**实证研究13点**

（文字稿，语音版参见“铭心而论”公众号）

陆 铭

“翅膀长在你的肩上，太在乎别人对于飞行姿势的批评，所以你飞不起来。”

——麦卡勒斯《伤心咖啡馆之歌》

最近十余年来，计量经济学的方法在中国内地高校得到了普及。在技术层面，研究的质量得到了极大的改善。但与此同时，实证研究却出现了三大弊病：第一，因为缺乏对于所研究问题的具体背景的了解，研究方法的适用性在不少研究中被忽视了，一些研究得到的政策结论，甚至南辕北辙；第二，研究普遍关注于系数的统计显著性及内生性的处理，其他一些在实证研究中非常重要的基本问题被忽略了；第三，因为追求方法的所谓现代性和科学性，一些无法用到最酷炫的技术的最基本事实和重要问题缺乏关注。

各位现在看到的这一系列讲座，目的是为了澄清一些实证研究中的相关误区。这些讲座不能取代实证研究方法本身的学习，事实上，计量经济学方法本身的学习材料已经非常容易找到，上海对外经贸大学和中国青年经济学家联谊会联联合办的实证研究暑期学校也致力于普及实证研究方法。

在各位听到和看到的这一系列讲座中，为了增加实际问题的理解，我举了一些例子，仍然是围绕着我自己的研究（参见文末目录），而较少涉及到其他研究，原因有两点：第一，我自己就处在城市和区域发展相关问题的最为核心的争论当中，所以这些例子有助于大家理解研究和政策讨论的关系，顺便也澄清一些相关的问题；第二，我对自己的研究中存在的问题最为了解，而对于讲座中提到的其他人的研究，则不宜过多贸然评论，相信各位能独立判断。

因此，各位听众和读者各取所需，有一分用，取一分，有十分用，则取十分。

**目录：**

1.  研究选题有没有价值观问题？这关乎学术的审美

2.  为什么要有参照系，并在边际上看问题？好的研究者不犯方向性错误

3.  中国真的能够为一般理论提供证据吗？你要检验的理论是个市场经济啊

4.  为什么制度背景是重要的？请确定你讲的故事不是发生在印度

5.  理论对于经验研究到底意味着什么？理论不是经验研究的遮羞布

6.  横截面数据可以用来干什么？一大堆问题，来不及秀技术

7.  面板数据到底何用？有时候，重要的就是“固定不变”的

8.  交互项有什么用？看差异的差异

9.  为什么要去做异质性分析？不要相信有普适的政策

10.  经济的显著性为什么要评估？宰牛何用杀鸡刀

11.  模型的解释力在什么情况下是重要的？别“拿着鸡毛当令箭”

12.  为什么要做实地调研？高校真的是象牙塔啊

13.  我们做的研究真的有政策含义吗？好吧，你说的对

**＃1研究选题有没有价值观问题？这关乎学术的审美**

 13点用来骂人的话，说的是一个人太疯狂，可是我有的时候觉得在中国的学术研究，特别是政策研究的相关领域里面，我们受到了一些既有的制度和激励机制的制约，恐怕我们的狂人是太多了，但是我们却缺少一些疯子，来突破既有的制度和激励机制对我们的制约。

 有的时候我们对于很多既有的研究，什么是好的研究的判断也出了很多问题，特别是对于青年学子来讲，很多时候我们在模仿一些所谓的优秀的研究的时候，恐怕忘记了很多学术研究，特别是政策研究本来所应该追求的价值，甚至产生了一些对于方法的误用，这就是我做这一系列语音它的一个原始的初衷。

 今天我们先来讲第一个问题，那就是选题，我要问的问题是研究的选题到底有没有价值观的问题。

 先从地方开始说起呢？先从一次我和一位非常资深的一位经济学家之间的讨论开始说起。大家都知道在今天的学术界，特别是高校里，我们来评价一位学者是否好，是否足够晋升到比较高的职位上，尤其是对那些从海外学成归国的海归来讲，我们的评价体系大量鼓励他们去追逐国际顶尖杂志的发表。

 在这样的制度框架之下会产生什么样的问题？那一次我在跟一位比较前辈的学者在讨论到这个问题的时候，我就问了这样的话，如果我现在在所有条件都同样的情况下，比如说同样有数据，同样有理论、同样有模型，但是我面临两个选题，一个是研究美国的黑人教育问题，另外一个是研究中国的留守儿童问题，那么在中国的环境之下，在时间有限的情况下，我先研究哪个问题呢？

 我想可能在今天的激励机制下，可能很多的选择都会去研究关于美国的黑人教育问题，而可能暂时把留守儿童的问题放在一边。其中的道理也很简单，因为美国的黑人教育问题可能对于美国的学术界来讲就更为关心一些，那么在发表上可能就相对来说更加容易发表在比较好的杂志上。而似乎中国农村留守儿童的问题，是一个中国特有的问题，所以在我举的这个例子里，研究显然是有价值观问题的。

 我们要问的一个问题是，我们做研究到底是为了发表，还是为了解决现实的重大问题，当然我这里主要讲的是那些实证的应用研究。

 我们中国的人口，在世界上占有大约五分之一的比重，对于我个人来讲这样一个庞大的人口，是一个足够庞大的群体，任何有关中国问题的研究恐怕都不是一件小事。但是我们中国的社会科学研究的确面临一个从传统的中国学术话语体系，逐渐走向世界，要跟全球的社会科学研究的话语体系去对接这么一个问题。

 于是问题就来了，在目前的英文杂志里，的确有国际所通行的所谓顶尖杂志以及其他一般杂志这样的区分，而大量的顶尖杂志其实是在美国出版的，当然也包括少量在欧洲出版。

 在美国学术研究里，的确存在所谓具有普遍意义的理论还有应用性的研究这样的区分。在应用研究当中，其实他们很多的应用研究是在研究美国的一些重要问题，就比如说我今天讲到的教育问题。其实美国的教育也有很强的特殊性，它的教育财政体制并不是世界通行的，具有普遍意义的教育财政体制，它的人口学特征也非常特殊，因为美国的历史大家都知道，产生了一个非常庞大的黑人群体。那么对于这样一个在长期美国历史上倍受歧视人群的教育的关注，就成为了美国社会一个非常重要的问题。在美国发表的学术论文，由于有一些应用性的问题，对美国来讲实在是太重要了，所以他们可能也会对有一些对美国非常重要的问题，给出一定的空间来发表它。

 到了中国，我觉得很多问题就有很多差异了。在中国研究里，当然有很多也的确是具有普遍意义的一些问题，再拿教育举例子，我自己也做有关教育的研究，比如说我最近的研究涉及到了人力资本外部性这样一个非常重要的问题，它讲的意思是说一个城市的教育水平如果提高了，怎么能够通过知识的溢出效应来提高整个城市的劳动生产率。另外还有我研究技能互补性，也就是高技能劳动者和低技能劳动者之间，是一个什么样的关系，是相互竞争的还是互补的？再比如说我的朋友圈里有一些陕西师范大学的朋友，他们最近有一些研究是关于0到3岁的儿童早期教育的一些问题。

 我刚刚举的这些例子都是有普遍意义的话题，但是有一些中国的问题就极为特殊。比如说我今天举到的例子，就是农村留守儿童的问题，关于农村留守儿童问题的产生，大家不难想到，其实是跟中国非常特殊的户籍制度是有关系的，大量的农村进城务工人员，他们在城市里，特别是在一线城市，很难将自己的孩子送到公立的学校去读书。尤其在最近这些年，随着北京、上海这样的城市加强对于人口规模的控制，外来人口把自己的孩子送到公立学校去读书——甚至民办学校去读书——的困难就越来越强了。

 这个问题是非常具有中国特色的问题，而留守儿童的问题到底有多严重，其实官方的统计和学者的统计口径也不太一样，官方的统计口径是统计的年龄在16岁以下的学龄儿童，而且他们统计的范围是，在父母双亲当中，只有一位离开孩子进城打工，而另外一位可能是和自己的孩子留在老家的，这样就不算留守儿童。

 但是对于学者的估计来讲，如果要是一位家长外出，另外一个家长留在身边，就把他们记入了留守儿童的统计范围，所以这两者的统计出来的数据是大相径庭，那么到底一个孩子在成长期是双亲都离开了，还是仅仅离开了一个，对于孩子后面的教育发展以及长大成人后的表现有什么影响，这本身就是非常重要的问题。

 甚至不同的统计口径所得出的留守儿童的规模到底有多大，随着时间的变化和大城市控制人口的政策，对留守儿童的数量影响到底有多严重，这个基本问题现在恐怕都搞不清楚。再包括留守儿童问题会产生怎样的社会后果，比如说是不是留守儿童长大以后的犯罪倾向会更严重，还有在公共政策领域，对于解决留守儿童的问题，到底是采取措施让更多的留守儿童跟随他们的父母进到城里去接受教育，还是鼓励他们的父母回到老家，通过加大在他们的老家进行的教育投入来解决问题，这些都是广泛存在争论的，而且值得研究的问题。

 我举的这些例子我想大家不难理解，很多对于中国非常特殊的问题的研究，放到国际学术界，恐怕那些英文杂志的编辑就会问一个问题，你所研究的这个问题，对于英语世界的读者来讲到底有多重要，他们可能会觉得这样的问题太具有特殊性了。但是对于中国的政策研究，包括推进中国的发展来讲，这些问题的重要性几乎是可以不用辩驳的。

 长期以来我个人认为，由于我们过于注重英文文献的发表，对于类似我刚刚讲的这样一些具有重大的中国实践意义的问题的关注，学术界恐怕是不够的。最近这些年从国家层面到学校层面，恐怕也注意到了学术研究跟实践脱节的问题，现在也有一些变化，比如说国家开始鼓励智库建设，但是当前存在的对于智库建设的一些鼓励，我觉得又产生了新的问题。

 比如说智库的研究和学术的研究到底是什么关系，如果把智库研究等同于不断的给政府写政策报告，用上级的领导批示作为对智库研究肯定性的一种指标，那么我们需要问的问题是，智库的研究是不是就不需要学术研究和扎实的数据分析来作为它的基础，是不是智库研究就可能很难产生批评性的研究成果？

 所以我们共同面临的问题在未来就是，如何在学术研究严谨的研究和政策研究之间找到某种最佳的平衡点和结合。

 可能很多人会问，我是一个个体，国家的政策变化和制度变化可能我影响不了，那么我自己该做什么样的选择？坊间存在这样一种说法，甚至有一些从海外回来的一些经济学家，我的前辈也采取这样的看法，那就是一个年轻学者在40岁甚至45岁前，在拿到学校的长聘教授职位之前，可以不用那么关心国家的政策研究和应用研究的问题，可以更多的在乎发表的学术论文的档次。等到拿到学校的长聘教授职位以后，再转来做中国问题的应用研究，我自己对于这样的看法一直是不太认同的，一个人如果要是在40岁甚至45岁之前，长期都从事跟中国没有关系的研究，恐怕这个人的知识结构到了中年之后就已经固化了。

我后面会讲到，在国际学术界发表的很多学术研究的论文，它所运用的理论框架，往往是一个标准市场经济的理论框架，而当前中国的很多问题其实不是产生于一个教科书式的标准的市场经济体制。所以如果长期的脱离中国的实践，对于形成一个固化的思维就会产生严重的问题，到了中年之后再进行扭转就非常困难了。

不仅如此，学术研究特别是在应用研究这个问题上，话题的前沿性的把握是非常重要的。一个人长期脱离中国实践的话，如果再想到中年以后试图去把握中国研究的问题的前沿性，我觉得恐怕也不是一件容易的事情。

 今天我第一讲讲到了关于研究选题的价值观的问题。打一个比方吧，现在流行音乐界也有类似的看法，在一个看脸的时代，还有人会在意唱功吗？我觉得这是一个多元价值的问题，有的人就是要去看脸，有的人就是在意唱功，但是我相信对于真正好的艺术和好的学术来讲，总会有人会去欣赏它，发光的并不一定都是金子，但是我仍然相信，虽然目前的评价体系并不是那么鼓励那么去做真正重要的，跟中国发展相关系的那些重大话题的研究，我仍然相信是金子就总会发光的。

**＃2 为什么要有参照系，并在边际上看问题？好的研究者不犯方向性错误**

 这一讲讲两个非常重要的思维方式，一个叫参照系，另外一个叫在边际上看问题。

 这两个思维方式虽然说是我们经济学家特别强调的，但是其实我不认为只有经济学才需要有参照系和边际上看问题的思维方式。其实我们在讨论任何社会科学问题，公共政策，甚至我们在平常的生活当中去讨论一些有关社会公共话题的时候，这两种思维方式都是非常重要的。

 首先我们就来谈参照系，今天我给大家举三个例子，来讨论什么是参照系，以及为什么参照系对我们讨论问题来说特别重要。

 先来谈中国经济的投资和消费的关系，很长一段时间以来，中国被全世界范围之内的经济学家都批评为投资过度了，而消费相对来说不足。于是这样的问题就引起了一些争论，有一些学者提出，中国的经济结构其实并没有那么坏，中国的投资和消费也没有那么多的问题。甚至有著名经济学家提出这样的观点，说你看中国仍然是一个发展中国家，如果我们把发展中国家跟发达国家去比较，比如说就拿中国和美国相比较而言，中国的人均资本的存量非常低，因为我们是发展中国家，而美国，人均资本拥有量要远远超过中国。别的不说，就说机场，美国整个国家的机场有几千个，而中国一千个都不到，仅从这样的指标看起来，似乎中国和美国的差距就非常大。所以就有学者说，中国还要持续加大投资，这句话看起来好像蛮有道理的，但是我想告诉大家，这里有一个参照系选择的一个错误问题。

 在我们讲到中国经济是不是投资过多，消费不足的时候，其实非常重要的一个参照系，是在宏观经济学当中的一个有关投资消费比率的黄金率这么一个理论。所谓黄金率就指这样一种状态，如果一个国家一直在投资，加大自己的产能，但是这个国家的消费却不足的话，那么这个国家投资就会出现投资回报下降的状态，说的通俗一点就是你的产品卖不出去，或者必须大减价来卖。这样的状况下，哪怕你今天的投资很高，看上去GDP增长速度非常快，但是到了未来，因为你大量的产品卖不出去，或者贱价来销售，结果就导致企业的投资回报很低，再进一步进行投资的能力就会下降。这样的话从长期来看，从经济的可持续发展角度来看，那么这个国家的投资就显得过度了，如果你通过促进消费来提高投资的回报率的话，那么有可能使得从长期来讲的经济增长率会更快。

