

· 科学论坛 ·

大数据时代下计量经济学若干重要发展方向*

汪寿阳¹ 洪永淼² 霍红³ 方颖⁴ 陈海强^{4**}

- (1. 中国科学院 数学与系统科学研究院, 中国科学院 预测科学研究中心, 北京 100037;
2. 康奈尔大学 经济系, 美国; 3. 国家自然科学基金委员会 管理科学部, 北京 100037;
4. 计量经济学教育部重点实验室(厦门大学), 厦门大学 王亚南经济研究院, 经济学院, 厦门 361005)

[摘要] 基于第206期“双清论坛”, 本文回顾并总结了大数据时代新的数据形式和数据变量之间新型复杂关系给计量经济学带来的机遇和挑战, 认为国内学者应该充分利用“互联网+”战略和数字经济快速发展背景下我国在大数据资源积累方面的优势, 在宏观数据复杂关系建模和预测、高维度和非结构化数据建模、时空数据建模、大数据政策评估计量分析以及现代网络生态环境下金融数据建模等领域加强研究, 力争取得原创性重大理论突破, 服务国家重大战略需求。

[关键词] 大数据时代; 复杂数据; 计量建模; 政策评估

DOI:10.16262/j.cnki.1000-8217.2019.04.013

改革开放以来, 我国经济建设不断取得重大成就, 人民生活水平持续改善, 我国社会主要矛盾已经转化为人民日益增长的美好生活需求和不平衡不充分的发展之间的矛盾。但另一方面, 世界政治、经济形势依然动荡不安, 一些发达国家的经济增长仍然没有恢复到经济危机前的水平, 新兴经济体的潜在风险巨大, 一系列“黑天鹅”事件在全球化背景下极易转化为世界性的政治经济危机。国内则正处于产业结构转型升级的关键阶段, 在外部环境不断变化下, 影子银行、房地产泡沫、地方融资平台、国有企业高杠杆等“灰犀牛”事件有可能在经济发展放缓的情况下转化为不同程度的系统性经济风险。为此, 非常有必要对我国经济发展和经济金融风险进行长期的跟踪、分析和预判, 对各项社会公共政策包括宏观政策、产业政策和金融政策等进行科学的量化评估与分析。因此, 针对当前国家经济和社会发展的重大需求, 大力发展计量经济学的前沿理论与方法, 提倡以科学方法对经济管理和公共政策评估开展量化研究, 对于推进国家治理体系和治理能力现代化, 守住不发生系统性经济金融风险底线具有重要意义。

计量经济学作为经济学领域的一个重要分支, 主要以社会经济活动的实际数据为素材, 以统计分析方法为手段, 以预测和识别因果关系为目标, 为经济管理的实证研究和量化分析提供理论基础和方法工具。由于社会经济数据的非实验性质, 利用计量建模和量化分析的科学研究方法显得尤其重要。一方面, 计量经济学通过对实际数据的科学分析, 为各种互为竞争的经济理论提供“证伪”的可能; 另一方面, 通过对经济管理模型的量化分析, 为政策实施和政策评估提供科学决策的基础。因此, 计量经济学在推动经济学科学化过程中发挥了不可替代的关键作用。虽然计量经济学最初以社会经济数据的分析为主, 但随着量化分析的重要性在各学科领域的日益凸显, 计量经济学的分析工具和分析方法已逐渐渗透到心理学、医疗管理、公共管理、金融工程、能源管理、社会学和政治学等各个领域, 成为发展非常活跃、应用日益普及的方法论学科和交叉学科。

在大数据时代, 新的数据形式和数据中变量间的新型复杂关系给计量经济学带来前所未有的挑战, 但另一方面也给计量经济学的发展和原创性的

收稿日期: 2019-03-14; 修回日期: 2019-04-11

* 本文根据第206期“双清论坛”的研讨整理。

** 通信作者, Email: hc335@xmu.edu.cn

理论突破带来千载难逢的机遇。在复杂数据计量建模方面我国与西方国家处在同一起跑线上。换言之,计量建模是目前国内经济学与管理学领域中少数能够进入国际前沿的学科之一,一批中国学者已经跻身国际学术前沿。

