

寻找只有“一只手”的经济学家： 探讨经济学实证研究的方法

甘犁

西南财经大学及美国德克萨斯农工大学

李奇

清华大学及美国德克萨斯农工大学

2006年3月

1. 寻找只有“一只手”的经济学家

美国前总统杜鲁门曾对他的经济智囊们说过：“能不能给我找一个‘只有一只手’的经济学家？所有的经济学家都说：从一方面而言应该是这样，从另一方面而言则应该是那样...”在英文里，从一方面而言是“on the one hand,”直译为“从一只手而言”。从另一方面而言在英文里则是“on the other hand”，直译为“从另一只手而言”。杜鲁门总统的意思是希望经济学家们能给出一个直截了当的回答，而不要模棱两可。

在西方，这则故事通常被人们拿来开经济学家的玩笑。比如，只有在经济学这个专业，两个持相反观点的人都可以得到诺贝尔奖，甚至两个持相反观点的人可以同时得到诺贝尔经济学奖（Myrdal 及Hayek, 1974年诺贝尔奖）。¹为什么西方的经济学家给人一种不够直截了当的感觉？

从好的一方面来讲，这说明了经济学家的严谨。经常的情况是，对某一件事情或某一个政策建议，通常没有一个清晰的答案，也就没有办法给出直接了当的政策建议。从不好的一方面来讲，这说明了经济学研究的水平还达不到一些自然科学，比如说物理学，的精确程度。这里就想讲讲这个不好的一方面：什么是高水平的经济学理论？到底困难在什么地方？

我们想首先举一个中学物理学的例子：牛顿（1642-1727）的万有引力定律。这里面的过程大家都比较熟悉。牛顿坐在一棵苹果树下，一个苹果掉在牛顿的头上。牛

¹据Milton Friedman 所讲述，1974年，也就是诺贝尔经济学奖开始颁发后的第五年，瑞典知识分子们非常想将诺贝尔奖颁给Myrdal——诺贝尔奖预订在最初5年内不颁发瑞典人。然而，因为Myrdal有非常严重的左倾政治倾向，他们认为应该找一个有强烈右倾政治倾向的人来平衡。这就是为什么Hayek获了奖。见 <http://www.peterjaworski.com/Friedman>.

顿被苹果落地的现象所吸引和困惑，他因此而创立了一个万有引力的模型。但是，这个模型要成为一个好的模型，甚至成为一个定律，必须满足三个条件。第一，万有引力的模型除了可以解释苹果及其它物品为什么落地之外，还必须能够解释为什么月亮不落地。也就是说，模型能够解释我们能够观测到的已知现象。第二，万有引力模型对苹果落地的解释与其它的模型不一样，而且实验和数据证明万有引力模型是正确的，其它的模型不正确。第三，万有引力模型能够预测到我们还不知道的现象。人们根据模型的预测，观测到以前还不知的现象。历史上，海王星和冥王星的发现，就是通过万有引力模型首先推算出它们的存在，甚至它们的轨迹，然后由天文学家观测到它们的存在。

牛顿的过人之处正在于他提出了“为什么苹果会掉到地面”这个看似简单却又十分深刻的问题。事实上，长期以来，人们猜测牛顿的推理思路其实是：既然这种能使苹果落地力量能够到达树顶那么高，那它也应该能到达比树顶更远的地方。比如说，它的作用力是否能达到月亮那么高呢？

在任何科学领域，提出一个创造性的问题后，首先必须将这个问题（例如“为什么苹果会掉到地上”）转化为一个概括和抽象的问题（例如“为什么所有物体都会掉到地面上”）。更重要的则是建立起一个模型或假设用于抽象问题的研究。在牛顿地心引力理论的例子中，这个模型将引力的作用联系到物体的质量上。这个模型还能够与物体的移动速度相联系。在这样一个模型建立后，如何求解它则是大学和研究院课程学习的实质所在。

大多数实证研究都发生在一个模型或假设的建立之后。实证研究回答的问题和理论研究有所不同。在地心引力理论的例子中，这个问题则是：这个模型怎样才能被确认为是一个定理？为了验证一个模型或假设的真实性，必须有几个步骤。