 那我们谈一谈对于美国的追赶，我们不否认中国相比美国来讲，我们人均资本拥有量还是一个比较低的水平，但是我们今天要讨论的问题是我们用多快的速度来追赶美国。如果我们急于求成的话，可以套用中国的古话叫“欲速则不达”，也就是说你今天投资多了，表面上看起来好像追赶的速度快了，但是当我们把时间拉长的时候，其实你的投资降低一点，消费上去一点，能够提高投资的回报，这样的话，反而从长时间来看，你追赶美国的速度会更快，这就是黄金率的道理。

 以黄金率作为参照系的话，中国的投资恐怕就长期以来太高了，产能出现了过剩的问题。如果要是我们选择错了参照系，就会对中国经济的经济失衡的状态，产生严重的错误的判断。

 我举的第二个例子是关于中国的制度和经济增长的关系。很多人都持有这样的观点，而且是一种非常流行的观点，因为它有非常著名的经济学家作为背书，说中国的经济制度，尤其是官员晋升激励的制度，是有利于中国经济增长的。因为大家都知道，中国的地方官员的考评和晋升，往往都是基于一些经济的表现，比如说经济增长率、招商引资还有税收。如果一个地方官员为了实现这样的目标来追求自己的晋升，那么你看到的结果就是我们的地方政府官员工作非常努力，大干快上，似乎这也的确带来了中国经济的快速增长，于是就有人把它总结为一种中国模式，通过实施官员的晋升激励考核，把它跟经济增长挂钩起来，来促进经济增长。

 但是在这一点上，又涉及到参照系问题了，如果我们用长期的可持续的经济增长作为一个社会的最佳状态，同时我们又以一个国家内部不同地区之间的资源最优配置，来作为一个最优状态的话。那么相对于这样的最优状态，中国目前所形成的情况，恐怕你最多只能说官员的激励考核制度有利于促进短期的和局部的经济增长。

 也就是说，恰恰是因为我们的地方政府有非常强的激励来加大本地的投资，结果在整个国家层面就出现了我前面所讲的产能过剩、投资过度的问题。

 与此同时，由于每一个地方都在最大化本地的投资和经济增长，结果就使得一些欠发达地区和人口持续流出的地方，反而出现了严重的投资过度的问题。也就是说官员激励和考核制度之下，所形成的经济增长最多也就是一种短期的、局部的经济增长，但是很难在时间上持续下去，并且在地区和地区之间形成了重复建设、产能过剩，甚至加剧了地区和地区之间的市场分割。

 所以一旦我们选择一个比较正确的参照系的话，那么就知道，如果我们进行制度的改革，比如说去修改地方政府官员的激励和考核制度，就可以让中国经济增长的质量能够有所提高，持续性也可以加强。

我给大家举的第三个例子涉及到区域间经济发展的差距，如果大家去看2004年以来中国地区之间——主要是指省之间——人均GDP的差距或者收入的差距，你会发现似乎中国的省和省之间出现了一些收敛的现象，所谓收敛就是指不同地区的人均GDP和人均收入的差距在缩小。于是就有很多人说，这是一种经济发展的收敛现象。真的是这样吗？大家如果去回顾一下我们在经济学当中所学的经济增长的收敛理论，他讲的是有两个国家，——注意国家和国家之间的劳动力是不能自由流动的，——在劳动力不能自由流动的情况下，在发达国家，逐渐的就会出现资本过剩，投资的回报率下降，这个时候发达国家的资本会考虑转移到发展中国家，跟发展中国家的劳动相结合来进行生产，而这个时候发展中国家正好可以利用他丰富的劳动力资源，结合发达国家转移过来的资本进行追赶，于是不同国家之间的收入差距就可以缩小。中国改革开放以来40年的历史，其实就是在我刚才所讲的理论当中，跟发达国家差距逐渐缩小的历史过程。

可是当我们讲中国地区间的收入差距的时候，大家就要小心了。在我们刚才所讲的理论当中一直在强调国家之间的劳动力是不能自由流动的，而在一个国家内部劳动力应该是可以流动的，尽管在中国当前的状态之下不是那么自由。但是对于参照系的选择来讲，我们就要以劳动力可以自由流动的状态作为资源配置最优化的参照系。

 那么中国出现的省和省之间的收入差距缩小，是因为在劳动力不能自由流动的情况下，资本转移到了欠发达地区吗？请注意，如果你这样认为的话，其实你是在说，在中国的一些东部沿海城市和大城市，已经出现了资本回报的下降，于是他们就转移到了投资回报更强的中西部地区和小城市。可是实际上的情况真的是这样吗？其实不是的，我在我的一系列研究当中告诉大家，其实2004年以后，所出现的欠发达地区经济增长速度更快的现象，主要是因为通过行政力量所左右的资源在空间上重新配置，加大对于中西部的投资而来的。而随着时间的推移，有大量的投资回报其实并不高，特别体现在基础设施、工业园和新城的建设上。当你加大投资的时候，当然你会看到欠发达地区的经济增长速度是快了，但是请注意，这种速度的加快，恐怕伴随着地区之间空间资源配置效率的下降，这是我最近一系列的研究在讲的问题。

 我们再回到前面所讲到的经济增长理论里面的收敛来看，大家要知道，在国家之间，发达国家由于出现投资回报率下降，把资源转移到发展中国家，这个时候不同国家之间的收入差距的缩小，和资源配置效率的提高，两者是并行不悖的。而在中国一个国家内部，通过加大对于欠发达地区的投资，如果这个投资是没有可持续性，也没有足够高的回报，那么尽管你在短期可以看到欠发达地区的经济增长率更快了，但是其实对于这个国家来讲资源配置效率是下降的。

 这又是一个例子告诉我们，我们不能只看数据，说好像数据看上去的现象跟理论所预测的是一致的，其实因为你的参照系没有理解清楚，就会对现实生活当中在出现什么样的情况出现严重的误判。

 一旦我们用正确的参照系讨论现状的时候，你就会知道，其实既有的制度改革有很大的空间。可是每当我们讨论改革这样的问题的时候，一定会有很多人提出这样的质疑，说现在不能这样，不能那样，好难，可是我想提醒大家，我们千万不能以现状来否定改革的必要性，否则一个社会就没法进步了。

 我在我那本比较畅销的《大国大城》的书里曾经讲过一句话，叫“理想不是用来实现的，而是用来指明方向的。”学术上所探讨的参照系，其实就是我们的理想状态，没有任何一个国家可以真正的实现那样一个理想的状态，但是如果一个国家可以持续的往那个理想状态迈进的话，那么对于整个国家的发展和社会的福利来讲，都是一个福音和进步。

 好的，讲完参照系问题以后，我们再来讲讲“从边际上看问题”。我们从一个小故事讲起，有一次我参加了一个学术研讨会，我在会上讲到了关于促进中国人口在地区之间自由移民这样一个问题的重要性，讲完之后就有一位前辈学者做了一个评论，说陆老师，你如果今天到欠发达地区看看就知道，有很多人是流动不起来的。比如说一些残疾人、少数民族，所以要对他们进行财政的补贴。这句话听起来似乎也是对的，但是我要告诉大家，我们在学术问题和公共政策讨论当中，特别注意，请大家采取在边际上讨论问题的办法。在我刚刚讲的例子里，有一部分人因为身体残疾，或者说他们是少数民族，语言跟汉族地区不通，这些问题会形成劳动力流动的障碍。那么即便你改革制度，他们的选择可能也不会发生变化，但是我们今天讨论的问题是，如果我们进行户籍制度改革，如果我们让外来人口的子女能够在城市，特别是大城市同样进入公立的学校读书，如果我们廉租房、公租房的提供不是按照户籍身份来的，而是按你是不是在当地缴纳税收，并且是低收入群体来，……这些制度改革了以后，就会有很多人选择离开老家进入到城市里居住和工作，所以对我们讨论公共政策问题来讲，最为重要的是，要去盯住那些有可能会受到制度变化和政策变化影响的人群来讨论问题，只要这个人群足够大，我们讨论的问题就一定是重要的问题，而且我们所相应要采取的政策措施也就会产生效果，因为它是影响最终的均衡状态的。

 换句话说，对于有些人群来讲，它不会影响经济政策变化以后的均衡状态，虽然这部分人群不会最终影响经济发展的均衡状态，但是公共政策上，当然要在乎他们的福利，比如说对那些不愿意流动出来的残疾人、少数民族地区的居民，加大财政补贴是没有问题的。但是由于我们采取边际上进行讨论问题的思维方式，显然促进劳动力在地区之间，跨地区、更为自由的流动，要更加重要，要远远重要。

 好的，今天我们给大家讲了参照系和边际上看问题两种思维方式，希望对大家有所启发，最后我用一句话来总结一下，那就是“聪明人不犯方向性的错误”。

**＃3 中国真的能够为一般理论提供证据吗？你要检验的理论是个市场经济啊**

 最近这些年来，在学术研究当中出现了一种趋势，那就是运用来自中国的数据，为经济理论提供证据。这种理论研究和经验研究之间的关系，就好象在物理学里存在理论物理和实验物理一样，实验物理能够为物理的理论提供证据，从而形成一种非常重要的学术贡献，那么如果中国的经验和中国的数据真的能够为经济理论提供经验证据的话，那么也是一种非常重要的学术贡献。

 但是问题是，中国真的能够为一般理论提供经验证据吗？不可否认的是，在很多场合之下是可以的，特别是那些比较微观的研究，如果涉及到人的行为，比如说用实验经济学的方法来检验一种行为理论，那么我认为中国人的行为和其他国家人的行为并没有本质上的区别，这个时候，用来自于中国的数据和案例来检验理论，我觉得并没有什么太大问题。

 但是如果对一个具体的问题来讲，涉及到一个国家的制度背景的话，恐怕就要看具体的情况，才能区分是不是真的能够把中国的证据当作一般理论的证据了。

 我这里特别要提醒大家，我们在学习过程中所学到的很多理论，有关于市场经济的理论。其实这样的市场经济理论是一个无摩擦的世界，其中恐怕很多理论都没有政府的管制，没有生产要素在不同地方、不同部门之间流动的障碍，这样一种世界，是一个无摩擦的市场经济。但是我经常跟我的同行和学生说，中国的起点是一个计划经济，在中国，存在着三个词都是以TION结尾的词，分别都是regulation、distortion和misallocation，中文可以翻做为管制、扭曲和错配。

 在这样一个存在着管制、扭曲和错配的经济体，如果我们盲目的去运用那些无摩擦的市场经济理论，用中国的数据去做检验，很有可能你得到的数据检验的结果，看上去和一般市场经济理论推导出来的结果是一样的，但是在实践应用上，就会产生严重的误解。

 今天我结合我自己的研究，来举三个例子。

 第一个例子就是最优城市规模理论。大家知道，一个城市的人口在增长的时候，一开始当城市人并不够多的时候，人口的增长会带来规模经济效应，劳动生产率会逐渐提高。但是过了某一个点之后，人口继续增长的话，那么就会带来规模不经济，会带来劳动生产率的下降，于是在理论上就会存在一个所谓“最优城市规模”。城市的劳动生产率会随着人口的增长先上升再下降，呈现出一条倒U型的曲线。于是就有一些研究用中国的数据去拟合这样一条倒U型曲线，结果还真发现存在这样一条倒U型曲线，有一些中国最大的城市，似乎已经到了城市规模继续增长，劳动生产率开始出现下降这样一个阶段。于是就有学者说，你看，中国有些城市已经超出了最优规模，于是支持政府采取管制的手段来控制城市的人口。

 可是我要提醒大家的是，在中国最大的城市里，实际上却存在着很多的管制，这些管制本身就是导致这些城市劳动生产率下降的因素。

 比如说在中国的大城市里，一直存在着基础设施和公共服务的短缺，之所以产生这样的原因，是因为这些城市在历史上是根据一个比较低估了的人口规模，来规划自己的基础设施和公共服务的。而现实中的城市人口增长的速度远远超过了当初的规划，而在这些城市又在进行人口的管制，其中对于低技能劳动者的管制，其实是严重的制约了低技能劳动力的供给， 于是这个城市的服务价格就会比较高，会导致这个城市的劳动生产率受到负面的影响。

 那么当你在这样一个存在着严重的管制的这样一个城市里，看到了一个所谓负面的城市规模的规模不经济，到底是因为人口增长太快导致的，还是因为政府的规划、基础设施和公共服务的供给不足导致的呢？

 我给大家解释了以后，大家恐怕就明白了，在这样的城市，如果真要缓解城市的规模不经济，恐怕是要增加基础设施和公共服务的投入，而不是简单的控制人口。

 我的第二个例子又要讲到中国地区之间的发展差距的收敛，大约是在2003年、2004年左右，中国省和省之间的人均GDP的差距开始出现一个拐点，之后差距是有所缩小的，于是很多人就认为，中国的区域经济发展验证了经济增长理论里的所谓“收敛”的理论。

 真的是这样吗？大家如果要是回顾一下收敛这样的概念就会知道，其实这个“收敛”是在经济增长理论里首先提出的，它比较适用的是国家和国家之间的经济发展差距的缩小。而在不同的国家之间，为什么会出现收敛呢？那是因为在发达国家，随着投资和资本积累，会出现所谓资本边际回报的下降，这个时候企业为了挣钱，就把自己的投资转向发展中国家。 而这时候发展中国家正好存在着大量的劳动力，于是就借助于自己的劳动力比较丰富的比较优势，利用来自发达国家的资本进行生产，于是实现相对快速的经济增长，从而追赶发达国家，不同国家之间的发展差距就有所缩小。

 但是我想再次提醒大家，国家之间的收敛理论是建立在一个非常重要的假设前提之下的，那就是国家之间的劳动力流动是存在障碍的。这个时候，当发达国家出现劳动力不足，资本边际回报下降的时候，他就把资本转移到发展中国家。但是对于中国来讲，一个统一的国家的不同地区之间，你就要注意，如果地区之间要实现经济发展的收敛的话，其实有一个办法，就是让欠发达地区的人口流动到发达的地区来。

 而在实际中的趋势是什么呢？就是中国一直存在着流动劳动力的障碍，而在大概2003年、2004年的时候，中国开始动用行政干预的力量，把大量政府可控的资源，包括转移支付，包括建设用地的指标投入到欠发达地区，在这些地方建工业园、新城还有基础设施，整个在中国的中西部形成了一个靠投资推动的经济增长方式。

