

ChatGPT 与大模型将对经济学研究范式 产生什么影响?

洪永淼^{1,2}, 汪寿阳^{2,3}

(1. 中国科学院数学与系统科学研究院, 北京 100190; 2. 中国科学院大学经济与管理学院, 数字经济监测预测预警与政策仿真教育部哲学社会科学实验室, 北京 100190; 3. 中国科学院预测科学研究中心, 北京 100190)

摘要 以 ChatGPT 为代表的大模型是通用生成式人工智能技术的一大突破, 不但对人类生产方式、生活方式与思维方式产生深刻影响, 也正在推动经济学研究范式的变革. 大模型是大数据与人工智能催生的一种新的系统分析方法, 适合于研究复杂人类经济社会系统. 本文首先讨论 ChatGPT 与大模型的主要特征以及发展范式, 特别是大模型如何破解“维数灾难”的方法论原理, 然后详细探讨 ChatGPT 与大模型将如何影响经济学研究范式, 包括从理性经济人假设到“人机结合”的人工智能经济人, 从孤立经济人假设到其行为可测度的社会经济人, 从宏观经济学和微观经济学的分离到两者的融合, 从定性分析和定量分析的对立到两者的统一, 从长期流行的经济学研究“小模型”范式到“大模型”范式, 以及计算机算法作为经济学日益重要的研究范式与研究方法等. 最后, 指出包括大模型在内的人工智能技术作为经济学乃至社会科学的研究方法的局限性.

关键词 大数据; 大模型; 研究范式; 维数灾难; 因果关系; 实证研究

收稿日期: 2023-12-09

基金项目: 国家自然科学基金“计量建模与经济政策研究”基础科学中心 (71988101)

Supported by National Natural Science Foundation of China (71988101)

作者简介: 洪永淼, 中国科学院数学与系统科学研究院, 中国科学院大学经济与管理学院, 中国科学院大学数字经济监测预测预警与政策仿真教育部哲学社会科学实验室, 研究方向: 计量经济学、时间序列分析、金融计量学、统计学, E-mail: ymhong@amss.ac.cn; 通信作者: 汪寿阳, 中国科学院预测科学研究中心, 中国科学院大学经济与管理学院, 中国科学院大学数字经济监测预测预警与政策仿真教育部哲学社会科学实验室, 研究方向: 金融系统工程、经济分析与预测, E-mail: sywang@amss.ac.cn.

本文的一些观点曾在 2023 年春季首届中国数字经济发展和治理学术年会、2023 大数据与统计科学交叉学术论坛、第十五届中国决策科学学术年会、第七届中国系统科学大会、第六届中国人文社会科学评价日暨期刊评价主任委员年会、大连海事大学 2023 经济与社会管理高端论坛、厦门大学公共事务学院与《公共管理学报》“AI 驱动的社会科学研究与公共治理新范式的构建”高端学术论坛、湖南大学经世致用讲坛、北京理工大学管理与经济学院明理讲坛、华中科技大学经济学院、广东工业大学经济学院、江西财经大学统计学院、江西师范大学财政金融学院、暨南大学经济学院、中国工商银行博士后科研工作站、中国光大银行等宣读。作者感谢与会者以及陈金灿、陈丽纯、陈茜玲、陈忠、程兵、董丰、李依桐、吕正浩、史九领、徐坚皓、徐卫超、薛润坡、郑挺国、张齐圣、周瑾等的建议与帮助。

How Will ChatGPT and Large Models Influence the Research Paradigm in Economics?

HONG Yongmiao^{1,2}, WANG Shouyang^{2,3}

- (1. Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China; 2. School of Economics and Management, and MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation, University of Chinese Academy of Sciences, Beijing 100190, China; 3. Center for Forecasting Science, Chinese Academy of Sciences, Beijing 100190, China)

Abstract Large models, exemplified by ChatGPT, represent a significant breakthrough in general generative artificial intelligence technology. Their far-reaching implications extend into diverse facets of human production, lifestyle, and cognitive processes, prompting a transformative paradigm shift in the realm of economic research. Originating from the convergence of big data and artificial intelligence, these large models introduce a novel approach to systemic analysis, particularly adept at scrutinizing intricate human economic and social systems. We first discuss the fundamental characteristics and development paradigms of ChatGPT and large models, focusing on how these models effectively tackle the methodological challenges posed by the “curse of dimensionality”. We then delve into how ChatGPT and large models will influence the paradigm of economic research. This includes a shift from the assumption of the rational economic man to an AI-driven “human-machine hybrid” economic agent, from the isolated economic individual to the socio-economic individual whose behaviors are measurable, from the separation of macroeconomics and microeconomics to their integration, from the separation of qualitative and quantitative analysis to their unification, and from the long-dominant “small-model” paradigm to a “large-model” paradigm in economic research. We also cover the increasing significance of computer algorithms as a prominent research paradigm and method in economics. Finally, we point out the limitations inherent in artificial intelligence technologies, including large models, when employed as a research method in economics and the broader social sciences.

Keywords big data; large models; research paradigm; curse of dimensionality; causality; empirical research

1 引言

随着人工智能等数字技术的迅速发展和广泛应用,人类的生产方式、生活方式正趋向智能化,人们足不出户便可以从事各种生产活动和消费活动.这种变化不仅影响经济主体特别是生产者、投资者、消费者以及政府的行为方式,也改变了宏观经济的运行模式,包括生产力的提高和生产关系的重塑(洪永淼和史九领,2024).最近人工智能取得了一系列突破性进展,尤其是以 ChatGPT 为代表的大语言模型的发展范式与广泛应用,正在深刻影响经济学研究范式的变革.

人类与自然界最根本的一个区别是人类具有自我意识和思维能力,包括形象思维和理性思维.思维是指人类接受、处理信息,在人脑中形成印象和观念,并进行推理和决策的心理认

知过程. 在大数据时代, 大数据可以视为人类经济社会现实活动在高维数据空间的一个映射, 大数据因此为人类认识世界、改造世界提供了一种新的手段与方式. 人工智能是基于大数据学习、模仿人类认知过程而自动实现认识世界和改造世界的计算机方法. 在人工智能领域, 机器学习的基本概念和思想早在 20 世纪 50 年代就已提出, 计算机算法与统计学的非参数方法在很大程度上是类似的. 但由于数据资源和算力算法等计算资源的限制, 这些方法在相当长时间里并没有得到广泛的应用. 大数据和人工智能紧密相连. 数据先于算法存在, 数据使机器智能化, 而智能化系统在实际应用中产生更多的新数据, 推动机器更加智能化.

大数据革命催生了一个新兴的学科, 即计算社会科学 (Lazer et al., 2009; Giles, 2012), 其最重要特征是数据驱动的研究范式, 旨在从人类经济社会活动产生的海量大数据中揭示人类经济社会复杂系统的本质特征与发展规律. 本文探讨以 ChatGPT 为代表的人工智能大模型的诞生、发展与应用将如何改变经济学与计量经济学的研究范式. 这里之所以将经济学与计量经济学并列, 是因为实证研究是现代经济学最主要的研究范式 (Angrist et al., 2017), 而计量经济学是经济学实证研究最主要的方法论. 随着数据可获得性的提升以及计算机技术的发展, 经济学研究在过去 40 多年里发生了“实证革命”的范式变革, 从原来以基于偏好、技术、禀赋、制度、行为等方面的基本假设的演绎推理为主的研究范式, 转变为以数据为基础、运用计量经济学等方法推断经济因果关系的研究范式, 这种研究范式就是实证研究. 传统的案例分析也属于实证研究, 但基于数据、运用计量经济学推断方法的实证研究, 其科学性和严谨性得到了显著提升. 大数据革命强化了经济学的实证研究范式 (洪永淼和汪寿阳, 2021a, 2021b).

作为经济学实证研究的主要方法论, 计量经济学的很多方法 (如因果推断方法) 已广泛应用于社会科学的很多其他领域, 如管理学、金融学、社会学、心理学、政治学、教育学、历史学等. 诺贝尔经济学奖获得者乔舒亚·安格里斯特 (Joshua Angrist) 认为, “应用计量经济学所考虑的问题和其他社会科学或者流行病学所考虑的问题并无本质上的区别. ……任何希望从数据中得到有用推断的人都可称为应用计量经济学家” (安格里斯特和皮施克, 2012, 第 2 页).

最近, 以 ChatGPT 为代表的通用生成式大模型技术的突破性进展, 正在深刻重塑人类生产方式、生活方式、思维方式以及社会治理方式, 也在深刻改变经济学的研究范式. 长期以来, 构建经济理论的基本方法论是经济学建模, 即在一系列基本假设基础上, 借助数学等逻辑工具进行演绎推理, 研究少数关键的经济变量如何影响经济主体的行为与宏观经济的运行; 与此相对应, 计量经济学建模大都以简约模型为主, 其中模型的函数形式大多给定, 模型的未知参数维数不大, 具有较强的经济可解释性. 最近, 洪永淼和汪寿阳 (2023)、汪寿阳等 (2023)、程兵 (2023) 对 ChatGPT 与大模型如何影响经济学、金融学以及社会科学的研究范式作了有益探讨. 本文的目的是进一步深入探讨以 ChatGPT 为代表的大语言模型的诞生、发展与应用将如何影响经济学与计量经济学的研究范式, 以及在人工智能时代如何推动经济学研究方法的创新. 大模型是大数据与人工智能催生的一种新的系统分析方法, 能够将定性知识与定量信息综合集成在一起, 特别适合于分析人类经济社会复杂系统.

本文第二节将介绍 ChatGPT 与大语言模型的主要特征. 第三节讨论 ChatGPT 与大语言模型的“规模至上”方法论如何破解经济学与计量经济学经常遇到的“维数灾难” (curse of

dimensionality). 第四节从理性经济人与人工智能经济人、孤立经济人与社会经济人、宏观经济学与微观经济学、定性分析与定量分析、小模型范式与大模型范式、计算机算法与计算经济学等几个重要关系入手, 深入、系统探讨 ChatGPT 与大模型对经济学研究范式的影响. 第五节讨论包括 ChatGPT 与大模型在内的人工智能技术作为经济学研究方法的局限性. 第六节是结论与展望.

2 ChatGPT 与大语言模型

为了探讨 ChatGPT 等模型对经济学与计量经济学研究范式的影响, 我们首先需要了解 ChatGPT 的主要特征. 为了打破数字科技巨头谷歌的垄断地位, 美国公开人工智能实验室 (Open AI) 于 2015 年正式成立, 起初定位为非营利性研究机构. 成立之后, Open AI 便致力于开发大语言模型. 2022 年 11 月 1 日, Open AI 正式推出 ChatGPT, 标志着通用生成式人工智能技术的一大突破. ChatGPT 是一种模仿人类认知方式的深度学习算法. 众所周知, 人类通过眼、耳、鼻、舌、身等感官接受信号, 这些信号被汇集传输到大脑中, 如果汇集信号强到一定程度, 便会激活大脑中一些区域的神经元进行分析处理并形成输出. 由于人类的认知思维具有一定的规律性, 这种认知思维方式与过程可以用数学模型如人工神经网络模型来模拟. 人工神经网络是一个基于所谓“激活函数” (activation function) 的非线性变换的数学模型, 包含很多未知参数, 这些参数可以通过数据进行估计 (White, 1992). 计算机科学通常不称之为“参数估计”, 而是称之为“模型训练”或“算法训练”. 在统计学与数据科学中, “模型”一般是指给定函数形式且未知参数维数较低的参数模型 (parametric model), 对未知参数维数很高且没有假设特定函数形式的一般数学模型, 通常称为算法 (Breiman, 2001), 也称为非参数模型 (non-parametric model). 本文将一般意义上的数学模型 (算法或非参数模型) 和统计学意义上的参数模型统称为“模型”, 并用“大模型”与“小模型”来区分它们.

