

计量中的事后检验：稳健性、安慰剂、异质性

Zircon

2026 年 5 月 26 日

目录

| | |
|------------------------------|----------|
| 1 安慰剂检验 | 2 |
| 1.1 什么是安慰剂检验 | 2 |
| 1.2 为什么要进行安慰剂检验 | 2 |
| 1.3 常见的安慰剂检验方法 | 2 |
| 1.3.1 改变政策发生时间 | 2 |
| 1.3.2 随机生成实验组 | 3 |
| 1.3.3 替换样本安慰剂检验 | 3 |
| 1.3.4 替换变量安慰剂检验 | 4 |
| 1.4 因果推断方法对应的安慰剂检验 | 4 |
| 2 异质性分析 | 4 |
| 2.1 为什么要做异质性分析 | 5 |
| 2.2 什么是好的异质性分析 | 5 |
| 2.2.1 强化因果关系 | 5 |
| 2.2.2 反弹琵琶 | 6 |
| 2.2.3 呼应经典文献 | 6 |
| 2.2.4 具有政策含义 | 7 |
| 2.2.5 紧扣核心变量 | 7 |
| 2.3 总结 | 8 |

| | |
|----------------|----------|
| 目录 | 2 |
| 3 稳健性检验 | 8 |
| 3.1 什么是稳健性检验 | 8 |
| 3.2 为什么要做稳健性检验 | 8 |
| 3.3 稳健性检验的方法 | 9 |
| 3.3.1 变量替换法 | 9 |
| 3.3.2 补充变量法 | 10 |
| 3.3.3 分样本回归法 | 11 |
| 3.3.4 调整样本期 | 11 |
| 3.3.5 滚动窗口法 | 12 |
| 3.3.6 改变样本容量法 | 12 |
| 3.3.7 内生性问题 | 12 |
| 3.3.8 其它方法 | 13 |
| 3.4 总结 | 14 |

1 安慰剂检验

1.1 什么是安慰剂检验

安慰剂效应 (placebo effect), 又名伪药效应、假药效应、代设剂效应, 是指病人虽然获得无效的治疗, 但却让其“预料”或“相信”治疗有效, 而让病患症状得到舒缓的现象。

随着因果推断方法在实证研究中的使用比例不断提升, 越来越多的文章也会进行安慰剂检验。其检验基本原理与医学中的安慰剂类似, 即使用假的政策发生时间或实验组进行分析, 以检验能否得到政策效应。如果依然得到了政策效应, 则表明基准回归中的政策效应并不可靠。进一步, 经济结果可能是由其他不可观测因素导致的, 而非关注的政策所产生。

1.2 为什么要进行安慰剂检验

说到底就是两件事。一件事是让审稿人相信你的因果识别是干净的——“基准回归出了显著系数, 你也得能回答”如果换一个伪政策、换一组对照、换一段样本期, 这个显著性会不会照样跑出来?”;跑出来了, 原本的结论就不再可信。另一件事是把故事讲圆——好的实证文章在主结果之外要有“反例”作支撑, 让读者在两边比较中相信你的识别。

无论用的是 DID、RDD 还是 IV, 安慰剂检验的思想可以统一概括为一句话:

核心思想

构造一个理论上不应该出现政策效应的“伪政策”, 看你的回归结果是否仍然显著。如果仍然显著, 就说明基准结果可能来自数据本身的某种偶然性, 而非真正的政策因果。

但有一处常见的陷阱要避免——

避免“为安慰剂而安慰剂”

DID 之后随机生成 10000 次实验组、画一张系数分布图, 看上去“完美”;但如果不能说清楚我担心的是哪一种偶然性, 所以用这个方法去排除它, 这张图就只是流程化的形式。方法可以照搬, 理由必须自己写。

1.3 常见的安慰剂检验方法

1.3.1 改变政策发生时间

通过将政策发生时间前置以进行安慰剂检验的方法在双重差分中十分常见。实际上, 这与我们经常看到的平行趋势检验是同一种方法。

前置政策发生时间在实际操作中也并不困难, 设定一个虚拟的政策发生年份代替真实的政策发生年份, 之后纳入回归即可。

林毅夫等 (2020) 使用双重差分法、三重差分法和合成控制法来识别消费券的发放效果, 并评估政府在助力经济复苏中的作用。但潜在的威胁是, 发券城市的市民本身就有更强的消费倾向。为此, 作

者进行了安慰剂检验，分别假设消费券发放的时间提前 30 天或者 15 天，然后考察这些虚拟的消费券发放是否会影响支付笔数。

关于采用改变政策发生时间进行安慰剂检验的方法，还可参考以下文献：

- 吕越, 陆毅, 吴嵩博, 王勇. “一带一路”倡议的对外投资促进效应——基于 2005—2016 年中国企业绿地投资的双重差分检验 [J]. 经济研究, 2019, 54(09):187-202.
- 徐现祥, 李书娟. 官员偏爱籍贯地的机制研究——基于资源转移的视角 [J]. 经济研究, 2019, 54(07):111-126.

