

Doi:10.11835/j.issn.1008-5831.jg.2020.10.002

欢迎按以下格式引用:何婧,邹潇.基于微博大数据的经济不确定性测度及其对宏观经济的影响研究[J].重庆大学学报(社会科学版),2022(5):61-72. Doi:10.11835/j.issn.1008-5831.jg.2020.10.002.



**Citation Format:** HE Jing,ZOU Xiao. Measuring economic uncertainty of China based on Weibo data and research on its impact on macroeconomy [J]. Journal of Chongqing University (Social Science Edition),2022(5):61-72. Doi:10.11835/j.issn.1008-5831.jg.2020.10.002.

# 基于微博大数据的经济不确定性测度及其对宏观经济的影响研究

何婧,邹潇

(西南财经大学 统计学院,四川 成都 611130)

**摘要:**经济不确定性是一个重要的经济参数,它反映了经济的运行状态与经济参与者的评估与预期不一致的程度,与宏观经济运行和政策之间存在密切联系。经济不确定性的增加可能会对经济运行产生负面影响,因此研究经济不确定性对理解经济运行中的波动、制定宏观调控政策、激发市场经济活力等具有重要的意义,量化测度经济不确定性就是其中最关键的一步。文章提出了一种基于微博大数据的经济不确定性指数的量化构建方法,主要思想是利用人们对经济形势的主观认知态度的不一致程度来衡量经济的不确定性。该方法能考虑到更多经济参与者对经济的主观看法,从新的角度量化了我国经济不确定性的程度,且数据量大,覆盖面广,更能实时地体现经济不确定性程度的变化。首先,利用长短记忆(LSTM)模型对人们实时发布的微博中对经济运行的主观认知态度的情感倾向进行分类,效果良好,节约大量人工判别成本,有利于经济不确定性指数的实时计算和预测。然后,通过计算公众对经济持乐观态度和悲观态度的比例差异来计算经济不确定性指数 EUI (Economic Uncertainty Index)。EUI 反映的是经济参与者对经济运行状态的情感不一致程度,其不一致程度越大,经济不确定性程度越大,反之则越小。实证分析结果显示:微博数据量与经济热点事件有直接联系,时效性强,微博数据对重大经济事件的发生具有良好的反映;基于微博大数据构建的 EUI 与经济发展趋势具有一致性,对重大事件发生都有即时的反映;EUI 与其他文献中常用的指数反映的经济不确定性程度和变化具有较高的一致性,例如 EUI 与 Baker 等编制的中国经济政策不确定性(EPU)指数变化趋势基本相似,且 EUI 的时效性更强;EUI 与市场波动性指数和股票收益率方差具有正相关性,与消费者信心指数和基金经理信心指数具有负相关性,符合实际经济意义。更进一步将 EUI 应用于对我国宏观经济的影响分析中,结果表明,经济不确定性对进出口和消费有显著的负向影响。

**基金项目:**国家自然科学基金青年科学基金项目“基于非正态假设下的高维协方差矩阵检验问题研究”(11701466)

**作者简介:**何婧,西南财经大学统计学院副教授,博士,Email: he\_jing@swufe.edu.cn

**致谢:**感谢西南财经大学数据科学与商业智能联合实验室的资助。

关键词:经济不确定性;微博;大数据;长短期记忆模型;向量自回归模型

中图分类号:F124;F224;G206

文献标志码:A

文章编号:1008-5831(2022)05-0061-12

## 一、研究意义和问题提出

经济不确定性是一个重要的经济参数,反映了经济的运行状态与经济参与者的评估与预期不一致的程度。经济不确定性与经济运行和经济政策之间存在密切联系。Bloom指出当经济发展较好时期,其不确定性相对较小;而当经济不景气时,经济不确定性普遍较高<sup>[1]</sup>。战争、石油价格上涨和金融危机等可能导致经济衰退的外部冲击通常也会增加不确定性。众多国内外学者的研究指出,经济不确定性的增加会影响经济政策的实施和微观主体的决策,对经济运行产生负面影响。例如Bloom等指出经济不确定性会影响公司层面的投资和就业,不确定性的增加将会导致GDP下降<sup>[2]</sup>。此外,不确定性的增加将会导致企业更加谨慎,这将使经济对刺激性政策的反应大幅下降。Jurado等发现宏观不确定性具有强烈的反周期性,不确定性的增加与实际经济活动的大幅下降直接相关<sup>[3]</sup>。王义中和宋敏研究了经济不确定性对资金需求和公司投资的影响,发现宏观经济不确定性的增加会影响资金需求渠道从而影响公司投资行为<sup>[4]</sup>。同时,对经济不确定性的预期将有利于流动性资金需求对公司投资的正向促进作用。因此,研究经济不确定性对理解经济运行中的波动、制定宏观调控政策、激发市场经济活力等具有重要的意义,量化测度经济不确定性就是其中最关键的一步。在新的宏观经济环境下,基于大数据的经济不确定性测量方法具有覆盖面广、经济参与者多和可以利用实时数据作预测等特点,具有广阔的发展空间。