从计量经济学与统计学的视角看, 神经网络是一种非参数统计模型. 可以证明, 当样本容量增加时, 如果参数维数随之增加, 则神经网络模型具备无限逼近满足一定正则条件的任何未知函数的能力, 这一性质被称为神经网络的“泛逼近性质” (universal approximation property). 计量经济学家哈尔伯特·怀特 (Halbert White) 对神经网络模型做出了重要的原创性基础理论贡献 (White, 1992).

神经网络是深度学习算法的基础模型. 深度学习实际上是神经网络模型的多层非线性变换. 简单的人工神经网络模型只包含一层非线性变换, 这足以一致估计满足一定正则条件的任何未知的非线性函数. 如果增加多层非线性变换 (可多达几十层、几百层甚至几千层), 模型参数维数将呈指数式增长, 模型复杂度越来越高, 可以刻画复杂系统的许多精细结构, 包括非线性、异质性与交互性等重要特征. 这样深度学习的整体逼近能力比一般的人工神经网络有很大的提升.

2017 年, 谷歌提出了一种名为变换器 (transformer) 的深度学习算法. 深度学习的基础结构是一个大语言模型, 输入为文本数据, 输出也是文本数据. ChatGPT 的第一个版本 GPT-1, 其参数数量为 1.17 亿, 这是非常庞大的数量. 在 GPT-2 版本中, 模型参数数量从 1.17 亿上升到 15 亿, 训练数据也增加了. 在 GPT-3 版本中, 参数数量达到 1750 亿个, 并使用大约 2/3 互联网数据、整个维基百科以及 2 个大型图书馆数据进行训练. 通过这些海量

文本数据的训练, GPT-3 能够执行各种复杂的工作任务. 2022 年 11 月, Open AI 正式推出了 ChatGPT, 即 GPT-3 的强化学习版本 GPT-3.5. 强化学习给人工智能设定一个目标函数, 并使该目标函数最大化, 其主要方式是通过人机对话, 当机器产生一定的输出时, 人类告诉它哪些是正确的, 哪些是错误的, 以此反馈信息给机器进行修正. 人机对话提供了源源不断的反馈, 让机器知道自己哪些地方做得好, 哪些地方需要改进, 极大改进了 ChatGPT 的学习效果. 2023 年 4 月, Open AI 进一步推出 GPT-4, 其重要突破是实现多模态输入, 训练数据从单一模态的文本数据变成多模态的文本与图像数据, 相应的参数维数也显著增加. ChatGPT 等大模型的训练之所以成为可能, 是计算机科学特别是算力与算法快速发展的结果.

ChatGPT 主要通过互联网公开大数据执行各种复杂任务. 众所周知, 大数据包括结构化数据和非结构化数据. 结构化数据如 GDP 时间序列数据, 每个季度有一个数值, 可以用矩阵形式表示; 而非结构化数据则不能用数字来表示, 如文本数据, 包括政府工作报告、上市公司财务报表、新闻媒介报道、微信微博聊天评论等. 除文本数据外, 非结构化数据还包括图像、音频与视频数据等, 比如医院拍摄的 X 光照片、地球卫星拍摄的遥感图片, 这些非结构化大数据包含了结构化数据所没有的大量有价值的信息. 例如, 语言是人们进行沟通交流、传递思想、表达情感的重要工具, 因此, 作为一种最主要的非结构化大数据, 文本数据包含了经济主体的很多情感等心理信息. 通过自然语言处理技术, 可从文本数据中提取许多有用的心理信息, 特别是构造可观测心理变量以进行定量分析. 有人估计, 非结构化数据占据了整个大数据的 80% 左右. 与结构化数据相比, 文本等非结构化数据都是高维数据, 需要使用大模型进行分析, 例如, 大语言模型可用于精确分析文本数据的语法结构和上下文语义.

由于训练数据包括几乎整个互联网的文本数据, ChatGPT 可视为一部迄今为止内容最全面、智能化程度最高的大百科全书, 在很多方面接近甚至超过人类的智力水平. ChatGPT 可以执行各式各样的工作, 如帮助人们进行编程, 修改程序. 在实证研究或计算机模拟仿真实验时, 编程是必不可少的, 在编程过程中, 即使是微小的错误也可能会花费一个人很长时间来检查纠正, 而 ChatGPT 在检查编程程序时能够提供快捷准确的帮助, 甚至直接帮助编写程序. ChatGPT 还可以帮助人们进行文献综述、英文语法检查、论文写作等. 比尔·盖茨 2023 年初在接受《福布斯》杂志采访时指出, ChatGPT 的诞生在人类技术革命的重要性方面类似于个人电脑和互联网的发明.

通过上述分析可知, ChatGPT 具有几个显著特点: 首先是智能性. 目前 ChatGPT 的智能相当于大学生的水平, 能够顺利通过考试并取得 A 等级的成绩. 虽然 ChatGPT 还存在许多缺陷, 包括有时会产生所谓的“幻觉”, 但目前全世界都在使用 ChatGPT 并为其提供免费训练, 因此 ChatGPT 改进、迭代与完善的速度非常快. 毫无疑问, ChatGPT 将逐渐接近人类的智能水平, 并在许多方面超越人类, 特别是对需要长时间重复的工作, 人类可能会感到疲倦, 但机器则完全没有这个问题. 其次, ChatGPT 是一种生成式人工智能技术, 其所生成的文字内容是大模型根据互联网大数据中单词词组同时出现的概率来预测生成, 这本质上是一种推测归纳的方法. 第三, ChatGPT 具有很广泛的通用性, 使用的信息是互联网公开信息, 可以执行各种任务, 包括生成结构化知识、提供解决问题的方案以及进行逻辑推理等, 帮助提升人类的决策能力与决策水平. ChatGPT 的这些重要特征都是基于互联网海量大数据与大语言模型而实现的.

3 ChatGPT 与大模型“模型至上”方法论

以 ChatGPT 为代表的大模型的发展理念是什么？尤其是开发 ChatGPT 所采用的方法论是什么？ChatGPT 的发展理念是大模型范式，其算法基础是大语言模型。在人工智能技术发展进程中，深度学习就已呈现大模型的“规模”特征，即参数个数非常庞大。如上文所述，深度学习是人工神经网络模型的多层非线性变换，参数维数取决于输入数据的维数、非线性变换的层数、以及每一个隐藏层的神经元数目等因素。深度学习算法每增加一层非线性变换，其参数数量就会呈指数级增长。早期许多深度学习算法（如卷积神经网络、循环神经网络、生成对抗网络、强化学习等）的参数数量就已达到几十万、几百万甚至几千万的数量级。

为什么 ChatGPT 需要大模型呢？ChatGPT 的输入是文本数据。文本数据本质上是高维数据。例如，常用的汉字数量有几千个，如果考虑它们各种可能的排列组合，可以想象这些文本数据的维度有多大。因此，文本数据本身就是一种高维复杂系统，对于这样的高维复杂系统，简单的小模型显然不足以胜任，必须采用高维复杂模型。人工智能科学家在实践中发现，以人工神经网络为基础的深度学习及其各种变种对文本数据以及其他非结构化数据（如图像、音频、视频等）具有很好的理解与预测能力。

过去，小模型主要针对结构化数据，如果要分析文本数据，则必须将小模型限定在某个特定领域（如金融），这样的文本数据的同质性比较高，小模型的拟合与预测效果会比较好。ChatGPT 的一个重要特征是通用性，即可以应用于各个不同领域以及执行各种工作任务。为了实现这种通用性，它所训练的文本数据必须涵盖各行各业。例如，如果希望 ChatGPT 能够回答关于政治、经济、金融、社会、法律、文化、历史、环境等各个领域的问题，那么必须使用这些不同领域的各种文本数据来训练 ChatGPT。一个语言模型，如果只让它回答特定领域的问题，那么模型参数维数无需像 ChatGPT 这样的大语言模型那么多，可以大幅减少，这种模型相对于通用生成式大语言模型，可视为小模型。小模型就像是专门针对特定应用场景而开发出来的“偏科机器”，其“举一反三”即泛化能力相对不足。所谓“泛化”能力，是指基于某个特定数据训练的模型或算法，对未见到的新数据的预测能力。从统计学视角看，如果专注于某个领域的的数据，那么这些数据具有较高的同质性，在这种情形下，模型无需过于复杂，模型参数维数不用太高，便可在某个特定领域达到不错的预测效果。若将不同领域的文本数据混合在一起，那么数据会呈现显著的异质性，此时若将一个训练好的小模型应用到其他场景的数据，可能会存在较大偏差，导致泛化能力不足。因此，发展通用生成式人工智能技术需要大模型技术。计算机科学家在实践中逐渐认识到，不断扩大模型规模，不断使用更多训练数据，是提高算法预测能力的一个非常有效的手段。大模型已成为当前人工智能的一个重要发展方向。

为什么以前没有考虑使用大模型？主要原因在于计算资源特别是算力与算法的限制。从计算机技术视角看，训练如此规模的大语言模型需要巨大的算力。例如，为了训练 ChatGPT，需要使用数千个 GPU 进行 24 小时不间断的数据输入，每次训练需要消耗约 450 万美元的电力成本，相当于开一辆汽车从地球到月球来回一圈。显然，ChatGPT 能够横空出世，是计算机技术特别是算力与算法迅猛发展的结果。如今，随着算力与算法技术的进步，训练大模型在技术层面上已不是问题。有人认为 ChatGPT 是工程意义上的技术进步，而不是科学意义

上的进步, 因为科学进步往往涉及理论与方法的创新. 但是, ChatGPT 不仅是一项具体技术的重大突破, 而且是人工智能发展的一个新方向、新理念、新范式. 以 ChatGPT 为代表的大模型的诞生、发展及日益广泛的应用, 正在深刻改变经济学与计量经济学的研究范式, 这种研究范式变革很可能导致经济学的理论创新.

大模型通过学习互联网上各种文本数据, 能够给出比小模型更准确的推断预测结果, 展现惊人的泛化能力, 生成的内容质量更好、更智能. 相比之下, 小模型的专用性使其在回答某些特定领域的问题表现更佳, 但对于通用性的应用, 小模型常常无法给出令人满意的答案. 从概率论与统计学视角看, 所有模型都有偏差. 所谓偏差, 是指模型估计量或预测值的平均值与真实函数值之间的差异, 如果这两者的差距越小, 则模型偏差越小, 随着模型复杂度的增加, 估计未知函数的偏差会越来越小. 因此, 增加模型复杂度可以减小模型偏差.

另一方面, 统计学与计量经济学期长期面临的一个根本问题是所谓的“维数灾难”. 在数据容量有限的条件下, 模型未知参数维数越多, 每个参数的估计越不精准, 估计误差越大, 这会导致模型过拟合, 影响其泛化能力. 统计学通常用均方误差 (mean squared error) 来测度预测的准确性. 均方误差可分解为方差与偏差平方之和. 因此, 即使模型偏差很小, 如果模型的未知参数数量非常多, 其估计方差也会变得很大, 导致预测不准确. 统计学家与计量经济学家在选择模型时通常会考虑偏差与方差之间的平衡. 事实上, 计算机科学家通过计算机模拟仿真实验以及实际应用, 发现了大模型在一定条件下会出现一种“双降” (double descent) 现象, 即当模型参数个数增加时, 以均方误差衡量的预测误差先呈现出下降趋势, 等模型复杂度达到一定临界点时, 预测误差便会上升, 这是统计学与计量经济学经典教科书介绍的 U-型曲线. 但是, 当模型参数维数继续增加并达到另一个更高临界值时, 模型预测误差会再次下降, 虽然第二次下降的速度相比第一次下降的速度会缓慢很多, 实证研究发现, 模型参数维数的指数增长, 才能换来预测误差的线性减少, 这就是所谓的“双降”现象 (Belkin et al., 2018; Nakkiran et al., 2021). 统计学家 Hastie et al. (2019) 也发现, 在一定假设条件下, 高维线性统计模型也会出现“双降”现象. “双降”现象打破了人们长期以来对模型复杂性的固化认知, 它使大家逐步认识到, 在通常的“小模型”空间之外, 还有一个以前从未发现的“大模型”空间, 那里呈现了与小模型完全不一样的规律性特征, 特别是当模型复杂度跨过一个很大的临界值后, 大模型便会呈现所谓的“涌现”能力, 这种涌现能力与大模型的“双降”现象密切相关. 事实上, 由于大模型训练数据是几乎整个互联网的文本数据, 这在一定程度上避免或减缓了维数灾难的问题. 大模型不仅具有较小的偏差, 还可以对每个参数进行较为精确的估计, 即使 ChatGPT 拥有 1750 亿个未知参数, 在整个互联网海量数据容量面前, 这个参数维数仍然不算大. 因此, 基于互联网海量大数据的大模型拥有较强的泛化能力.

