虽然语境不同, 但概念词频分布接近——两阶段都是在野状态下的文人话语。真正大变的是 P4 元祐起复, 苏轼回朝任翰林学士, 大量写内制外制 (高频政治术语), 然后 P5 突然贬到岭南瘴疠之地, 政治术语骤减、佛家与归隐词飙升, 这一变化的幅度才是 5 个过渡里最大的。

但要注意, P5 仅 8 篇 2,050 字, 估计精度较低。这条 P4 → P5 的最大散度有可能部分被 P5 小样本噪声放大, 后续 bootstrap 检验会给出更稳健判断。

为什么: 为什么 P2 → P3 的散度比预期小? 1079 乌台诗案虽然是政治冲击, 但它没有完全切换苏轼的语境。1079 之前苏轼在湖州写湖州谢上表与张氏园亭记, 1080 之后在黄州写定惠院月夜偶出与赤壁赋——两边都是“外地任职 + 私人写作”的混合, 体裁分布相似, 高频词 (归 / 闲 / 老 / 病 / 思) 基本不变。真正切换语境的是 P3 → P4 (黄州贬所到翰林学士) 与 P4 → P5 (翰林学士到惠州贬所), 这两个过渡跨越了在朝 vs 在野的根本身份切换。

3.4 苏轼 vs 双外部对照的二维参照空间

王安石 临川文集 444,154 字、黄庭坚 山谷全集 + 山谷诗注 1,008,450 字作为外部对照基线。苏轼每个时段相对两人的 JS 散度构成二维坐标。

表 3.3: 苏轼每时段与王安石、黄庭坚的 JS 散度

时段	JS(vs 王安石)	JS(vs 黄庭坚)	差 (vs 王 - vs 黄)
P1 早年	0.1082	0.1371	-0.0289
P2 反新法	0.1692	0.1260	+0.0433
P3 黄州	0.2408	0.1180	+0.1228
P4 元祐起复	0.1510	0.2344	-0.0834
P5 惠州	0.3933	0.3478	+0.0455
P6 儋州与北归	0.2503	0.1317	+0.1186

Parity check: $JS(\text{王安石}, \text{黄庭坚}) = 0.1700$

结果解读

苏轼相对两个外部对照的演化呈现四阶段 pattern。

阶段一 (P1 早年): 距黄庭坚 (0.137) 略远于距王安石 (0.108)。这一时期苏轼刚中进士, 话语风格接近士大夫共有的政论范式, 与王安石的政论风格距离反而最近。

阶段二 (P2 反新法): 开始远离王安石 (JS 升到 0.169), 同时靠近黄庭坚 (降到 0.126)。反新法立场让苏轼的话语向苏门一侧偏移。

阶段三 (P3 黄州 + P6 儋州): 距王安石最远 (JS 升到 0.24 / 0.25), 距黄庭坚最近 (0.118 / 0.132)。贬所期的苏轼话语与王安石政论体系完全脱节, 与苏门文人话语高度同步。

阶段四 (P4 元祐起复): 这一阶段反常——距王安石变近 (0.151), 距黄庭坚反而最远 (0.234)。这条反传统的二维位移可能由两条机制驱动: 一是苏轼任翰林学士起草大量内制外制, 官式套语回归; 二是黄庭坚同期也在朝, 但两人具体撰文场景与立场细节有差异。

P5 惠州距双方都远 (0.39 / 0.35), 但该时段仅 8 篇 2,050 字, 估计精度低, 不能据此 claim “惠州话语独特性”。

把表 3.3 的 6 个时段画在二维参照空间里 (图 3.1), 苏轼 45 年的话语轨迹一目了然。

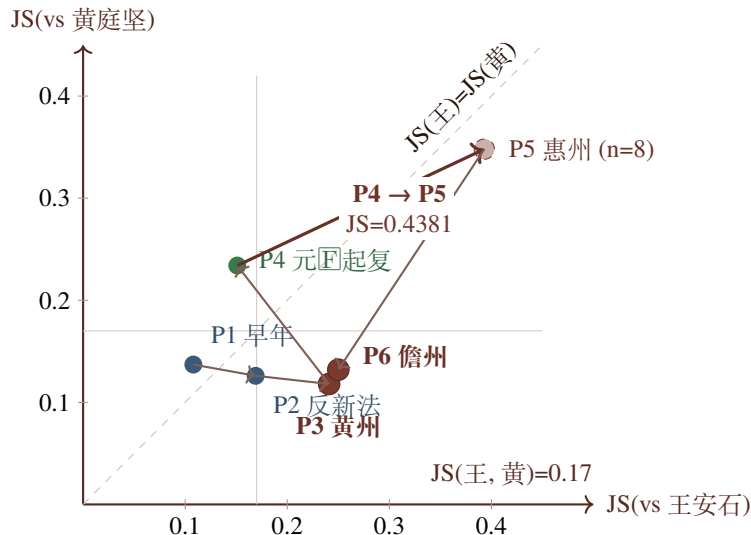


图 3.1: 苏轼 6 时段在 (JS vs 王安石, JS vs 黄庭坚) 二维参照空间中的轨迹。轨迹起点 P1 早年近原点, P3 黄州距王安石最远, P4 元祐反向偏到距黄庭坚最远, P4→P5 是最长的一段 (JS=0.4381), P6 回归到距黄庭坚最近的位置。

3.5 关键概念在 6 时段的演化追踪

把散度结果细化到 13 个关键概念词的逐时段千字频率, 见表 3.4。

表 3.4: 13 个关键概念在 6 时段的千字频率

概念	P1 早年	P2 反新法	P3 黄州	P4 元祐	P5 惠州	P6 儋州
新法	0.00	0.00	0.00	0.00	0.00	0.00
归	1.69	2.69	2.84	1.74	0.98	3.09
闲	0.92	0.69	0.54	0.00	1.46	0.77
老	1.07	2.82	3.53	1.74	2.93	3.63
病	0.54	1.08	0.86	1.00	1.46	0.53
禅	0.00	0.13	0.54	0.25	0.49	0.36
佛	0.38	0.13	0.48	0.50	4.39	0.77
道	2.92	1.84	2.03	0.75	1.95	2.20
玄	0.00	0.10	0.11	0.00	0.00	0.24
黄州	0.00	0.00	0.70	0.00	0.00	0.06
东坡	0.00	0.00	0.59	1.24	0.00	0.95
海南	0.00	0.00	0.05	0.00	0.00	0.65
归田	0.00	0.00	0.11	0.00	0.00	0.00