1.3.2 随机生成实验组

另一种常见的安慰剂检验的方式是**随机生成实验组**。

刘瑞明等 (2020) 根据中国文化体制改革的“准自然实验”，采用中国 283 个地级市 2002~2016 年间的的面板数据，评估了文化体制改革对地区旅游业发展的影响。对于文章的结论而言，一个可能的质疑是，人均国内旅游人次、人均国内旅游收入、人均国内外旅游总人次和人均国内外旅游总收入四个指标的统计显著可能来自于某些随机因素。为此，作者通过随机生成实验组的方式进行安慰剂检验，以判断文化体制改革的旅游促进效应是否是由其他随机性因素引起的。利用这一方法进行安慰剂检验主要目的是，**排除由其他随机因素造成的经济后果**，以得到更加可信的因果识别效应。通过随机抽取实验组，重复多次，提取安慰剂结果系数或值，然后将其绘制在图中，并观察真实的政策效应与安慰剂结果。当真实的政策效应与安慰剂检验结果显著不同时，可排除其他随机因素对结果的干扰。

1.3.3 替换样本安慰剂检验

替换样本进行安慰剂检验与随机生成实验组的方法较为相似。不同之处在于，随机生成实验组的安慰剂检验方法最终结果以图形展示，而替换样本安慰剂检验结果多以表格形式展示。在实际操作过程中，替换样本安慰剂检验不需要重复模拟，这在技术上显得容易一点，但在理论逻辑上更加严谨。比如，某政策颁布后，受政策影响的是污染行业，在因果识别后，可对非污染行业进行分析，探究是否存在政策效应（亦或对政策范围外的污染行业进行分析）。如果对于非污染行业依然存在所谓的政策效应，那么前文的分析并不可靠。例如：

张琦等 (2019) 以《环境空气质量标准 (2012)》的实施为准自然实验，采用双重差分法检验了新标准实施引致的官员动机变化对企业环保决策的影响。文章以 74 个试点城市以外的其他城市中重污染企业作为安慰剂样本，进行了证伪检验。

汪德华等 (2019) 基于 2013 年 CHIP 数据，运用截面数据双差法，评估了 20 世纪 90 年代中期二片地区“国家贫困地区义务教育工程”的政策效果。并以不受政策影响的样本作为分析对象，进行了安慰剂检验。

1.3.4 替换变量安慰剂检验

替换变量进行安慰剂检验主要分为**替换被解释变量**和**替换解释变量**。与稳健性检验有所不同的是，稳健性检验希望在替换变量后结果依然稳健，而**安慰剂检验希望替换变量后结果不再显著**。

首先，**替换被解释变量**。某项政策实施后，对特定经济活动会产生影响，但并不是对所有的经济活动都会产生影响。因此，将被解释变量替换为预期不会受到政策影响的变量进行安慰剂检验，以排除其他可能的干扰因素。例如陈林和万攀兵 (2019) 以双重差分法实证检验了《京都议定书》的政策效果。由于《京都议定书》未将 PM2.5 纳入减排考核目标，这诱使 CDM 项目实施方对 PM2.5 的防治有所忽略。可以预期，CDM 项目的实施并不会对以 PM2.5 为代表的常规空气污染物产生显著的减排效果。因此，以 PM2.5 为被解释变量进行安慰剂检验。

其次，**替换解释变量**。这一方法似乎没那么常用，或者说与前文的随机生成实验组和替换样本安慰剂检验有些类似之处。做法主要是将解释变量替换为看起来相似，但经济意义不同的变量。例如梁斌和冀慧 (2020) 在研究失业保险如何影响求职努力时，使用“有失业保险”的虚拟变量作为解释变量进行安慰剂检验。这主要是因为“有失业保险”不等于“领取失业保险金”。

1.4 因果推断方法对应的安慰剂检验

- 工具变量：**替换样本**；
- 双重差分：**改变政策发生时间与随机生成实验组**；
- 断点回归：**改变断点和查看控制变量是否有跳跃**；
- 合成控制：**将控制组看做实验组进行分析**。

2 异质性分析

有的老师曾经调侃道：

当代研究生写作有“三大法宝”：用 PSM 来解决内生性，用三步法来检验机制，**划分东中西部来做异质性分析**。虽然话语中饱含戏谑，却也说出了一部分实情，各类模板化的异质性分析在实证论文中确实有滥竽充数之势。

多位在学界颇有建树的老师在讲座中也提到了相似的观点：

异质性分析不是用来凑篇幅的，应该**做有意义的异质性**。

但是，对于如何做有意义的异质性分析，老师们却并没有细讲。因此，本文基于近年来发表在经济学顶刊上的部分做法，试图做一个粗浅的总结。当然，囿于自身水平，这种总结肯定是不全面的，还请读者见谅。