4 ChatGPT 与大模型对经济学研究范式的影响

大数据与人工智能对经济学、计量经济学以及相关学科产生了深刻影响. Varian (2014)、洪永淼和汪寿阳 (2021a, 2021b)、刘景江等 (2023) 对大数据以及机器学习如何影响经济学和管理学研究范式与研究方法进行了探讨. 例如, 海量实时或高频大数据可用于构建高频经济金融数据 (如日度消费者价格); 非结构化大数据特别是文本数据可用于构建经济金融心理变量, 如政策不确定性指数、投资者情绪指数等; 机器学习降维技术可用于识别、选择重要特征

或重要变量; 实时或高频大数据可用于实时预测, 等等. 最近, 洪永淼和汪寿阳 (2023)、汪寿阳等 (2023)、程兵 (2023) 初步探讨了以 ChatGPT 为代表的大模型对经济学、金融学研究范式的影响. 在这一节, 我们从六个方面深入、系统探讨 ChatGPT 与大模型对经济学研究范式可能产生的重要影响.

4.1 理性经济人和人工智能经济人

新古典经济学的一个基本假设是理性经济人, 即经济主体有一个目标函数, 会利用一切可以利用的信息与资源, 使自身利益最大化, 这是一种理性行为. 但是行为经济学研究表明, 人类经济行为在很多情形下并不满足完全理性假设, 存在着非理性因素与现象. 赫伯特·西蒙 (Herbert Simon) 提出有限理性理论 (bounded rationality), 并因此于 1978 年获得诺贝尔经济学奖. 所谓有限理性, 是指经济主体无法获取决策所需的所有信息, 也不清楚问题的所有错综复杂的关系和行动后果, 换言之, 经济主体不仅收集处理数据能力有限, 计算能力也有限. 因此, Simon (1976) 认为经济主体并非追求完全理性假设下的利益最大化, 而是满意即可. 西蒙的思想推动了行为经济学、行为金融学以及人工智能经济学的蓬勃发展. 行为经济学、行为金融学主要研究经济主体的非理性行为及其对经济的影响, 这些行为学科正在逐渐渗入到宏观经济学等其他学科 (Akerlof, 2002; Angeletos and Huo, 2021; Angeletos and Lian, 2022; Angeletos et al., 2021; Hommes, 2021; 那艺和贺京同, 2017). 随着 ChatGPT 的智能程度越来越接近人类的智能水平, 其通用性使人工智能能够替代人类很多工作. 如果人工智能能够辅助甚至替代经济主体决策, 将大幅提升经济主体的理性程度. 第一, 人工智能特别是 ChatGPT 拥有庞大信息集, 几乎是整个互联网数据, 而且这些信息的获取是低成本甚至几乎无成本的. 相反, 一个人每天能够获取信息的渠道与数量都有限. 第二, 经济主体的很多决策均是在不确定市场环境下进行的, 最优决策需要基于对未来经济走势的精准预测 (Granger, 1999). 人工智能特别是机器学习的预测精准性比传统计量经济学模型要高很多. 第三, 人工智能在信息处理与逻辑推理方面比人脑具有明显的优势, 但在与创造力密切相关的形象思维方面至少在目前还不能为人类提供太多的帮助. 人类智能与人工智能因此各有所长, 相辅相成, 如果能够实现“人机结合”, 将使人类具有更强的创造力和智能水平. 西蒙很早就认识到人工智能和管理决策科学这两个领域的内在联系. 钱学森等 (1990) 在提出定性分析与定量分析相结合的综合集成法以研究复杂系统时, 便主张采用人机结合的分析方法. 诺贝尔经济学奖获得者、宏观经济学理性预期学派的代表人物之一托马斯·萨金特 (Thomas Sargent) 也提出类似有限理性的有限理性预期 (bounded rational expectations). Sargent (1993) 提出以人工智能经济人取代理性经济人假设, 通过使用人工神经网络模型进行适应性学习, 帮助改进经济主体的预期, 从而缩小有限理性预期和完全理性预期之间的距离. 人工智能在很多方面可以辅助人的决策, 例如, 在高考领域, 利用人工智能可以帮助更好选择高考志愿; 在学术研究中, 可以利用 ChatGPT 辅助总结文献、准备 PPT、编写程序、撰写论文等; 在量化交易中, 利用算法预测市场走势, 确定交易策略; 在企业管理中, 人工智能可以帮助企业优化经营计划与提升管理效率. 在发挥政府作用方面, 人工智能可以帮助政府优化各种规划、计划、政策和改善市场机制设计, 弥补由外部性造成的市场失灵, 实现稀缺资源的有效配置. 人工智能还能改进政策评估的精准性. 人工智能具有比较精确的泛化能力,

可用于提高估计虚拟事实的精确度, 从而提升政策评估的准确性, 更好发挥政府在资源配置中的作用. 当然, 单靠人工智能不可能实现有效的计划经济, 因为人工智能无法获得所有个人的信息, 更无法解决更为关键的人的激励问题.

4.2 孤立经济人与社会经济人

除了理性假设外, 新古典经济学还假设经济人是孤立的, 不考虑社会关系对经济人的经济行为的影响. 在标准微观经济学教科书中, 经典消费者理论只有一个消费者, 在财务约束条件下实现效用最大化; 厂商理论只有一个生产者, 追求利润最大化或成本最小化. 不管是消费者理论还是厂商理论, 都是一个孤立的经济人, 只对市场价格信号作出反应.

20 世纪 30 年代, “凯恩斯革命” (Keynes, 1936) 催生了宏观经济学. 传统宏观经济学主要研究总量平衡, 如总供给和总需求的平衡问题. 新古典经济学将经济学优化理论与总量平衡分析方法结合起来 (Ramsey, 1928; Cass, 1965; Koopmans, 1963). 例如, 宏观经济学理论预期学派, 假设一个具有完全理性预期的代表性经济人 (如消费者或投资者), 在跨期财务约束条件下, 选择一条最优消费与投资路径, 使其一生总效用最大化. 代表性经济主体的假设掩盖了现实经济中存在大量异质性经济主体的事实 (如异质性偏好), 因此无法研究异质性经济主体的行为以及由此产生的异质性经济关系, 如政策对异质性经济主体的不同影响.

在一个完全竞争性市场经济中, 市场价格充分反映了各种市场信息, 每一个经济人都是价格的接受者, 只需要对市场价格信号作出反应, 即可获得最优决策. 但是, 在现实经济中, 除了经济人的异质性以外, 经济人之间的决策行为甚至他们的目标函数都会互相影响. 马克思指出, 人是一切社会关系的总和. 因此, 现实经济中的经济人是社会经济人, 其经济行为不仅受经济因素的影响, 而且还受各种社会关系的影响, 同时人的经济行为和经济因素也会影响其社会关系. 传统制度主义经济学派早就提出“社会人”假设, 认为人作为一种社会存在, 除了追求经济利益之外, 还有对安全、友情、尊重、归属等方面的社会需求; 人的经济决策, 必须建立在自身的经济条件、以及构成其社会生活重要组成部分的人际关系的基础之上. 换言之, 人的经济行为与其所处的社会环境密切相关, 需要从个人与社会环境的相互关系去理解人的经济行为.

在当今的互联网时代和数字经济时代, 人与人之间的关系网络化、平台化, 变得密切相关和日益复杂, 即所谓的超联 (hyper-connectedness). 平台经济的出现, 使生产者与消费者之间、生产者之间以及消费者之间的关系变得更加直接、更加密切, 这一方面大幅度减少了信息不对称性, 但另一方面更加紧密而复杂的相互关联也会产生新的风险.

在互联网时代, 受历史、文化、语言、宗族、地理、环境等因素的影响而形成人与人之间、群体与群体之间的各种社会网络, 深刻影响着人的经济行为. Shiller (2001, 2019) 指出, 投资者通过互联网与媒体传播而形成的广大投资者的情绪传染, 会造成金融市场泡沫. 例如, 比特币价格之所以在过去几年疯涨, 并不是因为经济基本面发生了重大变化, 更多的是因为投资者的心理因素所致, 是投资者对比特币故事的塑造引起广大投资者对比特币的追捧, 从而导致比特币价格疯涨. 另一个投资者情绪传染的实例是 2021 年初美国上市公司“游戏驿站”公司的股票事件, 大量“游戏驿站”散户投资者通过网络和媒体渲染而产生情绪共鸣并且采取一致行动, 使“游戏驿站”公司股票在短短一个月左右的时间里上涨 20 倍, 最终打败了

实力雄厚的对冲基金公司,显示了金融市场广大散户投资者情绪传染所迸发出的巨大威力。

过去,由于缺乏经济社会网络数据,定量测度经济行为和社会关系之间的关系非常困难,甚至是不可能的。如今,随着互联网和人工智能等数字技术的快速发展,各种社会网络数据的可获得性大大增强,这使关于经济行为和社会关系之间互动关系的深入研究成为可能,而且使经济学研究不再局限于经济领域本身,而是可以进行跨学科研究,从而推动社会经济学这样的交叉学科快速发展。在金融学领域,一门交叉学科——社会金融学 (social finance) 正在兴起。社会金融学融合了金融学、经济学、社会学以及环境科学等多学科的知识,将传统金融工具和社会、环境目标相结合,旨在通过金融投资等工具帮助解决一系列伦理、社会与环境问题,推动人类经济社会可持续发展,并产生积极的社会影响。目前,环境与社会治理 (ESG) 已成为国内外经济学、金融学等学科的一个热门研究领域。

十多年来,经济学家和社会科学工作者基于越来越多的经济社会网络数据,使用网络模型量化测度和研究各种经济社会关系,取得了不少进展 (Stachurski and Sargent, 2022)。由于人类经济社会是一个高维复杂系统,并且大量经济社会网络数据是非结构化大数据 (如文本、图像、音频与视频数据),因此需要采用定性与定量分析相结合的系统综合集成方法,而大模型包括多模态大模型正是这样一种可行的系统分析方法。