在此背景下,2018年8月9~10日,国家自然科学基金委员会(以下简称自然科学基金委)管理科学部、数理科学部、信息科学部与政策局联合召开了主题为“大数据时代计量经济学前沿理论、方法与应用”的第206期双清论坛,来自美国、香港等国家和地区的本领域知名专家以及自然科学基金委数理科学部、信息科学部和政策局等相关工作人员共50余名代表参加了论坛。论坛包括5个大会报告和3个分会场的主题小组报告。其中,美国南加州大学萧政教授、北京大学陈松蹊教授、香港城市大学马跃教授、中国科学院预测科学研究中心汪寿阳教授和美国康奈尔大学洪永淼教授分别作大会主题报告,分别围绕“宏观计量理论与应用”、“大数据理论、方法与应用”和“微观计量理论与应用”三个主题凝练和提出我国在该研究领域急需关注和解决的重要基础科学问题以及相应的跨学科资助模式,研讨今后5~10年的重点资助方向。

1 大数据时代下计量建模面临的机遇和挑战

大数据给已有计量经济学理论和方法所带来的深刻挑战主要体现在以下几个方面:

(一)大数据在具有海量信息优势的同时,又具有信息价值密度低的特点。在利用大数据进行计量经济分析时,给定样本中可供分析使用的变量维度会很高,甚至出现远远高于样本量的情况。如何更有效地筛选信息因而也成为大数据计量经济分析所面临的一个重要挑战。

(二)大数据一方面表现在可得数据形式的多元化,除了传统的结构化数据外,还包括文本数据、音频数据、视频数据等非结构化数据,甚至包括任何可以电子化记录的信息;另一方面,传统的结构化数据的形式也日益多元化,从简单的点数据,扩展到区间数据、符号数据和函数型数据等。因此如何对信息含量丰富、数据形式多层次化的区间数据、符号数据和函数型数据进行计量建模,是富有挑战性的研究工作,同时也对经济预测和政策监管有着重要的现实意义。在计量经济学理论与应用中,专门针对

区间数据和函数型数据建模的相应研究也是方兴未艾,引起学界和业界的更多关注。

(三)大数据不仅表现为数据形式的多样化,还主要表现为变量之间关系的复杂化。大数据时代数据特征变化加快,数据收集手段多样化,收集频率愈加密集,经济变量之间更容易表现出时变性、非线性和非平稳性的特点。在宏观经济数据和金融数据分析中,时变性、非线性和非平稳性日益成为主要的特征事实。已有的计量经济建模方法不能很好地刻画经济变量之间的复杂关系,从而严重制约了计量建模在宏观经济预测和实时监控中的有效性,这也是亟需解决的研究难点之一。

(四)大数据时代“互联网+”的迅猛发展也导致了新的网络型数据的产生。从微观个体的角度而言,以互联网为基础的社交媒体数据的产生对于研究个人行为,以及社交群体对于个人行为的影响提供了重要的数据基础。从宏观角度而言,随着金融科技的发展以及数据可得性的增加,以金融机构间复杂交易网络为基础的金融网络数据对于研究金融风险传染和金融风险管理具有重要的意义。但是,新兴的网络型数据给已有计量经济理论与方法带来了新的挑战,对于网络数据建模、网络形成的建模,以及网络稀疏性处理等关键问题都需要更深入的理论研究和更多的应用尝试。

2 计量建模现状与发展趋势

2.1 宏观数据复杂关系建模和预测

现代宏观数据建模和预测方法的发展趋势主要体现在如下四个方面。

首先,宏观经济数据多呈现出非平稳和时变等复杂特征。其中,非平稳性主要指数据的高持久性,即一阶序列相关性为1或接近于1。主流的宏观经济变量,如国内生产总值(GDP)、货币供应量、消费者价格指数(CPI)、利率以及汇率等都具有这个特征。经济学家通常采用单位根(unit root)或渐近单位根来描述以上变量的动态特征,但单位根仅局限于线性结构,这一缺陷使其在描述实际数据中的非平稳动态结构时往往力有不逮,当前时间序列计量经济学的一个重要研究方向是要提出新的更加广义的非平稳序列模型来描述这些复杂动态特征。而时变性主要指宏观经济数据结构出现时变,主要有两种表现形式,包括突变式结构变化和渐进式结构变化,不同的结构性变化对应着不同的模型假设,模型