随着互联网和信息技术的快速发展,普通民众可以通过各种自媒体平台发布自己的观点、事实和新闻等,形式丰富多样。新浪微博自2009年8月上线以来,已经成为我国最有影响力的自媒体平台之一。在微博平台上发布的有关经济的观点或评论可以直接反映经济参与者对经济的主观认知,覆盖面广,传播速度快。本文提出了一种基于微博大数据的经济不确定性指数构建方法,主要思想是利用人们实时发布的微博,判断其对经济运行的主观认知态度的不一致程度来衡量经济的不确定性。本文提出的方法能考虑到更多经济参与者对经济的主观看法,数据量巨大,更能实时地体现经济不确定性程度的变化,从一定程度上解决经典的经济不确定性测量方法中所不能解决的问题,为经济研究提供新的视角和新的发现。

## 二、文献回顾

李华杰等将现有文献中对经济不确定性的量化测度方法归纳为三类:代理指标替代法、主观感知截面数据偏差法和状态量估计法<sup>[5]</sup>。以国外研究文献为例,第一类代理指标替代法是指利用能反映经济状态的指标作为经济不确定性的替代指标。例如,Bloom利用由芝加哥期权交易所推出的股票市场波动率指数(VIX, Volatility Index)作为经济不确定性的替代指标,他发现在重大经济冲击事件后都会出现VIX指数的峰值,且该指数与其他经济不确定性代理指标之间存在很强的正相关性<sup>[6]</sup>。另一个常用的经济政策不确定性(EPU, Economic Policy Uncertainty)指数由Baker等<sup>[7]</sup>提出,利用与经济政策不确定性相关的报道文章数量作为经济不确定性的替代指标。Dzieliński则考虑利用Google上与经济相关的关键词搜索频率来构造经济不确定性指数<sup>[8]</sup>。这样做的好处在于

Google 关键词搜索频率远高于报纸上新闻报道的数量以及其中关键词出现的频率,且能反映公众个人的经济行为或对经济活动的看法。第二类主观感知截面数据偏差法则是利用微观截面数据(以调查数据为主)挖掘出经济参与者对经济的主观感受和预期,以个体差异的离散程度来衡量经济不确定性。例如 Bachmann 等<sup>[9]</sup>和 Scotti<sup>[10]</sup>分别利用彭博调查数据和企业调查数据来构造经济不确定性指标。第三类状态量估计法则是将经济不确定性作为宏观经济的内部状态量,通过状态空间方程估计状态值来获得经济不确定性。这类方法依赖于模型和观测数据的选取,对数据的质量要求高,数据处理量大,计算复杂。

由于经济环境不同和数据的可得性受限,上述经典经济不确定性指数构建方法难以直接应用于中国经济不确定性指数的构建。为了适应于中国经济研究,近年来对我国经济不确定性对宏观经济影响的研究取得了很多成果。黄宁和郭平基于 PVAR 模型研究经济不确定性对我国宏观经济的影响和地区差异<sup>[11]</sup>。金雪军等运用 FAVAR 模型研究经济不确定性对宏观经济的影响<sup>[12]</sup>。欧阳志刚等建立了时变 VAR 模型分析了经济不确定性对经济增长的影响<sup>[13]</sup>。许志伟和王文甫利用基于最大份额识别方法的结构 VAR 模型分析经济不确定性对主要宏观经济变量的影响<sup>[14]</sup>。这些文献研究结果都发现经济不确定性的增加也会对我国经济增长、投资、消费、价格、出口等带来短期的负向影响。然而,这些文献采用的都是 Baker 等<sup>[7]</sup>基于中国香港《南华早报》(英文)每日新闻中相关报道文章数的比例所构建的 EPU 指数作为我国经济不确定性的替代指标。为了更好地衡量我国的经济不确定性,Huang 和 Luk 选取了 10 种具有代表性的中文报纸,编制了新版中国经济政策不确定性指数,使其更符合我国的经济形势<sup>[15]</sup>。章上峰等用宏观经济景气指数作为我国经济不确定性的替代指标<sup>[16]</sup>。利用大数据时代的优势,还有很多文献构建了基于大数据的经济不确定性指数。马丹等收集了 60 个月度宏观经济指标和 4 个季度指标,运用含有潜变量的混频动态因子随机波动模型测算我国宏观经济的不确定性<sup>[17]</sup>。杜龙波提出利用谷歌检索趋势值和我国特有的百度搜索指数值来衡量我国经济不确定性的方法<sup>[18]</sup>。这些方法都从不同角度衡量了我国经济不确定性的程度。

上述方法大多是利用宏观经济指标或新闻报刊报道数比率来构建经济不确定性指数。本文利用微博大数据提出了一种新的经济不确定性指数构建方法,考虑了更多经济参与者的主观看法,比主观调查数据收集成本低、难度小,从新的角度量化了我国经济不确定性的程度。实证分析表明这种新的经济不确定性指数与已有文献中的经济不确定性指数具有较高的一致性,并且对重大事件的发生有更为即时的反映。

### 三、基于微博大数据的经济不确定性指数构建方法

本文的经济不确定性指数的构建主要基于微博大数据中人们对经济不确定性的看法的刻画。如前文所述,构建的经济不确定性指数的一类方法是基于新闻指数,即通过统计一些新闻媒体关于经济不确定性报道的文章数目或关键词出现频率来作为经济不确定性的测度。这类方法容易受到选取新闻媒体和关键词的主观性影响。而微博的受众群体广泛,利用微博大数据构造的经济不确定性指数能挖掘出更多经济参与者对我国经济的主观认知,可以在一定程度上改进这个问题。