首先，解释已经观察到的现象。这对任何假设或模型而言都是绝对最低的要求。所以万有引力模型必须能够解释为什么地球上所有的物体都会掉到地面。

第二点，测试不同模型的对错。为了确认某一个模型是正确的，则必须确认其它模型或假设是不正确的。举个例子，如果有其它对苹果落地的解释，那就要确认它们是错误的。在万有引力理论的例子中，我们很难去发现是否有其它模型可以解释为什么苹果会落地。如果有其它假设，那对实证研究者而言，一部分的挑战则来自于在理论上推导出对不同的模型或假设所有的不同推论。而这些不同的推论能够在实验或实际观察中得到证明。而这些不同的推论中，其中一个和多个能够在实验或实际观察中得到证实，而另外的推论则被推翻。当然，也存在所有被考虑的推论都是错误的情形。牛顿的万有引力理论指出，不是所有的物体都会掉到地上——这与物体移动的速度有关。比如说，万有引力模型包含了月亮就不会落到地面，和我们所观察到的现象是一致的。

第三点，则是模型的预见性，也是对模型的进一步测试。一个模型或假设应该能预测我们尚未知的现象或联系。事实上，这也是建立模型的主要目的之一。牛顿的万有引力理论再次成为一个最好的例子。应用它对海王星和冥王星存在的预测在随后都得到了证实。

上述几点对物理学家和其他自然科学而言，这些都是老生长谈，不值一提。但是，对经济学和其它社会科学而言，这仍然值得一提再提。因为经济学中普遍的问题是假设、学说、模型太多，而能够被证明的假设、学说、模型却太少。经济学中很少有一种假设、学说或模型能够经得满足上面这三个条件。一篇非常好的理论文章通常从一个实际的现象出发，提出一个理论模型去解释这个现象，然后用这个理论模型去解释其它已知的现象。通常，这样一篇文章也能够发表在很好的学术刊物上。但是，一个严肃的学者是不会把这种学说当作定律，因为这样的一篇文章仅仅算是满足了第一个条件。一个高水平的经济理论至少应满足以上所说的第一和第二两个条件。我们的理解是，经济学缺少的不是能够满足第一个条件的理论或模型，而是缺少满足第二个条件、更别说满足第三个条件的理论或模型。通常，对一个现象有不同的理论，而这些不同的理论有基于不同的假设。出现有“两只手”的经济学家也就不奇怪了。

2. 经济学发展的主要困难

当然，一个自然的问题是：为什么是这样的情况，难道经济学家们不如物理学家们聪明？事实是，验证一个经济学的理论通常比验证一个自然科学理论要困难得多。为了说明这个问题，我们举一个财政赤字的例子。

包括中国在内的世界上大多数国家的政府都有或多或少的财政赤字。长期而言，财政赤字减少一个国家的资本积累，从而影响一个国家长期的经济发展水平。但是，短期而言，财政赤字对经济发展的坏处却不是那么明显。实际数据中，我们可以发现高赤字通常伴随着高利率。我们可不可以说财政赤字对经济发展的坏处在于高利率呢？这里面的关键在于财政赤字会不会导致高利率。实际上，至少有三种假设或理论可以解释高赤字伴随着高利率的现象：

第一，财政赤字导致利率升高，原因是赤字减少了可以以贷款的资金。第二，高利率导致财政赤字，原因是政府必须为现有的负债付更多的利息。第三，高赤字和高利率都是由其它的原因造成的。比如，由于经济发展不好，国民总收入减少。政府税收也因此减少，从而导致财政赤字增加。另一方面，经济发展不好也导致总的投资风险增加。银行因此提高贷款利率以弥补风险。在这三种不同说法中，只有第一种说法讲财政赤字在短期对经济发展有坏处，其它两种说法都说错不在财政赤字。

这个问题的重要性不言而喻。到底哪一种理论是正确的？许多学者为此作出了很多努力，但经济学界对这个问题并没有形成一个一致意见。要验证哪种理论正确，当

然需要通过实际的数据。前面已经提到，高赤字通常伴随着高利率。但是，要搞清楚数据之间的因果关系，也就是我们通常说的“鸡生蛋、蛋生鸡”的问题，则非常困难。

以上所说的关于财政赤字和利率关系的三种理论都假设有明显的因果关系。第一种理论假设高利率是“因”，而财政赤字是“果”。第二种理论假设财政赤字是“因”，而高利率是“果”。最后，第三种理论则假设总的经济环境是“因”，而财政赤字和高利率都是“果”。