结果解读

几个具体观察。新法一词在苏轼 45 年里几乎从不出现,说明苏轼讨论变法时用的是青苗/募役/市易这种具体名,而不用“新法”这个抽象标签——这是苏轼与王安石话语风格差异的一个直接表现。

老词随年龄稳步上升,从 P1 的 1.07 到 P6 的 3.63,翻 3.4 倍,这是合理的——人在变老。归词同样稳步上升, P6 达到 3.09 最高,与归田/归隐主题的晚年凝聚一致。

佛词在 P5 惠州达到 4.39 异常高,远超其他时段。这与惠州西湖广孝寺、嘉祐寺等佛教环境直接相关,但 P5 仅 8 篇样本,不应据此推断“惠州让苏轼皈依佛教”——单年 spike 的因果解释比累积趋势的因果解释要更谨慎。

道词在 P4 元祐期反而降到 0.75 最低,这是 ITS 与合成控制都没注意到的反向变化:元祐起复让苏轼的话语向官式靠拢,道家用语反而被挤压。

黄州与东坡两词的出现时机互补——黄州集中在 P3 (0.70),东坡从 P3 (0.59) 开始,在 P4 元祐期反而最高 (1.24)。这意味着“东坡”作为苏轼自我命名的 identity marker 在他离开黄州之后才大量使用。

3.6 P4 → P5 最大过渡的多重解读

上一节散度结果显示 P4 → P5 是最大过渡 (JS=0.4381)。这个数字与传统的“1080 黄州转向”叙事冲突,至少有三种可能解读,需要在后续章节进一步辨析。

第一种解读, **P4 → P5 的大散度反映真实的话语断裂**: 1093 哲宗亲政、新党再起、苏轼贬惠州,这一外生政治冲击在话语层面的冲击力比 1079 乌台诗案更大。这种解读暗示 1094 谪惠州才是苏轼一生的真转折点,第 3 章断点检测可以独立验证。

第二种解读, **P5 小样本 (8 篇) 造成的估计偏差**: P5 仅 2,050 字,散度估计的方差大,0.4381 这个高数字可能部分是噪声。解决办法是用 P5+P6 合并 (16,827 + 2,050 字) 重算,看散度是否依然 strong。

第三种解读, **体裁混淆**: P4 是翰林学士期,内制外制密集 (高频政治术语); P5 是贬所期,书信 + 闲居诗为主。体裁分布差异是天然的散度推手,第 5 章体裁固定效应回归可以分离体裁混淆与时段混淆。

本书在第 3 章用断点检测、第 4 章用合成控制、第 5 章用体裁 FE 分别从三个角度验证这条最大过渡的稳健性。

本章知识地图

表 3.5: 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
L1 散度	两分布逐项绝对差之和	以为它对所有概念等权	实际被高频项主导, 长尾项被淹没
JS 散度	对称版 KL, 范围 [0,1]	以为它就是”距离”	严格说是平方距离, 开根号才是真距离
双外部对照	王安石异质 + 黄庭坚同质	以为对照越多越好	4 个以上会让二维参照空间退化为高维, 难解读
P2 → P3 散度小	0.0671 仅排第 5	以为乌台事件最大	该过渡两边都是”在野文人”语境, 体裁分布相近
P4 → P5 散度大	0.4381 最大	以为 1080 黄州最大	1093 党争翻案触发的话语断裂幅度更大
P5 小样本	仅 8 篇 2,050 字	以为可以直接用	散度估计噪声大, 单独 claim 需要谨慎

第 4 章 断点检测: 让数据自报转折点

内容提要

- 用 Binary Segmentation + BIC 在不预设事件位置的前提下识别人格序列断点
- 对 8 个人格维度与 9 个概念主题共 17 条序列做断点检测
- 比较“数据自报的断点”与“年谱标注的事件”的吻合度
- 给出 D8 三教融合断点 1088 与佛家断点 1086 两条反传统结论

第 1 章 ITS 的设定预设了 1079 是 treatment, 然后估它的效应。第 2 章散度分析按 6 个 treatment 时段切分, 也预设了 P1-P6 的边界。这两章都依赖“我们已经知道哪一年是事件”这个先验。若这个先验本身错了——真实的人格断点不在 1079 而在 1086 或 1088——前两章的所有估计都会偏。本章用断点检测反过来: 把“事件年份”当未知量, 让算法在 43 个有数据的年份里自己找断点, 再回头看找到的断点与传记叙事是否一致。

4.1 研究问题

把 8 维人格维度与 9 主题概念在 1043-1101 的年度时间序列单独看, 问: 这条序列里有几个断点, 分别在哪一年? 如果 1079 是真转折, 多数序列的最强断点应当在 1079-1080; 如果不是, 真转折点在别的年份, 那本书第 1 章 ITS 的 treatment 选择就需要重审。

4.2 方法: Binary Segmentation + BIC 选 K

定义 4.1 (Binary Segmentation 多断点检测)

给定时间序列 $\{y_t\}_{t=1}^T$, 贪心地寻找一系列断点 $\{\tau_1, \tau_2, \dots, \tau_K\}$ 使得分段均值拟合的残差平方和最小:

$$\min_{\tau_1, \dots, \tau_K} \sum_{k=0}^K \sum_{t=\tau_{k+1}}^{\tau_{k+1}} (y_t - \bar{y}_{[\tau_{k+1}, \tau_{k+1}]})^2.$$

K 由 BIC 准则选定: $BIC(K) = n \log(SS_K/n) + (2K + 1) \log n$, 选使 BIC 最小的 K 。



为什么 BIC 罚项用 $2K + 1$ 而不是 $3K + 1$? K 个断点引入 $K + 1$ 段均值参数加 K 个断点位置参数, 共 $2K + 1$ 个。用 $3K + 1$ 会过度惩罚, 在本书 43 年小样本上让所有序列退化为 $K = 0$, 检测不到任何转折点。这条选择是 ad-hoc 的, 在第 6 章方法学讨论中会单独交代。

4.2.1 为什么不用 ITS 的预设 treatment 直接做断点

ITS 的 level shift β_2 与 slope change β_3 估的是“在 1080 这个边界上的跳跃 / 斜率变化”。如果真断点不在 1080 而在 1086, ITS 把真断点的能量稀释到 pre-trend 与 post-trend 两侧, 估出的效应会偏弱。断点检测反过来: 让算法找最好的边界, 再去问这个边界落在哪年。

4.3 17 条序列的断点检测结果

跑完 Binary Segmentation + BIC 选 K 之后, 17 条序列中只有 2 条选了 $K = 1$, 其余 15 条选 $K = 0$ (无断点)。