同时,不同于通过调查问卷收集到的主观调查数据,微博文本数据中反映的公众对经济的主观认知是一个定性变量,无法通过计算其方差或标准差来构建经济不确定性指数。可将公众对我国

经济的看法划分为“正向”“中立”和“负向”三种情感。考虑到每个月人们发布微博的数量不同,对第  $t$  个月,分别计算带有正向情感和负向情感的微博占当月含有经济不确定性关键词的微博总数的比例,并记为  $\text{Rate}_{+,t}$  和  $\text{Rate}_{-,t}$ 。由不确定性的定义,公众对经济持乐观态度和悲观态度的比例差异越小,说明经济不确定性程度越大,反之则越小。基于上述分析,构建第  $t$  月的经济不确定性指数 (EUI, Economic Uncertainty Index) 如下:

$$\text{EUI}_t = 1 - |\text{Rate}_{+,t} - \text{Rate}_{-,t}| \quad (1)$$

由式(1)定义的经济不确定性指数取值在  $[0, 1]$  区间内,对不同月份直接可比。

为了计算每个月的经济不确定性指标 EUI,需要分别统计每个月带有正向情感和负向情感的微博数量。由于微博数据量巨大,直接判断每条微博的情感倾向并计数的成本过高,实现困难。考虑到微博数据中可能存在用词不规范和经常诞生新词汇等特性,本文采用长短期记忆 (LSTM, Long Short-Term Memory) 网络模型来对每条微博进行情感分类。LSTM 模型是循环神经网络模型的变体,具有良好的“记忆”能力,能学习并记忆上下文信息,在语言处理领域中具有独特的优势。对于情感分类问题,滕飞等<sup>[19]</sup>、於雯和周武能<sup>[20]</sup>等文献发现 LSTM 模型优于基于情感词典的分类方法和传统的机器学习分类方法。

LSTM 模型具有链式结构,其中每个传输单元的结构如图 1 所示。

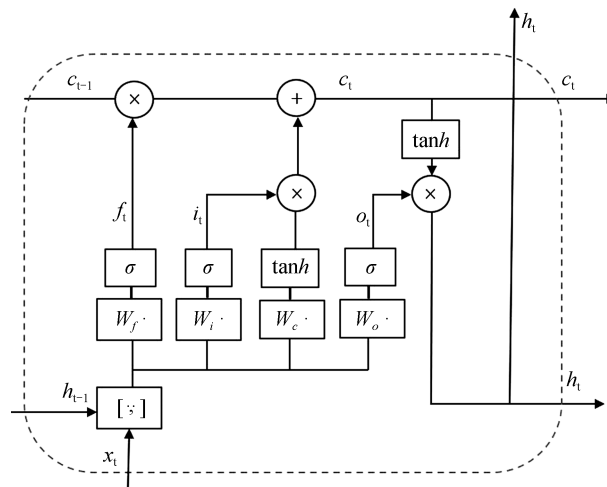


图 1 LSTM 模型的单元结构框图

其中:  $\mathbf{x}_t$  表示当前的输入数据,在本文中  $\mathbf{x}_t$  表示由微博短文本生成的词向量;  $h_t$  表示当前的隐藏单元的取值;  $c_t$  表示当前的记忆单元的取值。LSTM 模型利用记忆单元  $c_t$  用于保存信息,以及三种类型的门 (gate)——输入门、遗忘门和输出门——来控制信息的记忆和流动。分别用符号  $i, f, o$  表示输入门、遗忘门和输出门,  $\mathbf{W}$  和  $b$  表示权重矩阵和偏置项。  $\sigma$  表示 sigmoid 函数。由图 1 可知,每个门的作用及计算方式如下。

输入门 (input gate) 用于决定当前时刻网络的输入  $\mathbf{x}_t$  有多少保存到单元状态  $c_t$ , 其计算公式为

$$i_t = \sigma(\mathbf{W}_i \cdot [h_{t-1}, \mathbf{x}_t] + b_i)$$

其中  $[h_{t-1}, \mathbf{x}_t]$  表示把两个向量连接成一个更长的向量。

遗忘门 (forget gate) 用于决定上一时刻的单元状态  $c_{t-1}$  有多少保留到当前时刻  $c_t$ , 其计算公式为

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

输出门(output gate)用于控制记忆单元值  $c_t$  有多少输出到  $h_t$ , 其计算公式为

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

当前时刻下的  $c_t$  和  $h_t$  值则由以下公式计算:

$$c_t = i_t \odot \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) + f_t \odot c_{t-1}, h_t = o_t \odot \tanh(c_t)$$