4.3 宏观经济学和微观经济学

“凯恩斯革命”催生了宏观经济学,从此经济学有了宏观经济学与微观经济学之分。宏观经济学研究整个国民经济的运行规律,特别是宏观经济变量之间的关系,如菲利普斯曲线 (Phillips, 1958)、奥肯定律 (Okun, 1970) 以及泰勒规则 (Taylor, 1993) 等宏观经济定律。长期以来,宏观经济学与宏观计量经济学均是基于宏观经济变量进行建模,研究宏观经济变量之间的数量关系与逻辑关系,并据此解释宏观经济现象与预测宏观经济走向。这种经典宏观经济建模思想首先对微观经济数据进行加总,获得宏观经济变量,然后对宏观经济变量进行建模,其优点是可以得到简约模型,并具有较强的经济可解释性 (Glandon et al., 2023)。实证宏观经济学也经常使用因子模型,不管是拥有高维宏观经济数据还是大量微观经济数据,宏观经济学家经常假设存在维度不高的潜在共同因子驱动这些高维宏观或微观经济变量的变动 (Stock and Watson, 1996)。因子模型是降维的一个有效方法,从高维宏观经济数据或大量微观经济数据中提取的共同因子可用于解释或预测宏观经济走势,但这些因子本质上也是一种信息“加总” (aggregation)。先加总后建模的主要缺点是加总会使原始数据中很多信息失真或消失,例如微观经济行为的异质性、交互性、非线性、时变性、厚尾、跳跃等特征或现象,会因为加总而失真甚至消失 (Kaplan and Violante, 2018; Marcellino et al., 2003; Stock and Watson, 2016; 李戎等, 2022)。长期以来,人们一直批评宏观经济学缺乏微观经济学的基础,这一点与经典热力学与统计物理学具有严谨的微观基础存在巨大的差异。例如,经典热力学与统计物理学基于互相独立的微观粒子遵从布朗运动的假设,可推出一定温度下微观粒子运动速度的概率分布,即著名的麦克斯韦-玻尔兹曼分布。这个分布描述了在系统处于平衡时,处于任何速度范围内的粒子所占的比例,是系统温度的函数。二十年来,宏观经济学在构建微观理论基础方面取得了一定的进展。此外,经济学家还将行为经济学特别是一些非理性因素引入宏观经济学分析框架之中 (Akerlof, 2002, 2007),克服宏观经济学理性预期学派只

考虑理性预期这种理想情景的局限性,这在某种程度上回归到凯恩斯 (Keynes, 1936) 的“动物精神”和 Shiller (2001, 2019) 的“非理性繁荣”的分析框架。

然而,宏观经济学的微观基础仍有待新的突破,而大模型在这方面可以发挥重要作用。例如,假设整个宏观经济共有 N 个消费者,每个消费者在每个时期的消费为 C_{it} ,则整个宏观经济消费总额是 $C_t = \sum_{i=1}^N C_{it}$ 。宏观经济学通常直接研究宏观消费总量 C_t 的动态演变特征,如 Hall (1978) 的消费平滑理论或随机游走理论。在拥有海量微观动态大数据条件下,可以通过大量微观消费大数据研究宏观消费总量的动态特征。宏观经济变量与微观经济变量的动态特征可能存在显著差别。例如,Granger (1980) 证明,在一定条件下,大量具有“短记忆”动态特征的微观消费变量,在加总之后,宏观消费变量会呈现出“长记忆”动态特征。从概率论看,如果知道或者能够估计出大量微观消费变量的联合概率分布,则通过变量变换便可获得宏观消费总量的概率分布,从而获得总消费的动态演化特征。显然,这是一个高维消费概率模型,涉及到高维雅可比矩阵求逆和高维积分,这在算力和算法上将会面临很大挑战。如何基于大量微观数据研究宏观经济行为,是未来经济学与计量经济学的一个重要发展方向。最近,Hong et al. (2023) 提出基于一个大面板微观数据直接预测宏观经济变量的新方法,他们使用了 1500 个上市公司股票价格月度数据,基于随机森林等机器学习“大模型”,预测月度通货膨胀率。研究显示,相对于很多流行的宏观时间序列模型,基于大量微观数据的大模型对通货膨胀率具有显著优异的样本外预测能力,这说明大量微观数据所包含的异质性信息有助于预测通货膨胀。类似这样基于大量微观数据预测宏观经济变量的大模型方法,有望为构建宏观经济学的微观理论基础提供新的典型经验特征事实和新的洞见。可以预料,在大数据和大模型时代,经济学即将发展到了一个新阶段,即宏观经济学与微观经济学的界限将越来越不明显。

除了大量横截面微观经济数据之外,大数据也提供了很多高频甚至实时经济数据,高频或实时经济数据为及时把握宏观经济变化趋势提供了一种可能性。例如,Choi et al. (2022) 通过构建政府支出高频数据,研究宏观层面的财政政策对美国通货膨胀的影响。从统计学视角看,高频数据可用于构建更加有效率的低频统计量,例如利用日内收益率数据构建日度波动率 (Aït-Sahalia and Jacod, 2014),利用高频数据实时预测低频宏观经济变量(如月度通货膨胀率、季度 GDP 增长率等),实时监测宏观经济运行状况。以宏观经济实时预测 (nowcasting) 为例,实时预测是对当前、非常近的未来或最近的过去的宏观经济变量的预测 (Giannone et al., 2008)。统计部门的宏观经济数据发布通常存在时滞,因此不能及时获得宏观经济变化的数据,无法对当下的宏观经济形势做出及时、准确的判断。而实时预测能够在季度 GDP 数据公布前估测季度 GDP 总量或增长率,所利用的数据不局限于官方统计数据,还可包括高频金融数据、新闻媒体数据、搜索引擎数据等异源、异构、异频大数据,方法也不受制于传统的统计学与计量经济学模型。实时预测已被世界上许多机构特别是中央银行用于实时监测宏观经济指标的变化。高频或实时经济金融大数据的可获得性使得经济学的低频建模与高频建模的界线日益模糊了。这样,可以在一个统一的分析框架中研究“慢变量”(如低频宏观经济指标)和“快变量”(如高频金融变量)之间的动态相互关系。

4.4 定性分析与定量分析

长期以来, 由于历史等各种原因, 经济学的定性分析与定量分析在学术界一直存在争论. 自从有了大数据, 各种非结构化数据特别是文本数据所包含的有价值信息便引起了经济学家与计量经济学家的广泛关注. 众所周知, 语言是人类进行信息沟通与情感交流的主要工具, 因此文本数据包含结构化数据所没有的很多有价值信息, 特别是经济主体对政策变化与外来冲击的心理反应, 包括预期、情绪、情感等信息 (洪永淼等, 2023). Shiller (2019) 指出, 相对于结构化经济数据, 文本数据包含很多关于经济运行与经济发展的丰富信息. 因此, 从文本数据提取各类经济主体心理信息便成为一种有效方式, 这种方法比传统的心理数据构造方法 (如信心与预期统计调查数据、心理学实验测度法) 具有不少优势, 特别是在样本代表性和抽样频率等方面. 从文本数据提取心理因素的主要工具是自然语言处理技术, 目前经济学与其他社会科学常用的方法包括词频法, 词袋法, 主题法等 (Gentzkow and Kelly, 2019). 通过文本数据测度心理变量, 打破了定性分析和定量分析的界限, 使原来只能进行定性分析的很多经济学问题, 转变成可以进行定量分析. 目前, 从文本数据提取经济主体心理信息的常用方法简单可行, 但也存在一些缺陷, 如所构建的心理变量大多存在不可忽视的测度误差, 这些测度误差在回归分析时可能会造成估计偏差 (洪永淼等, 2023). 而且, 目前几乎所有基于文本数据的心理变量测度都是通过加总而得, 很多微观层面的心理异质性消失了. 举一个简单例子, 如果在一个文本数据中, 一半的人是悲观情绪, 另一半的人是乐观情绪, 则基于关键词加总而得到的情绪指数可能在整体上既不表示悲观也不表示乐观, 原有的异质性情绪消失了. 在这方面, 基于互联网海量文本数据训练的 ChatGPT 可用于构建微观层面 (如每个消费者、投资者) 的心理变量, 避免由于加总而导致异质性心理信息的损失. ChatGPT 是目前自然语言处理领域最先进的技术, 可借助其拥有的几乎整个互联网信息与知识, 比较准确解读文本数据的语法结构与上下文语义, 从而比较准确从文本数据中提取心理信息.

ChatGPT 是一个大语言模型, 拥有 1750 亿个参数, 但它的输入是文本数据, 输出也是文本数据, 换言之, 作为一种通用生成式人工智能技术, ChatGPT 主要用于定性分析, 但这种定性分析的基础却是一个超高维的极其复杂的大语言模型. ChatGPT 本身是定性分析与定量分析的高度统一, 在这里, 大语言模型的复杂度越高、训练的文本数据越多, ChatGPT 定性分析的准确性、逻辑性与智能性也就越高. 一段时间以来, 中国经济学界对经济学研究出现过度数学化、模型化的现象提出了一些担心与批评 (洪永淼和汪寿阳, 2020; 李志军和尚增健, 2020). 这些批评大多数是中肯的, 有助于在经济学研究中避免误用、滥用数学模型, 特别是避免只注重数学模型而轻视经济思想的不良倾向, 但这种批评不能被变为妨碍使用数学和模型的借口. ChatGPT 的诞生表明, 高度数学化、模型化与精准的定性分析是可以统一的. 钱学森等 (1990) 早就提倡发展定性分析与定量分析相结合的综合集成分析法, 以研究包括人类经济社会系统在内的复杂系统或复杂巨系统. 大模型范式本质上是大数据与人工智能催生的一种新的系统分析方法, 能够将定性分析与定量分析有机融合在一起, 非常适合于研究复杂经济社会系统.

4.5 小模型范式和大模型范式

经济理论构建大都通过经济学建模. 经济学建模的基本方法论是: 基于偏好、技术、资源、禀赋、制度、行为 (如预期) 等假设, 借助数学等逻辑工具的演绎推理, 研究少数关键经济因素对经济主体决策与宏观经济运行的影响. 在建模过程中聚焦少数主要经济变量而忽略其他次要因素的影响 (通过假设其他因素不变), 既是数学抽象的必然要求, 也是保障模型的经济可解释性的有效方法. 正如列宁 (1990, 第 142 页) 所指出的, “物质的抽象, 自然规律的抽象, 价值的抽象以及其他等等, 一句话, 一切科学的 (正确的、郑重的、非瞎说的) 抽象, 都更深刻、更正确、更完全地反映着自然”. 然而, 从预测角度看, 一个具有很强经济可解释性的简约经济学模型, 其预测能力不一定精准. 一个主要原因是影响现实经济社会活动的因素成千上万, 繁多复杂. 为了说明这一点, 我们考察以下预测模型:

$$Y_t = g(X_t) + \varepsilon_t, \quad (1)$$

其中 X_t 是可观测预测变量的集合, Y_t 是可观测因变量, ε_t 是所有不可观测的因素对 Y_t 的影响的总和. 为简单起见, 这里假设可观测变量集合和不可观测变量集合是可分的. 统计学家与计量经济学家经常称 $g(X_t)$ 为信号 (signal) 而称 ε_t 为“噪声” (noise), 模型 $g(X_t)$ 对 Y_t 的预测能力取决于“信噪比” (signal to noise ratio). 如果信号强, 则 $g(X_t)$ 预测能力强; 反之, 如果信号 $g(X_t)$ 相对噪声 ε_t 较弱, 则 $g(X_t)$ 的预测能力将较差. 在小数据场景下, 很多因素不可观测, 这些不可观测因素的影响都被归到噪声 ε_t 中, 因此信噪比较弱, $g(X_t)$ 的预测能力较差. 相反, 在大数据场景下, 可观测信息集显著扩大, 因此信噪比大幅提升, 这为改进预测精度提供了一个可能性, 而大模型正是实现这种可能性的一个重要方法与工具.