假设错误则会导致统计推断产生偏误,影响预测效果。因此,准确预测宏观经济走势往往依赖是否能够准确捕捉其中的时变结构。

其次,宏观数据建模还需要适当处理数据中的趋势成分。在对趋势数据进行回归分析时,经济学家往往采用差分法或滤波法去除数据中的趋势成分,使数据平稳化,之后再对其进行建模。但这些传统方法会导致数据中有效信息的大量流失,尤其在数据结构不清晰的情形下,破坏数据本身结构,带来其他计量经济学难题。Phillips^[1]、White 和 Granger^[2]为经济学家应如何选择和运用趋势回归模型提供了指导性意见,认为在对复杂时间趋势的建模中,由于现实中具体数据所包含的趋势结构是未知的,因而非参数建模方法是解决此问题的良方。在这个研究方向上,Gao 和 Hawthorne^[3]建立了一个半参数半线性(semi-parametric partially linear)形式的趋势回归模型来研究全球年度平均气温时间序列数据。Phillips 和 Hansen^[4]考察了在回归变量为一阶单整 I(1) 情况下,最小二乘估计量和工具变量估计量的大样本性质,并提出了完全调整的误差修正法(fully-modified error-correction method)来改善估计量的小样本性质及其极限分布中由于内生性问题造成的偏差,从而提高统计推断的准确率。Cai 和 Wang^[5]考察了含有非平稳变量的预测回归模型中的内生性问题,并借鉴了 Amihud 和 Hurvich^[6]中提出的投影法来分解自变量和随机误差项之间的内生关系。

第三,宏观数据建模需要关注对区间数据的分析并将其用于经济政策设计,比如在设定经济增长目标的时候,不应该简单设置为一个特定值,而是一个增长区间,这样有更多的空间来处理未来不确定因素。理论上,区间样本数据比点样本在一段时间内含有更加丰富的信息,利用区间样本的信息优势可以得到更加有效的统计推断。比如,在金融资产波动性的研究中,一段时间内的价格区间是波动性测度,比传统的用收盘价进行价格纪录包含了资产价格波动的不确定性。针对区间数据建模,现有研究多集中在线性模型,包括对区间的中点和宽度建立向量自回归模型、Space-time 自回归模型以及区间 Informetric 模型。最新的区间数据建模进展主要集中在区间非线性建模,区间数据模式识别(机器学习)以及广义区间(上下边界大小无约束)建模三方面。在非线性的模型方面,Rodrigues 和

Salish^[7]和 Sun 等^[8]提出了区间门限模型,可以刻画不同机制下区间的动态特征。而随着机器学习技术的发展,区间数据的非参数建模也得到了大力发展,该类数据驱动方法在区间预测上具有较大优势。广义区间建模可以用来处理现实中存在区间上下边界逆序的情况,比如男女工资区间,风险资产与无风险资产价格区间可能在时间序列方面发生边界的逆序改变,Han, Hong 和 Wang^[9]率先构建了广义区间时间序列模型的计量理论与系统性分析框架,提出了对于区间模型的最小距离参数估计方法,并建立了极限分布理论。

最后,宏观经济预测需要利用好模型加权综合技术。模型不确定性广泛存在于计量建模、分析和预测的各个领域,基于多模型的方法是解决模型不确定性的主要途径,其通过一定的权重对来自不同的估计和预测进行加权,同时在进行参数推断时通过权重的随机性来考虑模型不确定性。在这些研究中,权重的选择非常重要,早期的研究主要基于信息准则的加权方法,即利用 AIC 或 BIC 值的指数构造权重进行加权。最新的研究提出了一些新的准则,比如 Yang^[10]引入 ARM(Adaptive Regression by Mixing)方法,Hansen^[11]提出了 MMA(Mallows Model Averaging)方法,该理论开创了最优模型平均研究领域。Kuersteiner 和 Okui^[12]利用类似方法为两阶段最小二乘估计和信息极大似然估计选择最优工具变量。Hansen 和 Racine^[13]针对异方差情形,提出了 JMA(Jackknife Model Averaging)方法,上述方法均被证明在某些特定情形下能够大幅改进宏观数据预测。进一步,在考虑宏观经济数据非平稳和时变特征情形下,如何进一步完善权重设计并对相关预测进行统计检验是未来的重要研究问题。