而困难在於，在经济学和其它社会科学研究中，学者们不能像物理学家或其它自然科学家一样，设计并进行可以控制的实验。经济学家们不可能给不同的国家随机地安排赤字水平，然后观察它们的利率水平；也不可能给不同的国家随机地安排利率水平，然后观察它们的赤字水平。这种不可实验性是经济学、同时也是其它社会科学发展的最大困难，也是学者们最集中努力的地方。

当然，我们这里并不是说经济学中从来没有进行过任何的实验。这里举一个例子。1985年，美国伊利诺依州进行过真正意义上的实验。大家知道在美国一个人如果失业，他将会得到一定的失业救济金。政府当然希望失业的人能尽快地找到工作。问题是如果一个人努力地去找工作，他找到工作的机会是不是会更大。这个问题看似简单，实际上也不是那么简单。原因是找工作的压力应该能让人们付出最大的努力。具体而言，这个实验是，对那些来领失业救济金的人，随机地抽出一部分人，告诉他们如果他们在11个星期之内找到工作，那么他们将会得到500美元的奖金。这个实验的最后结果是，那些被告知有可能得到奖励的人，平均花了17.0个星期左右找到了新的工作，而其它人则花了18.3个星期才找到工作。这个实验证明了人们找工作时还不够努力。如果人们更努力，找到工作的机会会更大。（详见 Woodbury and Spiegelman 1987）

另举一个例子：美国兰德公司进行的关于医疗保险的一个实验。1971年，兰德公司用美国联邦政府的资金成立了一个保险公司。这个保险公司提供四种不同类型保险计划：个人须承担0%、25%、50%及95%的医疗费用，但每年每个人总的花费不超过1000美元。总共有5809个人参加了这个实验。每一个人被随机地安排了其中一种的保险计划。这个实验的目的是研究医疗保险对健康有怎样的关系。值得指出的是，这个关系并不是想象的那么简单，我们也不能直接比较有保险的人和没有保险的人之间的健康差异。原因是经济学中常提到的样本选择的问题：一种可能的情况是有保险的人身体健康比较差，因为这样的人更容易买保险；另一种可能的情况是有保险的人身体健康比较好，因为对一个经济条件更好的人来说，他既可以支付得起保险，同时健康状况也较好。兰德公司的这个实验发现，个人承担的医疗费用的比例越低，人们越容易去使用医疗服务。但是，个人的健康却与医疗保险的选择没有多大的关系。

中国过去25年的改革开放可以说是一个巨大的实验场，也取得了举世瞩目的巨大成就。现在回想起来，有一个方面也许可以做得更好。邓小平有“摸着石头过河”的对改革的著名描述。通常许多改革的方案是先在较小范围内试点，然后再全面推广。这是一个非常科学的办法。但是，在执行时，通常都是选择一些比较容易成功的地方或单位。也就是说，样本的选择并不随机。结果当然是实验的结果可能不具备推广性。当全面推广某一个改革方案时，就会出现许多没有料想到的情况，改革的效果也不如预期的好。就好像如果姚明能淌水过得了河，而我们其它人却不一定能过得去一样。

正是因为这样的困难，经济学界对许多问题尚未形成共识。即使一种假设、学说或模型在一个严肃的经济学杂志发表，也可能只说明它满足了上述的第一个条件。一个严肃的经济学家是不会、也不应该把这样的假设、学说或模型当作一种“真理”去宣传，尤其是不应该向外行人去宣传，因为外行人比较容易被说服。如果现实或政策需要经济学家们发表自己的观点，那么一个严肃的经济学家应该是有“两只手”的经济学家，在讲清楚自己的想法的同时也说明学术界没有共识的现状。

经济学当然需要也渴望能对现实生活中的经济问题有一个清晰一致的解释和答案，就像物理学或其它自然科学一样。如果经济学中许多理论也能像牛顿的万有引力定律一样满足前述的三个条件，经济学家们也能成为只有“一只手”的学者。但是，事实上是，绝大多数经济学理论只能说满足了“第一个条件”。在这样的情况下，经济学家们有“两只手”就变得很普遍了。如果某一个经济学家对某一个问题提出了非常肯定的回答，人们倒是应该问一问：对这个问题，有没有其它的假设或解释，有没有严格的实证研究证明其它的假设或解释是错误的。