统计学与计量经济学的一个基本建模原则是“保持模型简约” (Keep It Sophistically Simple) (Zellner et al., 2001). 任何一个模型需要有足够的去捕捉数据中有价值的信息, 在此前提下, 模型越简单越好. 这样既可减少模型偏差, 又可控制模型复杂度, 减少模型过拟合的可能性. 统计学与计量经济学一直存在建模方法论的争论, 例如 20 世纪统计学参数与非参数建模的方法论之争, 即统计学家罗纳德·费希尔 (Ronald Fisher) 的“小模型”与卡尔·皮尔逊 (Karl Pearson) 的“大模型”之争. 费希尔主张使用参数模型, 即函数形式已知的模型 (如线性回归模型), 其中未知参数的维数不大, 因此可称为“小模型”. 费希尔不鼓励使用非参数模型, 他认为非参数模型涉及很多参数, 可能会导致过度的估计误差, 出现过拟合现象. 皮尔逊则认为, 非参数模型在样本容量增加时, 其未知参数数量相应增加, 因此模型偏差会越来越小, 最终趋近于零; 但在参数模型中, 虽然参数数量较少, 如果模型错误设定, 则模型偏差将无法消除, 不管样本容量有多大, 因此皮尔逊不鼓励使用参数模型. 这两种建模方法一直存在争议, 深刻影响了统计学的发展. 有趣的是, 计量经济学也有类似的建模方法论之争, 即罗伯特·恩格尔 (Robert Engle) 的“小模型”与大卫·亨德里 (David Hendry) 的“大模型”之争. 诺贝尔经济学奖获得者恩格尔主张“从特殊到一般”的建模方法, 即从一个特定的小模型出发, 应用计量经济学诊断等方法拓展原有模型, 这是从简约模型到复杂模型的建模方法; 而伦敦政治经济学院计量经济学派代表人物亨德里则主张从“一般到特殊”的建模方法, 即从一般的大模型出发, 通过统计假设检验与经济理论约束等方法降低模型维数获得简约模型, 这是从复杂模型到简约模型的建模方法. 恩格尔的建模方法比较适合于小数据场景, 因

此模型复杂度不能太大。亨德里的大模型方法论则比较适合大数据场景,其最大优点是模型偏差较小,所获得的结论对模型的依赖性不会太大。但是,无论是统计学还是计量经济学,这些方法论之争背后都蕴藏着相同的思想,即获得一个既有较好经济可解释性又有较强泛化能力的简约模型。在统计学与计量经济学,获得简约模型的主要方法是降维。常用的经典降维方法有主成分分析法、因子模型、AIC 和 BIC 等模型选择准则、投影寻踪回归 (projection pursuit regression) 与单指标模型 (single index model) 等,所选择的最佳模型通常具有较低的复杂度和较强的预测能力。此外,也可通过经济理论约束条件进行降维,例如经济主体具有风险厌恶的特征,则其效用函数必定是凹函数。又如,如果市场有效性假说成立,那么所有历史信息都无法预测未来的回报率,因此预测模型的滞后项预测变量的系数都应该为零,这样可帮助简化模型。

最近,机器学习和统计学的交叉融合产生了一个称为统计学习的新领域,其中 Tibshirani (1996) 的套索 (LASSO) 回归是一个非常重要的降维方法。假设一个高维线性回归模型包含 p 个解释变量,共有 p 个未知参数,模型参数维数 p 比较大,甚至有可能出现 p 大于样本容量 n ,在后面这种情况下就无法用最小二乘法估计参数。但如果稀疏性假设成立,即 p 个参数中只有少数几个参数的数值不为零,则可通过引入一项惩罚项控制所有未知参数绝对值之和来估计未知参数。这样虽然会牺牲模型偏差,但能够显著降低估计误差与均方误差,最终识别出那少数几个系数不为零的重要变量,并将所有系数为零的变量全部剔除。这是 LASSO 降维的基本思想。在计量经济学与统计学,几乎所有降维预测方法都是基于类似的思想。简单地说,假设有一个预测模型,通过训练数据得到很好的拟合效果,训练数据既包含对样本外预测有用的信号,也包含对样本外预测无用的噪声(一般情况下,样本外数据中的噪声与训练数据中的信号和噪声之间没有相关性)。如果一个模型在训练时使用过多的参数,则所得到的模型不仅会捕捉有用的信号,还会捕捉到很多无用的噪声,这些包含在模型中的训练数据噪声对样本外预测不但没有帮助,反而会造成干扰。因此,要获得一个能够准确预测的模型,就需要在拥有足够强的信号的情况下避免捕捉太多的噪声。降维可以帮助解决这个问题,因为它可以控制模型复杂度,大幅减少噪声的影响,从而提高样本外预测的精准性。

长期以来,统计学与计量经济学所使用的数据大都是“小”数据。这里所说的“小”数据并不仅限于几十或几百个样本点,也可能包括几十万维度的文本数据。但是,相对于人类拥有的海量大数据来说,大量的大数据信息尚未被有效挖掘与充分利用。例如,Engle and Russell (1998) 基于高频金融交易数据提出了一个自回归条件久期模型。当时还没有大数据的概念,但金融学已经有了逐笔交易这样的高频大数据。Engle and Russell (1998) 拥有的数据容量达 100 多万,但只使用了其中不到 5% 的数据。为什么计量经济学一直使用小模型与小数据?首先,小模型容易解释,特别是每个参数都可以有丰富的经济含义。其次,现实数据量有限,无法支持大模型的训练。最后,计算资源受到约束,包括数据的存储容量、处理能力和运算速度等。因此,长期以来,小模型一直是统计学与计量经济学的主流方法论。

大量实证研究 (Kelly and Xiu, 2023) 表明,在预测经济与金融市场变化趋势方面,机器学习在很大程度上改进了传统计量经济学模型,但其预测准确性尚未达到令人满意的程度。一个根本原因是人类经济社会系统是一个高度复杂系统,受到成千上万的相互关联的变量的影响,涉及政治、经济、社会、法律、政策、技术、历史、文化、心理、环境等各种因素。与自

然界不同, 经济系统受人类心理影响大, 而人类心理本身存在很大的不确定性. 经济主体在面临不确定性时通常形成一种预期, 如果预期过高或过低, 就会产生悲观或乐观的情绪. 这样的情绪会影响人们的经济行为, 如影响投资选择和消费倾向等. 客观存在的经济社会活动影响人类心理, 而人类心理反过来又会影响经济主体的行为与经济社会系统的运行, 这种心理影响称为“反身性”. 此外, 经济主体存在明显的异质性, 如不同所有制的企业, 不同收入水平的消费者, 其行为特征包括对经济政策与外生性冲击的反应以及所受到的影响是不一样的. 最后, 由于偏好、技术、环境、人口、政策与制度等变化, 经济系统具有时变性, 时变性可以是突变, 也可以是缓慢变化. 反身性、异质性、交互性以及时变性对预测未来会产生重大影响. 为了显著改进经济金融预测, 可以考虑大模型范式. 由于其灵活性, 大模型可以容纳互相关联的高维变量, 刻画经济主体的异质性、变量之间的非线性与交互性, 以及模型参数的时变性, 从而大幅度降低模型误差, 提高预测精准度. 最近, Lopez-Lira and Tang (2023) 使用 ChatGPT 判断关于上市公司的新闻标题对上市公司股价是好消息、坏消息或无关消息, 并给出理由与情绪积分. 他们的研究发现, 这样构建的“ChatGPT 积分”与上市公司日收益率存在正相关关系, 而更基础的大模型如 GPT-1 和 GPT-2 等并不能精准预测收益率, 因此 ChatGPT 较强的股票收益率预测能力是由于使用更复杂的大模型即 GPT-3.5 而获得的结果.

事实上, 计量经济学与统计学已存在一些“大”模型. 这里所说的“大”模型是指模型所包含的未知参数维数比统计学与计量经济学常用模型的参数维数要大很多, 但可能远远小于 ChatGPT 与大语言模型的参数维数, 而且输入数据也不一定是文本数据, 可以是结构化数据. 在经济学与计量经济学, “大”模型并不少见. 例如, 预测科学领域的模型组合是一个大模型 (Sun et al., 2021, 2023). 假设有多个预测模型, 将这些模型线性组合成为一个大模型, 可以提供比较稳健的预测. 使用单一模型进行预测, 就像购买高回报与高风险股票一样, 而使用模型组合则类似于购买共同基金, 虽然回报率较低, 但更加稳健而风险较小. 由于每个预测模型都有一些未知参数, 如果将很多不同预测模型线性组合起来, 形成一个大模型, 则这样的大模型的参数维数就相当可观. 例如, 假设共有 m 个模型, 每个模型有 p 个未知参数, 则组合模型的参数维数 (包括组合权重系数在内) 将达到 $m(p+1)$. 如果允许参数与组合权重系数具有时变性, 则未知参数值的总目数达到 $m(p+1)T$, 将随时间样本容量 T 的增加而增加 (Cui et al., 2023a). 在金融计量经济学, 多元 GARCH 模型的估计由于维数灾难等原因是一个长期没有解决的计量经济学难题 (Engle, 2002). 假设有 m 个资产, 则这 m 个资产相对于历史信息集的条件方差-协方差是一个 $m \times m$ 维半正定对称动态矩阵. 当对这个矩阵进行多元 GARCH 建模时, 如果这个矩阵模型的每个元素包含 p 个未知参数, 则多元 GARCH 模型未知参数的维数将达到 $m(m+1)p/2$. 当 m 或 p 较大时, 未知参数的维数将接近甚至超过时间序列样本容量 T , 导致多元 GARCH 模型参数估计具有高度挑战性 (Cui et al., 2023b). 金融计量经济学家已提出各式各样的多元 GARCH 模型, 通过各种约束条件降低参数维数, 但仍然没有从根本上解决高维多元 GARCH 模型的参数估计问题.

经济学还有其他类型的大模型, 如可计算一般均衡模型. 假设存在 2000 种商品, 则需要 2000 个需求模型, 2000 个供应模型, 共 2000 个联立方程式. 如果每个需求或供给模型各有 10 个未知参数, 那么总共需要估计 40 万个参数. 在微观层面, Lanier et al. (2023) 考虑估计一个高维超市商品需求模型, 其中共有 4368 种商品, 每种商品用一个 24 维的未知特征向量

来刻画, 异质性消费者共分为 30 类, 每类消费者的效用函数均是二次型, 但不同类消费者的效用函数参数值不同, 这个大型需求模型共有 105549 个未知待估参数, 参数维数比较大, 但可避免商品加总导致的信息损失, 能够比较精准评估商品需求弹性、商品替代效应以及消费者福利如何随商品价格或税率变化而变化. 20 世纪六七十年代, 诺贝尔经济学奖获得者劳伦斯·克莱因 (Lawrence Klein) 推行一个名为“世界连接模型”(world link model) 的全球宏观经济模型. 到 20 世纪 90 年代, 该模型有 70 个国家参与, 共有 2 万个联立方程, 如果每个方程有 10 个未知参数, 则将拥有超过 20 万个未知参数. 另一个宏观计量经济学大模型是高维向量自回归 (vector autoregression, VAR) 模型与高维结构化向量自回归 (structural VAR, SVAR) 模型. VAR 与 SVAR 模型 (Sims, 1980) 是一个标准的宏观经济分析工具, 广泛应用于宏观经济实证研究与政策分析. 但是, 一个主要困难是 VAR 与 SVAR 模型的参数维数会以内生变量维数的平方而快速增加. 例如, 一个 100 维、含二阶滞后项的 VAR 模型, 将拥有至少 20000 个未知参数, 远远超过常见的宏观经济数据样本容量, 由于这个原因, VAR 与 SVAR 模型通常只用于刻画小型宏观经济系统, 如货币-产出-通货膨胀构成的小模型. 基于大量小型经济系统的研究使相关研究趋于碎片化, 研究结果呈现局部性与片面性, 无法刻画宏观经济系统的整体特征与复杂关联. 近年来, 一些研究使用因子增扩 (Bernanke et al., 2005)、贝叶斯压缩 (Bańbura et al., 2010; Chan, 2023) 以及正则化 (Basu and Michailidis, 2015) 等降维方法来估计高维 VAR 与 SVAR 模型, 从而可以涵盖高维宏观经济指标集.