 由于投资本身在短期里就是经济增长，当然你会看到在不同的省之间，人均GDP的差距有所缩小。但是关键的问题是，在人口不能自由流动的情况下，通过增加投资导致的结果，现在已经非常清楚了，那就是在欠发达地区出现了投资过剩、产能过剩、基础设施投资大量闲置，鬼城、空城出现，工业园出现闲置。

 那么这样的一种所谓的“收敛”，是真的收敛吗？大家不要忘记，在国家之间，劳动力不能自由流动的情况下，发达国家将资本转向发展中国家，它实际上是一种有效率的结果，而在一个国家的内部，阻碍劳动力自由流动，再通过政府行政力量引导资源投向人口流出的欠发达地区，这种做法实际上带来的是严重的资源配置的低效率和经济增长的不可持续性。

 如果不区分这一点，就会以为中国地区之间的经济发展差距的缩小，是为主流的经济增长理论的收敛理论提供了证据，并且在实践上会更加进一步的主张用行政力量来干预资源的地区间分布。

 我要举的第三个例子是一个著名的发展经济学理论，叫“刘易斯二元经济理论”和相应“刘易斯拐点”。根据刘易斯理论，在发展中国家存在大量的农业剩余劳动力，但是到了一定的经济发展阶段，就会出现一个所谓“刘易斯拐点”，这个时候农村的剩余劳动力已经被现代化的工业部门和城市部门吸收得差不多了，这个时候就会出现工资的快速上涨和劳动力的相对短缺现象，这个就叫“刘易斯拐点”。

 于是有经济学家就观察到，大概也是在2003年、2004年之后，中国开始出现了快速的工资上涨，和城市部门的劳动力短缺的现象。可是这个现象能作为刘易斯拐点的证据吗？同样因为我刚刚讲的道理，中国在存在比较严重的户籍制度对劳动力流动的障碍的情况下，又尤其存在比较大的城市对于人口规模进行管制的政策情况下，同时我们又加大了对于欠发达地区那些人口流出地区的转移支付和补贴，还有大量政府主导的工程建设。一方面是因为人口流入地不接纳，或者不充分的接纳外来人口，一方面在人口流出的那些人的老家，加大了投入，这时候当然向人口流入地转移的人口的增长速度就会降下来，就体现为工资的上涨和劳动力的短缺。

 可是大家想，这种拐点到底是一种政策的拐点还是真正的刘易斯意义上的拐点呢？要回答这个问题，其实并不困难，因为大家只要回顾一下刘易斯的二元经济理论就知道，在那个理论里是没有劳动力流动障碍的。

 所以如果中国真的出现的是一个劳动力已经真正短缺的刘易斯意义上的拐点的话，你就不会看到在中国存在着这样的一些现象，比如中国的城市化进程一直是慢于工业化进程的，中国城市化的水平相对于世界上处在类似发展阶段的其他国家来讲，低了10个百分点左右。 你也不会看到在2003年之后，中国持续出现的是劳动收入占比下降的这样的一种现象。你更加不会看到，在所谓刘易斯拐点真正到来的时候，中国的城乡间收入差距还在进一步的扩大，一直到2009年的时候，这种城乡间收入差距扩大的趋势才得到了遏制。

 那么我们为什么要严格的区分中国经济是不是“刘易斯拐点”呢？因为它的经济政策的含义是完全两样的，如果中国出现的工资上涨和劳动力短缺，特别是在东部沿海大城市的劳动力短缺，是因为一个制度性的因素的话，那么在经济政策上就应该主要通过去除劳动力流动的障碍，来缓解这样的一个趋势。但是如果真的中国是一个人口学意义上的刘易斯拐点的话，那么它就主要是一个人口数量的问题了。

 我今天举的这样一些例子是想提醒大家，特别是那些年轻的学生和青年学者，在研究中国问题的时候，一定要非常的小心，不要简单的、直接的把对于中国经济的研究，建立在自由市场经济的那些理论上，通过收集数据做一个回归模型，就以为自己得到的结论验证了所谓的主流的市场经济的理论。如果简单粗暴的这样做研究的话，很有可能就会得到一系列错误的结论，特别是在自由市场经济理论之下，得到的结论却用来指导中国实践，提出的政策建议是加强对于中国市场的管制，可是，实际上的情况却往往可能是中国现在存在的很多问题，已经是过度管制的结果了。

**＃4  为什么制度背景是重要的？请确定你讲的故事不是发生在印度**

 在很多场合，我都强调过，在做中国研究的时候，特别是那种具有重大的现实含义的研究——有的是政策研究——的时候，要特别注意中国制度背景的特殊性。如果不把制度背景交代清楚，那么对于中国的相关研究来讲，就可能出现一些误区，甚至可能对一些结果的解读就可能出现偏差。

 今天我来给大家举一个例子，其实这个例子我已经举了很多次了，但是由于它比较有代表性，我宁愿再重复一下。比如说曾经有过一篇论文，发表在一本非常不错的经济学杂志上，它讲的这个故事跟金融是否能够有效促进企业家创业有关。

 长期以来在经济学学术界对于金融作用有不同的看法，有的人认为金融对实体经济没有什么太多作用，它只是在分配资源。而另外一种观点就认为，金融其实对实体经济的发展是有作用的，如果能够提供一个证据证明金融能够促进企业家创业，那么它就是金融能够对实体经济产生真实作用的经验证据。

 于是在我讲到的这件论文里，作者讲了这样一个故事，用的是中国的数据。在上世纪90年代中期的时候，国有企业进行了一场住房制度改革，当时很多公有的住房就以非常低廉的价格，送给了（或者半送给了）国有企业的职工，于是作者就相应的发现，有大量来自于国有企业的员工，开始去做了创业，其中很多其实就是自我雇用形式的创业。

 有趣的是，在同样的时期，来自于非国有企业的人，似乎没有出现这种大量创业的现象。那么为什么会出现这样的反差呢？作者提供了一个解释，认为来自于国有企业的职工，因为经历了住房体制的改革，几乎免费或者低价的拿到了一套住房，于是他们就可以把这套住房拿到金融体系里去抵押贷款，然后再用这笔钱进行创业。如果这个故事是成立的，那么就可以解释为什么来自于非国有部门的人做不到这一点，因为他们没有享受到住房制度改革给他们的一个廉价的或者免费的资产所得，没有办法用房子去做抵押物进行贷款和创业。如果这个故事是成立的，就证明了金融其实有促进企业家创业的作用。

 故事真的是这样吗？作者在自己的文章当中也讲到了另外一种可能性，那就是，是不是因为90年代中期的时候，国有企业也出现了大量职工下岗的现象，于是职工下岗也会去创业、自我雇用。为了排除这样的可能性，作者讲了这样一段话，说国有企业的下岗这件事情，是在1996年之后才开始的，而住房体制改革大约在1994年前后就开始了，所以如果早在1994年的时候就看到有大量来自于国有企业的员工去进行了创业，就不是因为大规模的下岗所导致的，而是因为当时的住房体制改革导致了企业员工用房子去做抵押贷款。这样听起来似乎也挺有道理的，而且可能很多人甚至会认为，这个作者考虑问题还挺仔细的，想到了另外一种解释自己的研究发现的可能机制，而且去做了排除。

 但是我要指出的是，在这篇文章里，由于缺乏对于制度背景的了解，恐怕就出现了非常严重的逻辑漏洞。第一个制度背景是，其实在1990年代，1996年之后只是从上海开始成立了大量的再就业服务中心，开始帮助下岗员工再就业。但是其实并不是从1996年开始才有企业员工下岗这一回事，其实早在1992年、1993年、1994年这些更早的年份就已经出现了下岗员工的现象，而且规模还并不小，这些都是有统计数据的，大家只要查一查相关的年鉴就知道了。（也可以看我的《劳动与人力资源经济学》教材。）所以作者看到1994年的时候就出现国有企业的员工开始大量进行创业，转向自我雇用，其实很可能只不过是因为他们下岗了，找了一份新工作而已。

 更有意思的是，作者认为当时国有企业的员工会拿自己的房子去做抵押物来贷款创业，可是大家想一想，在1990年年代中期，真的中国的金融市场那么支持住房抵押贷款来进行创业吗？当时我读到这篇文章的时候，我人正在美国，于是我就迫不及待的打了一个电话给我在上海的朋友，我有朋友在上海的农业银行工作，我当时就问，全上海用房产来做抵押进行创业或者企业投资的有多少，当时是2011年，全上海的农业银行系统大约只有一万起，通过住房的抵押贷款来进行融资和创业的。

 大家想一想，在1990年代中期，当时严格说来中国根本就没有一个住房的市场，或者说住房市场发育才刚刚起步，更不要说在金融领域里用住房抵押贷款来进行投资这件事情了。如果大家稍微问一问今天年龄大约在60岁，甚至65岁这个年龄群体，他们就是当年经历过下岗的那个群体，大家去问一问他们，他们会不会把自己的房子用来抵押去创业，你就会知道了，这个事情就算能做，在90年代中期，这个人数也就是凤毛麟角。

 所以在我看来，经济学很难说是一门真正的科学，对同样一个数据分析的结果，其实基于不同的制度背景的理解，所能够得到的结论完全两样。如果忽略中国的制度背景，或者说对中国历史上的制度背景并不够了解的话，在我刚才所讲的故事里，你可能就认为中国的历史提供了一个金融能够帮助企业家进行创业的证据。而根据我对制度的理解，其实这只不过是因为有大量的国有企业员工当时下岗了，转而自我雇用，到底谁更对？其实也很难说我们到底有没有更丰富的证据来证明我是对的，而别人是错的，或者至少要有更多的证据来证明这一点才行。但是至少在没有更多的证据之前，根据我对历史的了解，我认为作者恐怕犯了一个错误。

 接下来我再给大家举另外一个例子，是来自于我自己的研究，我想用这个例子来说明，其实对于制度背景的了解，对于我们如何利用中国的制度背景来做研究，其实也是非常有帮助的。

 如果比较熟悉我的研究的朋友可能注意到，一段时间以来，我对中国房价的决定及其后果有一些非常有特点的看法。这件事情要从中国的土地政策开始说起，中国的土地制度非常有特点，因为在我们国家，土地是公有的而不是私有的。一个城市的土地供应，在多大程度上把土地可以由农业用途转为城市的用途，包括工业用途、住宅、商业、基础设施，其实都是受制于一个所谓“建设用地指标”这样一个制度。

 于是在2003年之后，中国的中央政府就开始使用建设用地指标这样一个调控手段，对于东部沿海地区，特别是东南沿海地区开始收紧城市的土地供应，相应来讲，他们就把大量的建设用地指标投向了中国的中西部，特别是中小城市。试图用这样的办法来平衡区域间的发展，当然这种平衡是不是一种真正的平衡，这是我在别的课程里（喜马拉雅《中国经济72讲》）里面已经详细分析过的。

 当时的逻辑是想说，如果我给了中国的中西部一些建设用地指标，他就可以用这个指标来建设工业园、新城来发展自己。结果在客观上，就导致了一个后果，那就是在东部沿海地区，特别是东南沿海地区，又尤其是在北京、上海这样的大城市出现了一个矛盾，那就是人口大量流入的同时，土地和住房的供应跟不上需求了，结果就出现了房价持续上涨的现象。

 同样在这个逻辑之下，我在2015年的《中国社会科学》上发表了一篇文章，我探讨了房价上涨对于劳动工资的影响，我大概讲了这样一个逻辑，那就是在东部沿海地区，收紧土地供应，同时人口又在流入，结果就会导致房价上涨，而房价上涨的时候，房价是老百姓生活当中非常重要的生活成本，当这个生活成本上升的时候，就会推动当地的工资上涨。在这样的逻辑之下我们就可以比较好的解释，为什么在2003年、2004年之后，中国的东部沿海地区出现了比较快速的劳动工资上涨的现象。

 如果要是不把这样的一个非常具有中国特色的逻辑给讲清楚的话，我想对于中国的房价和工资上涨的解读，恐怕就有其他的解释了。比如说很多人认为中国的房价高，特别是在北京和上海这样的地方房价高，体现的是房地产的泡沫。但是根据我的一些研究，我讲到，房地产的泡沫是否存在，永远是从需求方定义的，但是在中国，像北京和上海这样的大城市的房价高启，在很大程度上是因为政策控制了这些地方的土地供应所导致的，是一个住房的相对短缺现象。

 同样的道理，如果离开了我上述所讲的那些逻辑，当你看到中国东部的这些城市的工资上涨，恐怕你会把它理解为是一个劳动力短缺的现象，很多人就把这个当做是一种“刘易斯拐点”出现的现象，但是在我刚才所讲的逻辑里，它其实只是体现出了在土地供应收紧的时候，房价上涨导致了生活成本上涨而出现的一个结果，恐怕并不是劳动力短缺所导致的。

这个研究里还有一个细节，就是当我想讲房价上升会进一步推升工资水平的话，这个时候房价是一个内生变量，你既可以说房价影响了工资，也可以反过来说，工资上涨引起了房价上升。为了解决逻辑上的难题，经济学里就有一个办法，我们去找一个所谓“工具变量”，这个变量是一个外生的变量，它直接影响房价，但是不会直接影响工资，如果这个变量找到的话，就可以用来比较清楚的说明由房价影响工资这样的逻辑机制，那么这个工具变量是什么？在我的2015年《中国社会科学》这篇文章里面，我就用了城市层面的土地供应。因为根据我刚才所讲的制度背景，城市层面的土地供应是由中央政府（或上级政府）所决定的，它是建设用地指标这么一个配额所决定的。

我的文章发表以后，就有读者问我，包括一些同行，说陆老师你是不是做错了，因为影响房价的显然不是土地总的供应量，而是住宅用地的供应量。这听起来似乎也是对的，但是我要告诉大家，这里面又有一个小细节，其实是跟制度背景有关的，那就是在一个城市里，它的建设用地供应总量是由中央所决定的，它基本上不能被地方政府所控制，但是给定了一个地方的土地供应的总量，其中有多少用地用在工业用途，多少用在商业或者住宅的用途，这个比例关系其实是由地方政府所决定的。所以真正满足外生条件的，其实是土地供应的总量而不是住宅用地的数量。所以，如果不明白这个制度背景，就会把一个看似合理的一个变量（比如住宅用地）当做外生变量来使用，其实又犯了一个逻辑上的错误。