当然，在现实的经济活动中，决策者们往往需要经济学家给出一个较为肯定的方案或模型。我们当然不反对经济学家对模型与政策有自己的偏好。应该作的是明确指出个人偏好，并同时指出其它可能的模型或假设。换句话说，在大多数的情况下，因为我们对经济生活中的许多问题还没有清晰一致的解释和答案，经济学家应该是“有两只手的”学者。

3. 实证研究的主要目的和内容

当然，为了解决不能在实验室中进行实验的困难，学者们也发展了各种不同的办法。实验经济学是其中的一种。实验经济学是在实验室里让参加实验的人在模拟的经济环境中作出一些经济抉择。通常，这些抉择的好坏直接影响参加实验的人经济收益。但是，因为受实验成本的限制，个人的经济收益通常很小。模拟的经济环境或者不够实际，或者有些经济环境根本就不可能模拟。因此，实验经济学取得了一些成绩，其创始人(Vernon Smith)也获得了2002年诺贝尔经济学奖，但是，它对经济学的贡献却注定是有限的。

另一种办法是寻找“自然的实验”。这里我们列举两个例子。第一个例子是测试一个古典经济学中非常基本的关于工资的理论。这个理论认为如果劳动力的数目减少，而劳动力所用的固定资产保持不变，工资将会增加。虽然这个理论非常基本，要用数据来证实这个理论却并不容易。一方面，因为工资受到整个市场的影响，用微观(企业)的数据来测试这个理论不合适。另一方面，宏观上则很难发生劳动力变化而固定资产不变的现象——但是，历史上确实发生过至少一次：1348年发生在欧洲大陆的瘟疫。在几年时间内，欧洲大陆的人口因瘟疫减少了三分之一。因为当时主要的资产是土地，而土地则没有什么变化。那么，根据古典经济学理论，工资应该增加。事实是，瘟疫以后，工资基本加倍。(详见 Cipolla 1980)

另一个有名的例子是计算一个人的工资到底会因为多上一年学而增加多少。这个看似简单的问题，事实上却非常难回答。困难的地方在于，那些上学多的人，通常相对更聪明勤奋一些，家庭环境也更好一些。这些人即使不多上学，他们的工资也会高一些。因此，我们很难区分多上学的人多拿到的工资到底有多少是因为学校的教育，而又有多少是因为家庭环境的影响或天生的因素。我们不可能对不同的人随机地安排教育水平。但是，经济学家们发现，双胞胎们经常有不同教育水平，也有不同的收入水平。于是，经济学家假设双胞胎之间的聪明程度和家庭环境是一样的。比较双胞胎们的教育水平和收入水平的差距可以帮助我们发现到底多上一年学可以使工资增加多少。(详见 Ashenfelter and Krueger, 1994) 用“自然的实验”来验证经济理论或找出经济变量的关系，是一种非常有前途的研究方法。九十年代以来，由以普林斯顿的 Ashenfelter, Card, Angrist, 和 Krueger 劳动经济学家为代表，寻找“自然的实验”已成为经济学一个主流的研究方式。但是，由“自然”来进行的实验必然只能偶可遇见，绝大多数经济学理论是不可能通过“自然的实验”来验证的。

第三种办法，也是最常用的办法，是通过计量经济学中发展的一系列方法，去发现并验证因果关系。这就是我们常谈到的实证研究。

我们根据自己的研究经验总结出经济学实证研究的三个主要方面：(1) 测试不同模型的对错；(2) 找到经济变量中的一些关系；(3) 估计经济模型的结构。下面我们简单的讨论一下实证研究的这三个主要方面。

(1) 测试不同模型的对错

测试不同模型的对错是最为重要的、也是最为常见的实证研究。它主要由两个方面组成：第一，从模型中推导出能够用数据检验的结论。对不同的模型，必须能够推出不同的能够用数据检验的结论。第二、估计模型并检验数据是否与前述的模型的结论一致。