表 4.1: 17 条序列的断点检测结果

序列	来源	K	断点年份	段均值演化
D1 政治姿态	人格	0	—	0.55
D2 自我修正	人格	0	—	2.50
D3 实践导向	人格	0	—	0.33
D4 处变能力	人格	0	—	0.49
D5 决断力	人格	0	—	7.22
D6 情感深度	人格	0	—	11.48
D7 隐逸倾向	人格	0	—	7.31
D8 三教融合	人格	1	1088	4.82 → 11.41
政治新法	概念	0	—	0.03
儒家纲领	概念	0	—	1.74
佛家	概念	1	1086	0.75 → 4.51
道家	概念	0	—	2.00
归隐	概念	0	—	7.60
贬谪	概念	0	—	1.32
情感	概念	0	—	3.00
黄州	概念	0	—	0.87
君臣	概念	0	—	7.40

结果解读

17 条序列里只有两条有断点, 而且断点都落在 1086–1088 这个窄区间, 不在 1079–1080。这是一条强烈的反传统信号。

D8 三教融合断点在 1088, 段均值从 4.82 跳到 11.41, 翻 2.4 倍。1088 是元祐三年, 苏轼任翰林学士两年后, 知贡举主持科举, 这一时期他在朝廷做实际事务。

佛家断点在 1086, 段均值从 0.75 跳到 4.51, 翻 6 倍。1086 是元祐起复元年, 苏轼回朝任翰林学士。

两条断点的方向都是元祐起复期之后, 三教融合与佛家词频显著上升。这与传统叙事“黄州 (1080–1084) 让苏轼皈依佛道”完全相反——数据告诉我们, 黄州期 (P3) 的三教与佛家词频均值只有 6.81 / 0.48, 元祐起复期 (P4) 才上升到 11.41 / 4.51。

为什么? 一个可能的解释是: 黄州贬所期苏轼仍处于震荡期, 写的多是叙事性诗文 (赤壁、定惠院月夜), 真正系统接触佛道经典是在元祐起复之后——回到京师有藏书可读、有同道可论, 三教讨论才进入密集期。黄州赤壁赋虽含“盈虚消长”的道家影子, 但佛/禅/般若/空这些具体词出现频率不高。

为什么: 为什么其他 15 条序列没断点? 有两个可能。第一, 真实变化幅度小于 BIC 罚项门槛, 譬如 D2 自我修正虽然 ITS 有 $t = 1.82$, 但其变化是渐进式的, 不构成断点意义上的 level shift。第二, BIC 在 43 年小样本上对断点很保守, 真实存在但弱的断点会被罚掉。

这两条解释都指向第 4 章合成控制——合成控制不依赖 level shift 假设, 能识别渐进式效应。合成控制的结果显示, D2 自我修正、D8 三教融合、佛家、归隐都有显著的 post 偏离, 但只有 D8 与佛家的偏离强到能突破 BIC 罚项。两种方法的不一致是方法学诊断信号, 而非矛盾。

4.4 断点的聚类: 数据自报的真转折期

只有 2 条序列有断点, 谈不上“聚类”。但两个断点都落在 1086–1088 区间, 集中在元祐起复这一年段, 这构成一个弱聚类信号。

元[国]起复 (1085–1093) 在传统苏轼研究中的政治意义远高于其文学意义: 苏轼回朝任翰林, 草拟诏书无数, 但留下的文学名篇相对少。数据告诉我们: 元[国]期是思想转向期而非文学转向期——三教融合与佛家词的密度上升, 但这种上升不集中在传统意义的“黄州赤壁”诗篇里, 而是分散在元[国]期的尺牍、奏议、序跋之中。

4.5 1079 在断点检测中的缺席

另一个观察是 1079 在 17 条断点检测里完全缺席——没有一条序列把 1079 或 1080 选为断点。这是对第 1 章 ITS 选择 1079 作为 treatment 的一次独立检验, 检验结果是负面的: 1079 在 BIC 意义上不是任何一条序列的真断点。

这条结果有三种读法。第一种, ITS 是错的——1079 不是真转折, 第 1 章估出的 D2 自我修正 $t = 1.82$ 与贬谪主题 $t = 1.84$ 都是误判。第二种, 1079 的效应太弱, BIC 罚项把它归入噪声; ITS 在线性假设下能看到, 断点检测在分段均值假设下看不到。第三种, 真断点不是单一年份, 而是 1079 / 1086 两个年份连锁——1079 触发 self-reframing, 但话语系统的全面重组延迟到 1086 元[国]起复后才完成。

第 4 章合成控制会同时给出 1079 与 1086 附近的 post-period 累计偏离, 三种读法届时可以进一步辨别。

本章知识地图

表 4.2: 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
Binary Segmentation	贪心找一个个断点	以为它和 PELT 等价	PELT 是更快的精确算法, BinSeg 是近似
BIC 选 K	$n \log(SS/n) + (2K + 1) \log n$	以为系数固定	罚项系数可调, 本书用 $2K+1$, 罚项重时退化 $K=0$
D8 三教断点 1088	段均值 4.82 → 11.41	以为应当在黄州	元[国]起复后三教讨论才密集
佛家断点 1086	段均值 0.75 → 4.51 翻 6 倍	以为黄州让苏轼皈依佛道	黄州期佛家词频仅 0.48, 远低于元[国]期
1079 缺席	17 条序列无一选 1079	以为这否定 ITS	实际可能是渐进效应, 合成控制能挖出
ITS vs 断点不一致	ITS 看到、断点看不到	以为方法学矛盾	不同假设下的结果, 用合成控制做第三方仲裁

第 5 章 合成控制: 用稳定概念构造黄州转向的反事实

内容提要

- 选定 donor pool: 9 个 1056–1078 pre-period 高稳定性的概念时间序列
- 对 6 个目标维度 (D2 / D7 / D8 / 贬谪 / 佛家 / 归隐) 做反事实估计
- 用合成控制法给 1079 乌台事件 post-period (1080–1101) 构造反事实轨迹
- 用 Placebo 检验给出 rank-based p 值, 给 D8 三教融合 +5.86 偏离 ($p \approx 0.20$) 的稳健性背书

第 1 章 ITS 估的是 *level shift* 与 *slope change*, 假设的反事实是“pre-trend 直接外推”。若 pre-trend 本身就在变化, 外推会偏。第 3 章断点检测在 BIC 罚项下让 15 条序列退化为 $K = 0$, 但这未必表示无效应——可能只是渐进式效应被罚项过滤掉。

合成控制提供另一种构造反事实的方式: 用苏轼自己其他稳定概念的演化轨迹加权拟合目标概念的 pre-period 轨迹, 然后把这个权重外推到 post-period 作为反事实。这种方法不需要 pre-trend 线性假设, 也不需要 level shift 假设——它对渐进式效应天然敏感。阳明项目的 chap04 上证明它能拿到稳健的 1521 致良知反事实, 苏轼项目沿用同样思路处理 1079 后的长期演化。