其中  $\odot$  表示对应元素相乘。

LSTM 模型中的参数由反向传播算法迭代更新得到,最终输出微博短文本的情感分类。

## 四、实证分析

### (一) 数据采集

本文选取新浪微博 2013 年 1 月 1 日至 2019 年 6 月 30 日的数据用于构建我国月度经济不确定性指数 EUI。以“中国经济”“国内经济”“我国经济”“我国 GDP”为关键字进行搜索,利用 Python 爬取新浪微博全网中包含关键词的原始微博、微博文章和微博评论等数据,共计约 170 万条。与构建 EPU 指数方法用到的新闻报道文章不同,微博的参与者不仅限于官方媒体和新闻自媒体,更多的是普通民众,因此存在一些用语不规范或完全没有信息含量的微博短文本。本文旨在利用公众对经济形势的看法来构建经济不确定性指数,因此只去掉重复和无关内容,剩余有效数据共计超过 150 万条。将每月含有关键词的微博数量绘制折线图如图 2 所示。

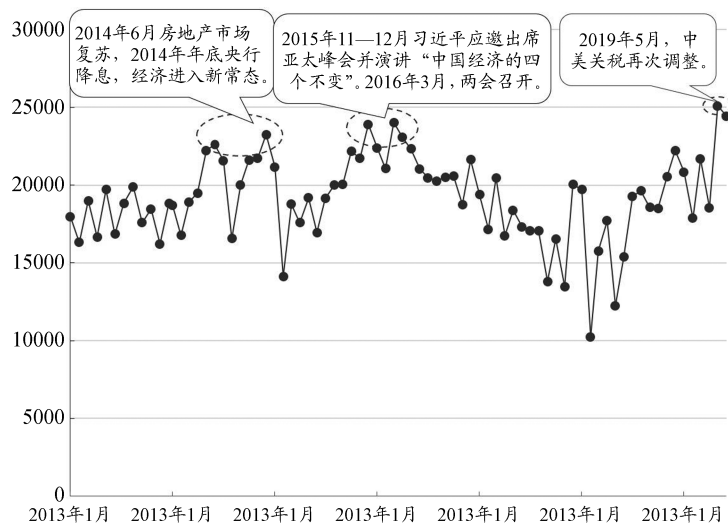


图 2 每月含有关键词的微博数据总量的折线图

正如前文所言,微博是一个为人们提供发布自己的观点、事实和新闻等的平台,每当有重大事件发生或经济形势发生变化时,更多的人会在微博平台上发表自己的观点或转发自己支持的观点。由图 2 可以看出,微博数据量有三个明显的波峰。2014 年 6 月起,住建部发布一系列限购松绑的新政策,全国房地产市场复苏。2014 年底央行降息,中央经济工作会议指出中国经济进入新常态。这几个重大事件的发生引起了人们的热烈讨论,微博数量迅速上升至波峰位置。2015 年 11 月,习近平总书记发表演讲,谈到中国经济四个“没有变”。12 月 4 日,《人民日报》刊发了四位专家的文章,阐释“四个没有变”的深刻内涵。2016 年 3 月,全国两会在北京召开,再次引发人们对中国经济的热烈讨论。2018 年 3 月起,随着中美贸易争端的不断升级,微博评论和转发数量呈现明显上升趋势。

2019年5月,中美关税再次调整,美国对华为实施技术限制,多家美国科技企业宣布中止为华为供应关键软件和零部件,国家发改委召开发布会回应当前经济热点,再次引发热议,微博数量达到最高峰。由此可见微博数据量与经济热点事件有直接联系,时效性非常强。正是考虑到这点,本文选取微博大数据用于构造经济不确定性指数。

## (二) 微博数据情感分类

为了计算由式(1)定义的经济不确定性指数 EUI,需要计算每月带有正向和负向情感的微博比例。由于数据量巨大,由人工标注每条微博的情感分类成本非常高,且耗时长,不利于实时预测。因此本文利用 LSTM 模型对微博数据的情感分类进行学习和预测。首先对 15 万条数据进行情感分类标注,将正向、中立和负向情感分别标记为 1, 0, -1。利用这 15 万条有标签的数据进行模型训练和测试,然后将训练得到的模型应用于余下的 135 万条数据,从而得到所有微博数据的情感分类标签。表 1 展示了部分爬取到的微博数据及其情感分类。

表 1 微博数据及其情感分类示例

时间	点赞	转发	评论	微博内容	作者	情感分类
2013年01月01日	3	0	0	中国经济虽有不平等,但很有韧性	行军蚁 ZMO	1
2016年01月28日	8	47	3	黄卫东:评索罗斯要打劫中国:在全球熊市中,中国将是他的“打劫对象”之一,中国经济很可能会硬着陆	张清同志	-1
2017年12月12日	8	6	3	满足美好生活需要的新供给占比持续增加,看好2018的中国经济	滕泰	1
2019年05月20日	180	1	16	事实上,贸易战对中国经济影响有限(百分之一之说太过随意)。至于,华为冲击,更是臆断(任先生已做说明)...	卢麒元	0

如表 1 所示,微博内容都以短文本的形式为主。我们首先需要将微博的短文本数据转化成可用于训练 LSTM 模型的数值型数据。本文采用常用的文本数据预处理方法,主要包括以下四个步骤。

(1) 提取文字信息:去掉微博内容中常常包含的符号、网址、表情等无用信息,仅保留文字。本文通过用正则化表达式制定的规则提取文字信息。

(2) 进行中文分词:把句子中的词进行独立划分,便于让计算机“理解”其含义。本文选用 BosonNLP 进行中文分词,它是目前中文分词工具中准确率最高的一种,对微博数据的分词准确率达到 93%。