 我今天给大家举的这两个例子，其实是要从做研究，还有如何做好应用性的政策研究角度给大家解释。应用性研究它首先是必须反映真实世界的，所以我强烈建议大家，为了做好研究，平常一定要养成一个习惯，那就是在看到一个数据，或者做了一个数据的结果的时候，不要在心里总想着那些具有一般意义的理论。尽管在经济学里已经形成了一个风潮，就是希望我们用国别的数据来验证所谓一般的理论，但是对于一项应用性研究，首先要求的是它要反映真实世界。

 这里就要扣这一次课程的主题了，我希望大家在做研究的时候，多多去做调研，问一问在中国真实发生的制度背景到底是什么。甚至大家也不妨问这样一个问题，把你所做的研究，称之为是印度的研究是不是也可以，哪怕你用的数据是中国的数据，是不是我也可以说，同样这个结果其实我也可以认为是在印度发生的。养成了这样一个习惯，经常问问自己，恐怕就比较容易把自己的研究更多的、有意识的建立在中国真实发生的制度背景基础之上，而不会犯那些常识性的错误，或者违背真实世界的错误。

＃5  **理论对于经验研究到底意味着什么？理论不是经验研究的遮羞布**

 在我们这一系列的讲座当中，我们讲到了经验研究，主要是用数据和事实来说话的那些研究。那么相比之下，理论模型则主要来讲一个逻辑关系，特别是对于经济学研究来讲，理论模型其实特指是用数学来写的一些模型。

 现在数学方法的应用，不仅在经济学当中非常流行，而且越来越多的在政治学、社会学这样的其他社会科学门类当中也在使用数学的工具。由于数学工具的流行，就给很多的读者和年轻学生朋友们造成了一个印象，一个研究如果离开了数学就不好。是不是那些没有数学来表述的模型的文章就不好呢？既然这样的话，我们又实际上看到了很多其实没有数学模型的文章，这又是怎么回事呢？

 我今天所讲的话题里要讲到理论的重要性，以及应用理论模型的时候出现的一些误区。关于理论模型对于实证研究的重要性，我想就不用再多说了，正好非常巧的是，美国的克拉克大学的华人学者张俊富教授最近有一篇文章，题目就叫理论模型有多重要。在这篇文章当中他讲的已经非常清楚了，哪怕是对于实证的研究、经验的研究，理论模型有的时候都是非常重要的。比如说对于有一些实证研究来说，可以用一个理论模型来讲述他研究的目的和意义。再比如说有一些非常重要的经验研究发现的现象，是基于数字和统计的。那么这时候如果有一个理论模型，可以把这些现象背后的逻辑能够有一个非常严谨的表述和解释。再比如在做实证研究的时候，我们经常要用一些回归分析技术，那么对于这样一些回归模型，函数是怎么设定的，变量是怎么选取的，如果有一个理论模型加以说明的话，那也会好很多。 最后再比如说，在社会科学研究里，我们都希望能够对不同变量之间的因果关系和机制能够有所解释，那么这个时候，数学模型也能够帮助我们更加清楚的认识到因果关系和它们的作用机制。

 在讲明了理论模型、数学模型对于经验研究的重要性后，是不是每一篇实证的研究都需要去写一个数学模型呢？这个其实也不一定，从而也就可以解释为什么哪怕是在国际、国内的顶尖杂志上发表的很多文章，其实它里面也并不一定有一个数学模型。这至少有两种情况。

 第一种就是变量和变量之间的作用机制其实非常直接。举一个例子来说，在经济学和社会科学的研究当中，有一个概念非常重要叫peer effects，翻成中文我把它叫“同群效应”，也就是说他要研究一个人的行为是不是会受到周围其他人的影响。这件事情在理论上非常简单，他主要的问题来自于人和人之间的行为是否相互影响，是不是有经验证据。所以如果要做一个经验研究去验证同群效应是否存在，或者说有多大的话，那么大多数的研究恐怕就不会去写一个数学模型把这个事情搞的特别复杂。

 第二种情况是关于有一些问题，已经有太多的理论了。有些理论说X和Y这两个变量之间是正相关的，有的理论说它们是负相关的，那么到底是正相关还是负相关，就需要通过实证研究、经验研究用数据进行检验，这个时候理论和经验研究之间的关系，就好比在物理学里面讲的理论物理和实验物理之间的关系。做实验物理的学者，并不一定要去发明一个新的物理理论，而只需要在实验室里对于既有的物理理论进行检验就可以了。在社会科学里，如果已经有的理论非常多，只需要去检验一下到底哪个更对，那么这个时候就并不一定需要去建立一个新的理论模型。

 比如说以前我自己做过关于收入差距和经济增长之间的关系，这个问题的理论已经非常多了，我自己的工作也并不是要去提供一个关于收入差距和经济增长之间关系的新的理论机制，而只是要回答一下两者之间的关系到底是正的还是负的，以及他们作用的相互渠道。而且我还用了中国的数据，这样的话就可以为我们理解在中国经济当中这两个变量之间的关系增加一个新的证据，这个时候我们也就没有去写一个理论的模型来说这件事。

 再接下来我就要回答关于理论研究和经验研究之间的关系有几个疑问，也是大家经常会存在的困惑，我谈一点自己的看法。

 首先，很多人可能误认为，如果要是有一个文章，他是既有理论，又有数据和实证研究，就一定好于没有写理论的实证研究。其实这是一个很大的误解。对于社会科学研究来讲，如果我们能够用一个实证的方法，比如说大家都知道有一些实证的方法，像工具变量法，像双差分的方法，断点估计这样的方法，都能够比较好的识别两个变量之间的因果关系，那么相比之下，如果有一篇文章他写了一个数学模型，但是在实证方面就做得不够好的话，恐怕就未见得是比纯粹的经验研究更加能够去说明变量之间的因果关系。

 第二种误解，经常认为一篇文章如果没有一个理论在文章前面，就是没有理论。这句话本身就是有问题的，其实我们经常讲理论模型，在狭义上来讲，当然是指用数学方法写的理论模型，其实在理论这个词上面，本质上来讲，它要表述的其实就是一种逻辑关系。而逻辑关系所构成的理论，并不一定要用数学来表述。比如说在我前面所讲到的同群效应这样一个例子当中，这个逻辑关系就并不一定需要借助一个数学模型写出来。而且其实哪怕对于一个纯粹的经验研究来说，模型的形式、变量的选取，其实都是依赖于理论逻辑的，不能随便的构建实证的模型，包括选取变量。并不是说没有数学模型，就是没有理论。

 第三种误解，经常有人认为，我们看一篇论文的时候，往往这篇论文里是先有一个理论模型，然后再出现经验证据去验证理论模型的逻辑机制，很多人就以为，在实际做研究的时候，都是要先写一个理论，然后再去找数据。其实在研究中往往不是这样的。一个有经验的研究者，往往在心里面先要预估一下变量之间的理论逻辑关系，这个时候未见得他要用数学去写这个逻辑关系，而是要先去看看数据，如果数据最后得到验证了，然后再用数学把这个理论逻辑写下来。

 实际研究当中经常发生的情况是，在看数据的时候，恐怕会发现，数据所呈现的结论，跟事先所想到的理论逻辑不太一致。于是我们就要去想，为什么出现这样的情况，在你想的时候，其实你的心里就是在建立另外一套替代性的理论逻辑，然后再去看数据，直到心里所想的理论逻辑和数据最后呈现的结果一致，才停下来，然后你再用数学把这个理论逻辑写下来。

 所以不能单纯的简单的理解，在做研究的时候都是一定先有理论再有经验研究的，而实际上这两者是互动的。在写数学模型这个意义上，并不一定是先要写数学，然后再去看实际的数据，但是如果我们把理论理解为是一种逻辑的话，那么一定是先有理论逻辑，再去看经验的结果的。

 最后，我想讲一个看法，有的时候还会有这样一种误解，我觉得是非常严重的，那就是有些人可能觉得当我在X和Y这样两个变量之间的逻辑关系在数据上看不太清楚的时候，我就去写一个理论模型来说明它，这样做是不是一个好的研究？我觉得不是。因为在数据上X和Y这两个变量之间的关系，其实是有可能存在多个理论逻辑的，如果你离开数据本身，就只用一个数学模型来写X和Y之间为什么会发生这样的关系，除非你这是一个重大的新的发现，或者你的理论模型本身所讲的这个机制，也是一个非常重要的新的理论机制，这个时候，这篇文章本身其实它是一个理论文章，经验研究只是一个装点。

 但是如果这些条件都不吻合，仅仅作为一个经验研究，你要讲一个逻辑，然后数据本身没有讲清楚，只是借助一个数学模型来讲的话，我看就不行了。因为就像我之前所说的那样，数据上呈现出来的两个变量之间的关系，可能有N种解释，我怎么知道你所讲的那种解释就是那一个正确的、最好的理解变量之间关系的解释呢？有没有可能存在其他的机制或者其他的渠道呢？

 在这个意义上我就用一句话来作为这一节的结尾，那就是，理论是重要的，但是对于那些迷信理论，甚至以为理论可以取代经验研究的学生和青年朋友来讲，请不要忘记，理论不是经验研究的遮羞布。

**＃6  横截面数据可以用来干什么？一大堆问题，来不及秀技术**

 我先从我自己的一项研究开始说起，这项研究是我和陈斌开、钟宁桦一起做的。

 大家知道一段时间以来，中国经济有一个非常重要的现象，那就是消费占GDP的比率一直不高，而且在相当长的一段时间里，这个比率或者说消费率还持续下降。这个现象非常重要，因为它表明了中国经济存在一些结构性失衡的一种现象，如果经济增长当中靠消费来拉动的比率不高，并且持续下降的话，那么经济增长就更多的依赖投资和出口来拉动。这不仅是一种中国经济国内结构失衡的一种表现，甚至成为中美贸易争端的一个非常重要的背景，因为中国经济一直在大量出口，形成了国际贸易的一些失衡的现象。

 那么对于这样一个非常重要的现象，如何进行解释呢？如何中国经济的消费率一直低，甚至有些年份还在下降呢？于是就需要找一些解释。那么我们自己的研究是怎么解释的呢？我们就从户籍身份和消费的关系开始入手。我们观察到，在城市里，有大量的外来人口，这些外来人口没有本地的户籍身份，他们的消费如果比一般的城市居民要来得少的话，那么，由于外来人口在持续的增长，就可以解释为什么中国经济的消费率比较低，而且持续在下降了。

 根据我们的研究，我们的确发现，在所有其他的因素都一样的情况下，一个城市的外来移民由于缺乏本地城镇户籍，他的消费就会比同样是生活在一个城市的城市本地居民要低17%-21%。而这样的人口数量有多少？大约有2.6亿左右，平均每6个中国人，就有一个属于我们今天所讲的，在城市里生活和居住，但是却没有本地城镇户籍的这样一个人群。

 这样我们就解释了中国的消费率一直很低，并且持续下降的现象。在政策上这个研究就有非常重要的意义，它成为推进中国户籍制度改革的一个非常重要的实证依据，尤其是当中国经济的内需一直不足时，我们的研究说明，如果在未来的城镇化过程中，能够持续不断的给外来人口本地户籍身份和平等的待遇，那么就可以利用这样的城镇化进程，来启动内需。在当前的国际环境日益有些恶化，中国和其他的以美国为代表的国家之间的经贸关系出现了一些争端的时候，启动内需对于未来得中国经济发展来讲，可谓意义重大。

 我讲了我们自己研究的重要性，我们的研究有没有缺陷呢？有的，其中最为重要的一点，我们的研究用了所谓的叫截面数据，截面数据的意思就是说，给定一个既定的时间点，我们所用的数据是在同样的时间点里的不同的人之间的差异，来看一个变量和另外一个变量之间的关系。但是这里就会碰到一个问题，我们想说的故事是说，没有城镇户籍的外来人口的消费要比城镇居民来得低，但是这里的户籍身份到底代表着什么呢？我们想说的是，户籍身份代表的是制度歧视，但是从逻辑上来讲，其实当一个人没有城镇户籍的时候，他很可能在其他方面，也有一些跟城镇居民之间的差异。比如说消费的偏好，这一点就很难观察到，当你观察到一个农民工朋友，他不大到饭店里吃饭的时候，也很可能是因为这个农民工他不喜欢到外面吃饭，而不是因为他想省钱。也就是说，如果我们所看到的户籍身份带来的消费差异，仅仅是因为偏好和文化的差异的话，那么即便你把户籍给了这个农民工朋友，他的消费行为也不会发生变化，那么相应来说政策的改革就没有效果。

 那么既然存在这样的问题，我们怎么办呢？如果说从技术上来解决类似这样的问题，当然有一些办法，我们学过计量经济学的人都知道，有一些办法的确可以克服截面数据模型所存在的那些问题，比如说我举两个例子。

 第一例子是用面板数据，所谓面板数据就是我们持续去跟踪一个个体，在不同的时间点里，我们能够看到同样一群人他的行为变化，具体到今天我们讲的户籍和消费的关系里，如果我们能够看到有些人今天没有户籍，而明天他就有了，我们来观察同样这个人，在其他一些因素，比如说偏好、习惯、文化都没有发生变化的时候，仅仅因为户籍身份发生了变化，我们来看一下他的消费有没有发生变化。这个时候这种户籍对于消费的影响就不是因为类似于偏好或者习惯和文化这样的因素了。但是遗憾的是，理想当中的数据在实际研究的生活当中，其实是没有的，即使有这样的面板数据，在这个面板数据里，发生了户籍身份变化的样本，也非常少，而且即便是那些发生了户籍身份变化的样本，他们的户籍身份的变化也不是随机的，而有可能是一些其他的因素导致的，这个时候你看到的户籍身份的变化，对于消费的影响恐怕也不仅仅是因为户籍发生了变化，而可能是因为其他的原因导致的，这就很遗憾了。

 还有一种可能想到的办法就是利用实验的办法，我们的文章投稿到一些杂志上的时候，的确就有审稿人建议我们用实验的方法，这里我想吐个槽。什么叫实验方法？就是如果我们有办法对一群人进行一个随机的实验，改变这个人群当中一部分人的特征，我们来看这个特征的改变以后，对我们所关心的一些现象，比如说消费，产生什么样的影响，那么这个时候由于这个实验是可控的、随机的，我们就能够识别出来我们想看到的这个因素会产生什么样的影响。

 但是我们所研究的问题，是户籍身份对于消费的影响，请大家想想看，在我们实际生活当中，存在着这样的随机的实验来改变人们的户籍身份吗？恐怕没有。那么我们可以自己去做这样一个随机实验吗？不要开玩笑了，如果我真有本事这样做的话，我就发财了，那意味着我可以随机的给一部分外来农民工发户籍，而且就算我有这样的本事，这样的研究根本过不了所谓的道德审查。我凭什么动用一个个人的权利，就给一部分人户籍身份，而另外一部分人没有这样的户籍身份，而且这个户籍身份的给和不给，要关系到一大群人一辈子的福利。