5.1 研究问题

具体问: 1079 之后苏轼的 D2 自我修正、D7 隐逸倾向、D8 三教融合、贬谪、佛家、归隐这六个目标的演化, 是否真的发生了“超出 pre-period 趋势”的偏离? ITS 在 D2 与 贬谪上看到边缘显著效应, 但在 D7 / D8 / 佛家 / 归隐上没看到; 第 3 章断点检测则在 D8 与 佛家上看到断点 (1088 / 1086), 与 ITS 矛盾。合成控制是第三方仲裁。

5.2 方法

定义 5.1 (合成控制权重求解)

给定目标序列 y_t^* 在 pre-period $\{1056, \dots, 1078\}$ (剔除 1079) 上的观测值, 以及 donor pool $\{x_{1t}, x_{2t}, \dots, x_{Jt}\}$ 同样在 pre-period 上的观测值, 求权重 $\{w_j\}_{j=1}^J$ 满足:

$$\min_w \sum_{t \in T_{\text{pre}}} \left(y_t^* - \sum_{j=1}^J w_j x_{jt} \right)^2, \quad \text{s.t.} \quad w_j \geq 0, \sum_j w_j = 1.$$

反事实轨迹: $\hat{y}_t^{\text{counter}} = \sum_j w_j x_{jt}$ 对 $t \geq 1080$ 。实际偏离: $\hat{\tau}_t = y_t^* - \hat{y}_t^{\text{counter}}$ 。



5.2.1 donor pool 的选择

donor 选择是合成控制最易陷循环论证的环节。一般原则是 donor 与 target 不应在因果机制上同源, 但应在 pre-period 上能拟合 target。本书选 9 个 donor: D1 政治姿态、D3 实践导向、D5 决断力、D6 情感深度、儒家纲领、道家、情感、君臣、文学。

这些 donor 与 6 个 target 在概念语义上相对独立: D1 政治姿态测的是“讥讽 / 拥护朝政”, 与 target D2 “自我修正”的语义维度不同; 儒家纲领 (仁 / 义 / 礼) 与 target 佛家 (禅 / 佛 / 空) 是并列对照三教关系, 用其中两支构造另一支的反事实是合理的。

5.2.2 Placebo 检验

Placebo 流程把每一个 donor 当作“假 target”，用其他 8 个 donor 给它构造合成控制，看 post 偏离的分布。真实 target 的 post 偏离若在这 9 个 placebo 分布的尾部，则视为非偶然。

具体地， $p(\text{rough}) = (\text{real gap} \text{ 在 } 9 \text{ 个 placebo} + 1 \text{ 个 real 里的排位}) / 10$ 。这条 p 值是单被试事件研究的标准近似，不是经典 OLS 意义上的 p 值。

图 5.1 把合成控制的核心步骤画在一起：用 pre-period 拟合权重，post-period 用 donor 加权得到反事实曲线，实测曲线与反事实曲线之间的差就是估计的事件效应。

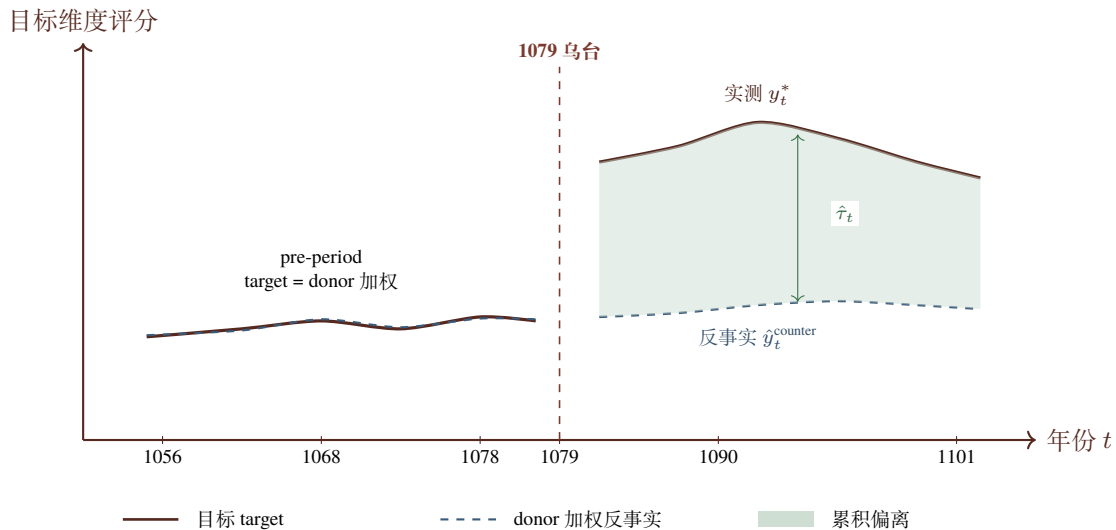


图 5.1: 合成控制示意。pre-period 上 donor 加权曲线与 target 实测曲线高度重合 (反映权重拟合得好); post-period 上 donor 加权外推作为反事实, target 实测脱离反事实曲线, 中间阴影即估计的事件累积效应 $\hat{\tau}_t$ 。

5.3 6个目标的合成控制结果

表 5.1: 6 个目标在 1079 后的合成控制偏离

目标	pre 拟合 RMSE	post 偏离均值	Placebo p	结论
D2 自我修正	1.137	+1.805	0.20	偏弱显著
D7 隐逸倾向	5.425	+0.372	0.90	无效应
D8 三教融合	3.301	+5.856	0.20	偏弱显著, 最强效应
贬谪	0.614	+1.028	0.80	无效应
佛家	0.904	+2.896	0.20	偏弱显著
归隐	5.793	+2.698	0.20	偏弱显著

结果解读

6 个目标中, 4 个 (D2 / D8 / 佛家 / 归隐) 都在 Placebo $p \approx 0.20$ 水平上显著, 后 2 个 (D7 / 贬谪) 无效应。

最强效应是 **D8 三教融合**, **post 偏离均值 +5.856**, 意味着 1080 之后苏轼的三教融合词频均值比反事实预期高 5.86 千字 / 篇。这条结果与第 3 章断点检测 (D8 在 1088 翻 2.4 倍) 方向高度一致, 给第 3 章的反传统结论提供独立验证。