2.2 高维非结构化数据建模与应用

高维非结构化数据分析被广泛地应用到包括经济分析、市场营销、电子商务、信号处理等各个研究领域。英国皇家学会(2017)的机器学习研究报告指出 91% 的全球数据是在过去 5 年中创造出来的,在这个大数据的时代,包括众多的数据来源、有政府公告、社交网络、交易数据、GPS 导航等等。现在全球每天创造数据的速度是每天 250 亿 Gb 的惊人速度。

在处理大数据时,面临的挑战是主要有两个。首先 80% 以上的数据是非结构的(所谓的非结构数据是指没有具体的数据模型来描述的,它是随机

和模糊的);其次,数据是高维和非线性的。在大数据集中,单个数据的技术价值和经济价值是非常小的,为了从大数据中提取和挖掘到一个有用的信息和结构,必须至少使用该数据集的一个子集,从而使得这个问题变成一个高维(例如500—50000维)和非线性的数据建模问题。在许多预测问题中,虽然拥有巨量的大数据,但经常是仅拥有数个和少量相关数据集,如某些特殊病人的特殊病症数据。

针对高维非结构化数据,一个重要工具就是机器学习(machine learning)技术(也称为人工智能技术)。机器学习的过程本质是一个从数据中建立数学模型的过程,它包括有监督和无监督的学习(supervised and unsupervised learning)、强化学习(enforcement learning)、转移学习(transfer learning)、表示学习(representation learning)、深度学习(deep learning)等。由于这其中的主要贡献者来自于人工智能研究专家,因此被称为机器学习,他们试图强调和追求让机器像人一样去分析和处理自然世界的各类问题的目的。在过去十多年中,以深度学习(深度神经网络^[14])为代表的机器学习方法取得了令人瞩目的理论和应用成果,为从大数据(主要是从各类非结构数据中)中抽取有用信息和服务提供了强有力的工具,如图像和人脸识别、声音的识别和处理、自然语言处理和理解、知识脸谱图等各项应用。

英国皇家学会在2017年机器学习战略报告中提出,为了能够完成安全关键(safety-critical)的机器学习模型须满足以下两个必要条件:首先是模型的可证实性,即在许多应用中,特别是安全关键的应用,一个机器学习系统做出的决策和预测必须是拥有很高标准水平上的可证明性质;其次是模型的稳健性,主要指模型处理数据异常性的稳定性,例如还是以英法机器翻译模型为例,一旦出现一个新英语或法语单词,是否需要将上述的38.9亿个参数完全修改?如果是一个稳健系统,应该是不需要做大的或没有参数改正的。未来,针对这两个问题的研究将是计量经济建模的重要研究领域。

2.3 时空数据建模与应用

随着社会经济的发展,数据的社会网络和空间相关特性越来越强。在经济、金融、管理、社会学各个领域,除了变量之间有相关性,同一变量的不同样本也普遍存在着依赖关系,这种依赖关系叫做社会

网络(social network)或者空间相关性(spatial dependence)。在不同的领域里面,又可以叫做溢出效应(spillover effect)、同辈效应(peer effect)、邻近效应(neighborhood effect)、策略互动(strategic interaction)。如何在大数据环境下有效利用这种社会网络和空间相关的信息是当前计量经济学的发展方向之一。

要对社会网络进行建模,首先需要理解社会网络的形成和发展。社会网络的形成除了有随机因素的存在,更多的是选择的结果。Currarini等^[15]研究了基于个体类型和类型相关的收益形成的朋友网络,解释了现有网络数据中存在着的一些特点,并用来解释美国学生在社交行为中的种族选择行为。在动态发展的角度上,König等^[16]使用网络动态形成机制来解释现有网络数据中的层次性(nestedness)。Jackson和Rogers^[17]利用网络动态形成理论来研究网络中的节点(node)是如何通过随机认识和局域化的搜寻来相互认识的。Behrens等^[18]运用贸易数据和重力模型,研究国家之间的贸易额这个网络数据到底是受哪些因素影响的;通过考虑国家之间的互动性,这些因素对贸易网络结构的影响得到了修正。