 所以我通过刚刚讲的这两个看上去能够解决问题的办法，说明了，对于一个特定的研究问题来讲，恐怕在这个问题仍然是一个非常重要的问题的时候，能够使用的数据真的就只是截面的数据。

 当然在实证研究当中，也存在一些其他的高科技手段，比如说非常流行的双差分模型，断点估计，这些东西都是非常依赖于数据所生成的制度背景。往往在一些特定的问题里也没法使用。

 那么既然这些高科技手段都没法使用，我们是不是就不要去做那些重要的问题？我认为是不行的，对于一些重要的政策问题，或者说应用问题，如果我们能够用截面的数据来做出变量之间的关系，总比我们对这样的一些重要的问题毫无认识要来得好得多。

 当然我这样说我也不希望给大家造成一种误解，那就是在做研究的时候，永远强调我做的研究非常重要，所以你不要管，我就用截面数据，我随便报告几个方程就OK了，我觉得也不是这样。如果大家有机会去看一下我们发表的户籍和消费之间的关系的文章，我们这个文章有早期版本和晚期版本两个版本。比较早期的中文版本我们发表在《经济研究》2010年，后来我们有一个用了新的数据和新的估计方法的一个版本，发表在英文的World Development上面，大家如果有机会去看一下我们的文章就会发现，我们虽然是用的截面回归的一个分析，但是我们仍然做了很大的努力，来尽量的排除是不是有可能因为其他的机制来影响我们的结果。

 我这里也举一两个例子，比如说大家都会关心,是不是我们看到的外来人口少消费的现象，是因为文化的差异？换句话说，是因为外来移民不喜欢多消费，因为他们跟城镇居民的文化不一样。如果这个逻辑是对的，我们可以想两个办法来解决，第一我们在拿外来人口跟本地城镇人口做比较的时候，我们不去比较那些出生在城市的城市居民，而去比较那些出生在农村，但是现在已经是城里户籍的这样一些居民，如果我们相信出生在很大程度上是决定了一个人的文化和习惯的话，那么将外来人口跟出生在农村，并且已经获得城镇户籍的居民进行比较，就更加可以相信这种差异是来自于户籍制度造成的差异，而不是文化差异。

 另外一个办法，我们就去观察一下那些在城市里居住时间比较长的人，如果真的是因为没有本地城镇户籍的人在文化上有差异的话，你应该相信随着时间的推移，他在城市里居住的时间越长，他越会像一个城市居民。但是我可以告诉大家，这一点在数据里是看不出来的，不管你在城市里居住多长时间，只要你是一个没有本地城镇户籍的外来人口，你的消费就是要比本地的城镇居民要来得更低。通过这样的一种比较，我们也基本上可以排除这是因为文化差异所导致的。

 所以通过今天我讲的这个例子，我给大家讲了在有一些非常重要的实证的应用研究当中，不要过于片面的追求高科技，有一大堆非常重要的问题，来不及秀技术。类似像这样的问题实在是太多了，比如说在我长期关注的其他问题，包括留守儿童的问题，包括一段时间以来，由于我们的超大城市抬高了外来孩子进入本地公立学校的门槛，造成了一些返乡的儿童。这样重要的问题，具有很强的政策含义，但是如果我们要想研究留守儿童和城市孩子之间有什么差别，我们想看一看返乡的儿童，是不是跟那些留在城市的儿童形成了一些差异，这些问题，恐怕当我们去研究它的时候，能够收集到的数据，在很大程度上就是一个截面的数据。

 在这个意义上，我觉得中国真的是有一大堆的重要的政策问题，来不及等到数据已经完全完备了再去研究，也来不及去秀我们高科技的技术。但是同时非常遗憾的是，现在我们的学术界，特别是一些年轻的学生，恐怕很多时候都首先看到了一篇文章的技术，却忽略了重要的问题。

**＃7  面板数据到底何用？有时候，重要的就是“固定不变”的**

 今天的话题是面板数据不能解决什么问题？我们先从一个例子开始讲起，现在大家都非常关心城市的经济增长，特别是对于那些欠发达地区，很多人出了很多的主意，希望对欠发达地区的发展能够起到一些促进作用。

 比如说有人提出要促进欠发达地区的投资，要给予他们优惠的政策。还有人认为领导人的能力非常重要，所以要把比较有能力的领导人安排到欠发达地区，去当当地的行政长官。还有人认为一个地方的行政级别如果高的话，可以带来很多的资源，所以把一些原来属于县的地区，如果提升为城市的话，也可以促进当地的经济发展。

 但是所有这些因素讲完了之后，大家可能忽略了一个非常重要的，决定一个地区或者一个城市经济发展潜力的重要因素，那就是地理。在我之前做过的研究里，我从两个维度来捕捉了影响一个地方经济发展的地理因素，一个就是到沿海大港口的距离，另外一个就是到区域性的中心大城市的距离。在我的研究当中，如果我们把各种影响经济发展的因素全部都放在模型例如的话，那么在影响一个地方经济的那些因素当中，大概到港口的距离和到大城市的距离，可以解释城市与城市经济增长差异的大约25%左右。而所有的解释经济增长因素全部放在一起，大概可以解释城市之间经济增长差异的50%，而剩下的一半我们就不知道是什么样的因素导致的了，也有可能是因为一些随机的度量误差的因素都是有可能的，反正我们不知道。

 那么，在我们已经可以知道的可以解释城市间增长差异的因素当中，到港口的距离和到大城市的距离这两个地理因素，解释了大约一半。这样就足以证明地理这个因素在决定一个城市的经济增长当中所起到的作用，几乎是决定性的，当然我们也不是说它是唯一决定性的，只是说它的重要性要远远远远的超过那些其他影响经济增长的因素。

 举了这个例子之后，我们再来讨论一下面板数据到底可以用来解释什么问题。首先要解释一下什么是面板数据，面板数据同时有截面的维度，也有时间的维度，比如说我们观察一些城市，如果只在一年当中去观察这些城市的话，他所形成的数据就是截面数据。但是如果我们在若干年当中，对同一群城市反复的进行观察，那么这个时候它就具有了时间的维度，就构成了一个叫“面板数据”的数据集。

 通常来讲，如果我们想在截面上看解释变量对于被解释变量的影响，可能就不能很好的去分析一个经济问题，比如说我们想知道投资对于经济增长的影响，我们看到了投资能够促进经济增长，这个时候我们是不是就能够把投资对于经济增长影响的系数，当作投资对于增长影响大小的度量呢？恐怕不一定的，因为一个地方投资多的时候，往往是由一些其他因素导致的，这个时候你就不能把投资这个变量之前的系数，完成理解为是投资对于增长的影响的大小。

 这个例子就是非常著名的所谓叫“遗漏变量的偏误”，那么如果我们要是有一个面板数据的话，我们就可以去看解释变量的变化如何影响了被解释变量的变化。仍然拿投资对于经济增长的影响作为例子，我们可以在面板数据里去看，同样的地方的投资的变化如何影响当地增长的变化，这个时候我们假设那些影响投资的因素是不随时间变化的，比如说像地理这样的因素，那么给定地理这样的因素以后，如果一个地方投资增长的快，它的经济增长率就会更高的话，那么我们就大概可以把这样的一个影响的程度，当作是投资对于经济增长的影响了。

 但是在我刚刚举的这个例子里，请千万注意一个问题，如果我们的研究问题本身是想知道类似于像地理这样的因素，对于经济增长有多少影响的时候，面板数据恐怕就无能为力了，因为地理这样的因素，通常是不随时间变化的，由于它没有变化，所以你就很难通过解释变量的变化，来看被解释变量相应发生什么样的改变。

 用经济学的术语来讲，那就是类似于像地理这样的因素，其实是不随时间变化的固定效应，而面板数据的长处，恰恰是控制住那些固定效应之后，看其他可变的因素对于被解释变量的影响。

 在这个例子当中，我提醒我的朋友们，特别是年轻的学生，不要盲目的相信某种特定的方法，而是要看到这个特定的方法是不是有助于你回答你想关心的这些问题，甚至有没有可能当你去使用一个方法的时候，你可以看到一个变量对另外一个变量的影响，但是，是不是有可能当你这样做的时候，遗漏掉了更为重要的影响因素，而那个影响因素可能在面板数据里就是不随时间变化的。

 我们再举一个例子，在经济学文献当中有一个问题叫“资源诅咒”，它的意思就是说在不同的国家，有一些国家的经济增长恐怕在长期里是不可持续的，有一类国家就有这样的情况，那就是一些资源非常丰富的国家，由于这些国家有丰富的资源，比如说石油，于是这些国家的制度建设就不是很完善，往往到现在他们仍然都是一些威权国家，还有这些国家可能腐败也非常严重。同时，由于资源实在太丰富了，大家很容易挣钱，结果这样的国家就不太重视教育和科技的发展。所以，从长期来看，这样一些国家的经济发展的表现也不太好，这就叫“资源诅咒”，表面上看起来拥有资源是件好事，但是从比较长的时间来看，这似乎又成了一件坏事。

 那么为了研究资源诅咒这样的一个故事是不是存在，于是就有人用中国年度的省级或市级面板数据来研究资源诅咒。但是请大家注意一下，在资源诅咒的这组文献当中，其实他研究的本身就是一个长期的问题，他不是要研究今年跟去年相比较而言，我这里多了一些石油资源，于是今年是不是就会比去年的制度更差，腐败更严重，更加不重视教育和科技。而是说资源诅咒这个问题，本身就是一个长期的经济发展问题，是要看至少10年，甚至几十年的时间跨度里，有些国家有比较丰富的自然资源，是不是会导致这些国家制度不好、腐败严重，以及不重视教育和科技。

 所以对于这样一个特定的经济学研究的问题来讲，恐怕面板数据也不是一个好的研究这个问题的数据来源，而是说它可能比较适合在比较长的时间里，用截面的方式来看国家和国家之间的差异，是不是资源的丰腴程度，导致了这些国家的制度差异、腐败差异等等。

 在今天我举的这两个例子里，我想告诉大家，面板数据可以解决一些问题，但是它不能解决所有的问题。我建议大家在做研究的时候，还是要回到你所要研究的问题本身。有的时候对于一些特定的问题来讲，它的重要的问题本身就是截面的差异，比如说今天所举到的资源诅咒的例子。而另外一些问题，本身重要的问题就是那些固定不变的因素，到底怎样影响到了我们关心的被解释变量，比如说地理因素如何影响到了城市的经济发展。

 所以我最后就用一句话来结束这一节，那就是，我们在研究一个问题的时候，首先要多想想，多想想，再多想想，要把自己想关心的问题和什么是重要的问题想清楚，然后再去想用什么样的数据和方法来研究我们的问题，而不是反过来让研究的方法和数据束缚我们的思想。

**＃8  交互项有什么用？看差异的差异**

 很多时候我们在讨论一些具体的现实问题的时候，就涉及到解释变量对被解释变量的影响是如何相互依赖的。我来举几个例子，比如说在城市发展当中，很多人都认为当城市长大了，特别是那些比较大的产业结构比较高端的城市，只需要高技能劳动者，而不需要低技能劳动者。那么高技能劳动者和低技能劳动者之间的关系到底是怎样的？从经济理论角度来讲，有一种说法叫“技能互补性”，也就是说，当高技能劳动力增加的时候就会带来很多对于低技能劳动者的需求，来进行生产上的辅助性的工作，这时候增加雇佣一些低技能劳动力，其实是可以提高高技能劳动力的劳动生产率的。这个时候高技能劳动力对于生产的影响，其实就取决于低技能者的数量。

 同样的道理，我刚才所说的高技能劳动力和低技能劳动力之间的相互依赖关系，其实也存在于本地劳动力和外来劳动力之间。因为往往本地劳动力平均来讲它的教育水平比较高，而外来劳动力的平均教育水平要相对来说低一点，他们所从事的工作岗位也有非常大的差异。比如说在我们的城市生活当中，大量的环卫工人、送快递的、做早餐的，其实都是外来劳动力，这时候本地劳动力的生产力或者他们的生活质量也是取决于外来劳动力的数量，换句话说本地劳动力和外来劳动力之间，恐怕也是相互依赖的关系。

 还有一个例子涉及到推迟退休年龄这样的问题，大家都知道中国现在出现了比较明显的老龄化的趋势，那么为了缓解养老金的压力，如果我们可以推迟退休年龄，这个时候就可以延长大家工作的时间，从而缓解劳动力供给不足和老龄化的问题。但是这个时候大家可能会担心一种问题，那就是推迟退休年龄是不是会导致老年人口仍然占着工作岗位，从而导致年轻人失业，对于这个问题本身又取决于，相对来说比较老年的人口和年轻的劳动力的关系，是相互替代的还是互补的。如果担心推迟退休年龄会导致失业，本质上来说，就是认为老年人和年轻人之间是相互替代的，你有工作了，我就没工作了。但其实可能还有另外一种可能性，那就是老年人和年轻人其实是互补的，一方面，老年人有经验，他们可以通过传帮带这样的做法，来提高年轻人的劳动生产率。反过来说，年轻劳动力如果从事一些辅助性的岗位，也有可能提高老年人的工作效率。那么他们到底是互相替代的还是互补的，本身就对于我们理解推迟退休年龄这样的公共政策会导致什么样的后果，具有非常重大的意义。

 对于我刚刚讲的这些例子来说，在实证研究当中，都需要使用到交互项，也就是说在方程右边的解释变量里，可能我们需要把两个决定被解释变量的影响因素进行相乘，来看这样一个交互项或者交叉项的系数是正还是负。比如说在我刚刚所讲的几个例子里，如果高低技能劳动者，本地和外来劳动者，老年人和年轻人之间是互补的，那么它们之间的交互项对于产出的影响的系数，就应该是正的，如果是互替的，它们就应该是负的。

 再比如说在我之前做的一个研究里，我研究了新城建设的问题，当时我发现，在中国大量建设了很多新城，而这些新城往往建得密度非常低，同时这些新城又建设在距离当地的老城非常远的地方。于是我们发现，如果一个地方新城建设的密度非常低，同时又建得非常远的话，那这个地方的投资效率就比较低，从而带来债务负债率比较高这样的现象。当时我就问了一个问题，如果一个新城建得离老城非常远，在这种情况下，同时如果这个新城建设的密度比较高一点，是不是可以缓解远距离对于债务这种负面的影响呢？于是我们在研究中也构造了一个密度和距离之间的交互项，这样我就可以回答是不是密度增加，可以有助于缓解距离远这样新城建设当中的负面影响，结果的确如此。