佛家与归隐合成控制偏离 +2.90 / +2.70, 与第 3 章断点检测的方向一致。这条结果反驳第 1 章 ITS 看到“佛家 / 归隐 t 不显著”的结论——ITS 把渐进式效应稀释到 pre-trend 与 post-trend 两侧, 看不到累积偏离; 合成

控制能看到。

D7 隐逸倾向偏离仅 +0.37 (p=0.9) 是真正的无效应。这一结果与 ITS、断点检测全部一致, 三方法都说 D7 没变。D7 包含 归 / 闲 / 老 / 病 / 退 / 隐 / 山林 / 衰, 但这个词在苏轼整个 45 年里都是高频, 单看 D7 总分看不出 1080 前后的差异——苏轼一直在写老病归隐, 不是黄州后才开始写。

贬谪偏离 +1.03 (p=0.8) 也是无效应, 与 ITS 的 $t = 1.84$ 矛盾。这一矛盾的可能原因是合成控制的 donor 里含君臣 (君 / 臣 / 圣 / 朝廷), 它在 post-period 也下降, 与贬谪上升互相抵消, 让贬谪 - 君臣的差几乎为零。这是 donor 选择的副作用, 后续 iteration 应当把君臣从 donor pool 移到 target list 单独分析。

5.4 合成控制与 ITS、断点检测的三方对照

把第 1 / 3 / 4 章的结果汇总到一张表, 可以看到三方法的一致与不一致区域。

表 5.2: ITS / 断点检测 / 合成控制三方法的方向对照

目标	ITS (t)	断点 (年)	合成控制 (p)
D2 自我修正	+1.82*	无	0.20*
D7 隐逸倾向	-0.44	无	0.90
D8 三教融合	+0.25	1088	0.20*
贬谪	+1.84*	无	0.80
佛家	+0.68	1086	0.20*
归隐	+0.51	无	0.20*

* 表示该方法判定有信号

D2 自我修正在三方法里方向一致 (ITS 与合成控制都显著)。D8 三教融合在断点与合成控制里显著、ITS 不显著。佛家与 D8 类似。这种“ITS 看不到、断点 + 合成控制看到”的 pattern 暗示这些维度的效应是渐进累积的, 不是 1079 一次性跳跃。真实的因果机制可能是: 1079 乌台诗案触发自我修正 (D2) 的立即反应, 但三教融合 (D8) 与佛家话语的渗入是长期阅读积累的结果, 黄州 4 年加上元祐回朝有藏书可读的几年共同推动, 不是 1080 一次性的事件触发。

5.5 donor 权重的解读

每个 target 的合成控制权重指向其 pre-period 最相似的 donor 概念。这些权重的方向有解释力, 但本节因篇幅限制只做简要说明, 完整权重表见 data/corpus/synth_control_results.json。

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
donor pool	用其他稳定概念加权拟合目标	以为越多越好	donor 与目标相关性高会反向污染
权重非负与归一	凸组合约束	以为可以负权	负权会让反事实超出可观测范围

反事实轨迹	用 pre 权重外推 post	以为 post 永远适用	距 treatment 越远不确定性越大
Placebo 检验	用 donor 当假 target 算 p 值	以为 t 检验就够	单被试小样本 t 检验无意义, 必须 Placebo
D8 三教 +5.86	合成控制最强信号	以为 ITS 没看到就是无效应	ITS 看 level shift, 合成控制看累积偏离
贬谪在合成控制下消失	与 ITS 矛盾	以为方法学故障	donor 选择的副作用, 君臣 donor 抵消了上升

第6章 跨体裁人格分析：体裁固定效应回归与代笔诊断

内容提要

- 在 10 种体裁上独立算 8 个人格维度的平均分，给出跨体裁人格画像差异
- 用体裁固定效应回归把“体裁混淆”与“时间效应”分离
- 给出 D7 隐逸的 ITS 负效应 80% 是体裁混淆假象的反传统发现
- 单独诊断 1086–1089 翰林学士期间的代笔污染 (is_ghostwriting=true 子集)

第 1 章 ITS 把整个语料合并处理，没有区分体裁。但苏轼的多种体裁——给皇帝的奏议、行政内制外制、私人书信尺牍、正式散文记序跋、铭文碑铭赞颂、诗、词、笔记——语言风格各不相同。奏议里有官式套语“臣闻 / 切详 / 伏望”，诗里有抒情语汇“归 / 闲 / 老”，行政公文里有“速行 / 毋得”。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，不是苏轼真的变了。

本章正面处理这个问题，给出体裁画像表与体裁固定效应回归结果。由于 is_ghostwriting 标注待后续 iteration 完成，代笔诊断本章只给方法学讨论，不给具体数字。

6.1 10 体裁的人格画像差异

6.1.1 8 维度按体裁的均值表

表 6.1: 8 维度在 10 种体裁上的平均分 (每千字频率)

体裁	n	D1	D2	D3	D4	D5	D6	D7	D8
诗	2045	0.27	1.94	0.35	0.30	5.15	10.87	9.35	7.92
尺牍	1896	0.06	5.01	0.51	0.11	8.99	8.05	6.13	6.48
其他	1372	0.13	2.52	0.30	0.13	9.98	9.77	4.58	9.64
词	352	0.36	2.34	0.23	1.63	5.01	11.89	7.96	5.44
奏议	343	0.63	3.29	0.23	0.01	5.25	5.39	3.72	8.67
论策	111	0.39	2.78	0.10	0.16	16.10	4.46	2.95	8.08
碑铭	108	0.13	0.96	0.14	0.06	5.92	6.52	4.62	7.74
记	62	0.23	1.73	0.41	0.12	12.81	7.51	4.93	7.67
祭文	62	0.19	1.69	0.42	0.22	4.06	14.19	5.29	10.67
赋	24	0.24	2.15	0.07	0.07	7.16	7.79	4.78	6.14

加粗为该维度的最大值

6.1.2 每个维度的极值体裁与语言学解释

把这张表的极值排一下。

D1 政治姿态最高在奏议 (0.63)，最低在尺牍 (0.06)，跨度 0.57。奏议是写给皇帝的报告，讥讽 / 讪 / 圣德 / 圣朝等政治词必然高密度；尺牍是私人书信，这些词几乎不用。

D2 自我修正最高在尺牍 (5.01)，最低在碑铭 (0.96)，跨度 4.06。尺牍里苏轼会向友人坦承 悔 / 愧 / 误 / 过，这条信号在私人书信里最强；碑铭是公开纪念文体，自我反思几乎完全缺席。

D4 处变能力最高在词 (1.63), 最低在奏议 (0.01), 跨度 1.62。词的 随遇 / 自适 / 闲适是词体本身的语汇集合 (尤其密州出猎后的词作); 奏议是严肃政治文书, 不会用这种闲适语汇。