2.1 为什么要做异质性分析

首先，我们为什么要做异质性呢？这主要源自两方面的原因：

一方面是因为辛普森悖论。我们以下图为例，横轴是服药剂量，纵轴是病人健康水平，蓝色散点表示年轻人，红色散点表示老年人。

可以发现，无论是对年轻人还是老年人来说，服药都能缓解病情，优化健康。但是，当我们把两类人放到一起时，却发现一个很荒谬的结论：药吃的越多，健康水平越差。这种组内趋势与整体趋势相异甚至相反的现象，就是著名的辛普森悖论。

这种现象在经济学里面也并不鲜见。以时隔五年发表在《金融研究》上的两篇文章为例，潘越等 (2017) 研究发现，方言多样性越高的地区，上市公司的创新产出越高。但是，张杰和王文凯 (2022) 却发现，方言多样性会降低企业的创新投入。两篇文章的解释变量都是使用徐现祥老师的数据，而且工具变量都是地形坡度。两篇文章最大的差异就是前者使用的是上市公司，而后者使用的是工企库。相比于上市公司，工企库内的企业规模明显要小很多。样本选择范围的不同可能是两篇文章结论迥异的重要原因。

另一方面是因为没有普适的政策。以陆铭老师的工作论文为例，他们研究了中国 2004 年大规模关闭开发区对企业生产率的影响，发现这一政策使得沿海企业的 TFP 降低了 9.62%，但是对其他区域的企业影响却不显著 (Chen 等 2019, NBER)。这说明开发区政策仅对沿海企业有效，对内陆企业的影响有限。同样地，邵帅等 (2019, 经济研究) 在研究中国的城市化推进与雾霾污染时，发现城市化进程加剧了雾霾污染。但是，异质性分析则表明，紧凑集约型的城市化模式对雾霾污染有显著的促降作用，而规模扩张型的城市化模式则不利于抑制雾霾污染。

综合上述文章可以发现，异质性分析有时候可能会得出不同、甚至是相反的结果，这也是异质性分析有趣的地方，也是其价值所在。

2.2 什么是好的异质性分析

江艇 (2022, 中国工业经济) 指出：一篇因果推断经验研究文章的重点永远是正确识别处理变量对结果变量的因果关系。因此，文章中的每一字每一句都应该为这一目标服务。异质性分析也不例外。所谓“兵无常势，水无常形，能因敌变化而取胜者，谓之神”（《孙子兵法·虚实篇》），异质性分析也一样，并没有一定的章法，重要的是与论文本身相契合。

2.2.1 强化因果关系

例如，Muralidharan 和 Prakash (2017, AEJ) 评估了印度的免费自行车政策对女生入学率的影响，发现该政策可以有效增加女性的入学率，其异质性分析部分则进一步分析了离学校距离对政策效应的影响。自行车作为一种交通工具，有其局限性：短途完全可以走路，自行车节约不了多少时间；长途则只能搭车，自行车意义不大。因此我们可以预期，距离与政策效应之间应该更可能呈倒 U 型分布。实证结果也如作者所料，自行车在家离学校 5-13 公里时作用最大。这一异质性结果无疑进一步强化了作者所讲的故事。

2.2.2 反弹琵琶

另外一种强化因果关系的方式是“反弹琵琶”，即提出一个竞争性假说，再放入模型中予以验证。若竞争性假说不能（完全）解释既定的因果关系，则说明基准结果很可能是对的。这一过程被形象的称之为“赛马”（Horse Race）。

例如，Brown (2011, JPE) 试图验证，在锦标赛机制中，超级明星会降低其他选手的努力程度（因为有超级明星后，其他选手可能沦为陪跑）。但这实证时，却存在一种竞争性的假说：超级明星参加的可能都是比较难的赛事，其他选手发挥更差可以由赛事难度解释，而非超级明星效应。作者的解决方法是将选手们按排名分为高、中、低三档。如果选手的水平下降可以由赛事难度解释的话，那么可以预期，低水平选手会受到更大的影响，即系数更大。但回归结果却刚好相反，高水平选手受到的影响更大且在统计上显著，低水平选手影响小且统计上不显著。实证结果与竞争性假说完全相反，从而进一步强化了既定的故事。

相关的做法还有：朱孟楠和徐云娇 (2022, 金融研究), Coviello 等 (2022, JPE), Mirenda 等 (2022, AER) 等。

2.2.3 呼应经典文献

首先，异质性分析最重要的莫过于有**经典文献支撑**，现有文献也多以此展开分析。例如劳动经济学中的性别差异、行业差异与城乡差异等 (刘子兰等 2019, 经济研究; Xu, 2019, JPubE); 公司金融中的企业性质、企业规模等 (张璇等, 2017; Mao 等, 2022, USUE); 城市经济学中的城市规模、城市行政级别等 (陈诗一和陈登科, 2018; Chen 等, 2021, JEBO)。