 在上面这样一些例子当中，我都讲了一些关于交互项的实际运用的例子，但是我今天所想讲的另外一个意思可能更为重要，那就是交互项的巧妙运用，有的时候能够帮我们在没有很好的因果关系识别的方法的时候，能够帮我们靠近因果识别。

 比如说在之前我举过一个例子，那就是户籍与消费之间的关系，我们当时想说的一个道理就是，在城市里居住但是却没有本地城镇户籍的外来劳动力，他会挣钱，但是相对来说消费就不会消费那么多。换句话说，外来人口这样一个户籍身份，会对消费不利，但是可能你会反问我，这种农村户籍或者说外来人口户籍，它到底是一个制度的影响还是一个文化的因素？也就是说是不是因为农村居民有某种特定的文化或者消费习惯，他进到城里来，他也不太习惯增加消费。

 这个问题其实是合理的，当时为了回答这个问题，我们就想了这样一个逻辑，如果农村户籍的身份，真的只代表文化的话，那么我们应该看到，当这些人在城市里居住的时间越来越长的时候，那么农村户籍对于消费的负面影响应该会逐渐的变小，因为它有一个在城市里逐渐适应城市文化的这样一个过程。所以在做这个研究的时候，就可以在方程的右边放入农村户籍身份乘以在城市里居住时间这样一个因素，构成一个交叉项或者交互项，来看它的系数会怎样。

 最后我再举一个例子，最近我和上海对外经贸大学的王丹利老师做了一个研究，我们想研究在中国农村教育有的时候水平比较低是由什么样的因素导致的？其实我们想讲，在有些地方可能具有这样一种传统文化，那就是当宗族的势力比较强的时候，这个地方很容易产生宗族之间的冲突，如果宗族冲突非常激烈的话，就需要进行一些武力的斗争，比如说我们会发现有一些现象叫做“械斗”。我们发现在一个地方械斗比较严重的话，这个地方就会产生武力的冲突，而大家就会少投资于教育，从而导致教育水平会比较偏低。

 但这其中会有一个问题产生了，我们所看到的械斗这样一个现象，到底是不是真的代表了宗族的冲突，还是由于一些其他的因素所导致的，比如说是不是因为在械斗比较严重的地方，同时也是资源比较贫乏的地方呢。为了回答这样的问题我们想了很多办法，其中有一个办法是跟交互项的使用有关的，那就是在逻辑上，如果械斗的确代表的是一种宗族冲突，大家可以想，这种影响主要会影响男性，因为男性是参与械斗的主力，女性不大会参与械斗的。与此同时，这种影响估计主要会产生在农村地区，而对于城市地区来讲，恐怕就不大会通过械斗来进行宗族之间的冲突，从而影响教育。所以我们就可以把械斗发生的度量指标和是否在农村地区以及个体是男性还是女性这样的变量进行交乘，从而观察这些交互项的作用，来看它是不是真正的是因为宗族冲突的作用影响了教育。

其实我讲到这里，如果一些熟悉微观实证研究方法的听众朋友就已经反应过来了，其实在我们使用交互项的时候，其实本质上就是在看difference in difference，或者说差异的差异。比如说在刚才械斗这样一个例子里，我们就是想看械斗对于教育的影响是否在农村和城市之间存在差异，是否在男性和女性之间存在差异，这个本质上其实就是一个双差分的分析。

我要说的一个意思就是，交叉项可以帮助我们解释变量之间的相互依赖性，同时也可以近似的被认为是一个双差分的分析，来接近因果的识别。

**＃9  为什么要去做异质性分析？不要相信有普适的政策**

 这节的话题是为什么要做异质性分析？我们先从一个具体的例子开始谈起，那就是开发区政策，大家都知道在中国经济改革开放以来，沿海地区，特别是东南沿海地区，获得了相对好的、比较快速度的发展，尤其是长三角和珠三角。

 相对来讲在中国的中西部，制造业的发展、经济发展的速度总体上来讲不尽如人意，于是就有了这样一种理解，认为中国的东部沿海地区，特别是东南沿海地区，他们的经济发展是因为享受了一些优惠的政策，其中就包括了开发区政策。因为你如果回顾中国早期改革开放的历史的话就知道，中国的一些经济特区，比较大的经济技术开发区等等，都首先是从沿海地区，特别是东南沿海地区开始实施的。

 既然是这样的话，是不是可以把同样的开发区政策，在广大的中西部进行实施，这样的话就可以促进中西部的发展呢，很多人就这样想的。可是如果要是你真的是这样相信的话，或者说像有一些有关开发区的研究所得到的结论那样，开发区的政策的确可以促进经济发展，甚至有的研究就直接认为开发区政策可以促进欠发达地区的经济发展，那么我们就要问一个问题了，如果你真的相信开发区政策是无条件的可以应用在所有的地方的话，那么为什么我们今天在中西部的广大地区可以看到这样的现象，那就是开发区遍地开花，——以工业园为例，现在几乎每一个县都有一个以上的工业园，——如果开发区政策、工业园政策真的是无条件有用的话，那为什么看到在中西部大量的开发区现在出现闲置呢？为什么企业不搬迁到中西部的这些开发区去，然后把开发区填满呢？

我们在之前的一项研究里就说了这样一个道理，其实开发区政策是否能够促进当地的发展，实际上是取决于一个地方的地理条件的。中国改革开放以来，特别是上世纪90年代中期以后，开始全面发展开放经济，以及出口导向的制造业，这个时候，沿海地区因为拥有大港口，在国际贸易里的国际贸易成本可以借助海运降到比较低的水平，于是又在这个基础上形成了产业的集聚效应。所以这个时候，如果你给一个当地经济发展的优惠政策，它就可以借助于集聚效应，能够更加有效的进行工业的发展。

而相反，在中国的广大中西部地区，它恰恰是远离沿海大港口的地区，运输成本比较高，同时因为当地的产业发展，没有形成集群，所以它的集聚效应也不够高。也就是说如果你把同样的开发区政策，放在远离大港口的地方，本身的产业规模又比较小，人口密度又不够高，甚至有一些中小城市，它本身就远离大的经济集聚的中心城市，交通基础设施条件也不够好，那么当你把经济的资源和政策集中在这些地方的时候，其实它就不会产生在沿海地区同样的促进经济发展的作用。换句话说，我们往往把一些政策想用来促进欠发达地区的发展，但是欠发达地区的一些地理条件、自然条件，往往又会局限这种政策的作用，这就是一个悖论。

 也是因为同样的道理，在我的研究当中，还去考察了新城建设，中国现在很多地方都建有当地的新城，希望通过城市的扩张，新城的建设搞房地产开发、基础设施建设来推动当地的经济发展。其实类似像这样的政策是否能够成功也是有条件的，在我的研究当中我告诉大家，如果一个新城建设在人口流出地，那么人口在流出，城市的建设却在扩张，这个时候这样的投资就比较无效率，甚至有可能因为大量的新城建设，依赖于地方政府的负债，结果导致增加负债的负面效果。

 而这样人口流出地又是在什么地方呢？往往又同时是在中国的中西部，特别是一些中西部的中小城市，最后给他带来了巨大的债务负担，又尤其是当当地做新城建设的时候，把新城建设的特别大，远远超过实际需求，还有建设的地方又远离当地所在的地级市的市中心的时候，这样的新城建设效率就更加低，推动经济发展的作用就更小，带来的债务负担就更大。

 在我今天举的例子里面我都提醒大家，做经济分析，特别是实证的研究的时候，要注重做一些异质性的分析，也就是说政策的效果往往是取决于其他的条件的，而在现实生活当中我们往往发现，一些自上而下推进的政策往往都有一刀切的嫌疑，往往会忽略政策实施的有效性的条件。而地方政府在学习其他地方经验的时候，也往往会照搬其他地方的政策，也会忽略自己恐怕缺乏其他地方获得成功的条件。这些现实情况就提醒我们，在做政策的应用分析的时候，特别是要多做一点思考。

 首先要做一些理论机制的分析和思考，来想一想是不是有可能在不同的样本里，不同的地区会产生同样政策效果的差异性。但是我这样说，也同时提醒大家，在做实证研究的时候，不要过于机械的分组，比如说你看现在有关区域经济的一些研究，都会对自己的样本进行沿海、内地或者东中西部的分组，可是你要进一步的问自己，在你的研究当中东中西到底代表了什么，到底它代表的是一个地理的条件、气候的条件，或者说它代表了到沿海大港口的距离，还是说它代表了人口流入还是人口流出，为什么会有这样的政策分析的差异性呢？这些问题不能不回答，不能只是简单的分分组结束了，否则你根本不知道自己这样的分组可以带来什么样的含义，以及为什么会这样。

 同样，我还会提到另外一个我们在实证研究当中碰到的情况，特别是对于很多的初学者来讲，在做研究的时候，会在起步的时候碰到自己想要的解释变量对被解释变量的影响在统计上不够显著的问题，这个时候很多人会着急。但是我想大家有的时候要冷静一下，你看到一个因素对一个被解释的因素的影响在统计上不显著，可能只是因为你看到的是一个平均效应，是所有样本的平均效应，但是其实可能你的X对于Y的影响其实只是在某一个子样本里面存在显著性，那么当你看总体上的平均效应的时候，它就可能不显著了。甚至有可能在第一个子样本里它的效应是正的，而到了另外一个样本里它的效应是负的，这个时候当你在加总的样本里去看平均效应的时候，正效应和负效应就会相互的抵消，以使得你看到的总体效应不显著。

 所以我特别提醒年轻的学生和刚刚起步的青年学者，在你们做实证研究的时候，特别是当碰到效应不显著的情况的时候，不妨看一看在不同子样本里面情况的差异，当然不要忘记事先做一些理论、逻辑的分析，为什么会可能存在这样的差异性。

好的，我用一句话来总结，那就是不要去相信会有普适的政策，尤其是在中国这样地域辽阔的大国家。

**＃10  经济的显著性为什么要评估？宰牛何用杀鸡刀**

 我们还是从一个现实的话题开始讲起。大家都知道我是研究城市相关问题的，在最近这些年，有一个跟城市相关的经济政策的讨论，引起了非常多的争论，那就是中国的一些大城市是不是人口已经太多了，所以需要管制城市的人口规模。

 我们先不讨论管制城市人口规模这件事情本身是不是符合全世界普遍存在的惯例，我们今天仅仅从研究的角度来讨论这个问题。在那些支持要对城市人口规模进行管制的那些人那里，一个非常重要的原因就是所谓“城市病”。他们认为城市人口规模增加了，就会引起很多的成本，比如说城市的拥堵和污染这样的问题，于是城市的生活质量就会下降。这个观点听起来似乎是有一点道理的，但是很遗憾的是，当我们去检索文献的时候却发现，基本上没有什么像样的研究，用数据、实证的方法来回答中国城市人口规模和拥堵、污染之间的关系这样的问题。于是我和我的团队就做了一组研究，来考察城市人口规模和拥堵、污染这样的城市病之间的关系，我们分别来说一说。

我们先来说拥堵问题。我们在自己的研究里有一个度量指标，那就是每一个城市的人均通勤时间。通过数据的估计我们发现，城市人口规模越大，人均的通勤时间的确是更加长了。但是仅仅在统计意义上这两者有正相关性，是不是意味着人口增加就真的严重的增加了城市的通勤时间呢？我们发现，这两者的相关系数其实是非常小的。根据我们的研究，如果一个城市的人口规模增加一倍，人均的通勤时间仅仅增加大约2.2分钟。而且有意思的是，这个系数不管是在中国的数据里，还是我们查的有关美国的相应的文献，都差不多。

 大家都知道通勤时间跟两个因素有关，第一个因素就是通勤的距离，比如说我们每天上班的时候，要走很远的距离，第二个因素就是拥堵，也就是说同样的一个距离上面，我们要花的时间更多，车堵在那里开不动。在大城市里，通常我们都容易理解，我们可能通勤的距离会比较长，但是拥堵这个因素需要单独进行分析。现在好在有一些大数据，比如说高德公司就公布了每一个城市的拥堵指数，所谓的拥堵指数你可以这么理解，就是在早高峰的时候，我们去测量一下一段路行驶的时间，然后我们在半夜12点的时候，路上没车的时候，我们也大概测试一下同样的距离我们需要花多少时间。这两个时间比一下，我们就知道了，一个城市在拥堵的时候所花的时间，是通畅的时候所花的时间大约多少倍，这个就叫拥堵指数。

 根据我们的研究，的确在比较大的城市，拥堵的程度是要高一点，但是高多少呢？我要告诉大家，在中国规模最大的城市，平均来看，这个拥堵指数大约是1.7多一点，而在中国规模最小的城市，平均来看拥堵指数也是1.6左右。换句话说，在中国最大的城市，平均来说，拥堵的时候在路上所花的时间是完全通畅的时候的大约1.7倍多，而在哪怕最小的城市，拥堵的时候所花的时间，也要是通畅时候的1.6倍左右。

 通过我刚刚的这个讲解，其实就要告诉大家，虽然在统计学意义上，大的城市的确有更拥堵的现象，上班的时候大家会花更长的时间，但是这个大城市和小城市之间的差别，或者说人口规模和城市拥堵之间关系的强弱，其实这个关系是非常小的。

 接下来我们再来看看人口规模和污染之间的关系，在中国的城市统计年鉴上，有八种城市的污染物，所以我们可以很容易的去看一看人口规模和污染之间的数量关系，在我们的研究中我们发现，在八个污染物当中，其中有三个是完全跟人口规模没有关系的，有四个人口规模和污染物之间的关系，是在统计上显著的，但是相关的系数其实并不大，也就是说随着人口规模的增加，污染物的排放的数量的增加程度其实没有那么大。而仅仅只有一个污染物，它跟人口规模之间是有比较明显的相关关系的，那就是生活废水。但是这一点也不奇怪，因为如果要是一个城市人口增加，生活废水的增长很小，那倒反而应该觉得很奇怪了才对，但是即便对生活废水来讲，人口规模增加一个单位所带来的生活废水的增加幅度，仅仅是0.85，它是小于一的，也就是说其实生活废水的排放是存在着规模经济性的。更有意思的是，当我们在数量估计中，去同时控制一个城市的一些经济发展相关的变量以后，那么人口规模和污染物之间的相关性全都在统计上变得不显著了。