(3) 去停用词:去掉短文本中的停用词,例如文本中使用过于频繁的词“我”“今天”“是”等和没有实际语义的词如语气词“啊”“吧”“呢”,助词“着”“了”“过”,连接词“那么”“下面”“之后”等。本文采用的停用词表来源于哈工大停用词表、四川大学机器智能停用词库和百度停用词表等,通过去除重复词汇,选用了约 1 800 个停用词。

(4) 词向量化:将第 3 步中分好的词转化成词向量,作为 LSTM 模型的输入向量。本文采用目前最常用的由 Google 公司开发的词向量化工具 Word2vec。

上述数据预处理的 4 个步骤中均用 Python 软件予以实现。然后训练 LSTM 模型进行情感分类。

本文共有 15 万条有情感标签的数据,用其中的 12 万条数据训练模型,3 万条数据用于测试,以选定合适的参数。最终,模型中的主要参数设定如表 2 所示。

表 2 LSTM 模型参数的设置\

参数	含义	取值
Max length	训练句子的最大输入长度	50
Word2Vec Dimension	词向量维度	200
Batch Size	每批次训练的数据量	64
Iteration	训练过程的迭代次数	5 000
LSTM Units	LSTM 层内的 LSTM 单元数	24
Test Size	测试集所占训练数据比例	0.2

利用上述训练好的模型,将其余的 135 万条无标签数据代入,即可得到每条微博的情感分类。利用 LSTM 模型对微博数据的情感进行分类的方法有利于实时计算和预测经济不确定性指数。

### (三) 基于微博数据的经济不确定性指数 EUI

统计所有当月情感分类为“1”和“-1”的微博数据占当月(含有关键词的)微博总数的比例,代入式(1)计算 EUI。将计算得到的 2013 年 1 月至 2019 年 6 月的 EUI 绘制折线图(图 3)。

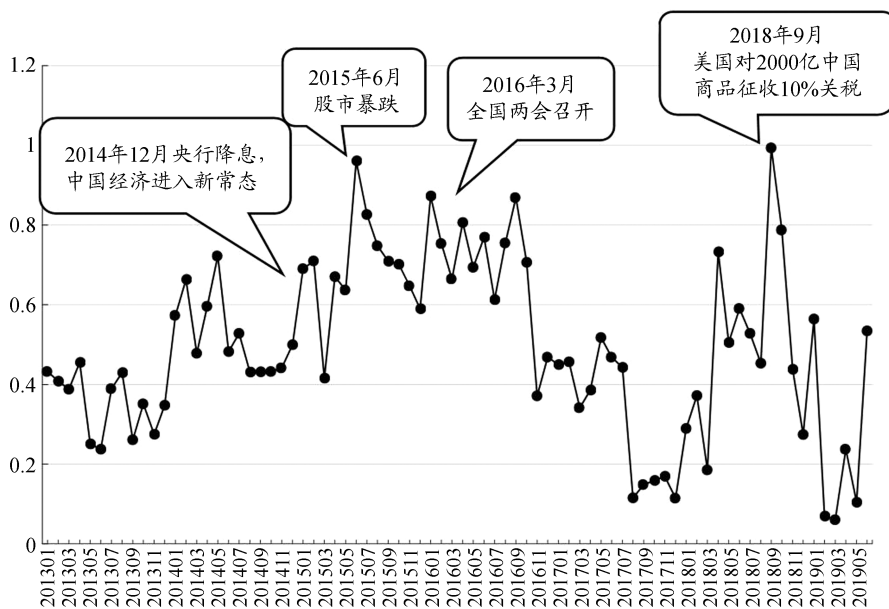


图 3 经济不确定性指数 EUI 的折线图

对比图 2 和图 3,可以看出 EUI 的趋势和峰值位置与微博数量有相似之处。例如,2014 年年中和年末房地产市场限购限贷松绑政策的发布,中央经济工作会议指出中国经济进入新常态、2015 年 6 月股市暴跌、2016 年年初全国两会召开,在这些经济大事件发生的时点附近,EUI 均处于波峰位置。2017 年全国经济处于平稳时期,因此 EUI 值相对较低。2018 年 3 月起,随着中美贸易摩擦的不断升级,EUI 值连续攀升。同年 9 月美国对价值 2 000 亿美元的中国输美商品征收 10% 的关税,中美贸易顺差达到历史最高值,有可能进一步加剧中美之间的贸易摩擦,EUI 达到峰值。同时,也有一些引起微博热议的事件,对应的 EUI 值并不高,例如 2015 年 11 月,习近平总书记发表演讲中国经济

四个“没有变”和2019年5月中美贸易争端的再次升级。这是因为对这些事件的微博评论的情感极性大多相同。因此,利用微博数据的情感分类构造经济不确定性指数比直接用微博数量更能反映经济不确定性的定义。

#### (四) 与其他经济不确定性指数的相关性分析

现有文献中最常用的经济不确定性指数为 Baker 等<sup>[7]</sup> 编制的中国经济政策不确定性(EPU)指数。中国 EPU 指数由网站 <http://www.policyuncertainty.com/> 提供。为了比较 EUI 与 EPU 指数,本文将 EPU 指数以 1995 年 1 月为基准转化为定基指数。将 2013 年 1 月至 2019 年 6 月的 EUI 和定基 EPU 指数绘制成折线图,如图 4 所示。