其次,社会网络建模往往涉及大量时空变量,因此往往需要快速、有效的方法来对网络数据和空间数据变量进行筛选,海量数据提供了更多的信息,但也为如何进行有效分析带来了挑战。当前这方面的发展还刚刚起步,还有许多重要研究问题悬而未决,比如发展适用于时空网络模型的有效变量选择方法。

第三,时空数据一个重要研究方法是空间面板数据建模。早期的空间面板数据模型主要集中于模型的估计,如Yu^[19]和Su^[20],对空间面板模型的统计推断研究较少。Baltagi等^[21]首次从理论上对如何检测面板数据中存在空间相关性展开了研究。Kapoor等^[22]针对静态随机效应空间面板数据的矩估计(method of moments estimation)提出了相关渐近理论。未来,进一步针对高维、动态固定效应的空间面板模型的估计和检验还是重要的理论问题。

2.4 大数据政策评估计量经济分析

因果推断以及政策评估是经济学研究的重要核心问题,处理效应模型(treatment effect model)在因果推断和政策评估中发挥着重要的作用,其中平均

处理效应(Average Treatment Effect, ATE)被常用来衡量政策对所有个体的平均效果。当前,对于平均处理效应的识别和估计主要沿着三条脉络展开,分别是匹配法、工具变量法和断点回归法。

匹配法(matching)主要指通过使用一系列可观测的个体特征对个体进行匹配,使用特征相似但又接受了不同处理的个体作为彼此的反事实,从而得到因果推断。在外生性或者无混淆分配(unconfoundedness)的假设下,该方法可以用来识别平均处理效应。Rosenbaum和Rubin^[23,24]进一步指出:当个体特征的维数很高时,无须控制所有个体特征,通过倾向得分(propensity score)的匹配,足以达到消除偏误的目的。匹配法的思路在Dehejia^[25]和Heckman^[26]的文章中得到了进一步引申和发展。

工具变量方法指实验组的设定存在一定的内生性,因此无混淆分配(unconfoundedness)的假设难以成立,此时,识别平均处理效应不仅需要控制变量,还需要工具变量的协助。工具变量要求与结果变量独立,但是又与政策变量高度相关。采用工具变量方法识别处理效应的研究始于Angrist^[27]、Angrist^[28]等。Angrist和Imbens^[27]提出了局部平均处理效应(local ATE)的概念,回答了存在不同工具变量的情形下,如何建立基于不同的工具变量识别出的平均处理效应之间的区别和联系。

断点回归方法(Regression Discontinuity Design, RDD)往往应用于在某些场合,个体是否受到政策或者某事件的影响完全由某一个连续的驱动变量(running variable)所决定,即个体受到政策或者某事件影响的概率在某一断点处不连续。在这种情况下, Lee和Lemieux^[29]提出可以通过该驱动变量对断点处的平均处理效应进行推断。断点回归方法不仅常用于截面数据,目前也扩展到了时间序列数据和空间数据,并且可以结合双差分的方法。例如, Fu和Viard^[30]将断点回归方法运用到了高频的时空数据,研究了北京交通限行对空气污染和经济活动的影响,其中在检验交通限行政策对不同区位的空气监测站点空气质量的影响时,结合了RDD和DID的方法。Fu和Gu^[31]进一步将RDD运用到高频时间数据,并结合双差分的方法,研究了我国重大节假日期间高速公路免费通行政策给城市带来的空气污染。

大数据背景下,如何利用大数据来展开因果推断以及政策评估成为一个重要研究方向。此外,大数据也使得我们能够进一步研究政策效应的异质性。比如,政策对结果变量分布尾端人群的影响程度往往不同于对结果变量分布中段人群的影响,同时,众多有关社会福利、平等性问题的回答都有赖于研究者了解政策效应在整个受众人群中的分布情况。因此,在用处理效应模型评估经济政策时,研究人员越来越重视分析政策的分位数/分布处理效应(QTE/DTE),而不仅仅着眼于其平均处理效应(ATE)。比如, Firpo等^[32]考虑了如何估计当外生解释变量的无条件分布发生微小的反事实变化时,对被解释变量无条件分布的影响。Chernozhukov等^[33]通过建立条件分布回归(Conditional Distribution Regression, CDR)来估计解释变量的无条件分布发生任意大小改变,对被解释变量的无条件分布的影响。Powell^[34]提出了一种称为广义分位数回归(generalized quantile regression)的估计方法,通过矩估计方法来估计无条件或条件分位数处理效应。