 我今天举到的关于人口规模和拥堵、污染之间关系的实证研究，其实是想借着这样的例子告诉大家，大家在做研究的时候，不要只顾着做回归的分析，然后看到有一颗星、两颗星、三颗星所标识的统计显著性就结束了，一定要记住看一看经济的显著性。特别是对于那些跟政策相关性非常大的问题，一定要看经济的显著性。因为在你做政策建议的时候，你说要去做一个政策还是不做一个政策，其实本质上都要涉及到成本收益分析。

 就比如说在我们今天所讲到的拥堵和人口规模之间的关系，这个问题上，如果你主张通过控制人口来缓解拥堵，仅仅是基于在统计上两者有一些相关性的话，那么你就会犯一个极大的错误，因为通过控制人口所达到的减缓拥堵的效果其实是非常微弱。但是反过来，你这样去做的时候，带来的成本是非常大的，比如说当那些被你管制了的人口没有留在像北京、上海这样的大城市，离开这个城市的时候，它的收入是会有大幅度的下降，那这笔损失，其实不管是对个人还是对国家来讲，都是一个巨大的损失。与此同时，同样根据实证研究，如果在一个大城市里管制人口规模的话，那么这个大城市的生活成本就会上升，给每一个留在大城市的人，所带来的生活的不便利性也是加强的。所以这两者相比较而言，其实通过管制城市人口来缓解拥堵问题，可以说是得不偿失，而且是远远远远远远的得不偿失。

 所以我提醒大家，必须在研究的时候看一看经济的显著性，而不只是看统计意义上的显著性。提政策建议的时候，也一定要非常非常小心，要注重成本收益分析，以使得一个政策建议不仅建立在科学的基础之上，而且建立在收益和成本的比较的基础之上。

好的，最后我用一个中国的古话稍微修改一下，来形容这里我们所讲的话题，那就是“宰牛何用杀鸡刀”。

**＃11  模型的解释力在什么情况下是重要的？别“拿着鸡毛当令箭”**

 模型的解释力度在实证研究当中我们通常会称之为模型的拟合优度，在计量经济学里它是用一个指标来表现的，那就是R平方值。

 今天我们还是从一个具体的政策争论开始讲起，大家都知道我是研究关于城市方面的一些经济学问题的，在中国城市发展当中有一个问题有更多的争论，它关系到中国的城市是怎么长大，一种理解就是认为中国城市的规模长大了，主要是由一些市场的力量所导致的。

 当然与之相反也有一些学者认为，中国城市的长大，在很大程度上是因为行政力量导致的，换言之就是，如果要是城市长大是因为行政力量导致的，那么这种结果就不是最好的结果，从政策含义上来讲，可能就需要通过相反的一个反作用力，通过行政力量管制城市的规模。

 当然这里我必须要说明一下，今天我讲的话题里，所讲到的城市规模的长大还是长小，主要是讲人口规模不是在讲土地，如果说是土地的话，当然了，中国的城市规模是由行政力量决定的，这点是没有争议的，我主要讲的是人口规模。

 那么在展开和这里的话题有关的分析之前，我们先不谈研究本身的事情，先说一说简单的背景。绝大多数的朋友我相信都可能不知道一个基本的事实，那就是中国的城市这个定义，如果放在国际视野里，其实它不是一个城市而是一组城市。中国的城市有很多种级别，大到直辖市，小到地级市和县级市，它有很多级别。中国的这些级别的城市，其实如果跟国际上其他国家，主要是指欧美和日本这样的国家去相比的话，其实我们的县级市甚至比县级市还要小的地理单位，才相当于国外的城市的概念。

 而中国地级市，更不要说直辖市，严格说来都不是一个城市，而是一组城市，在这个意义上，其实我们把中国的城市人口去跟别的国家的城市人口去做比较，就犯了一个基本的错误，那就是比较对象在统计口径上和尺度上，是完全不可比的。也正是因为这样的原因，我们常常都认为中国的城市人口规模已经太大了，其实都是和中国的城市的定义跟国外都不一样有关系。

 不仅如此，还要请大家注意到，中国是一个人口大国，截至目前为止，中国的人口都是世界上的第一大国，除了印度的人口跟中国可以有点比较之外，其他世界上绝大多数国家的人口要比中国小得多。

 而从国土面积来讲，虽然中国的国土面积也比较大，但是有将近一半的国土面积其实是不宜于居住的。也就是说中国是一个人地关系比较紧张的国家，那么在这样的国家发展城市，特别是大城市，因为它的人口密度比较高，所以就可以有效的节约用地，这些是我今天讲这个话题一些知识的背景。

 我们接下来再来讲关于研究的问题。我给大家讲一个我自己的研究，大家如果去看我的一本书叫《空间的力量》里，我们研究了中国城市的土地利用效率，也就是说每平方公里上面GDP的产出，当然这里的GDP没有包括农业，仅仅包括制造业和服务业。在我们的模型里，有一个非常重要的地理变量就是一个城市到沿海大港口的距离。我反复强调，因为中国发展制造业，特别是出口导向型的制造业，所以距离港口的远近就关系到一个城市做国际贸易的成本。在我的研究当中，仅仅到大港口的距离这样一个变量，大约就可以解释城市和城市之间土地利用效率差别的大约24%左右。而所有的解释变量，大约超过10个解释变量全部放到模型里，对模型的解释力度其实大约是在50%左右。换句话说，在我们所解释的土地利用效率的城市间差别里，有一半就来自于一个变量，那就是到沿海大港口的距离，而所有其他十几个变量加在一起，也只不过就能解释另外一半。

 刚才我讲的土地利用效率还是城市和城市之间在水平之上的差别，如果我们看长期经济增长率的差别，大约有六分之一模型可以解释的部分就来自于地理的变量，其中有两个，一个是到沿海大港口的距离，另外一个是到临近的大城市的距离。在这个意义上，虽然我们不能把我所讲的东西称之为叫地理决定论，但是我想说的是，地理因素在决定中国城市经济增长的表现的时候，它是一个绝对重要的因素。

 我们讲完地理和市场的因素之外，我们再来看看行政力量的作用，有文献就把行政级别当做是一个解释城市经济增长差异的一个因素，大家如果去看文献的话就会发现，这个因素对于解释城市经济增长表现差异的贡献，如果用R平方来表示的话，它的贡献并不是很大。即便如此我们来讨论一下，如果我举一个例子，是说我要在一个省里建博物馆或者国际机场，我们把这样的公共设施建在哪里？我想几乎是毫无争议的，这样的设施可能是建在省会城市，而省会城市又恰恰是一个行政级别相对来说比较高的城市。那么在我刚刚举的例子里，建省博物馆或者建国际机场建在省会，到底是因为行政级别的效应，还是因为规模经济的效应呢？换句话说，即便我们用行政级别这样一个指标去度量行政力量，其实其中捕捉到的部分，有可能是仍然来自于市场经济的规模经济效应。

 然后我们再来讨论另外一个跟历史有关的问题，今天我们在讨论到类似于像北京这样的城市的发展的时候，我们总是说北京集聚了全国过多的资源。这句话如果放在历史上是对的，因为在1949年新中国成立之后，中央政府的确动用行政力量把全国的一些优质资源，特别是医疗和教育集中到了北京。一直到今天，北京拥有的高校的数量和质量在全国都是遥遥领先的，在这个意义上来讲，北京的发展的确得益于历史上的行政力量。但是如果给定了这样一个历史的起点，北京已经集聚了大量的高技能劳动力和创新的能力，那么之后这种资源和人口和经济活动，向北京的集中，其实在很大程度上已经是一个市场的力量。

 所以我想提醒大家，在提政策建议的时候，如果我们基于自己的实证研究，之前我讲到过系数的大小，经济的显著性这样的因素是非常重要的，而在这里，我想告诉大家，其实模型的解释度也是非常重要的。当一个实证研究有两个解释变量，它的政策含义是相反的时候，请大家要特别注意R平方，或者说这个因素的贡献度。如果你仅仅关注了一个相对来说比较次要的解释因素，根据这个来提政策建议，恐怕就非常危险了。

 做过实证研究的朋友可能会问我，R平方是不是我们做实证研究的时候应该被追求的呢？这个要看你怎么看。如果你做实证研究的目的是一个纯学术的研究，也就是说你关心的问题是你的核心解释变量的影响的大小，又尤其是当你关心的核心变量它是一个随机的、外生的变量的时候，那么这个时候你不需要太多的关注R平方这样的指标。

 有的时候虽然你关心的核心的解释变量不是一个外生的，但是你有比较好的识别方法来识别因果关系，比如说我们反复提到的工具变量法、双差分的方法、断点估计这样的一些方法，如果你有这样的估计方法的话，那么R平方对于你所关注的问题来讲，也不是很重要。

 这也是为什么在很多现在发表的一些学术论文里，R平方都低于10%，甚至低于1%都是有可能的，那就是因为你关注的问题本身仅仅集中于某一个因素的影响大小。但是我们有时候做实证研究，不光是为了发表一篇学术论文，而是为了得到一些政策含义，或者说有一些文章发表的时候是学术论文，但是我们在用这些学术论文的时候，我们所希望得到的是一些政策含义，这个时候就碰到我今天所讲的问题了，请大家关注一下R平方，或者说模型的解释度。

 特别是当不同的变量之间的政策含义，它是相反的时候，我们就要比较不同因素对于被解释变量的解释度的大小。在这个意义上，我特别提醒做实证研究，特别是希望基于自己的实证研究得到一些政策含义的学生或者青年学者，要注意我今天所讲的问题。否则的话我同样用一句中国的古话结束这个话题，有可能你认为自己是在基于实证研究提出政策建议，但是你在做的事情其实是“拿着鸡毛当令箭”。

**＃12  为什么要做实地调研？高校真的是象牙塔啊**

 有一年冬天，我在一些朋友的组织下，一起到了中国西部一个省份里面的欠发达地区去做了为期一周的调研，这次调研只是我众多的调研活动当中给我印象最深的一次，我自己的感触也非常深。

 任何的事物都是分两方面的，在我们这次调研中的主题就是政府帮助欠发达地区的政策，包括扶贫政策产生了什么样的效果。从一方面来讲是积极的，的确我们看到现在中国政府对于贫困家庭，特别是贫困家庭孩子的教育有各种各样的补贴和扶持政策，这样就在很多程度上能够保证贫困家庭的孩子不会辍学。

 但是在另一方面，我们的调研过程中也发现了一些问题，比如说在扶贫工作中有大量的投入，人力和物力的投入，比如说建档立卡。很多基层的扶贫干部，整天的工作就陷入到了填报各种各样的表格里去了，有些表格出现一点错误，就要全部推翻重新来填。结果实际上我们很多基层干部工作的精力，没有有效的发挥在真正帮助大家发展经济这方面。

 还有一些项目，由于是政府推动的，缺乏市场的信息，最后导致生产出来的产品没有市场的需求，这样的政策也没有起到有效的作用，甚至导致了当地居民的一些反感。

 也有一些政策在短期里看上去似乎是有效的，比如说把原来居住在深山老林里面的居民搬出来，搬到政府新建的一些扶贫点上，然后在附近建一个厂，这些政策在短期里的确可以帮助这些居民离开那些发展条件和居住条件非常差的地方，到能够获得收入的地方，但是长期怎么办呢？有一些扶贫点的建设，距离当地的县城，更不要说城市的中心都非常远。所以这样的新建扶贫点恐怕过不了多久，当地的居民仍然是收入最低，最贫困的居民。

 我讲的这样一些现象，是为了围绕我自己调研活动来谈谈研究。在之前也讲到过，现在随着教学的改进，大家对于实证研究各种所谓先进的方法都已经充分掌握了，然后就用这些方法去做一些所谓非常标准化的研究。比如说在我今天所谈到的这个话题里，我就连续看到一些研究，去研究那些帮助欠发达地区的政策，包括扶贫的政策起到了什么样的效果，有一些研究就用到了所谓的断点估计或者双差分的估计，这些比较流行的而且似乎是比较标准的实证研究的方法。大量的结果也得到了结论，就是帮助欠发达地区的政策，包括扶贫政策是有效的，那么这里我们就要讨论一下跟研究相关的一些问题了。

 我们的讨论不是对于方法或者对于具体问题具体研究的讨论，我们分下面四个方面来谈问题。

 首先，我觉得很多的文章没有去深入的讨论我们的政策到底是“有效的”，还是“有效率的”这样一个问题。在最近我连续跟中国非常著名的有关收入分配和贫困问题的经济学家李实教授有几次深入的交谈，我非常钦佩李老师在他的报告里就讨论了有效和有效率的问题，相应的如果用英文词就是effective还是efficient。

 我来举几个例子，比如说如果你要去评估帮助欠发达地区发展的政策，发现相对于对照组来讲，得到政策的地方经济增长高了80%，那你会说这样的政策是有效的，是effective的。但是有没有看过投入呢？如果为了获得这80%的更高的经济增长，投入增加了120%，包括投资、银行贷款、政府的债务，那么明显投入的增长速度要超过经济增长的增长速度，那么这个时候我们能说这样的政策就真的是有效了吗？除非你完全用结果导向，不考虑成本和收益的分析。

 第二，那有人会说，我们所讲的有效可能在短期里不一定有回报，但是长期会有回报，比如说基础设施建设和公共服务的提供，往往都是在长期里获得回报的。那么问题来了，长期是多长呢？我们很多政策的实施已经快20年了，如果你看到我们新建的工业园、基础设施、新城到今天都仍然是在闲置状况，而且就像我反复所讲的，那些闲置的工业园、基础设施还有新城，大量新建在人口持续流出的地方，那么你认为要多少年才能获得所谓长期的回报呢？

第三，可能有人要说了，我们不能完全考虑经济，我们还是要考虑社会收益，比如说在一些政策之下，老百姓满意了，这是一个很大的政策的社会收益。真的是这样吗？当你持这样的观点的时候，或者我们在讨论来自于别人的观点的时候，我真的想问一句，真的是这样吗？大家调研过吗？去问过老百姓的感受吗？在我的调研里我问过类似这样的问题，有的时候我得到的回答的确是正面的，更加满意了，但并不总是如此。