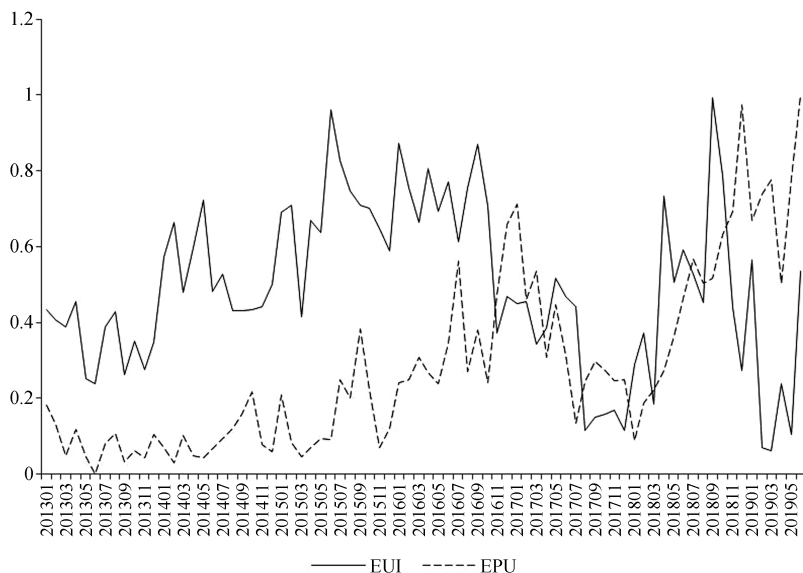


图 4 经济不确定性指数 EUI 与经济政策不确定性指数 EPU 的对比图

由图 4 显示,EUI 和 EPU 指数的变化趋势基本相似,且在重大经济事件发生的时点附近都会出现峰值。这说明 EUI 能够较好地反映经济的变化趋势。如图 3 中显示出的 2014 年年中和年末、2015 年 6 月、2016 年年初和 2018 年 9 月。相比而言,EPU 指数普遍会比 EUI 的峰值滞后 2~3 个月。这与两种指数的定义直接相关。EPU 指数是通过报纸上的新闻报道收集的与“经济”“政策”和“不确定性”相关的信息,而 EUI 是通过微博这种新兴的网络自媒体平台收集的相关信息。因此相比于 EPU 指数,EUI 更具有时效性。

为进一步说明本文所构造的 EUI 指数的合理性,参考杜龙波<sup>[18]</sup> 的做法,将 EUI 与其他衡量经济不确定性的指数进行比较。其中应用最广泛的经济不确定性指数是芝加哥期权期货交易所使用的市场波动性指数(VIX)和股票收益率的方差(VP),以及通过问卷调查的方式来直接表示公众对经济形势的态度构造的消费者信心指数(CCI)和基金经理信心指数(FMI)。表 3 中总结了各个指数的均值、方差和相关系数。VIX 指数来自网站 <https://stockcharts.com/>,VP 和 FMI 来自 RESET 金融研究数据库,CCI 来源于国家统计局网站。

表 3 中的结果显示,EUI 与 VIX 和 VP 都呈正相关关系(相关系数分别为 0.49 和 0.35)。但它们也有显著的不同,VIX 和 VP 更多地反映了金融市场的波动,EUI 则更多地反映了经济参与者对宏观经济的主观认知。将 EUI 与 CCI 和 FMI 相比,它们的相关系数均为负。CCI 和 FMI 均反映的是经济参与者对当前经济的评价和对未来经济的预期,是预测经济走势和股票市场运行方向的重要先行指标,



值越大表示预期越乐观。当经济不确定性高时,CCI 和 FMI 相对较低,符合实际经济意义。

表 3 其他衡量经济不确定性指数的描述性统计

变量名称	EUI	VIX	VP	CCI	FMI
	描述性统计量				
样本均值	0.52	23.23	17.22	79.23	73.44
样本标准差	0.24	13.46	14.33	3.25	4.86
	相关系数矩阵				
	1.00	0.49*	0.35*	-0.20*	-0.48*
	-	1.00	0.39*	-0.20	-0.55*
	-	-	1.00	-0.12	-0.32*
	-	-	-	1.00	0.12
	-	-	-	-	1.00

注:“-”表示空值,“\*”表示在 10%的有效水平上显著。

#### (五) 经济不确定性对宏观经济的影响分析

本节利用构建的经济不确定性指标 EUI,建立向量自回归(VAR)模型来分析经济不确定性对宏观经济的影响。选择投资、消费、进出口以及物价水平作为主要宏观变量的代表,变量含义和符号由表 4 具体给出。数据均来自于我国国家统计局数据网站,时间跨度从 2013 年 1 月至 2019 年 6 月,共计 78 条月度数据。

表 4 变量符号表示及其含义

符号	变量含义
EUI	经济不确定性指数
IR	固定资产投资额累计同比增长率
CR	社会消费品零售总额同比增长率
IER	进出口总值同比增长率
CPIR	居民消费价格指数同比增长率