尽管这些计量经济学大模型早已存在, 但由于数据量不足以及算力与算法的限制, 它们并没有得到广泛应用. 随着大数据的可获得性与计算机技术的快速发展, 现在已到了可以应用大模型的时候. 为此, 需要构建统一的大规模数据库, 尤其是利用人工智能技术来整合异源、异构、异频等各类数据. 另外, 为了训练这些大模型, 需要大力加强计算基础设施建设, 确保拥有足够的计算资源, 特别是算力与算法.

4.6 计算机算法与计算经济学

计算机技术特别是算法的发展, 为经济学研究提供了一种新方法与新工具. 李拉亚 (2014) 对计算机算法作为研究工具而影响经济学研究范式以及计算机作为技术而影响生产力与生产关系, 作了系统的论述与颇具前瞻性的推测. 前文提到, 计算社会科学基于大数据, 使用人工智能技术方法挖掘变量之间的逻辑关系, 这是一种数据密集型的研究范式, 其中算法发挥关键作用. 在计算社会科学诞生之前, 就已存在计算经济学、计算金融学等学科. 例如, 计算经济学一个著名的模型是可计算一般均衡模型. 早在 19 世纪 70 年代, 经济学“边际革命”代表人物之一莱昂·瓦尔拉斯 (Walras, 1874) 便提出了一般均衡论, Wald (1936) 给出一般均衡存在的第一个数学证明, 后来 Arrow and Debreu (1954) 提供了更为严格的证明. 这些证明将一般均衡的存在性视为一个不动点问题, 然后运用数学不动点定理给予证明. 这些数学证明建立了严谨的一般均衡理论, 但并没有给出如何求解一般均衡价格的方法. Scarf (1967) 开创性提出的不动点算法, 能够将不动点计算出来, 这不仅推动经济数学的发展, 也奠定可计算一般均衡理论的坚实基础, 使一般均衡论具有广泛的实际应用价值. 可计算一般均衡模型常用于政策评估.

众所周知, 数学在经济学研究中扮演着重要的方法与工具作用, 然而当人类经济社会系

统变得过于复杂时, 数学无法刻画经济主体的行为, 更谈不上求解结果, 尤其对于高维复杂系统, 数学往往力不从心. 在这种情形下, 可以采用计算机模拟仿真的方法与技术研究经济行为. 在这个领域, 一个广泛流行的模拟方法是基于经济主体的计算经济学模型, 这个方法综合应用经济学、仿生学和计算机技术 (特别是算法) 来研究经济主体的行为特征与复杂经济系统的运行规律, 其中一个著名例子是基于行为主体的模型 (agent-based model, ABM), 这个模型使用计算机模拟仿真技术, 研究遵从简单规则的大量经济主体的行为如何产生复杂的宏观经济特征 (如泡沫).

随着大数据与大模型日益广泛的应用, 对算法特别是有效算法的需求越来越大. 在计量经济学, 理论计量经济学家长期专注研究计量经济学模型与方法的数学性质, 特别是概率性质与统计性质, 如统计估计量与检验量的一致性、有效性、渐近分布等大样本理论, 对计算机算法的研究相对较少, 也不擅长算法的研究, 这是计量经济学与经济学在大数据与大模型时代的一个明显短板. 在 20 世纪 90 年代, 许多应用经济学家将人工神经网络模型应用于经济金融预测. 尽管大多数实证结果都具有样本内统计显著性, 但人工神经网络模型的样本外预测能力却比较差, 特别是如果采用一些模型选择准则, 如 AIC 或 BIC 准则, 通常会发现简单的线性回归模型的样本外预测会比人工神经网络模型要好得多. 现在看来, 有几方面的原因: 首先, 大多数计量经济学家擅长数学推导, 但不擅长计算机算法. 其次, 在 90 年代, 许多经济金融数据大都是小数据, 还没有可用的海量大数据, 甚至大数据概念也不存在. 最后, 最根本的一个原因, 是算力与算法的限制. 在 20 世纪 90 年代, 计算机的算力无法支持使用海量数据训练人工神经网络模型. 因此, 那时的人工神经网络模型在经济金融预测的表现不理想, 这并不是因为该模型不好, 而是由于当时数据、算法、算力等都存在许多客观限制. 可以预计, 以算法为核心的计算经济学与计算计量经济学将进入一个快速发展时期.

计算机科学一个重要的研究范式是定量经验研究, 即通过基于算法的计算机模拟仿真实验与实际应用的方式研究计算机算法的性质及其应用. 这种经验研究范式推动了算法、大模型计算机科学的快速发展, 这值得经济学家和计量经济学家参考与借鉴.

5 人工智能与 ChatGPT 的局限性

目前, 人工智能特别是以 ChatGPT 为代表的大模型备受推崇, 但这并不意味着包括 ChatGPT 在内的各种人工智能技术不存在缺点与不足. 洪永淼和汪寿阳 (2023) 对此进行了讨论. 这里, 我们进一步深入阐述人工智能作为研究方法与研究工具在经济学研究中的局限性.

首先, 人工智能是模仿人类认知过程而发展起来的一种数字技术, 它具有卓越的理性思维能力, 可以模拟人类的认知思维过程, 但却无法感知人类的直觉和情感, 如幸福、快乐、疼痛与悲伤等. 我们常说人工智能可以“理解”文本数据, 但实际上它并没有人的意识或理解能力. 例如, 当一个人听到“中国”这个词时, 便可能会想起它作为一个国家拥有的 960 万平方公里的领土面积、56 个民族、14 亿人口, 以及 5000 年的悠久历史. 但人工智能却无法像人类那样理解“中国”的内涵. 人工智能看到“中国”这个词时, 它用一个高维向量来刻画, 并从大量的文本数据中寻找出现该词以及相关单词或词组的概率分布, 据此进行预测. 毫无疑问, 在理性思维方面, 人工智能可能会超过人类, 但在感性思维方面, 它是有所欠缺的. 正如

ChatGPT 在接受《时代周刊》采访时 (Perrigo, 2022) 所说: “作为大语言模型, 我只是一个机器学习模型, 没有人类的知觉或意识. 我没有思想、感觉或过往, 无法做出决策或独立判断”. “我们无法理解所生成词汇的上下文语境或含义, 只能根据给定的训练数据, 根据某些单词或词组一起出现的概率生成文本”. 由于人工智能只具备预测能力, 而没有人类那样的思维意识、批判性思维和想象力, 所以经济学家和计量经济学家从事的创造性研究工作不可能被人工智能替代, 特别是人工智能 (包括 ChatGPT) 无法从现有数据推断出任何重大原创性的学术成果, 它只能在人类已有的海量知识和信息库中产生各种排列组合.

人类学习与机器学习还有一个重要的区别: 由于人工智能是基于概率预测, 因此需要足够多的数据才能获得较好的学习能力, 但人的学习能力不一定要经过多次重复才能获得. 例如, 一个小孩用手去触摸一个烧开冒气的水壶, 他只要摸过一次就不会再摸第二次, 不管水壶的形状、颜色是什么. 相反地, 机器需要“看到”很多水壶的照片或视频, 才会得到不要去触摸烧开的水壶的结论. 正如美国语言学家诺姆·乔姆斯基 (Chomsky et al., 2023) 所指出的, “人脑与 ChatGPT 之类的工具不同, 它不是一个笨拙的模式匹配统计引擎, 先狂塞进数百 TB 级数据, 再推测出可能性最大的对话答复或某个科学问题的答案. 相反, 人脑是一个非常高效甚至优雅的系统, 只需要少量信息即可运作; 它不推断数据点之间的直接关联, 而是创造解释.”

第二, 经济学实证研究最主要的目的是因果推断. 长期以来, 人类的思维方式一直是通过逻辑推断, 特别是因果推断, 来认识世界和改造世界. 所谓因果推断, 是指在控制所有其他解释变量不变的条件下, 观察某一解释变量的变化是否会引起因变量的变化. 例如, 只让政策变量发生变化, 观察 GDP 是否会随之变化. 实验方法, 特别是随机控制实验, 是识别经济学因果关系的最直观、最有力的方法. 但遗憾的是, 在人类经济社会活动中, 绝大部分的经济数据是基于观测而得, 并非精心设计的实验结果.

人工智能可以用于因果推断 (Athey, 2019; Athey and Imbens, 2019). 由于其出色的泛化能力, 人工智能能够比较精确地估测虚拟事实, 从而显著提升因果推断与政策评估的有效性与精准性. 但是, 大数据特别是经济大数据几乎都是观测数据, 包括大模型在内的人工智能算法事实上是统计学方法, 而基于观测数据、运用统计学与计量经济学方法进行因果推断本质上是一种统计关系推断 (Leamer, 1983). 这种统计关系需要在增加很多额外假设的条件下才可以被解释为经济学因果关系. 如果放弃这些假设, 或者这些假设不成立, 这种关系只能视为一种统计关系. 基于观测数据的人工智能因果关系本质上只是在计算机 (即电脑) 中两个变量之间的统计关联, 即两个事件或两个变量同时或先后出现的概率非常高, 计算机基于大数据能够捕捉到这种统计关联, 并将其推断为因果关系. 因此, 人工智能因果关系类似于英国哲学家休谟所定义的因果关系. 休谟认为, 每次太阳升起, 石头变热, 这两个现象日复一日共现在人的脑海中, 形成了一种惯常的联系, 这种联系被休谟称为因果关系. 因此, 人工智能因果关系和经济学因果关系并非完全一致. 要识别经济学因果关系, 不能仅仅依靠基于计量经济学与人工智能的因果推断方法, 还必须引入实验经济学 (如随机控制实验) 等方法. 如果要将基于观测数据的人工智能因果关系解释为经济学因果关系, 则需要经济理论的指导, 而且, 这种因果关系解释也只是一种可能性, 即人工智能因果关系与经济学因果关系是相容的或不互相排斥, 但并不意味人工智能因果关系就一定是经济学因果关系.

第三, 人工智能特别是大模型正在推动经济学与社会科学研究范式的变革, 特别是从模型驱动范式到数据驱动范式的转变, 但仍需要与经济理论相结合, 增强算法的经济可解释性. 任何模型都是建立在各种假设基础上, 模型驱动范式所获得的结论很可能具有模型依赖性, 致使不同模型可能导致不同的结论 (Brezna et al., 2022). 这就是为什么那么多经济学实证研究都要考虑所谓的稳健性检验 (robustness check), 即检验在不同的模型条件下是否能够得到相同或类似的结论. 相反地, 数据驱动范式试图通过使用与具体模型无关的算法, 通过算法直接从数据中获得经济变量之间的逻辑关系, 以得到与具体模型无关的稳健结论. 但是, 数据驱动范式所依赖的算法特别是大模型, 大多是“黑箱”, 缺乏经济可解释性, 这是人工智能特别是大模型最大的一个弱点. 数据驱动范式必须与经济理论相结合, 才能拥有经济可解释性. 因此, 数据驱动模式可获得比较稳健的结论, 但它并不能取代经济理论的指导.

第四, 人工智能和以 ChatGPT 为代表的大模型没有改变经济学乃至社会科学实证研究的本质特征, 即从样本推断总体性质的归纳范式. 人工智能特别是大模型强化了以数据为基础的实证研究范式, 但这种人工智能归纳推理在实践中依然可能会出现错误, 类似于统计学的第一类错误和第二类错误 (洪永淼, 2021). 需要注意的是, 虽然 ChatGPT 和其他大模型使用了海量大数据甚至是几乎整个互联网的文本数据, 样本容量极其庞大, 但这并非全样本. 人类经济社会发展是一个漫长的历史过程, 现有互联网大数据虽然样本容量庞大, 却仅代表这个历史过程的一部分. 此外, 人类经济社会发展在各个历史阶段会呈现出一些独特的特征, 导致经济运行规律具有历史性或时变性特征, 这使人工智能在推断与预测人类经济社会未来变化趋势仍然面临巨大的挑战, 人工智能没有改变经济学作为一门历史科学的本质特征.