更重要的是，我这里需要提醒大家，普通的群众当他说他更满意的时候，他往往是跟过去相比，所以只要比过去好了，他就更满意了。但是我们学者要做什么样的事情？我们学者要讨论的是，和一个更加理想的参照系比，有没有更好的办法达到更好的政策效果，或者说让老百姓满意了，要不要更满意。如果这样思考问题你就会发现，在很多时候其实我们在不同的帮助欠发达地区发展的政策（或者说扶贫政策）当中，相对来讲，促进移民的政策是一个更好的政策。有学者曾经跟我这样开玩笑说，其实最好的扶贫政策就是帮贫困家庭买几张车票去打工就可以了，因为打工收入远远高于贫困线。

 一讲到这个问题可能有的朋友就要说，那你们恐怕调研不充分吧，现在在农村很多人都不愿意动，有的人是家里面因为有老人或者生病的家庭成员也不能动。说到这个问题，又回到我在之前所讲的一件事，就是我们谈任何问题要在边际上讨论问题，我们今天讨论问题的关键，不在于对于根本不愿意动，或者缺乏迁移能力的家庭我们应该怎么办，对于这些家庭进行补贴，我觉得是必要的。我们今天要讨论的是在边际上，如果我们换一种政策，如果我们鼓励能够移动的那些居民出门去打工，甚至在外地定居，在经济发展条件、就业机会更充分的地方长期定居，这样的政策是不是一个更加有效的扶贫和促进欠发达地区那些人口进行发展的政策。

 不仅如此，如果大家真的去做过调研，可能也会发现我们现在做的一些政策，其实也造成了一些让老百姓不满意的状况。比如说在扶贫政策里，我们通常是划一条贫困线，在这个贫困线之上你就不是贫困家庭，在这个贫困线之下就是贫困家庭。结果一条线就决定了很多其实经济状况差不了太多的家庭的命运，在扶贫线之下的家庭就得到了各种各样的补贴，结果却导致了事实上的一些不公正的结果，有些老百姓对此也颇有一些不满意。

 第四方面我们又谈回研究，如果我们真正对有一些我们研究的对象做过深入的调研的话你就会发现，其实我们在用那些所谓标准的实证研究的方法的时候，现实情况恐怕并不适用于这些所谓标准的方法。

 比如说我们想找一个对照组说那个组是没有得到扶贫政策的，然后我们去研究得到扶贫政策的这些地方，是不是能够得到扶贫政策的正面效果，但是你恐怕要知道，其实一个地区是否成为贫困县有很多人为因素，有一些地方为了能够获得来自上级政府的财政转移支付，即便他的收入水平已经脱贫了，仍然瞒报自己的收入，能够保持自己的贫困县的身份，来获得政府的补贴。这个时候是或者不是获得扶贫政策的标准，其实就不是外生的，而是人为控制的。

 道理类似，比如说当我们去评估欠发达地区的工业园政策、开发区政策是不是能够有效的起到帮助企业发展和当地经济发展作用的时候，你要知道，其实进不进工业园本身也不是随机或者外生的，非常明确的是各个地方的政府在工业园招商的时候，其实都有自己的标准。几乎所有的地方政府在工业园里，都希望招商引资的时候能够有更多的大项目，投资多税收的贡献大，单位面积的产出大，有些地方就直接把这样的一些要求写到他的招商引资的政策里。所以其实一个企业进不进工业园，本身就不是随机的或者外生的，当你看到工业园政策对企业发展有用的时候，恐怕那是因为进工业园的政策本身就被更好的企业享受到了，而这个现象往往在那些招商引资比较困难的地方，是普遍存在的。

 我讲完了今天这样的一些结合调研的感受，我特别想提醒现在的青年学生和学者，在做研究、数据分析、回归模型之前，一定要多读文献，这里的文献不光是研究性的论文，因为很多大家现在读到的研究性论文本身可能就是有问题的，是对现实情况不够了解所导致的一些可能带有错误结论的。多读文献恐怕要包括多读调研报告，多读一些新闻报道或者政府的文件。更重要的是，我希望大家能够更多的到实地去做调研，或者换句话说要用脚做学问，多跑跑，多看看，用自己所看到的，听到东西来检验自己的结论是不是正确。

 如果你做过这样的工作，恐怕你就不会非常轻易的得到某一条政策对帮助欠发达地区的穷人能够产生正面的效果这样的结论了。套用以前毛主席曾经说过的一句话，没有调查就没有发言权，我真的很担心今天我们的高校是越来越像象牙塔了。

**＃13 我们做的研究真的有政策含义吗？好吧，你说的对**

 通常我们在做实证研究的时候，在论文的最后一个部分都会谈一谈它的政策含义，但是我们讲的政策含义真的有那么正确的政策含义吗？

 我大概要讲三种情况。第一种，也是从一个例子开始讲起。大约10多年前，有一次我参加政府内部的政策咨询会，讨论上海经济的发展，就有一位当时与会的学者提出这样的政策建议，说上海应该加大创新推动经济发展这样的一些相关投入。10多年过去了，就在不久前的另外一次也是一个内部的政策讨论会里，又有一位学者在谈到面对当前的国际贸易冲突的时候，也认为应该通过推动创新来推动中国经济的持续发展。

这些建议在10多年前一直到今天一直在被提出，我当然不是说通过创新来推动中国经济的发展这个政策是错的，但是我想告诉大家的是，创新推动经济发展这一件事情可能是大家都已经知道的事情，而当前在中国可能面临的问题是，政府已经有大量的创新投入，用来支持各种企业、科研院校进行创新，但是资金的使用效率恐怕没有很好的评估。甚至有研究发现，我们一些企业为了能够争取到来自于政府的财政资金的扶持，就虚报自己的研发投入，结果成了骗钱行为了。

跟这个例子有一点类似的是，在一些研讨会上，经常看到大家有这样的一种政策建议，比如说做了一个跟教育有关的研究就说，政策含义是政府应该加大对于教育的投入。也同样是这样的道理，现在关于教育的重要性和加大教育投入，这一点我想可能没有人有异议，但是关键的问题是，在今天我们讨论教育政策的时候，讨论的是到底是什么样的人群缺少教育投入，投什么教育，是职业教育、基础教育、技能教育还是高等教育？投在哪里，是更多的投入在人口流入地还是投在人口流出地？这些问题才是当前的教育政策需要讨论的问题。

 我要举的第二种情况是要说到没有一般均衡的思想，其实一般均衡的思想就是一个全面看问题的这么一个研究问题、思考问题的思路。我也举两个最近碰到的研究的例子，第一个研究是在一次会议上，有一位年轻学者的研究发现，农村人口大量向外迁移之后，农村地区的公共品投入相应的减少了，所以他得到的政策建议就是要鼓励农村人口回到农村，减少农村公共品投入变少的趋势。另外一个是研究一些扶贫政策的效果，发现得到扶贫政策扶持的那些地方，促进了当地的产业结构转型，制造业的比重有所提升。

在这两个研究里，其实共同存在这样的问题，就是我们所提出的这样一些政策建议，到底有没有有效的改善人民的福利。比如说如果农村人口减少，是为了追求更高的收入和更好的就业机会，那么难道我们应该通过鼓励农村人口回到农村，仅仅是为了增加农村的公共品提供吗？谁从这样的政策中得益了呢？谁又可能因为这样的政策而获得损失呢？

当我们讲到政策的时候，请大家千万注意，我们不能只看到一个政策建议的直接作用，而应该考虑到它的代价。而代价又有两种，一种是直接实施这个政策所带来的成本，比如说投资本身就需要投入，大量的政府推动的一些项目，本身还需要一个庞大的公务员系统去做工作，这些都是直接成本。不仅如此，一些政策还有一些间接成本，其中最主要的间接成本就是同样的这笔钱，在用到你所推崇的这个用途的时候，那么它就没有被有效的用到别的用途上去，这实际上就是一个机会成本的概念。

 但是我们很多人在提出政策建议的时候，却忘了这一点。在涉及到农村地区人口流出的问题的时候，我们往往忘记了其实为了达到社会福利最大化，提高农民收入的政策效果，有一个非常简单的政策，就是我反复讲到的促进那些愿意外出打工的人外出打工，而不是在他们的老家增加偏离比较优势的产业投入，比如说制造业。

 第三种其实就跟我刚刚讲的缺乏一般均衡思维有关系，那就是没有从全社会福利最大化的角度来思考问题。比如说最近这些年，有一些研究讨论了高铁开通以后所产生的后果，其中我就读到过这样的研究，说高铁开通了以后，加剧了人口向中心大城市的集中，由于很多人认为这种效果叫“虹吸效应”，所以相应的就认为高铁开通可能加大了地区之间的差距。所以在政策建议部分就提出，政府对于开通高铁这样的政策要谨慎，言下之意，似乎高铁开通是使得欠发达地区的或者是中小城市的人受损了。其实大家想一想，增加地区之间的连通性，降低贸易成本，降低人员流动的成本，总的来讲是增加全社会福利的，反过来讲，如果我们不进行高铁的连通，阻碍的人口向中心大城市的集中，谁得益了呢？难道是通过人口留在中小城市，社会福利是更大的呢？

与此类似，我要稍微提一下北京大学国家发展研究院曾经做的一个关于吉林研究的吉林报告，曾经引起了全社会的广泛关注。在这个报告中有一个核心的观点，认为吉林应该发展劳动密集型的产业。对于这个报告本身有很多讨论，我也不多加评述，但是我只讲一个观点，就算你静态的、局部的看，吉林这样一个地方存在劳动力多这样一个局面，似乎看起来好像吉林就应该发展劳动密集型的产业，比如说纺织业，但是你放在全国的角度来看，把纺织业放在吉林发展，还是放在安徽的长江沿线发展，更加有利于社会福利的最大化呢？吉林的人口一定要在吉林呆着来发展劳动密集型的产业吗？还是说要在全国范围之内，把劳动密集型的产业放在最最能够有效生产，最最能够低成本出口的地方呢？这个实际上又是一个局部看问题还是最大化全社会福利的这样一个问题。

到此为止，我讲了大概有三个方面的问题，来讨论我们的研究如何和政策含义挂钩起来。在实际生活当中，其实我们还经常会碰到有这样的一些政策研究，往往是在为政策的结论找注脚，甚至可能存在先有结论再有研究这样的情况，对这种情况，我就不多加评论了，因为这本身已经超出了科学研究的范畴。

 我们做研究经常喜欢得出一些建议，所以在我讲到研究如何得到正确的政策建议的时候，我也本身想给研究者提几个方面的建议：

 第一，我们要了解社会需求，要知道政策讨论的前沿和争论点到底在哪里，以免我们提出的政策建议，看上去似乎是对的，但是其实恐怕也是大家都知道的事情。

 第二，就是我刚才所说的，我们要有全面的、一般均衡的思维方式，不能只看到局部。也许对于很多的政策目标来讲，换一个思维方式，换一个投资的方向和结构，得到的结果要更好。

 第三，就是一个研究的立场问题，我建议大家多多站在最大化全社会福利的立场上去讨论问题。当然，我这里所讲的全社会福利最大化的立场主要是指国家的福利，因为涉及到跨国的问题，在全世界范围之内讨论问题的话恐怕就很难了，因为在全世界范围之内，每一个国家都有自己的独立的政策立场，每一个国家都要最大化自己的国家利益，而全球的社会福利最大化是一个遥遥无期的事情。但是在一个国家内部讨论问题的话，就需要讨论如何最大化本国国民的全体社会福利的问题了。

＃ 结语

 我最后想讲几句话，来进行整个课程内容的总结。很多时候，我们在课程里所讲的实证研究，实际上讲的就是以现代的实证经济学的研究方法，或者说计量经济学的研究方法，来做实证研究的一些相关问题。虽然我举的很多例子都是经济学当中的一些例子，其实这些例子所得到的启示，也可以推广到社会学、政治学、人口学这样的一些其他同样运用计量经济学方法来进行研究的一些兄弟学科。

最后我想讲一句话来送给大家，那就是，我们做的这些用计量经济学研究的方法，它本身实际上做的是一个实证的研究，是一个社会科学的研究，它必须要讲究（经济学的）理论和逻辑。在这个意义上，——

我们做的是社会科学研究，而不是统计学。

**# 本文所提到的一些陆铭团队的研究如下：**

陆铭，2017，《空间的力量：地理、政治与城市发展(第2版)》，格致出版社和上海人民出版社。

Glaeser, Edward L. and Ming Lu, 2018, “Human-Capital Externalities in China,” NBER working paper 24925.

Chen, Binkai, Ming Lu, Christopher Timmins, and Kuanhu Xiang, 2019, “Spatial Misallocation: Evaluating Place-Based Policies Using a Natural Experiment in China,” NBER working paper 26148.

Lu, Ming and Danli Wang, 2019, “(From) Kungfu’s Curse: the Effect of Clan Conflicts on Education and Its Evolution,” working paper, Shanghai Jiaotong University and Shanghai University of International Business and Economics.

李杰伟、陆铭，2018，《城市人多添堵？——人口与通勤的实证研究和中美比较》，《世界经济文汇》，第6期，1-16页。

郑怡林、陆铭，2018,《大城市更不环保吗？基于规模效应和同群效应的分析》，《复旦学报》，第1期，133-143页。

常晨、陆铭，2017,《新城之殇——密度、距离与债务》，《经济学（季刊）》，第16卷第 4 期，1621-1642页。

Han, Libin and Ming Lu, 2017, “Housing Prices and Investment: An Assessment of China's Inland-Favoring Land Supply Policies,” Journal of the Asia Pacific Economy, 22(1), 106-121.

梁文泉、陆铭，2016,《后工业化时代的城市：城市规模影响服务业人力资本外部性的微观证据》，《经济研究》第12期，98-111页。

Liang, Wenquan, Ming Lu, and Hang Zhang, 2016, “Housing Prices Raise Wages: Estimating the Unexpected Effects of Land Supply Regulation in China,” Journal of Housing Economics, Vol. 33, September, pp. 70-81.

Ming Lu and Kuanhu Xiang, 2016, “Great Turning: How Has Chinese Economy Been Trapped in an Efficiency-and-Balance Tradeoff?” Asian Economic Papers, 15(1), 25-50.

陆铭、张航、梁文泉，2015,《偏向中西部的土地供应如何推升了东部工资》，《中国社会科学》第5期，59-83页。

Chen, Binkai, Ming Lu, and Ninghua Zhong, 2015, “How Urban Segregation Distorts Chinese Migrants' Consumption,” World Development, [Vol. 70](http://www.sciencedirect.com/science/journal/0305750X/70/supp/C), June, pp. 133–146.

陈斌开、陆铭、钟宁桦，2010，《户籍制约下的居民消费》，《经济研究（消费金融专辑）》，62-71页。