利用 ADF 检验对表 4 中的五个变量进行平稳性检验,结果说明 EUI、IR 和 IER 均为平稳序列,CR 和 CPIR 为趋势平稳序列。将去掉 CR 和 CPIR 序列中的固定时间趋势后得到的序列分别记为 CR\_dt 和 CPIR\_dt,重新检验得知均为平稳序列。

为了探究宏观经济变量之间的相互影响关系,首先对上述变量进行格兰杰(Granger)因果关系检验。从表 5 中的结果可以看出,在 10%的显著性水平下,经济不确定性指数 EUI 是社会消费品零售总额增长率的格兰杰原因,并且和进出口总值增长率互为格兰杰原因。

表 5 格兰杰因果检验结果

原假设	F 值	P 值
CR_DT 不是 CPIR_DT 的格兰杰原因	3.24	0.08
EUI 不是 CR_DT 的格兰杰原因	3.15	0.08
CR_DT 不是 IER 的格兰杰原因	5.63	0.02
IER 不是 EUI 的格兰杰原因	6.44	0.01
EUI 不是 IER 的格兰杰原因	7.94	0.01

根据 AIC、BIC 信息准则和对模型结果的综合评价,选择滞后期为 1 的向量自回归模型 VAR(1)。根据模型的估计结果,得到特征方程的特征根均位于单位圆内,说明所构建的 VAR(1) 模型稳定,由此进一步进行脉冲响应分析。给经济不确定性指数 EUI 一个标准差单位的正向冲击,可以得到图 5 中的脉冲响应函数。

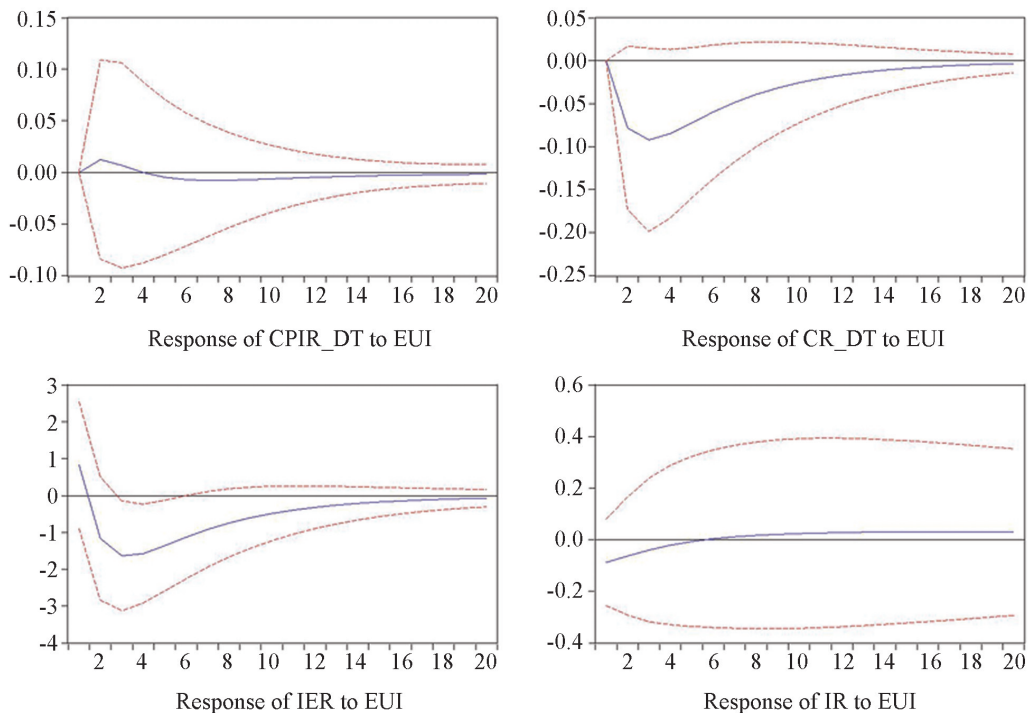


图 5 经济不确定性指数对宏观经济变量的脉冲响应函数

其中横轴表示各经济变量对不确定性冲击的响应的滞后期数(单位:月),纵轴表示响应的数值大小。实线表示各宏观经济变量在不同滞后期数响应的程度,虚线表示正负两倍的标准差偏离范围。由图 5 中的结果可知,当经济不确定性受到一个标准差单位的正向冲击后,进出口和消费的反应程度是最大的,价格和固定资产投资的反应程度较小。并且,经济不确定性对进出口和消费都造成了负面的影响。进出口对不确定性冲击的响应在第一个月就出现,小幅度上升后从第二个月开始迅速下降,到第三个月达到最低值后持续缓慢地回升,在 20 个月后逐渐趋于零。这说明不确定性对进出口贸易的负面影响在前三个月最明显,这样的负面影响大概将持续两年。这与鲁晓东和刘京军<sup>[21]</sup>等文献中得到的结论一致。而消费对不确定性冲击的响应有一个月的滞后,从第二个月开始迅速下降,同样到第三个月达到最低值后,经过两年左右逐渐回归至正常水平。