由于男女两性在生理结构、认知水平等方面的广泛差异，分性别讨论不仅可以**男女平等**，还有明确的**政策含义**，这成了很多劳经文献的标准动作。例如，Chen 等 (2020, AER) 在分析 Send-down movement 对地区教育的影响时，性别分组发现这一提升效应对女学生影响更大，说明该政策不仅提高了农村地区的教育水平，还缩小了性别教育差距。男女生之间的差异还体现在**行为模式**上。Cai 等 (2019, RSTtat) 在研究压力环境下的个体反应时，发现男性有着更好的抗压能力，在压力环境下表现更好，有时候甚至还能逆风翻盘，而女性则更可能发挥失常。Egebark 等 (2021, JPubE) 在分析相亲市场上的互动行为时，也发现了两性之间的差异：男性更有可能对学历较低的相亲女性做出回应，而受过大学教育的女性则会拒绝比自己学历低的男性。

在公司层面的研究中，国有企业与非国有企业的差异也是一个老生常谈的话题。相比于非国有企业，**国企具有鲜明的行政色彩**，利润最大化未必是其核心追求。Knutsson 和 Tyrefors (2022, QJE) 在研究救护车服务的公、私营差异时，就发现私营企业会绕开劳动法，让员工加更多的班，而且其救护人员的专业水平也低于公共部门。此外，国企与行政机关联系更为紧密，议价能力更强，**很多激励措施对国企未必有效**。例如，Hering 和 Poncet (2014, JEEM) 在研究中国的环境规制政策（两控区）对企业出口的影响时，就发现规制效应仅存在于民营企业，对国企的影响不显著。He 等 (2020, QJE) 在研究水污染规制对企业全要素生产率的影响时，发现规制显著降低了企业生产率，但这种效应主要是由内资私有企业驱动的，对国企和外企影响有限。Mao 等 (2022, USUE) 也发现了相似的结果，环境规制对国企的能源消费（煤炭）无显著影响。

相似的研究还有：刘行和赵晓阳 (2019, 经济研究), 田彬彬和范子英 (2018, 经济研究), 龙玉等 (2017, 经济研究), Helm (2020, RES) 等。

2.2.4 具有政策含义

如果说, 在国外的情境下, 能够对理论做出回应才是一篇好的论文, 那么在国内的情境下, 这一标准就变成了**能够对国家政策做出回应**。中文论文的选题有着很明显的政策导向, 有些期刊甚至明确要求有一定篇幅的政策建议。因此, 如果能基于异质性分析引出针对性的政策含义, 无疑能为你的文章增光添彩。

地区禀赋是影响政策成效的重要原因, 以此为支点通常能延伸出针对性的政策建议。例如, 林毅夫等 (2020, 经济研究) 在研究中国政府消费券政策的经济效应时, 发现在转移支付比例高 (财政实力弱) 的城市, 消费券对消费的刺激作用更弱。因而建议对于自身财政实力不足的地区, 应该允许其使用上级的转移支付, 或增加财政赤字来支持消费券发放。

个体/家庭禀赋的差异, 也会影响干预的成效。Borowiecki (2022, JPE) 在研究作曲家师承对作曲风格的影响时, 发现了一个很扎心的结论。如果作曲家的老师水平越高, 学生与老师的风格越相似, 其成为行业顶尖的可能性也越大。而如果作曲家的老师水平越次, 学生与老师的风格越像, 反倒会成为职业生涯的拖累。这提醒我们, 应该向高水平学者学习, 见贤思齐焉。

相似文献还有: 叶祥松和刘敬 (2020, 经济研究), 周茂等 (2018, 中国工业经济), 周京奎等 (2019, 经济研究), Demirci 等 (2019, JFE) 等

2.2.5 紧扣核心变量

还有一种比较经典的做法是紧扣文章的被解释变量或解释变量, 进行异质性分析, 这样做的好处是能够进一步深化文章的主题。较多的文献是从**解释变量**出发。例如 Borowiecki (2017, RES) 在研究情绪对作曲家创造力的影响时, 发现主要是负面情绪激发了作曲家的创造力。在使用文本衡量各类情感时, 划分积极或消极情感, 也是各类研究的标准动作了 (Vosoughi 等, 2018, Science)。Helm (2020, RES) 在分析贸易冲击对集聚经济溢出效应的影响时, 将贸易冲击按照产业类型划分为来自高科技产业的冲击和来自低科技产业的冲击, 结果发现集聚溢出主要是由高科技行业的贸易冲击引起的。

也有部分文献从**被解释变量**出发。例如, Gebresilasse (2023, JDE) 在农村道路建设和农业推广对作物生产率的影响时, 异质性分析中将作物划分为谷物与非谷物, 发现政策主要对谷物类的作用有积极影响。刘瑞明等 (2020, 经济研究) 在研究文化体制改革对地区旅游经济发展时, 即基于地区的旅游资源禀赋和旅游公共服务水平进行异质性分析, 发现地区旅游公共服务水平越好, 文化体制改革对旅游业的发展作用越强。