2.5 现代网络生态环境下金融数据建模与应用

近年来,世界金融业的混业经营、互联网化趋势日益明显。金融体系逐渐从银行主导向各类金融机构相互持有股权债权、业务合作与市场渗透、资产价值相互关联等方向发展,金融体系呈现出错综复杂的金融网络特征。在这些新背景下,对金融风险的建模日趋复杂,需要从以下方向发展更多先进计量建模方法。

首先,金融机构的系统性风险网络化传染特征愈加明显,随着金融工具的创新速度不断加快,金融衍生品和表外业务不断增加,一家金融机构风险爆发的多米诺骨牌效应会愈加明显。网络分析方法为系统风险传染研究提供了新的工具,通过将金融机构视为网络的节点,将金融机构之间的资产负债关系视为网络的链条,网络分析能够直接而形象地刻画出金融系统的内在关联,并基于此分析金融机构之间的互动与关联。因此,未来对系统性风险建模需要完善针对网络模型的计量理论。

其次,金融机构和金融产品的结构问题以及结构性风险管理变得愈发重要。互联网金融具有“成本低、效率高、覆盖广”的特点对人类传统的金融行为模式产生重大影响。互联网金融行业的竞争

也已从单一业务层面转入多维度的模式竞争,结构性理财新模式愈发重要,从网络视角出发考察银行间的结构关系对于理解和预防金融风险具有重要意义。而大数据目前是互联网金融发展的重要驱动力,如何提取互联网大数据相关信息并应用于风险建模是未来金融计量发展的一个重要研究方向。

第三,当前依然缺少对金融风险进行准确度量 and 识别的计量模型和方法。已有金融机构系统性风险监测度量方法主要有以下类型。第一是综合指标法,即根据历史数据分析,找到影响金融机构系统性风险的预警指标,并通过统计方法对指标进行汇总,构建起预警指标体系,从而反映金融体系风险的综合状况。有代表性的综合指标法有 Acemoglu^[35] 提出的高阶级联效应, Adrian 和 Brunnermeier^[36] 提出的 CoVaR, IMF 提出的金融稳健指标, Illing 和 Liu^[37] 构建的金融压力指数和荷兰央行的金融稳定状况指数。第二是 Kumar 等^[38] 提出的 Simple Logit 模型,通过将 KLR 模型和 FR 模型进行综合,提高了对货币危机的预警水平。第三是通过研究风险传染性和金融机构关联程度进而度量系统性风险的方法,包括网络分析法(network analysis approach)、共同风险模型法(Co-Risk)、困境依赖矩阵法(distress dependence matrix)和违约强度模型法(default intensity model)。第四是评估系统性风险损失及损失概率的方法,包括在险价值和条件在险价值(VaR, CoVaR)、边际期望损失法(marginal expected shortfall)和系统性期望损失法(systemic expected shortfall)、美国联储局和 BIS 提出的困境保保费法(distressed insurance premium)等。第五是国际组织以及各国中央银行、金融监管机构分别开发的监测预警工具,如欧洲央行采用的改进的综合指数法(CISS)、IMF 系统性风险早期预警系统(EWE)、英国系统性机构风险评估系统(RAMSD)等。

在大数据时代背景下,未来如何结合网络生态环境对金融风险进行准确建模,对管理和防范金融系统性风险具有重要意义。

3 未来 5~10 年大数据背景下计量建模目标及资助重点

3.1 总体目标

十九大报告中明确指出,“推动互联网、大数据、

人工智能和实体经济深度融合”,数字中国和智慧社会成为我国的重要发展目标。在大数据时代,可以预见我国无论在数字资源的总量、广度和深度等方面都居于世界领先水平,从而为大数据计量经济学领域的发展提供了最重要的基础。而改革开放以来 40 余年的发展与积累,大数据计量经济学领域在国内的发展几乎和国际发展站在同一条起跑线上,通过计量经济学、统计学、数学、计算机科学、信息科学、经济学和管理学等领域的交叉融合,大数据计量经济学领域也最有可能取得突破性的具有国际影响力的理论成果与应用成果。总体目标包括两部分:

(一) 学术和学科方面的预期突破进展。在大数据计量经济学领域取得具有原创性和引领性的学术成果,培养一批国际一流的学术人才,形成重要国际学术影响;形成一批基础理论研究和应用研究相结合的重要研究成果,服务国家重大战略需求,培育具有重要影响力的国家智库,以量化分析为核心,大力推进经济学和管理学的学科发展。

(二) 政策应用方面的预期突破进展。立足中国制度背景,从中国实际问题出发,发展适合中国国情的大数据政策评估理论与方法,将政策分析与评估方法运用到具体的经济政策中,为科学决策和提高国家治理能力服务;基于大数据计量建模和预测预警的方法,提高经济、金融风险防控的能力,为新时期守住不发生系统性风险的底线提供科学支撑和政策建议。

3.2 资助重点

本次双清论坛与会专家经过深入研讨,认为应该专注于大数据时代计量经济学理论、方法与应用的前沿发展,以计量经济学科为主,结合统计学、信息与计算机科学、经济管理、公共管理、金融工程等多学科,并建议未来 5~10 年应着重围绕以下 5 个领域:

(一) 宏观数据复杂关系建模和预测的理论、方法与应用;

(二) 高维度和非结构化数据建模的理论、方法与应用;

(三) 时空数据建模理论、方法与应用;

(四) 大数据政策评估计量分析的理论、方法与应用;

(五) 现代网络生态环境下金融数据建模的理论、方法与应用。

4 结 语

通过对大数据时代计量建模发展趋势的深入探讨,本文梳理和聚焦大数据时代计量经济理论发展的瓶颈问题与前沿方向,认为通过多学科的交叉融合,有利于为计量经济理论在大数据时代的理论和方法突破打下坚实的基础,有助于促进大数据计量经济学方法在经济管理、信息科学、医学管理等领域创新性应用,为科学决策和提高国家治理能力现代化提供科学支撑。

参 考 文 献

- [1] Phillips PCB. The mysteries of trend. *Macroeconomic Review*, 2010, 9(2): 82—89.
- [2] White H, Granger CW. Consideration of trends in time series. *Journal of Time Series Econometrics*, 2011, 3(1): 1—40.
- [3] Gao J, Hawthorne K. Semiparametric estimation and testing of the trend of temperature series. *The Econometrics Journal*, 2006, 9(2): 332—355.
- [4] Phillips PCB, Hansen BE. Statistical inference in instrumental variables regression with I(1) processes. *The Review of Economic Studies*, 1990, 57(1): 99—125.
- [5] Cai ZW, Wang YF. Testing predictive regression models with nonstationary regressors. *Journal of Econometrics* 2014, 178(1): 4—14.
- [6] Amihud Y, Hurvich CM. Predictive regressions: a reduced-bias estimation method. *Journal of Financial and Quantitative Analysis*, 2004, 39(4): 813—841.
- [7] Rodrigues PM, Salish N. Modeling and forecasting interval time series with threshold models. *Advances in data analysis and classification*, 2015, 9(1): 41—57.
- [8] Sun YY, Han A, Hong YM, et al. Threshold autoregressive interval models for interval-valued time series data. *Journal of Econometrics*, 2018, 206(2): 414—446.
- [9] Han A, Hong Y, Wang S. “Autoregressive conditional models for interval-valued time series data”. Cornell University, manuscript. 2018.
- [10] Yang YH. Adaptive regression by mixing. *Journal of the American Statistical Association*, 2001, 96(454): 574—586.
- [11] Hansen BE. Least squares model averaging. *Econometrica*, 2007, 75(4): 1175—1189.
- [12] Kuersteiner G, Okui R. Constructing optimal instruments by first stage prediction averaging. *Econometrica*, 2010, 78(2): 697—718.
- [13] Hansen BE, Racine JS. “Jackknife model averaging”. *Journal of Econometrics*, 2012, 167(1): 38—46.
- [14] Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015, 521(7553): 436—444.
- [15] Currarini S, Jackson MO, Pin P. An economic model of friendship: homophily, minorities, and segregation. *Econometrica*, 2009, 77(4): 1003—1045.
- [16] König MD, Tessone C, Zenou Y. Nestedness in networks: a theoretical model and some applications. *Theoretical Economics*, 2014, 9(3): 695—752.
- [17] Jackson MO, Rogers BW. meeting strangers and friends of friends: how random are socially generated networks? *American Economic Review*, 2007, 97(3): 890—915.
- [18] Behrens K, Ertur C, Koch W. Dual gravity: using spatial econometrics to control for multilateral resistance. *Journal of Applied Econometrics*, 2012, 27(5): 773—794.
- [19] Yu JH, Robert de Jong, Lee LF. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics*, 2008, 146(1): 118—134.
- [20] Su LJ, Yang ZL. QML estimation of dynamic panel data models with spatial errors. *Journal of Econometrics*, 2015, 185(1): 230—258.
- [21] Baltagi BH, Song SH, Jung BC, et al. Testing for serial correlation, spatial autocorrelation and random effects using panel data. *Journal of Econometrics*, 2007, 140(1): 5—51.
- [22] Kapoor M, Kelejian HH, Prucha IR. Panel data models with spatially correlated error components. *Journal of Econometrics*, 2007, 140(1): 97—130.
- [23] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 1983, 70(1): 41—55.
- [24] Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 1985, 79(387): 516—524.
- [25] Dehejia RH, Wahba S. Causal effects in nonexperimental studies: reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, 1999, 94(448): 1053—1062.
- [26] Heckman JJ, Ichimura H, Todd PE. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies*, 1997, 64(4): 605—54.