要从模型中推导出能够用数据检验的结论并不是一件简单的事，而要从不同的模型中推导出既不同而又能用数据检验的结论，则是一件不容易的事。通

常，一个实证研究人员对模型的理解也许要超过模型的创立者。为了说明这个问题，这里我们首先举一个海王星的发现的例子，然后再举一个经济学的例子。

历史上，海王星在被正式发现之前，曾在历史上被天文学家记录过好几次。但是，人们认为那是一颗恒星，而不是一颗行星。根据牛顿的万有引力定律，Le Verrier 于 1846 年七、八月份年计算出海王星的运行轨迹。他写了一封信给德国天文学家 Galle。Galle 于 9 月 23 日接到那封信。当天晚上，他与他的助手 Heinrich d'Arrest 根据 Le Verrier 计算出的轨迹，只用了不到 30 分钟的时间就发现了海王星。这个发现被誉为十九世纪最伟大的科学发现之一，也在科学史上留下了有趣的一个片段。在我们看来，Le Verrier 及 Galle 都是作实证研究的学者。Le Verrier 的工作完全是在牛顿的万有引力定律基础上的理论推导。而实证研究的很大一部分工作是理论方面的推导和讨论。

我们再举一个经济学的例子。根据个人破产法，一个人可以申请破产。如果得到批准，他以前的所有负债将不复存在。当然，他的绝大多数财产也被拿来偿还他的债务，一般能留下来的只是他的住宅。设立个人破产法的目的是给那些运气不好的人一个新的开始，也在某种程度上，鼓励了冒风险的企业家行为。这两方面的作用都是社会所乐见的。但是，设立个人破产法也可能会导致人们滥用该法：人们有可能在借钱消费时就准备将来宣布破产。很明显，我们现在的问题是人们是否会滥用该法。

理论上，个人破产有两个可能的假说。第一个假说被称为“不幸事件”假说——即当人们遇到不幸事件，如离婚、医疗问题、或失业时，他们宣告个人破产。第二个假说就是所谓的“策略性择时”假说。这个假说是人们获取多于必需的负债用于消费，并宣布破产。这种行为就滥用了个人破产法。

事实表明，一个人如果有比较高的负债水平就有较大可能性宣布破产。用一个简单的模型来表示：

$$\text{个人破产的可能性} = \alpha_0 + \alpha_1 * \text{负债水平} + \alpha_2 * X,$$

在此 X 是一系列影响一个人宣布破产可能性的控制变量。如果 α_1 参数是正数，则表示较高的负债导致较高的破产可能性。

然而，这个简单的模型并不能用于区分两种假说，因为这两种假说都隐含着较高的负债水平和较高的破产可能性相关联。为了测试哪个假说是正确的，我们必须从两个假说中推导出两个不同的但可以通过数据来验证的推论。在这个特定的例子中，通过对模型的理论上的推导，人们发现，这项测试主要在测试个人负债水平的外生性。该测试的基本想法是：如果“不幸事件”假

说是正确的，则负债水平受到外部不幸事件的影响，从而负债水平是外生变量。如果假说是战略性择时，则一个人决定是否申请破产和他(她)的负债水平的高低是同时发生的。在后一种情况下，负债水平是内生变量。这样，一项测试两种可能的假说，就成了测试负债水平是否是外生变量。在实践中通常用 Hausman 测试来验证负债水平的外生性。而 Hausman 测试的最大困难在于找到一组合适的外生变量。(详见 Gan and Sabarwal 2006)

当找到不同的模型或假说拥有的不同的推论之后，下一步就是找到一系列数据，然后通过计量经济学中发展的一系列方法来验证到底哪种推论是正确的。很明显，一个模型是否可被验证取决于是否有合适的数据。找到可以被验证的推论和找到合适的数据通常是相辅相成的。

因为大多数经济学模型都含有因果关系，所以在实证研究中一个至关重要也是最具挑战性的部分就是建立因果关系。让我们通过以下简单的线性模型把讨论具体化：

$$y = X \beta + \varepsilon$$

在此，X 是一系列影响结果变量 y 的变量。 β 是我们感兴趣 是参数。为了在没有附加信息的条件下估计这个模型，我们需要一个非常重要的前提假设就是 X 和 ε 的 协方差为零：

$$\text{Cov}(X, \varepsilon) = 0,$$

如果满足上述的条件，则X就是外生变量；相反，X就是内生变量。在绝大多数实证研究的文章中，讨论X变量的外生性是不可或缺的一步。事实上，如果一篇实证研究的文章不讨论X变量的外生性，或者对外生性的讨论不够使人信服，那么，这篇文章很难通得过匿名评稿人。