相似做法的文章还有: 臧文斌等 (2020, 经济研究), 宋弘和陆毅 (2020, 经济研究), Chen 等 (2022, EJ) 等。

2.3 总结

综合上文可以发现，异质性分析并没有定式，贵在**言之有物**。而且异质性分析也并非文章所必须。目前也有越来越多的文章不做异质性分析，转为将精力放到更具深度的**进一步分析**，试图拓展文章的纵深，挖掘出更有价值的信息。

3 稳健性检验

3.1 什么是稳健性检验

稳健性检验考察的是**评价方法和指标解释能力的强壮性**（robustness），也就是当改变某些参数时，评价方法和指标是否仍然对评价结果保持一个比较**一致、稳定**的解释。简单来说，当我们得出一个结论时，需要通过一系列方法来验证所得的结论是否可靠。当我们改变了一些条件或者假设发现所得结论依然不变，那么我们的结论就是稳健的，反之，所得结论有待商榷，我们需要找出使结论发生改变的原因并进行解释。

在较早的文献中，一般很少涉及稳健性检验，但近年来，大家对稳健性检验的重视程度越来越高，这也体现了大家对所得结论准确性的要求越来越高。做好稳健性检验，是使结论得到广泛接受的重要步骤之一。遗憾的是，目前关于如果做稳健性检验并没有统一的标准，也没有一个明确的说明告诉我们在文章中我们到底应该要从哪些角度去做稳健性检验。因此，每篇文章根据自己的研究目的不同，稳健性检验的角度也会大不相同。比如当你的文章着重于研究方法的设计时，稳健性检验则应该更多关注于研究方法成立的前提条件和假设；而当你的文章数据处理时，则应该更多的关注于数据本身的稳健性。

3.2 为什么要做稳健性检验

Seattle University 的 Nick Huntington-Klein 教授在他的文章 Robustness Tests: What, Why, and How 中写到：

当我们在课上学习到一个新方法时，老师会不断强调每个方法都有自己的假设和前提条件，而稳健性检验就是针对这些假设的。我们想要知道如果其中一个假设或者前提条件改变时，我们所得的结论是否依然可靠，这就是稳健性检验存在的意义。每当我们做稳健性检验时，我们应该思考以下问题：

- 我的研究假设是 A.
- 如果 A 不成立，那么我的结果 B 就可能出现有偏的估计 (可能估计值过高/过低/标准误过小/等等...)
- 我认为 A 在我的检验中可能不成立，因为 C 或者，D 是判断 A 是否成立的条件；
- 又如，D 是另外一种计量方法但是并没有 A 这个假设前提。
- 如果我们发现 A 不成立，那么我们则应该在稳健性检验中用 E 方法重新检验。

举一个简单的例子，假如我们现在准备研究政权的更替对于经济发展的影响，我们建立了一个简单的 OLS 回归模型将经济发展作为被解释变量，政权的更替作为核心解释变量进行估计：

- 我的分析假设是扰动项均值独立于所有解释变量，即变量外生，不受内部因素的影响，不存在遗漏变量的问题
- 如果存在遗漏变量问题，那么在回归中政权的更替这一变量的估计值就会过高或过低（取决于遗漏了哪些变量）
- 我认为我们这个分析中存在遗漏变量问题，因为政权的更替通常会伴随着暴力事件的增加，而暴力事件的增加则会影响经济的发展，所以暴力事件是我们在随机扰动项中没有控制的变量
- 那么，增加暴力事件这一变量作为控制变量是我可以进行的稳健性检验之一。
- 如果我们发现，增加了这一控制变量之后，使得我的结果与原先的结果完全不同，那么我们之前的结果则是不稳健的，我们应该加入这一变量进行重新估计。

本例中所提及的稳健性检验方法就是我们下文将要介绍的「补充变量法」。

3.3 稳健性检验的方法

稳健性检验保证的是同一个核心结论：换一份相似的数据、换一段样本期、换一种变量测量方式，文章的结论仍然能立得住。换言之，结论不该是某一个数据切片、某一种测量口径、某一组样本范围下才出现的”巧合”。

实践上，单一稳健性检验通常不够说服力，作者会同时使用多种方法相互佐证。下面四种是文献中最常见的——**变量替换法、加入遗漏变量、改变样本范围、改变模型设定**——每种方法在覆盖的”潜在质疑维度”上各有侧重。

3.3.1 变量替换法

在我们进行分析时，常常会选择自己最熟悉或者偏好的方法测量一个变量，而实际上**一个变量的测量方法有很多种**，我们根据以往文献研究或者依照自己数据可获得性选择的测量方法往往无法保证结论的可靠性。因此，在文献中，作者都会将变量替换法作为稳健性检验的方法之一。变量替换法包括替换因变量，替换主要自变量以及放宽变量条件等角度。

替换因变量 周京奎 (2019) 在研究农业生产率和农村家庭的人力资本积累关系时发现随着农业生产率提高，农村家庭倾向于进行教育投资，进而提升了家庭人力资本积累。在本文中作者首先采用家庭教育支出和家庭学杂费支出来衡量教育投资。在随后的稳健性检验章节中，作者将被解释变量替换为家庭教育支出占当年家庭收入的比例，考察农业生产率对教育支出占比的影响，进一步验证了农业生产率对人力资本投资影响的稳健性。