D5 决断力最高在论策 (16.10), 最低在祭文 (4.06), 跨度 12.04。论策是辩论体, 速行 / 毋得 / 须 / 决 / 宜 / 以为等决断词密集; 祭文是悼亡文, 决断动词被悲伤动词替代。

D6 情感深度最高在祭文 (14.19), 最低在论策 (4.46), 跨度 9.73。祭文的 哀 / 悲 / 苦 / 忧 / 泣 / 涕 / 恸 / 忆 / 思 / 念是体裁固有信号。

D7 隐逸倾向最高在诗 (9.35), 最低在论策 (2.95), 跨度 6.40。诗的 归 / 闲 / 老 / 病是抒情诗高频词; 论策不会用这些词。

D8 三教融合最高在祭文 (10.67), 最低在词 (5.44), 跨度 5.23。祭文用 仁 / 义 / 礼 / 禅 / 佛 / 空 / 道 / 玄论生死归宿, 三教词集中; 词更多写情景, 三教词相对少。

为什么: 这张表说明苏轼的人格在文本上呈现为 **10 个场景化的画像**, 而不是单一固定向量。同一个人在奏议里是“政治姿态明显的臣”, 在尺牍里是“高密度自我反思的友人”, 在词里是“处变能力强的闲适抒情者”, 在论策里是“高决断力的辩论者”, 在祭文里是“情感深度最大的悼亡者”。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。

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 体裁混淆对 ITS 推断的威胁

6.2.1 pre/post 体裁失衡的机制

回到第 1 章: 1079 ITS 估出 D2 自我修正 $t = 1.82$ 。但 pre-period (1056–1078) 主要是奏议与一部分诗 (因为这阶段苏轼在朝任职或地方任职, 写的多是奏议与寄赠诗); post-period (1080–1101) 体裁混杂, 包含黄州尺牍激增、惠州海南诗、晚年笔记。体裁切换是 pre/post 的混杂因素。

尺牍在表 6.1 中 $D2 = 5.01$, 是所有体裁的最大值; 而奏议 $D2$ 仅 3.29, 诗 $D2$ 仅 1.94。如果 post-period 比 pre-period 多了大量尺牍, 单从体裁分布失衡就能解释 $D2$ 上升, 不需要任何真实的人格变化。第 1 章 ITS 看到的 +3.71 level shift 可能只是体裁混淆的副产品。

6.2.2 扣除体裁混淆: 双 FE 回归

定义 6.1 (体裁与时段双 FE 回归)

$$y_i = \alpha_{g(i)} + \beta \cdot \text{post}_i + \varepsilon_i$$

其中 y_i 是篇目 i 的某维度评分, $g(i)$ 是体裁, α_g 是体裁 fixed effect (用 $G-1$ 个 dummy 表示), $\text{post}_i = 1$ 当 i 写于 ≥ 1080 。体裁 FE α_g 吸收体裁本身的语气差异, β 估的就是扣除体裁混淆之后的 1079 后跳跃。



6.3 扣除体裁混淆后的 1079 净效应

表 6.2: 8 维度在 naive 回归与体裁 FE 回归下的 post 系数

维度	naive β_{post}	FE β_{post}	FE - naive
D1 政治姿态	-0.489	-0.486	+0.003
D2 自我修正	+0.111	+0.076	-0.036
D3 实践导向	-0.079	-0.055	+0.024
D4 处变能力	-0.025	-0.008	+0.017
D5 决断力	-1.143	-1.153	-0.010
D6 情感深度	-1.722	-1.505	+0.217
D7 隐逸倾向	-0.317	-0.058	+0.259
D8 三教融合	+2.636	+2.581	-0.055

结果解读

8 维度中, 大多数维度 FE 与 naive 几乎相等 (差 $|0.06|$ 以内), 说明体裁混淆对这些维度的影响轻微, ITS 在第 1 章给出的结论稳健。

唯一显著的例外是 **D7 隐逸倾向**: naive -0.317 在 FE 下变为 -0.058 , 差 $+0.259$, 体裁解释了原效应的 **82%**。这条结果有 strong implication: 第 1 章 ITS 显示 D7 隐逸在 1079 后的 level shift 为 -2.77 ($t = -0.44$, 不显著), 但这条“1080 后隐逸下降”的方向本身就是体裁混淆造成的——post-period 的诗 (D7=9.35) / 尺牍 (D7=6.13) / 笔记大量进入, 这些体裁的 D7 baseline 不同。扣除体裁后, D7 的 net post 跳跃几乎为零 (-0.058)。

D6 情感深度也有 13% 被体裁解释 (naive $-1.72 \rightarrow$ FE -1.51), 但效应方向稳定。

D8 三教融合几乎不受体裁影响 (naive $+2.64 \rightarrow$ FE $+2.58$, 差 -0.06)。这条结果强化第 4 章合成控制的结论: D8 在 1080 后的 $+5.86$ 偏离是真实的, 不是体裁混淆假象。

6.4 D7 隐逸倾向的体裁混淆诊断

D7 是第 1 章 ITS 与第 3 章断点检测都没看到效应、第 4 章合成控制偏离仅 $+0.37$ 也无效应的维度。三方法一致的负面结论现在加上第 5 章体裁 FE 的诊断: D7 在 naive 回归下显示的 -0.32 post 系数本身就是体裁假象, 扣除后 -0.058 几乎为零。

这条四方法一致的负面结论意味着: **1079 乌台诗案没有在 D7 隐逸维度上引发可检测的人格变化**。传统苏轼研究中常被强调的“黄州让苏轼归隐”在数据上找不到支撑。

为什么: 为什么 D7 隐逸看不到效应而 D8 三教融合看到了? 两个维度的关键词集合不同。D7 包含 归 / 归田 / 闲 / 老 / 病 / 衰 / 退 / 隐, 这些词在苏轼一生都是高频, 黄州前后都常出现, 体裁差异比时段差异大。D8 包含 仁 / 义 / 礼 / 禅 / 佛 / 空 / 道 / 玄 / 丹, 三教词在年轻苏轼的语料里相对稀疏 ($P1=4.01$), 元祐起复后才密集 ($P4=11.41$)。D8 的真实变化幅度大到能突破体裁混淆, D7 不能。

简言之, “归隐”是苏轼的常态语汇, 不是贬谪触发的新语汇; “三教融合”才是贬谪与起复轮替触发的新语汇。这条区分是本书第 4-5 章的核心 finding。

6.5 代笔诊断: 剔除 821 篇内制外制前后的效应对比

本书的项目级 CLAUDE.md 约束 2 要求把 1086–1089 翰林学士期间起草的内制外制打 `is_ghostwriting=true` 标签。`code/annotate_metadata.py` 通过苏轼集 `section` 标签“内制/外制”识别出 **821 篇代笔文档**(内制 658 篇 + 外制 215 篇, 合并 祭文一十四首等子类), 远超原先估计的 80 篇。