当X是内生变量时，则必须存在一组外生变量Z（也就是所谓的工具变量，the instrumental variables），满足条件 $\text{Cov}(X, Z) \neq 0$ 和 $\text{Cov}(Z, \varepsilon) = 0$ 。在这两个条件中，第二个条件实际上极难通过数学的办法验证。前述的寻找“自然的实验”，就是在寻找令人信服的工具变量Z满足 $\text{Cov}(Z, \varepsilon) = 0$ 。

当工具变量的外生性不再受到怀疑以后，另一个”弱工具变量问题”（the Weak Instrument Problem）则变得明显起来。用数学来表述，也就是 $\text{Cov}(X, Z) \approx 0$ 。如果此条件满足，建立在工具变量基础上的模型估计方法就不再适用。当此问题在九十年代中期被发现以后，它迅速变成计量经济学理论研究的中心问题之一。根据Stock, Wright and Yogo(2002)的综述文章，目前，文献中对”弱工具变量问题”有以下几点认识：

(a) 如果工具变量较弱，则基于GMM方法(包括线性及非线性的回归分析)得到的参数估计一般是非正态分布，并且标准的假设检验和置信区间都是不可靠的。

(b) 实证研究者们经常会面对”弱工具变量问题”。寻找外生工具变量是一项困难的工作。同时，一个工具变量的外生性越好，有时这工具变量也越弱。比如，有时研究人员会用过去发生的事件来作工具变量。如果事件发生的时间越久，工具变量的外生性就越好。但是，工具变量也同时会更弱。

(c) 弱工具变量问题不是一个小样本的问题。根据 Bound, Jaeger, and Baker (1995)，当样本含量增加到 329000 时——可不是一个小样本含量——弱工具变量问题仍然存在。

(d) 有比传统的 GMM 更好的方法来解决弱工具变量问题。

(2) 找到经济变量中的一些关系

实证研究的第二个研究方面是建立经济变量之间的关系。任何人都可以通过运用回归分析来找到一个经济变量和另一些经济变量的关系。在我们看来，只是找出一个经济变量和其他经济变量的关系是不太会激起研究者的兴趣的，因为它并不能系统地帮助我们理解个人和组织的经济行为。

在寻找经济变量之间的关系中，最关键的部分是确立需要寻找的关系的重要性。在通常情况下，一种关系是否有意义来自于对经济学理论的讨论。

一般来说，这一部分的研究可以分为两个类别：

a. 获得某一种关系的精确量值。

例 1：教育回报。这个问题讨论的是每增加一年的教育能够增加多少工资。从问题本身来讲，它是一个很有意思的问题，因为该数据有利于我们理解人们如何选择受教育年限。然而，此项数据比其他经济关系更吸引人之处在于它是一个“自然实验”文献的陈列柜。前面提到用双胞胎来作“自然的试验”。研究人员还用过出生的日子，家与社区大学的距离等。

例 2：劳动力供给的工资弹性。即工资每增加 1%会导致多少劳动力供给的增加。此项数据在评价税收政策时是非常重要的。比方说，根据主张裁减税收的美国总统如里根、布什等，一项重要的讨论就是裁

减税收后，人们的税后收入会增加。这样一来，人们愿意延长工作时间，同时会生产更多的产品。整体税收收入也是有可能增长的。这种现象被称为 Laffer 曲线。这项方案提供了一条无痛苦的减税途径：即个人因低税率而受益；而政府的税收也因税收总量的提高而受益。里根将这项方案推广，布什同样用这项理由作为他 2000 和 2003 年裁减税收的基础。

不难看出，这项方案取决于劳动力供给弹性对税后工资的敏感性。如果劳动力供给弹性强，则此项方案就可行。很多文献表明对于男性来说，劳动力供给是非弹性的。也就是说，不论他们获得多少工资，他们都会全时工作。对于女性来讲，情况则是很不同的。大量且较复杂的文献来研究女性劳动力供给弹性。到目前为止也没有一个统一的意见到底女性劳动力供给弹性是多少。

