



朱小黄：不确定性与数据重构



文/意见领袖专栏作家 朱小黄



在异度均衡理论的研究中，需要深入讨论不确定性在人类文明秩序尤其是在经济生活中所具有的意义。随着这一研究的深入，引起了笔者对数据、经济计量模型的逻辑与运用的思考，并得出了在不确定性认知框架下，现有数据必须重构的结果，本文试图加以阐述。

一、深刻认识世界的不确定性本质

到目前为止，我们对不确定性的认识还有待深入，因为人类整体上还是生活在偶然性之中的。经济研究、风险研究等多方面的研究，主要应是寻找各种偶然性的成因条件或可能的后果。不确定性是风险管理的理论来源，现代风险管理主要是对不确定性，即赖特所说的可计量的不确定性进行管理，但如果提升到不确定性的层面上来看的话，不可计量的不确定性在广义上是否也是风险管理的应该考虑的问题。

看不见的手作为现代经济学的起点，它的哲学背景其实是基于不确定性的，是通过市场价格让市场自发地形成均衡状态，而不是通过自以为是的行政手段。其实计划经济的弊端根本上就在于它是以确定性为前提而设计的经济运行规则。

世界的本质为什么是不确定性的呢？我们知道这个世界的本源是无序的，热力学第二定律，就是熵增定律，说明所有的事情都是在熵增，熵增就是一个无序的结果。无序的物理世界决定了世界的不确定性本质。所谓人类文明，就是人类通过自律，建立秩序、建立规则、产生知识。通过规则的建立和遵循来减少熵增，本质上就是人类在对抗不确定性方面的努力。

除了物理学上的依据之外，人性的差异也是一个重要的不确定性的来源。世界上没有两个相同的人，就像没有两片相同的树叶一样。人性的差异导致了人类行为的高度不确定性。人性的自负是人性差异的深层根源，而自负是基因所决定的。

人类对宇宙自然的认知总的来看未知远远大于已知。由于宇宙与自然的无穷，人类的认知能力永远只是浩瀚之一角，对人类而言，已知的越多，则未知的越大。这就在客观上决定了人类面临着未知所挟带的不确定性。人类认知的能力永远是有限的，人类文明的规则边界也永远是有限的，因此人的行为的不确定性永远存在。

传统知识的传承和运用的偏差带来的不确定性也是惊人的。知识的传承是有偏差的，离世界的本源永远会存在或大或小的偏差度。《人类进化史》的作者也讲到，人类变成现在的这个样子，找不到必然的道理和根据，都是偶然性的产物。这种偏差可以体现在概率论的计算公式中。任何事情发生的概率，其实质就是与常识判断的偏差。在现代工业文明社会，由于知识与技术运用越来越广泛，认知偏差愈来愈多，事物发生偏差的概率也越来越大。

在不确定性的框架下，我们应该清醒地意识到：人类其实生活在偶然性之中。必然性的自由王国是可望不可及的。正如某位经济学家所言：身处偶然性之中的人类却在梦想过着必然的生活。这是一个方向性的认知偏差。

不确定性的分布方式和形态是多样化的，其中许多事物的变化具有较长的时间跨度，具有相对的稳定性，这就为人类生存和文明发展提供了窗口，产生了各种自然科学并积累成知识。而另一些事物尤其是人的行为，则是偶然性的，如果失去秩序的约束，走向无序才是本质。

二、不确定性框架下的数据分类和重构

人类的行为产生数据，人们的决策依赖于数据，数据的重要性不言而喻。在实际经济生活中，运用数据观察现状、预测未来，是宏观经济、微观经济和各种交易活动中的常态。各种数据分析、数据样本的确定，经济模型中的变量和常量的选择和计算，都基于数据。

近年来数字经济蓬勃发展，数据应用的广度和深度在不断拓展，导致数据及其应用模型的风险已成为影响世界的重要风险现象。因此，数据的使用理应慎之又慎，且应建立在坚实的理论认知基础上。但是当前数据使用过程中，尤其在计量模型的运用上，拿来主义颇多，缺乏对数据根源和性质归属的底层思考，在实际经济计量中，造成了两类问题，一种是历史数据和边际数据混杂在一起，将许多偶然性的、不可重复（即对未来并无表达意义）的数据用来建立模型预测未来；另一种是将许多自然界产生的可重复数据应用到预测人类行为上，作为经济预测模型、金融风险模型和智能模型的样本，产生风马牛不相及的问题。这两类问题愈演愈烈，导致了大量的模型失真或算力浪费现象的产生。由于世界的本质是不确定性的，在不确定性的框架下，万事万物具有偶然性的特征，也有相对确定性的特征，在自然科学领域，由于相对稳定的自然环境，产生了自然科学的诸多文明和算法，所以自然科学产生的数据基本上是可重复的、可验证的。这类数据对未来具有指导性，而人文领域中的许多现象，都与人的行为有关，具有偶然性的特征，除了人类文明规则指导下的行为，其他的行为都很难

重复和验证。基于这样的不确定性判断，我们将人类全部可以采集到的数据划分为历史数据和边际数据（即终端数据、实时数据），这两类数据都混杂着可重复的数据和不可重复的数据，即必然性数据和偶然性数据。前者对观测未来无效，后者对观察现状有用。厘清从客观世界到数据采集的过程，并在此基础上按照不确定性的原理对数据进行重构是当务之急。

物质领域里的规则相对稳定、时间较长，例如岩石风化、火山爆发，这就给像牛顿定律这样的规律留下了时间窗口，在这个时间窗口中，实验是可重复的，数据就是可重复的。而在人文科学领域则不然，很多行为是无法重复的，既然是不能重复的，那么凭什么拿这些不可重复的数据来建立模型预测未来？因此，现有的数据要重新定义、分类，只把那些在不确定性前提下具有重复空间的数据拿来使用，才能指导未来。

经济社会每个时期会有当时的文明秩序，而这个秩序在短期之内不会改变，在这种时期内的数据是可以重复，可以观测未来的。这种相对确定下的数据我们可以视为必然性数据，是可以重复验证的。但是，很多不可重复、不可实验的数据拿来建模、提炼因子、计算未来，是不合理的，这

预览已结束，完整报告链接和二维码如下：

https://www.yunbaogao.cn/report/index/report?reportId=1_44083