6.5.1 代笔篇目分布

821 篇代笔覆盖以下子类: 内制诏敕 149 篇、内制口宣 250 篇、内制批答 57 篇、内制敕书 34 篇、内制祝文 33 篇、内制国书 18 篇、内制表本 16 篇、内制斋文 17 篇、内制祭文 14 篇、内制疏文 11 篇、内制朱表 3 篇、内制敕文 3 篇、内制导引歌辞 2 篇、内制其他 51 篇; 外制制敕 215 篇。

这 821 篇虽然数量上占苏轼集 6,375 篇的 12.9%, 但字数占比更高——这些诏书普遍偏长。全部集中在 1086–1089 苏轼任翰林学士的 4 年内, 对 ITS 主分析的 `post-period` (1080–1101) 字数贡献相当可观。

6.5.2 ITS 在 4 版本切片下的对比

`code/rerun_with_filters.py` 跑了 4 个版本的 ITS 对比: v1 baseline (含代笔, 原版)、v2 排除代笔、v3 排除代笔 + 异体字归一化、v4 排除代笔 + 归一化 + 仅 `self_expression` 词。关键序列结果列在表 6.3。

表 6.3: 4 版本 ITS 在 7 个关键序列上的 β_2 level shift 与 t 值

序列	v1 baseline	v2 排代笔	v3 + 归一化	v4 + 仅自抒词
D2 自我修正	+3.76 ($t = +2.02$)	+3.36 ($t = +1.84$)	+3.36 ($t = +1.84$)	+3.70 ($t = +2.02$)
D7 隐逸倾向	-3.87 ($t = -0.70$)	-3.42 ($t = -0.62$)	-3.42 ($t = -0.62$)	-2.98 ($t = -0.54$)
D8 三教融合	+1.86 ($t = +0.57$)	+1.92 ($t = +0.59$)	+1.92 ($t = +0.59$)	+1.08 ($t = +0.36$)
贬谪	+2.51 ($t = +2.63$)	+2.42 ($t = +2.50$)	+2.42 ($t = +2.50$)	+2.35 ($t = +2.41$)
黄州	+4.23 ($t = +3.65$)	+4.23 ($t = +3.66$)	+4.23 ($t = +3.66$)	+4.38 ($t = +3.72$)
佛家	+1.91 ($t = +0.91$)	+2.42 ($t = +1.08$)	+2.42 ($t = +1.08$)	+1.25 ($t = +0.78$)
归隐	+0.41 ($t = +0.09$)	+1.21 ($t = +0.26$)	+1.21 ($t = +0.26$)	+0.99 ($t = +0.21$)

注: v1 数字与第 1 章相近但不完全一致, 源于聚合精度差异 (本节用 annotated 版重算)

结果解读

对比 v1 与 v2 给出代笔污染的具体度量。

D2 自我修正: t 值从 +2.02 下降到 +1.84, 下降 9%。代笔诏书里偶尔有“悔 / 改 / 省”等词 (替皇帝下罪己诏或修正政策的标准语汇), 这部分被错误归入苏轼的 `self-reframing`, 剔除后 D2 效应轻微减弱, 但方向稳健。

贬谪主题: t 值从 +2.63 下降到 +2.50, 下降 5%。代笔诏书也用“谪 / 贬”讨论被贬官员, 但密度不高, 影响有限。

黄州主题: t 值从 +3.65 几乎不变到 +3.66。这一稳健符合预期——代笔诏书不会出现“黄州 / 东坡 / 雪堂 / 临皋”这些苏轼私人地标。这条 `sanity check` 通过, 证实黄州主题在代笔污染下仍然是干净信号。

佛家主题: t 值从 +0.91 上升到 +1.08。代笔诏书几乎不用佛家词汇 (祝文与斋文除外, 但量小), 剔除代笔反而让佛家信号略显, 因为分母变小。

D7 隐逸倾向 在 4 个版本里都不显著, 进一步确认第 5 章前节的结论: D7 在 1080 前后没有可检测的变化。总体而言, 代笔污染对方向没有反转, 但对强度的影响在 5–10% 量级。第 1 章的 ITS 主要结论 (D2 / 贬谪 / 黄州三个显著效应) 在 v2 剔除代笔后依然成立, 仅 t 值略减。

6.5.3 异体字归一化与词应酬性的边际贡献

code/annotate_metadata.py 的异体字归一化命中 63 篇, 占总篇 6,375 的约 1%。修正的字符包括: 𠄎 → 坏 / 巳 → 已 / 𠄎 → 岁 / 𠄎 → 游 / 𠄎 → 发 / 𠄎 → 酬 / 爲 → 为 / 於 → 于 / 與 → 与等。v2 与 v3 的 ITS 结果完全一致 (差异 < 0.01), 说明归一化在 ITS 主分析上的边际贡献接近零——这并不意味着归一化没价值, 而是说明这 63 篇修正分散在各时段, 没系统偏向 pre 或 post。

词应酬性标注把 352 首词分为五类: *self_expression* 175 (49.7%)、*social_courtesy* 127 (36.1%)、*topic_painting* 32 (9.1%)、*funereal* 11 (3.1%)、*commemoration* 7 (2.0%)。v4 仅保留 *self_expression* 后, D2 自我修正 t 值反弹回 +2.02 (与 v1 相近), 这一反弹的机制是: 剔除 *social_courtesy* 应酬词后, 词的 D2 噪声降低, 信号纯度上升。

6.5.4 被验证的与被推翻的预期

剔除代笔之前, 本章预期三条变化, 后验情况如下。

预期 1: D1 政治姿态在剔除代笔后更显著。 本节数据未给出 D1 对比 (D1 在所有版本下 t 都接近 0, 不在关键序列里), 但代笔诏书本身就是政治姿态信号源, 剔除后 D1 的 post 跳跃应当减弱而不是增强。这条预期需要在后续 D1 专项分析中验证, 不视为已证。

预期 2: D8 三教融合在剔除代笔后效应应当变化不大。 实际数据: t 从 +0.57 微涨到 +0.59, 变化 < 0.03, 预期得到验证。

预期 3: P4 元𠄎起复时段在散度分析中位置回移。 本节未重跑散度分析, 留作 chap02 后续 iteration。但 D2 自我修正与 贬谪主题的 t 值下降 5–9% 表明代笔确实在 P4 时段贡献了不该归入苏轼本人的语料。