以上两个例子中有一件事情是一致的：如果研究者在论文中估计一些参数的数值，这些数值则是非常重要的。

- b. 发现两个变量间的关系。这种类型的研究实际上非常困难：已发现的关系必须是令人惊讶的。

例 1：一个最好的例子就是 Kuznet 关于消费函数实证研究的发现。根据 Kuznet 的消费函数，高收入将会导致：(1) 个人之间 消费水平的增高和 (2) 整体收入用于消费部分的减少。逐渐地，由于平均收入水平的显著增长，我们可以预料到人们将收入用于消费的部分会降低。然而，Kuznet 发现收入中用于消费的部分并没有改变。这是一个非常重要的实证发现，它直接导致了 Friedman 关于永久收入假说和 Modigliani 的生命周期假说的发现。Kuznet, Friedman 和 Modigliani 都获得了诺贝尔奖。其中，Friedman 和 Modigliani 关于此问题的贡献是他们获得诺贝尔奖的重要原因。

例 2：解释国际贸易

(i) 国际贸易中的相对优势理论：每个国家应出口那些在自给自足经济下相对低成本的货物。基于相对不同的自有的资源结构, Hechscher-Ohlin 模型成为教科书和学术论文公认的基本模型。在此，每个国家都利用它相对密集和相对丰富的因素去生产具有相对优势的产品。

(ii) 然而，自六十年代起，一系列实证论文表明最快的贸易增长存在于具有非常相似的天赋因素的工业国家之间。欧洲经济共同体将并不是将自有资源结构互补的国家组织在一起。很多此类的贸易扩张并

没有引发各国内部的分配冲突。最后，在很多新兴工业中，任何国家都很难清楚地指出相对优势。许多生产和贸易模式的形成只是机遇问题；实际上，存在大量的相似产品之间的双边贸易，如汽车。

(iii) 这些实证研究非常重要，因为它们就现有的 Heckscher-Ohlin 理论提出了重要的挑战。它们直接导致了一系列新的关于国际贸易的理论的兴起。这些新国际贸易理论被称作“新国际贸易”(New International Trade)，Paul Krugman 是这个领域的先驱者。

例 3：房屋交易量和房屋价格成正比关系。当房屋价格降低时，交易量同样降低。实际上，当房屋价格降低 10%，交易量将降低 40%。为什么这个关系使研究者感兴趣呢？因为新古典住房理论表明房屋价格和股票价格相似，是房屋所能提供的服务的贴现值。到目前为止，我们还不能肯定为什么当价格降低时交易量减少的原因，但是已经有几篇论文试着去解释这个原因。(详见 Stein 1995)

从前面所讲的例子来看，很明显如果新发现的一种关系和我们的共同观点或直觉，尤其是众所周知的共同观点或直觉相悖，这种关系就十分容易引起研究者的兴趣。但是，为了理解在文献中的共同观点或直觉，我们必须全面掌握该文献，并且对理论有深刻的认识。这并不是一个能够容易达到的要求。

目前，文献中寻找变量之间的关系已经不再局限于线性关系。非参数办法(non-parametric analysis) 已经成为实证研究的一个常用的方法。Li and Racine (2006) 是最新出版的一本关于这类研究方法的总结。

(3) 估计经济模型的结构

这种方法运用经济模型并估计该模型的参数，包括两大类的模型：一类是研究多个人或多个组织的相互竞争，集中在工业组织 (Industrial Organization) 这个领域。另一类则是研究有代表性的一个人或一个组织的最优决策者时，该人或该组织如何同时考虑将来和现在。我们许多的抉择，包括上学、工作、结婚、小孩、退休等，都同时对现在和将来有重要的影响。基本上，对这样的问题，我们目前只有一种模型，也就是动态优化模型 (Inter-temporal Choice Model)。我们现在的问题不是判断哪个模型是正确的，而是我们只有一个可靠的模型，我们所需要做的是确定该模型有多准确。

例 1：Hurd (1989) 研究了遗赠动机问题。他研究的背景是在美国有很大一部分财富(根据一些估计，达到财富的 40%) 是通过一代一代的遗赠。这是

否表明人们由于爱惜他们的下一代而自愿放弃自己一部分的消费？答案是：不一定。因为固定年金 (Annuity) 市场不够发达，也因为老年人面临不确定的死亡。虽然老年人知道他们将在一定的年龄去世，但是，他们也有可能活得很长寿。为了能够有足够的钱来支撑他们可能的很长的寿命（尽管活得很长的可能性不大），他们不会肆意消费而是把钱存起来。