第五, 人工智能以及 ChatGPT 等大模型所使用的数据是来源于互联网的开源数据. 互联网开源数据中存在大量不真实和虚假的信息, 也包含很多违背当今社会伦理与道德规范的内容, 还可能出现样本选择偏差的情况. 例如, 如果有关某个事件的文本数据只有少数几篇文章, 而这些文章根本无法代表社会大多数人的观点, 只能代表少数人的意见, 当人工智能分析这些数据时, 很可能将这些个别人的意见视为是代表性意见, 导致偏见甚至“幻觉”的出现. 互联网大数据也可能存在“样本选择偏差”问题, 如“数字鸿沟”导致的样本偏差. 互联网大数据无法反映不上网或很少上网的群体的经济行为信息, 这个群体可能包括了年长者、西部地区和农村地区的群众, 数字鸿沟在这些情况下是不可避免的. 因此, ChatGPT 提供的各种答案, 其真实性、客观性以及代表性无法得到保证, 其根本原因就在于, 它所使用的开源数据的真实性与代表性无法进行验证. 由于数据质量原因而产生的风险, 可称为数据风险. 随着数字经济的蓬勃发展, 数据成为关键的生产要素, 因此必须防范由于数据可靠性问题而引起的数据风险. 数据可靠性与数据质量是约束包括大模型在内的人工智能可靠性的最基本因素.

第六, 除了大数据可靠性问题引起的数据风险外, 还存在由于模型或算法本身的泛化能力问题而产生的模型风险或算法风险. 造成模型风险或算法风险的原因很多, 包括大模型的训练算法与训练时间、算法本身的有效性等, 但最主要的一个原因是数据漂移 (data drift) 和模型漂移 (model drift) 所引起的模型或算法泛化能力下降的问题. 技术进步、人口结构变化、偏好变化、政策变化以及诸如新冠疫情大流行和地缘政治军事冲突等重大外生冲击, 都会引起经济主体的预期与行为的变化, 从而导致经济结构与经济关系发生变化. 上述变化

首先会引起大数据的分布变化,这种变化可能是突变式的,也可能是渐进式的,或者两者兼之。比如,在文本数据中,随着经济社会快速变化与发展,很多新的网络语言不断出现,而一些旧网络词语也会被赋予新含义,从计量经济学与统计学视角看,这些互联网文本数据具有显著的时变性或平稳性,这种数据时变性称为数据漂移。由于漂移数据与原先用于训练模型的数据可能存在显著差异,数据漂移会导致模型或算法的样本外预测能力即泛化能力的弱化。除了引起数据漂移之外,技术、人口、偏好、政策变化以及外生冲击等因素还会引起经济变量之间的关系发生变化。由于模型或算法是基于已知数据特别是历史数据进行训练,模型无法刻画经济关系的新变化,从而影响模型或算法的泛化能力,这种现象称为模型漂移。尤其是大模型,由于其训练数据量庞大、训练时间长、训练成本高,无法经常或及时更新训练,在刻画变量之间关系的新变化方面存在一定的时滞,模型漂移的程度可能会更加严重,导致大模型的泛化能力显著下降。目前,ChatGPT 和大模型正在日益广泛地应用于经济社会各个领域,由数据漂移和模型漂移而产生的模型风险或算法风险正在成为一种日益重要的新风险。这种模型风险或算法风险不仅对经济主体的行为和经济运行产生深远的影响,也对经济学乃至社会科学研究范式的变革具有重要影响。

6 总结与展望

我们正处于大经济、大科技、大数据、大模型的新时代。数字技术革命与新一轮工业革命对人类生产方式、生活方式与思维方式产生重大变革,也在深刻改变社会治理方式。数字经济特别是大科技平台等新经济形态的出现,与中国超大经济规模的优势相结合,催生了新的经济运行模式与运行规律。例如,很多数字技术包括各种算法,一旦开发出来并投入市场,其使用的边际成本便几乎为零。数字经济每时每刻产生海量经济社会大数据,其中互联网和移动互联网的兴起与广泛应用,涌现出海量的消费和流通大数据,而当下物联网和工业互联网的蓬勃发展,正在产生海量生产与供应端大数据,这些大数据既是数字经济关键的生产要素,也为经济学、管理学乃至社会科学研究提供大量素材。挖掘、处理、分析海量大数据的主要工具是人工智能技术。以 ChatGPT 为代表的大模型是人工智能在自然语言处理技术领域的一大突破,对经济学与社会科学研究范式将产生深远的影响。

长期以来,经济学理论构建与创新主要是通过经济学建模与计量经济学建模。经济学建模大多建立在关于偏好、技术、人口、资源、制度、行为等因素的假设基础之上,通过数学等逻辑工具的演绎推理,聚焦少数关键经济变量,研究经济主体行为与经济运行规律。这种小模型范式能够聚焦所研究问题的最本质的特征与内在联系,模型的经济可解释性也比较强。但是,由于人类经济社会系统是一个高度复杂系统,其影响因素成千上万,且它们之间的关系错综复杂,并且具有显著的异质性、交互性、非线性、时变性等特征。为了更好刻画复杂经济社会系统的动态演变规律,精准预测其未来发展变化趋势,应该采用系统分析方法。作为大数据与人工智能时代的一种新的系统分析方式,大模型可以弥补小模型范式的不足与局限性,为经济学与社会科学研究提供新的研究范式与研究方法。

热力学与统计物理学的发展历史为我们提供了有益的启示。在 19 世纪和 20 世纪之交,整个经典物理学理论大厦已构建起来,物理学家曾欢呼物理学晴朗的天空只飘着两朵乌云,其中之一是热辐射问题。物理学家威廉·维恩 (Wilhelm Wien) 通过实验结果得到维恩公式,

该公式可以很好拟合短波范围的黑体辐射数据,但在长波范围内拟合效果不佳.另外两位物理学家瑞利勋爵 (Lord Rayleigh) 和詹姆斯·金斯 (James Jeans) 根据经典电动力学和统计力学推导出热平衡辐射能量分布公式,该公式适用于长波范围的黑体辐射数据,但在短波范围内拟合效果不佳.后来,马克斯·普朗克 (Max Planck) 提出了黑体辐射定律的通用公式,既适用于短波范围也适用于长波范围,完美地解释两种波长范围内的实验数据.长期以来,经济学与计量经济学的实证研究几乎只使用小模型.在大数据与人工智能时代,应该尝试大模型范式,探索以 ChatGPT 为代表的大模型对经济学研究范式可能带来的影响.应当指出,强调大模型并不意味着小模型就不重要,因为它们适用范围不同,就像分别适合于短波、长波的黑体辐射公式那样.我们应该将小模型范式和大模型范式辩证统一起来,以获得对复杂经济社会系统运行规律的深刻认识.

参 考 文 献

- 乔舒亚·安格里斯特, 约翰-施特芬·皮施克, (2012). 基本无害的计量经济学: 实证研究者指南 [M]. 郎金焕, 李井奎, 译. 上海: 格致出版社.
- Angrist J D, Pischke J S, (2012). Mostly Harmless Econometrics: An Empiricist's Companion[M]. Shanghai: Truth & Wisdom Press.
- 程兵, (2023). 以 ChatGPT 为代表的大语言模型打开了经济学和其他社会科学研究范式的巨大新空间 [J]. 计量经济学报, 3(3): 589-614.
- Cheng B, (2023). Artificial Intelligence Generative Content (AIGC) Including ChatGPT Opens a New Big Paradigm Space of Economics and Social Science Research[J]. China Journal of Econometrics, 3(3): 589-614.
- 洪永淼, (2021). 概率论与统计学 [M]. 第 2 版. 北京: 中国统计出版社.
- Hong Y M, (2021). Probability and Statistics for Economists[M]. 2nd ed. Beijing: China Statistics Press.
- 洪永淼, 史九领, (2024). 人工智能的政治经济学分析 [J]. 学术月刊, 56(656): 43-59.
- Hong Y M, Shi J L, (2024). Political Economy Analysis of Artificial Intelligence[J]. Academic Monthly, 56(656): 43-59.
- 洪永淼, 汪寿阳, (2020). 数学、模型与经济思想 [J]. 管理世界, (10): 15-27.
- Hong Y M, Wang S Y, (2020). Mathematics, Model and Economic Thought[J]. Journal of Management World, (10): 15-27.
- 洪永淼, 汪寿阳, (2021a). 大数据、机器学习与统计学: 挑战与机遇 [J]. 计量经济学报, 1(1): 17-35.
- Hong Y M, Wang S Y, (2021a). Big Data, Machine Learning and Statistics: Challenges and Opportunities[J]. China Journal of Econometrics, 1(1): 17-35.
- 洪永淼, 汪寿阳, (2021b). 大数据如何改变经济学研究范式?[J]. 管理世界, (10): 40-55.
- Hong Y M, Wang S Y, (2021b). How is Big Data Changing Economic Research Paradigms?[J]. Journal of Management World, (10): 40-55.
- 洪永淼, 汪寿阳, (2023). 人工智能新近发展及其对经济学研究范式的影响 [J]. 中国科学院院刊, 38(3): 353-357.
- Hong Y M, Wang S Y, (2023). Impacts of Cutting-edge Artificial Intelligence on Economic Research Paradigm[J]. Bulletin of Chinese Academy of Sciences, 38(3): 353-357.

- 洪永淼, 刘倬奇, 薛润坡, (2023). 政府与市场心理因素的经济影响及其测度 [J]. 管理世界, (3): 30–51.
Hong Y M, Liu F Q, Xue J P, (2023). Measurement of Psychological Factors and Their Economic Impact in the Relationship between Government and Market[J]. Journal of Management World, (3): 30–51.
- 李拉亚, (2014). 经济学计算机化研究进展 [J]. 经济学动态, (1): 130–137.
Li L Y, (2014). Progress in Computational Economics Research[J]. Economic Perspectives, (1): 130–137.
- 李戎, 刘岩, 彭俞超, 许志伟, 薛润坡, (2022). 动态随机一般均衡模型在中国的研究进展与展望 [J]. 经济学(季刊), (6): 1829–1846.
Li R, Liu Y, Peng Y C, Xu Z W, Xue J P, (2022). Dynamic Stochastic General Equilibrium Research in China: A Review and Prospect[J]. China Economic Quarterly, (6): 1829–1846.
- 李志军, 尚增健, (2020). 学者的初心与使命: 学术研究与论文写作中的“数学化”“模型化”反思 [M]. 北京: 经济管理出版社.
Li Z J, Shang Z J, (2020). The Original Aspiration and Mission of Scholars: Reflections on “Mathematization” and “Modeling” in Academic Research and Writing[M]. Beijing: Economy & Management Publishing House.
- 列宁, (1990). 列宁全集(第55卷, 黑格尔〈逻辑学〉一书摘要)[M]. 北京: 人民出版社.
Lenin V I, (1990). Collected Works of Lenin (Vol.55, Excerpts from Hegel’s <The Science of Logic>)[M]. Beijing: People’s Publishing House.
- 刘景江, 郑畅然, 洪永淼, (2023). 机器学习如何赋能管理学研究? ——国内外前沿综述和未来展望 [J]. 管理世界, (9): 191–216.
Liu J J, Zheng C R, Hong Y M, (2023). How Can Machine Learning Empower Management Research? A Domestic-Foreign Frontier Review and Future Prospects[J]. Journal of Management World, (9): 191–216.
- 那艺, 贺京同, (2017). 从“宏观经济学的麻烦”看行为宏观经济学的兴起与发展 [J]. 经济学动态, (7): 100–114.
Na Y, He J T, (2017). On the Rise and Development of Behavioral Macroeconomics from “The Trouble with Macroeconomics”[J]. Economic Perspectives, (7): 100–114.
- 钱学森, 于景元, 戴汝为, (1990). 一个科学新领域——开放的复杂巨系统及其方法论 [J]. 自然杂志, (1): 3–10.
Qian X S, Yu J Y, Dai R W, (1990). A New Discipline of Science — The Study of Open Complex Giant System and Its Methodology[J]. Chinese Journal of Nature, (1): 3–10.
- 汪寿阳, 李明琛, 杨昆, 林文灿, 姜尚荣, 等, (2023). ChatGPT+金融: 八个值得关注的研究方向与问题 [J]. 管理评论, (4): 3–11.
Wang S Y, Li M C, Yang K, Lin W C, Jiang S R, et al. (2023). ChatGPT+Finance: Eight Noteworthy Research Directions and Issues[J]. Management Review, (4): 3–11.
- Aït-Sahalia Y, Jacod J, (2014). High-Frequency Financial Econometrics (Illustrated Edition)[M]. Princeton: Princeton University Press.
- Akerlof G A, (2002). Behavioral Macroeconomics and Macroeconomic Behavior[J]. American Economic Review, 92(3): 411–433.
- Akerlof G A, (2007). The Missing Motivation in Macroeconomics[J]. American Economic Review, 97(1): 5–36.
- Angeletos G M, Huo Z, (2021). Myopia and Anchoring[J]. American Economic Review, 111(4): 1166–

1200.