谭远发 (2015) 研究父母政治资本如何影响子女工资溢价的影响时, 考虑到实际工资与保留工资正相关, 因此将正文中子女的实际工资替换为保留工资进行稳健性检验。李春涛 (2020) 研究金融科技发展对企业创新的影响时将企业的专利申请数量作为反映了企业的创新产出水平的衡量标准之一, 随后作者进一步运用企业研发支出总额占销售收入的比例更替企业创新的度量指标进行稳健性检验。

这里需要注意的一点是, 除了替换因变量, 学者有时还会**对因变量进行一些修正**, 比如王雄元 (2019) 在检验国际贸易增加如何影响企业创新行为时考虑到未取自然对数的专利申请量数据为离散型变量, 且其分布中存在大量 0 值, 可能不符合正态分布的假定, 因此采用泊松模型回归处理被解释变量非正态分布问题。

替换自变量 蔡晓慧 (2016) 在研究地方政府基础设施和企业技术创新关系时, 正文部分讨论中使用的地方政府基础设施的数据来自于金戈 (2016) 估算的省级基础设施资本存量数据, 而在稳健性检验中采用了**地级市市辖区道路密度**代表**基础设施资本存量**。因为道路交通是重要的基础设施, 也是企业通过扩大市场规模取得规模经济的前提, 道路交通的密度在一定程度上也反应了基础设施的基本存量。

替换自变量的文章比比皆是, 可参考 刘怡 (2017); 李卫兵 (2019); 董香书 (2012); 周颖刚 (2019); 申广军 (2017); 孙传旺 (2019); 顾夏铭 (2018); 梁斌 (2020); 于斌斌 (2015); 刘啟仁 (2020)。

放宽因变量或自变量条件 除了替换自变量与因变量外, 学者有时还会对因变量或自变量的选择条件进行放宽, 例如陈仕华 (2015) 在研究国企高管政治晋升对企业并购行为的影响时, 对被解释变量的衡量主要是基于董事长或总经理是否调任政府部门职位来判定高管政治晋升, 考虑到董事长或总经理升任集团层面的董事长或总经理, 或者升任集团层面的党委或党组书记时, 国企高管的行政级别也得到了提升, 因此在稳健性检验部分借鉴王曾等 (2014) 的测量方法, 将高管职位变更去向出现以下情况时均视为晋升: 平级或者更高级别的政府部门职位、集团层面的董事长或总经理、集团层面的党委或党组书记。以此替代变量进行测试。

3.3.2 补充变量法

在上文讲述稳健性检验时, 我们曾举到一个例子, 当探讨政权的更替对于经济发展的影响时, 我们会产生遗漏变量的问题, 而遗漏变量问题是我们大多数研究中都会遇到的问题, 我们只能尽可能多的在模型中加入我们能想到的以及之前文献研究过的对我们结果可能产生影响的变量。因此, **控制变量法**和之前的**变量替换法**几乎成为每篇文献中都会使用到的稳健性检验方法。

加入遗漏变量 除了前文所举的例子以外, 梁斌 (2020) 在探讨失业保险金对失业者求职努力的影响时, 将失业者在日志日搜寻工作的小时数作为因变量, 失业者领取到的失业保险金作为自变量, 并控制了个体特征变量以及家庭特征变量, 加入了省份虚拟变量后, 在稳健性检验部分提出, 失业保险金对失业者来说是确定性的收入, 因此本文预期厌恶风险的失业者 (risk-aversion) 更可能领取失业保险金, 也更可能为了日后稳定的收入而积极寻求工作, 因此又将风险这一变量纳入了考量。

类似的加入更多控制变量的文章可以参考蔡晓慧 (2016); 陈仕华 (2015); 张龙鹏 (2016); 李春涛 (2020)

加入各类虚拟变量 “加入遗漏变量”不仅指补回连续型协变量，也包括**控制其他层面的固定效应**，比如施炳展 (2020) 在研究互联网对制造业企业分工水平的影响时提到，在前文中作者只控制了年份固定效应和企业固定效应，虽然大多数企业并不会更换省份和行业，但是这种可能性是客观存在的，因此如果不加入省份和行业固定效应，有可能遗漏省份和行业层面不随时间改变的重要变量，从而使估计结果有偏和不一致。为了避免这一问题，作者在保留年份和企业固定效应的基础上，进一步加入了省份和行业固定效应。

类似的文章可以参考柳光强 (2018)；孙传旺 (2019)；罗勇根 (2019)

3.3.3 分样本回归法

不同的样本对于所得的结果具有不同的敏感性，因此在稳健性检验时，也常常进行分样本回归，常见的分类方法用按照人口规模分类，按照地理位置分类，按照城乡分类，按照性别不同分类等等。