理解是否有遗赠动机是很重要的，因为它影响到国家的储蓄。如果固定年金市场完善地建立了，人们将不再需要为可能的很长寿而过分储蓄，这样国家储蓄率将会降低。

定量估计遗赠动机却是十分困难的。我们无法观测到遗赠动机，也无法简单地相信通过问卷调查而得到的人们的遗赠动机。一种办法是我们得到一段时间内人们储蓄和消费量，并且将那些有孩子的人的储蓄和消费量和那些没有孩子的人做比较。用于预测储蓄和消费的模型被称作生命周期模型。

在此领域另一个重要的论文是Keane和Wolpin (1997) 的关于青年男性16岁以后的教育和职业选择。这些青年男性面临继续学习、找工作和找什么样的工作的选择。通过选择一种而放弃另一种决定了他不同的途径。怎样去分析这类问题呢？传统的回归分析似乎并不合适。Keane和Wolpin (1997) 运用由Gourieroux, Monfort, and Renault (1993) 发展的一套方法分析此问题。

这种方法被称作间接估计法 (Indirect Inference)。对于任何一组给定的结构参数，研究人员能够运用模拟的方法预测该模型的结果。典型的统计估计法找到一组参数使模型的预测结果和观测结果相符。然而，当可能发生的结果的数量较大时，因为存在高维数值积分问题，预测每种结果发生的可能性从数值计算上将变得非常难，或者在现有的计算机技术下变得不可能。间接估计法首先选择一个简单的描述性的统计模型。该方法不需要直接地使预测结果和观测结果相符，而是间接地使从模拟数据得来的描述模型的系数和观测数据得来的系数相一致。通过这样做，可以避免结果预测可能性所需要的高维数值积分问题。这种方法从很大程度上减少了估计模型的计算量。即使利用现代计算机，估计一个模型，如Keane和Wolpin (1997) 仍然要几个星期甚至几个月的时间。

4. 结论

我们认为，现代经济学不缺少理论，缺乏的是对理论模型的验证。自然科学通常可以通过实验去验证它们的理论模型，而经济学却很难做得到。因此，经济学中对模型或假设的验证必须用现有的可观测得到的数据来实现。这就是实证研究。

实证研究的主要目的是为经济理论服务。其主要的研究内容是去验证不同的经济模型和假设的对错。

因为数据主要出自于现实，数据之间的关系，尤其是它们之间的因果关系，往往非常困难。因此，仔细考虑因果关系是实证研究中必不可少的一个环节。

经济学的这种不可试验性，导致了实证研究通常成为经济学发展的瓶颈。因此，我们经常发现自己处在对经济问题不能给出清晰回答的境地，这也是经济学家们通常都有“两只手”的原因所在。

参考文献：

- Ashenfelter, O. and Krueger, A. (1994). “Estimating the Returns to Schooling Using a New Sample of Twins,” American Economic Review, 84, pp. 1157-1173.
- Bound, J, D. A. Jaeger, and R. Baker, (1999). “Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variables Is Weak,” Journal of the American Statistical Association, 90, 443-460
- Cipolla, Carlo M. (1980). “Before the Industrial Revolution: European Society and Economy, 1000-1700”, 2nd edition. (New York: Norton, 1980), 200-202.
- Gan, Li and Tarun and Sabarwal (2006). “A Simple Test of Adverse Events and Strategic Timing Theories of Consumer Bankruptcy.” “NBER Working Paper 11763.
- Gourieroux, C., A. Monfort, and E. Renault (1993), “Indirect Inference,” Journal of Applied Econometrics 8: S85-S118.
- Hurd, Michael (1989), “Mortality Risk and Bequests,” Econometrica, 57 (July, 1989): 779-813
- Keane, Michael P., and K. Wolpin (1997), “The Career Decisions of Young Men,” Journal of Political Economy 105: 473-522.
- Li, Qi, and Jeff Racine (2006), “Nonparametric Econometrics: Theory and Practice.” “Princeton University Press.
- Stein, Jeremy (1995), “Prices and Trading Volume in the Housing Market: A Model with Down-Payment Effects.” The Quarterly Journal of Economics, 110(2) (May, 1995): 379-406.
- Stock, James, Jonathan Wright, and Motohiro Yogo, “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments.” Journal of Business and Economic Statistics Vol 20, Number 4.
- Woodbury, Stephen A. and Robert G. Spiegelman (1987). “Bonuses to Workers and Employers to Reduce Unemployment: Randomized Trials in Illinois,” American Economic Review 77 (September 1987): 513-530.