- [27] Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. *Econometrica*, 1994, 62(2): 467—475.
- [28] Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. National Bureau of Economic Research. Technical Working Paper 136. 1993.
- [29] Lee DS, Thomas L. Regression discontinuity designs in economics. *Journal of Economic Literature*, 2010, 48 (2): 281—355.
- [30] Viard VB, Fu SH. The effect of Beijing's driving restrictions on pollution and economic activity. *Journal of Public Economics*, 2015, 125: 98—115.
- [31] Fu SH, Gu YZ. Highway toll and air pollution: evidence from Chinese cities. *Journal of Environmental Economics and Management*, 2017, 83: 32—49.
- [32] Firpo S, Fortin NM, Lemieux T. Unconditional quantile regressions. *Econometrica*, 2009, 77(3): 953—973.
- [33] Chernozhukov V, Fernandez-Val I, Melly B. Inference on counterfactual distributions. *Econometrica*, 2013, 81(6): 2205—2268.
- [34] Powell D. Quantile treatment effects in the presence of covariates. working paper. 2016.
- [35] Acemoglu D, Carvalho VM, Ozdaglar A, et al. The network origins of aggregate fluctuations. *Econometrica*, 2012, 80(5): 1977—2016.
- [36] Adrian T, Brunnermeier MK. “CoVaR”. *American economic review*, 2016, 106(7): 1705—1741.
- [37] Illing M, Liu Y. An index of financial stress for Canada. Bank of Canada, Staff Working Papers, 2003, 29(03-14).
- [38] Kumar M, Moorthy U, Perraudin W. Predicting emerging market currency crashes. *Journal of Empirical Finance*, 2003, 10(4): 427—454.

Several influential research directions of econometrics theory in big data era

Wang Shouyang¹ Hong Yongmiao² Huo Hong³ Fang Ying⁴ Chen Haiqiang⁴

(1. *Academy of Mathematics and Systems Science, Center of Forecasting Science, Chinese Academy of Sciences, Beijing 100037;*

2. *Department of Economics, Cornell University, U. S. A. ;*

3. *Management Science Faculty, National Natural Science Foundation of China, Beijing 100037;*

4. *Key Laboratory of Econometrics (Xiamen University), Ministry of Education,*

Wang Yanan Institute for Studies in Economics, School of Economics, Xiamen University, Xiamen 361005)

Abstract As a review of the 206th “Shuangqing Forum”, this article provides a summary about the opportunities and challenges for econometrics brought by new data form and complex relationship among variables in big data era. Forum participants reach a consensus that Chinese scholar should fully utilize the advantage of big data endowment under the background of the “Internet Plus” national strategy and fast development of digital economy, try to achieve certain significant and original theoretical breakthrough in the following areas, such as forecasting and modelling of Macro-level data, modelling of high dimensional and non-structural data, modelling of time series and spatial data, policy evaluation based on big data and financial data modelling with modern network.

Key words big data era; complex data; econometrics modelling; policy evaluations