比如，刘怡 (2017) 在研究婚姻匹配对代际流动性的影响时提出婚姻匹配是中国代际传递的重要机制，尤其是对女性而言，父代收入通过婚配市场作用于子代配偶的个人收入，形成代际传递，影响子代家庭收入。在稳健性检验中，作者根据子代的城乡分布，将子代样本划分为城镇和乡村样本，比较分析城镇和乡村地区的代际流动性及其婚姻匹配机制在代际传递中的影响，结果发现，城镇地区多依赖于婚姻匹配机制，而农村地区侧重于人力资本投资。

3.3.4 调整样本期

当我们在所得的整个数据集范围内进行分析时，常常会发现改变不同的时间段，得到的结论可能会完全不同。也许某一结论在某一时间段内得到的结果符合我们的预期，而当我们往后退 10 年，或者往前推 10 年再次回归，就会发现得到的结论完全不同！因此，选择正确的研究时间段也显得十分重要。在稳健性检验中，我们可以通过扩宽时间长度或者缩短时间长度来检验我们的结论。

扩展时间窗口 仇童伟 (2019) 在研究宗族代理人对村庄地权变更的影响时在第一个稳健性检验方法中提到，村庄的丧葬习俗表征了社区开放程度，在原文中采用了 2012-2014 的数据，而在稳健性检验中补充采用 1990-2014 年村庄丧葬习俗进行了处理。因为与仅采用 2012-2014 年丧葬习俗相比，采用 6 个时期的丧葬习俗可以规避单一时期测量造成的误差。

缩短时间窗口 李卫兵 (2019) 在研究空气污染对企业生产率的影响时在稳健性检验部分提到该文选定的样本期为 1998-2013 年，而大部分基于中国工业企业数据库进行研究的文献主要利用 1998-2007 年的企业数据，虽然该文对某些缺失的数据根据相关的会计准则进行了补齐处理，为避免处理后的数据干扰实证结果，作者将样本调整为 1998-2007 年，并重新进行 RD 估计。

缩短时间窗口的另一个好处是可以**排除其他政策的影响**，比如王雄元 (2019) 在研究“一带一路”如何影响企业创新行为的研究中提到，中国于 2013 年正式提出“一带一路”倡议，因此在样本仅保留 2013 年及以后开通“中欧班列”的样本有助于将本文的研究统一置于“一带一路”倡议的背景下，排除可能的其他政策干扰。(注：另一种排除同时期其他政策的影响的影响是通过**控制同时期政策带来的影响**，比如齐绍洲 (2018) 在研究排污权交易试点政策是否诱发了企业绿色创新文章时提到，排污

费征收政策与排污权交易试点政策并行，我们可以通过需要控制排污费征收政策对企业绿色创新的影响，进一步提炼排污权交易试点政策对企业绿色创新的因果关系。)

类似的缩短时间窗口的文章包括何欣 (2016)；孙传旺 (2019)

3.3.5 滚动窗口法

陈冬华 (2018) 在研究产业政策对股价同步性影响文章中提出，产业政策的影响是一个循序渐进的过程，因此在稳健性检验部分基于滚动窗口的实证研究方法对产业政策进行了动态研究。

3.3.6 改变样本容量法

当我们选择好了时间之后，同时也要确定我们的**样本是否最能体现我们所研究的问题**，同时样本中有没有**极端值**会影响我们的结果。因此，在稳健性检验中，我们需要将个别离群值剔除，或者在样本中选择最适合我们研究目的样本来检验我们的结论是否依然稳健。

选择子样本 鞠雪楠 (2020) 在研究跨境电商平台克服了哪些贸易成本时提出在跨境电商出口贸易中，中国向各个国家 (地区) 出口的分布并不均衡。其中，美国是中国最大的出口目的地；中国香港和新加坡是全世界重要的转口贸易地区，中国向这两个地区的出口可能也有转而向其他国家出口。为了确保实证分析的结论不受特定国家 (地区) 和转口贸易的影响，本文给出了剔除这三个国家以及地区的样本之后的实证分析结果。

缩尾处理 在处理离群值时，我们要进行缩尾处理，陈强远 (2019) 在研究中国技术创新主要激励政策对企业技术创新质量和数量的影响时提到，由于控制变量如资产收益率与负债比率的测算存在极端值，尽管上文已对资产收益率与负债比率进行了 5% 分位上双边缩尾。但为了进一步验证前文结论的稳健性，接下来本文对企业的资产收益率与负债比率进行了 1% 分位上双边缩尾处理。

扩充样本容量 除了剔除部分样本进行回归之外，我们依然可以通过增加样本来进行稳健性检验。比如原文中只采用了省会城市进行分析，在稳健性检验部分则可以将样本扩大到所有地级市城市，这一方法有时也被称为**降低数据维度**。