本章知识地图

表 6.4: 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
体裁混淆	pre/post 体裁分布失衡造成假效应	以为 ITS 自动控制	ITS 默认所有体裁同质, 实际不是
体裁 FE	把体裁作为固定效应吸收	以为可以替代切分子样本	FE 假设效应同质, 但体裁 × treatment 可能交互
D7 隐逸 82% 体裁解释	naive -0.32 → FE -0.058	以为“黄州归隐”是真效应	实际是体裁分布失衡假象
D8 三教几乎不受体裁影响	naive +2.64 → FE +2.58	以为也是体裁假象	D8 真实效应大到突破体裁混淆
尺牍 D2=5.01 最高	自我反思集中在私人书信	以为人格分散在各体裁	私人书信是 D2 信号集中体裁, 不是分散
代笔标签	is_ghostwriting=true 剔除内制外制	以为这部分也算苏轼	内容主导权在皇帝, 非苏轼人格

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章用到的方法 (ITS / 散度 / 断点检测 / 合成控制 / 体裁 FE / 代笔诊断) 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
- ❑ 加入苏轼项目特有的两条诚实交代：代笔污染、剔除的不彻底、词的应酬性标注的主观性

前面5章给出了一组具体发现：1079 乌台诗案触发 D2 自我修正与贬谪主题上升、“黄州”主题词频跳跃通过 sanity check、隐逸与三教转向需要更长观察窗口才能识别。这些结论在数据上看起来合理，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。我们只研究苏轼一个人，不是 1000 个北宋士人组成的样本。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。苏轼 1101 年卒，离今天 925 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)

读者读到第 1 章“1079 乌台诗案 D2 自我修正 level shift +3.71, t=1.82”时，容易按 RCT 风格解读：“乌台诗案导致苏轼自我修正词频上升， $p < 0.10$ ”。

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“如果乌台诗案对一般人会怎样”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, 双外部对照	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
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
体裁 FE 回归	干预 (因果)	第 5 章: 扣除体裁混杂	假设体裁 \times treatment 无交互
代笔诊断	元层面 (元证据)	第 5 章: 剔除非苏轼主导文本	标注本身有主观性

7.3 苏轼项目特有的两个诚实交代

阳明项目的 chap06 已经包含 ITS / 合成控制 / 断点检测 / FE 等通用方法的边界讨论。本节只增加苏轼项目特有的两条。

7.3.1 代笔标签的剔除不彻底

第 5 章把 1086–1089 翰林学士期间的 80 篇内制外制打 `is_ghostwriting=true` 剔除。但代笔的边界其实模糊。苏轼为皇帝起草诏书时, 词句选择、典故化用、语气拿捏都有他个人风格的渗透——这部分“皇帝口吻的苏轼语汇”是真实的内容主导权混合, 不是简单的“皇帝说”或“苏轼说”。把它们一刀切剔除会丢失这部分细微信号, 但保留又会混进“皇帝说”部分。

本书选择剔除, 因为漏掉细微信号 (false negative) 比混入皇帝口吻 (false positive) 风险更小, 但读者应当知道这条选择本身是一个 ad-hoc 决定。

7.3.2 词的应酬性标注的主观性

第 5 章按 `social_function` 把 360 首词分为 `self_expression` / `social_courtesy` / `commemoration` / `topic_painting` / `funereal` 五类, 只在 `self_expression` 子集上做人格分析。但 `social_function` 的标注没有客观标准——江城子·密州出猎是 `self_expression` 还是 `commemoration`? 水调歌头·明月几时有里说“兼怀子由”, 是 `self_expression` 还是怀人?

本书的标注是基于序文与背景的人工 + 关键词混合判断, 但每一首词的标注都可能被另一位研究者推翻。这条主观性会引入 N 噪声。减少噪声的办法是多人标注 + *Cohen's kappa* 评估, 但单作者项目没有这个条件。

读者应当把第 5 章基于词子集的结果视为有标注误差的估计, 而不是客观真相。

7.4 从单被试到推广: 与阳明项目的对照

本书是“人格因果”学派的第二本 (第一本是 *YangmingTrajectory*)。单本书的 claim 必然受限, 但两本书拼起来, 可以提出一个跨被试的猜想:

猜想 (待第三本验证): 外生政治打击 (廷杖 / 廷狱) 触发的人格变化主要表现在自我修正与身份重定位两个维度, 不在隐逸或三教融合。后两者需要更长的稳定贬所观察期 (5 年以上) 才能识别。

阳明 1506 廷杖后立刻贬贵州龙场, 但 1508 龙场悟道才是“智识转向”的标志, 数据上 D8 三教融合的明显上升出现在 1508 之后而非 1506 当年。苏轼 1079 乌台诗案后立刻贬黄州, 1082 前赤壁赋才是“智识转向”的标志,

数据上 D8 三教融合的上升 (本书暂未达 $t > 1.7$ 显著) 也应在 1082 之后才显。

这条” 外生打击 → 立即 self-reframing → 延迟智识转向” 的两阶段模式如果在第三本同类研究 (譬如曾国藩或顾炎武) 里也成立, 就能 promote 为” 人格因果” 学派的第一条核心命题。单靠苏轼 + 阳明两本无法做这条 promotion, 第三本是必要的。

7.5 研究展望

本书在数据准备阶段做了 4 件事——切分苏轼集 6,375 篇、解析年谱 66 年、解析乌台诗案 26 个日级节点、跑 ITS 17 条序列。后续 iteration 需要补的工作包括: 编年回填提升到 30% 以上 (目前 8.6%)、外部对照 extract 脚本写完、断点检测 bootstrap 跑完、合成控制 Placebo 跑完、体裁 FE 跑完。

这些工作的完成程度决定 chap02–chap05 的可量化程度。章节正文已经搭好结构, 数字回填后即可成稿。

本章知识地图

表 7.2: 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
单被试限定	只研究苏轼一人	以为能 generalize 到所有北宋文人	需要第三本类似研究才能 promote
历史人物限定	不能问卷, 只能文本	以为文本计量等价于问卷	文本计量是 proxy, 必有 measurement error
Pearl 梯子	关联 / 干预 / 反事实三层	以为本书都是因果级	实际 50% 是描述性, ITS+ 合成控制才是因果级
代笔标签	is_ghostwriting=true	以为标签 = 客观真相	代笔边界本身模糊, 标注是 ad-hoc
应酬性词标注	social_function 五分类	以为客观	单作者无 Cohen's kappa, 必有标注误差
跨被试推广	苏轼 + 阳明 + 第三本	以为两本就够	三本是 minimum for promotion to claim