- Angeletos G M, Huo Z, Sastry K A, (2021). Imperfect Macroeconomic Expectations: Evidence and Theory[J]. *NBER Macroeconomics Annual*, 35: 1–86.
- Angeletos G M, Lian C, (2022). Confidence and the Propagation of Demand Shocks[J]. *The Review of Economic Studies*, 89(3): 1085–1119.
- Angrist J, Azoulay P, Ellison G, Hill R, Lu S F, (2017). Economic Research Evolves: Fields and Styles[J]. *American Economic Review*, 107(5): 293–297.
- Arrow K J, Debreu G, (1874). Existence of an Equilibrium for a Competitive Economy[J]. *Econometrica*, 22: 265–290.
- Athey S, (2019). The Impact of Machine Learning on Economics[M]// *Economics of Artificial Intelligence: An Agenda*. Chicago: University of Chicago Press: 507–547.
- Athey S, Imbens G W, (2019). Machine Learning Methods that Economists Should Know About[J]. *Annual Review of Economics*, 11: 685–725.
- Bañbura M, Giannone D, Reichlin L, (2010). Large Bayesian Vector Auto Regressions[J]. *Journal of Applied Econometrics*, 25(1): 71–92.
- Basu S, Michailidis G, (2015). Regularized Estimation in Sparse High-dimensional Time Series Models[J]. *Annals of Statistics*, 43(4): 1535–1567.
- Belkin M, Hsu D, Ma S, Mandal S, (2018). Reconciling Modern Machine-Learning Practice and the Classical Bias-variance Trade-off[J]. *Proceedings of the National Academy of Sciences*, 116: 15849–15854.
- Bernanke B S, Boivin J, Eliasz P, (2005). Measuring the Effects of Monetary Policy: A Factor-augmented Vector Autoregressive (FAVAR) Approach[J]. *Quarterly Journal of Economics*, 120(1): 387–422.
- Breiman L, (2001). Statistical Modeling: The Two Cultures (With Comments and a Rejoinder by the Author)[J]. *Statistical Science*, 16(3): 199–231.
- Breznau N, Rinke E M, Wuttke A, Nguyen H H, Adem M, et al. (2022). Observing Many Researchers Using the Same Data and Hypothesis Reveals a Hidden Universe of Uncertainty[J]. *Proceedings of the National Academy of Sciences*, 119(44): e2203150119.
- Cass D, (1965). Optimum Growth in an Aggregative Model of Capital Accumulation[J]. *The Review of Economic Studies*, 32(3): 233–240.
- Chan J C, (2023). Large Hybrid Time-Varying Parameter VARs[J]. *Journal of Business & Economic Statistics*, 41(3): 890–905.
- Choi S, Shin J, Yoo S Y, (2022). Are Government Spending Shocks Inflationary at the Zero Lower Bound? New Evidence from Daily Data[J]. *Journal of Economic Dynamics and Control*, 139: 104423.
- Chomsky N, Roberts I, Watumull J, (2023). Noam Chomsky: The False Promise of ChatGPT[J]. *The New York Times*, 8.
- Cui L, Feng G, Hong Y, (2023a). Regularized GMM for Time-Varying Models with Applications to Asset Pricing[J]. *International Economic Review*: 10.1111/iere.12678.
- Cui L, Hong Y, Yao J, (2023b). Estimating High-Dimensional Multivariate GARCH Models: A Nuclear Norm-based Regularized QMLE Approach[R]. Working Paper.
- Engle R, (2002). New Frontiers for ARCH Models[J]. *Journal of Applied Econometrics*, 17(5): 425–446.
- Engle R F, Russell J R, (1998). Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data[J]. *Econometrica*, 66(5): 1127–1162.
- Gentzkow M, Kelly B, Taddy M, (2019). Text as Data[J]. *Journal of Economic Literature*, 57(3): 535–574.

- Giannone D, Reichlin L, Small D, (2008). Nowcasting: The Real-time Informational Content of Macroeconomic Data[J]. *Journal of Monetary Economics*, 55(4): 665–676.
- Giles J, (2012). Making the Links[J]. *Nature*, 488(7412): 448–450.
- Glandon P J, Kuttner K, Mazumder S, Stroup C, (2023). Macroeconomic Research, Present and Past[J]. *Journal of Economic Literature*, 61(3): 1088–1126.
- Granger C W, (1980). Long Memory Relationships and the Aggregation of Dynamic Models[J]. *Journal of Econometrics*, 14(2): 227–238.
- Granger C W, (1999). *Empirical Modeling in Economics: Specification and Evaluation*[M]. Cambridge: Cambridge University Press.
- Hall R E, (1978). Stochastic Implications of the Life Cycle-permanent Income Hypothesis: Theory and Evidence[J]. *Journal of Political Economy*, 86(6): 971–987.
- Hastie T J, Montanari A, Rosset S, Tibshirani R J, (2019). Surprises in High-Dimensional Ridgeless Least Squares Interpolation[J]. *Annals of Statistics*, 50(2): 949–986.
- Hommes C, (2021). Behavioral and Experimental Macroeconomics and Policy Analysis: A Complex Systems Approach[J]. *Journal of Economic Literature*, 59(1): 149–219.
- Hong Y, Huang N, Wang Y, Zhao Z, (2023). Forecasting Inflation Rates Using a Large Panel of Individual Stock Prices[R]. Working Paper.
- Kaplan G, Violante G L, (2018). Microeconomic Heterogeneity and Macroeconomic Shocks[J]. *Journal of Economic Perspectives*, 32(3): 167–194.
- Kelly B, Xiu D, (2023). Financial Machine Learning[J]. *Foundations and Trends in Finance*, 13(3–4): 205–363.
- Keynes J M, (1936). *The General Theory of Employment, Interest, and Money*[M]. London: Macmillan.
- Koopmans T, (1963). On the Concept of Optimal Economic Growth[J]. *The Econometric Approach to Development Planning*: 38.
- Lanier J, Large J, Quah J, (2023). Estimating Very Large Demand Systems[R]. Working Paper, Institute for New Economic Thinking at the Oxford Martin School, University of Oxford.
- Lazer D, Pentland A, Adamic L, Aral S, Barabási A L, et al. (2009). *Computational Social Science*[J]. *Science*, 323(5915): 721–723.
- Leamer E E, (1983). Let's Take the Con Out of Econometrics[J]. *American Economic Review*, 73(1): 31–43.
- Lopez-Lira A, Tang Y, (2023). Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models[R]. Working Paper.
- Marcellino M, Stock J H, Watson M W, (2003). Macroeconomic Forecasting in the Euro Area: Country Specific Versus Area-wide Information[J]. *European Economic Review*, 47(1): 1–18.
- Nakkiran P, Kaplun G, Bansal Y, Yang T, Barak B, et al. (2021). Deep Double Descent: Where Bigger Models and More Data Hurt[J]. *Journal of Statistical Mechanics: Theory and Experiment*, (12): 124003.
- Okun A M, (1970). *The Political Economy of Prosperity (First Edition)*[M]. Manhattan: W. W. Norton & Company.
- Perrigo B, (2022). AI Chatbots are Getting Better. But an Interview with ChatGPT Reveals Their Limits[M]. New York: Time.
- Phillips A W, (1958). The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957[J]. *Economica*, 25(100): 283–299.

- Ramsey F P, (1928). A Mathematical Theory of Saving[J]. *The Economic Journal*, 38(152): 543–559.
- Sargent T J, (1993). *Bounded Rationality in Macroeconomics: The Arne Ryde Memorial Lectures*[M]. Oxford: Oxford University Press.
- Scarf H E, (1967). On the Computation of Equilibrium Prices[C]// *Cowles Foundation Discussion Papers* 232, Cowles Foundation for Research in Economics, Yale University.
- Shiller R J, (2001). Irrational Exuberance[J]. *Philosophy and Public Policy Quarterly*, 20(1): 18–23.
- Shiller R J, (2019). *Narrative Economics: How Stories Go Viral and Drive Major Economic Events*. Princeton: Princeton University Press.
- Simon H A, (1976). *Administrative Behavior: A Study of Decision-making Processes in Administrative Organization*[M]. 3rd ed. New York: Free Press.
- Sims C A, (1980). Macroeconomics and Reality[J]. *Econometrica*: 1–48.
- Stachurski J, Sargent T J, (2022). *Economic Networks: Theory and Computation*[M]. *QuantEcon Book* I. Manuscript. <https://networks.quantecon.org/>.
- Stock J H, Watson M W, (1996). Evidence on Structural Instability in Macroeconomic Time Series Relations[J]. *Journal of Business and Economic Statistics*, 14(1): 11–30.
- Stock J H, Watson M W, (2016). Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics[M]// Taylor J B, Uhlig H. *Handbook of Macroeconomics*. Amsterdam: Elsevier: 415–525.
- Sun Y, Hong Y, Lee T H, Wang S, Zhang X, (2021). Time-varying Model Averaging[J]. *Journal of Econometrics*, 222(2): 974–992.
- Sun Y, Hong Y, Wang S, Zhang X, (2023). Penalized Time-varying Model Averaging[J]. *Journal of Econometrics*, 235(2): 1355–1377.
- Taylor J B, (1993). Discretion Versus Policy Rules in Practice[J]. *Carnegie-Rochester Conference Series on Public Policy*, 39: 195–214.
- Tibshirani R, (1996). Regression Shrinkage and Selection via the Lasso[J]. *Journal of Royal Statistical Society Series B: Statistical Methodology*, 58(1): 267–288.
- Varian Hal R, (2014). Big Data: New Tricks for Econometrics[J]. *Journal of Economic Perspectives*, 28(2): 3–28.
- Wald A, (1936). Über Einige Gleichungssysteme der Mathematischen Ökonomie[J]. *Zeitschrift für Nationalökonomie/Journal of Economics*, 7: 637–670.
- Walras L, (1874). *Elements of Pure Economics*[M]. London: Routledge.
- White H, (1992). *Artificial Neural Networks: Approximation and Learning Theory*[M]. Oxford: Blackwell.
- Zellner A, Keuzenkamp H A, McAleer M, (2001). *Simplicity, Inference and Modelling: Keeping It Sophisticatedly Simple*[M]. Cambridge: Cambridge University Press.