比如李卫兵 (2019) 在研究空气污染对企业生产率的影响时提到，本文提取的 PM2.5 排放浓度来源于城市层面，同时由于大样本选择下更易带来显著的回归结果，为了证明回归结果的准确性，我们参考江艇等 (2018) 的处理方法计算出城市层面的 TFP，将区域层面的数据降低至城市层面。(注：除了降低数据维度，我们同样可以提高数据维度，比如铁瑛 (2019) 在人口结构变动的的影响时多个个体维度进行调整，分别加总至企业维度和城市维度进行稳健性分析)

3.3.7 内生性问题

内生性问题是每个文章都要考虑到的问题。在处理内生性问题时，我们通常采用以下几种方法进行稳健性检验：

工具变量法 工具变量是解决内生性问题的一个重要方法,比如施炳展(2020)选择了中国建国初期各省份人均函件数量作为省份层面企业互联网普及率的工具变量,选择一个合适的工具变量可以对整个研究都有重要的影响,但同时也是十分困难的,我们可以通过大量的文献阅读积累来选择最合适本文研究的工具变量。

类似的利用工具变量克服反向因果关系的文献可以参考蔡栋梁(2018);周京奎(2019);梁斌(2020);刘啟仁(2020);张龙鹏(2016);罗勇根(2019)。

加入滞后变量 部分研究也会将自变量的滞后一期或者两期变量纳入模型中来解决内生性问题,比如孙传旺(2019)在研究交通基础设施与城市空气污染的关系时除了控制核心解释变量的内生性偏误,我们还担心其他控制变量也可能存在潜在的内生性问题。为了检验结果稳健并排除这一种担忧,将其他所有控制变量滞后一期;黄健柏(2015)到工业用地价格扭曲对企业过度投资的影响可能存在更长的时滞效应,把回归模型中的工业用地价格扭曲程度变量替换为滞后两期项,重新进行回归分析;李春涛(2020)考虑到创新投入也是影响专利产出的重要因素,本文在控制变量中加入企业创新投入的指标,并采用研发支出总额占销售收入之比来度量。由于创新投入对创新产出的影响具有时滞性,本文使用滞后一期的创新投入指标。

类似的文章可以参考顾夏铭(2018);刘啟仁(2020)。

样本自选择问题 陈强远(2019)在研究中国技术创新主要激励政策对企业技术创新质量和数量的影响中提到,高新技术企业认定等技术创新激励政策可能存在自选择问题,即企业整体绩效较好的企业更容易享受优惠政策,这可能导致估计结果存在偏误。为了解决这一问题,文章采用 Heckman 两步法进行了稳健性检验。类似的文章包括蔡晓慧(2016);周颖刚(2019)等。

注:因为内生性问题十分重要,也有一些文章不将其作为稳健性检验的一部分,而是作为正文当中的一部分,比如高晶晶(2019);韩永辉(2017);余吉祥(2019)

3.3.8 其它方法

验证前提条件 正如前文提到,稳健性检验就是为了检验回归方法中的前提条件是否满足,比如吕越(2019)在采用双重差分法研究“一带一路”倡议的投资对对外投资的影响时检验了 DID 的方法成立的条件,包括安慰剂检验,平行趋势检验等等,类似的文章周茂(2019);朱晓文(2019);梁斌(2020);陈冬华(2018)

同样李卫兵(2019)也在使用 RD 估计时,辅助进行了 RD 检验的有效性检验。;类似文章还有梁若冰(2016)

模型替换法 在上文中提到的蔡晓慧(2016)这篇文章中,作者依次在正文中采用线性概率模型进行研究后,在稳健性检验部分又依次采用 Logit 模型、Probit 模型进行估计基础设施对企业是否投入研发的影响;同样施炳展(2020)考虑到线性回归模型潜在的模型设定偏误,以面板 Tobit 模型替换线性回归模型后重新进行了回归;李春涛(2020)认为本文使用的专利数量有大量的零值,存在截尾

数据的特征，因此使用 Tobit 模型进一步检验金融科技发展对企业创新的影响；祝树金 (2020) 用断点回归能较好的识别因果关系，这里使用这种方法对前文的 DID 回归进行稳健性检验。

更换新的数据源 何兴强 (2019) 在探讨房价收入比对家庭消费房产财富效应的影响时，为了增强研究结论的稳健性，分别使用了调查数据、宏观数据、和不同的家庭调查数据重新估计本文的主要回归。这种方法对于数据的要求较高，因此使用频率较低

3.4 总结

分享一下 Cristobal Young (2015) 在针对稳健性检验时提出的一段话作为总结：

学者总是在努力能够通过他的文章采用无懈可击的证据来讲述一个“完美”的故事，但实际上我们必须承认，不稳健的结论有时可以引发我们更多深入的思考，也许一个重大的发现就隐藏在我们不稳健的结果背后。在稳健性检验时，我们需要更多的耐心来面对我们不稳健的结果，同时我们也需要更多的动力来揭秘不稳健结果背后隐藏的秘密。

因此，最后希望大家在面对不稳健的结果时，不要感到无措或者恐慌，静下心来思考一下背后的原因，这才是研究的意义所在。