进一步地,我们利用方差分解分析由经济不确定性冲击而引起的宏观经济变量变动的贡献程度,结果如表 6 所示。除了各宏观经济变量自身变化的贡献外,经济不确定性对各宏观经济变量变动的贡献程度最大的是进出口,截至第十二个月,方差贡献率超过了 12%;对消费的贡献程度次之,大约为 8.2%。经济不确定性对价格和固定资产投资的贡献最小,均不超过 1%。这与脉冲响应分析的结果一致。

表6 不确定性对各宏观经济变量变动的贡献度

时期	CPIR_dt	CR_dt	IER	IR
1	0.000	0.000	1.203	1.178
2	0.048	1.811	2.761	0.908
3	0.056	3.787	5.638	0.694
4	0.053	5.314	8.002	0.546
5	0.058	6.366	9.622	0.446
6	0.067	7.061	10.672	0.379
7	0.085	7.513	11.341	0.335
8	0.099	7.805	11.765	0.306
9	0.111	7.995	12.035	0.288
10	0.121	8.118	12.206	0.276
11	0.128	8.197	12.315	0.268
12	0.134	8.249	12.382	0.264

## 五、结论与启示

本文提出一种基于微博大数据的我国经济不确定性指数构建方法,主要思想是利用人们对经济形势的主观认知态度的不一致程度来衡量经济的不确定性。考虑到微博是我国最有影响力的自媒体平台之一,微博观点或评论可以直接反映经济参与者对经济的主观认知,覆盖面广,传播速度快,本文构建的经济不确定性指数能考虑到更多经济参与者对经济的主观看法,更能实时地体现经济不确定性程度的变化,为经济研究提供新的视角和新的发现。由于数据量巨大,本文采用了LSTM模型对微博数据的情感倾向进行判断,有利于经济不确定性指数的实时计算和预测。

从结果来看,对于重大事件的发生,本文构建的经济不确定性指数EUI都有即时的反映,如2014年年中和年末房地产市场限购限贷松绑政策的发布,中央经济工作会议指出中国经济进入新常态、2015年6月股市暴跌、2016年年初全国两会召开,和2018年9月中美贸易摩擦加剧,这些大事件的发生引发人们在微博上的热烈讨论,基于微博评论数据构建的EUI在时间发生的同月达到峰值。将EUI与其他常用的经济不确定性指数进行对比,发现其与不确定性代理指数VIX、VP具有较强的正相关性,与经济政策不确定指数EPU具有较强的一致性,且EUI的时效性更强。最后,本文通过建立向量自回归模型,研究了基于EUI衡量的经济不确定性对我国宏观经济的影响。结果表明,经济不确定性对进出口和消费有显著的负向影响。进出口对不确定性的响应在第一个月就出现,小幅度上升后迅速下降,到第三个月达到最低值后,大约持续两年时间。消费对不确定性冲击的响应有一个月的滞后,从第二个月开始迅速下降,同样到第三个月达到最低值后持续两年左右逐渐回归至正常水平。

### 参考文献:

- [1] BLOOM N. Fluctuations in uncertainty[J]. *Journal of Economic Perspectives*, 2014, 28(2): 153-176.
- [2] BLOOM N, FLOETOTTO M, JAIMOVICH N, et al. Really uncertain business cycles[J]. *Econometrica*, 2018, 86(3): 1031-1065.
- [3] JURADO K, LUDVIGSON S C, NG S. Measuring uncertainty[J]. *American Economic Review*, 2015, 105(3): 1177-1216.
- [4] 王义中, 宋敏. 宏观经济不确定性、资金需求与公司投资[J]. *经济研究*, 2014(2): 4-17.
- [5] 李华杰, 史丹, 马丽梅. 经济不确定性的量化测度研究: 前沿进展与理论综述[J]. *统计研究*, 2018(1): 117-128.
- [6] BLOOM N. The impact of uncertainty shocks[J]. *Econometrica*, 2009, 77(3): 623-685.
- [7] BAKER S R, BLOOM N, DAVIS S J. Measuring economic policy uncertainty[J]. *The Quarterly Journal of Economics*, 2016,

- 131(4):1593-1636.
- [8] DZIELINSKI M. Measuring economic uncertainty and its impact on the stock market[J]. *Finance Research Letters*, 2012, 9(3):167-175.
- [9] BACHMANN R, ELSTNER S, SIMS E R. Uncertainty and economic activity: Evidence from business survey data[J]. *American Economic Journal: Macroeconomics*, 2013, 5(2):217-249.
- [10] SCOTTI C. Surprise and uncertainty indexes: Real-time aggregation of real-activity macro-surprises[J]. *Journal of Monetary Economics*, 2016, 82:1-19.
- [11] 黄宁, 郭平. 经济政策不确定性对宏观经济的影响及其区域差异: 基于省级面板数据的 PVAR 模型分析[J]. *财经科学*, 2015(6):61-70.
- [12] 金雪军, 钟意, 王义中. 政策不确定性的宏观经济后果[J]. *经济理论与经济管理*, 2014(2):17-26.
- [13] 欧阳志刚, 何富美, 薛龙. 经济政策不确定性, 双轮驱动与经济增长[J]. *系统工程理论与实践*, 2019(4):986-1000.
- [14] 许志伟, 王文甫. 经济政策不确定性对宏观经济的影响: 基于实证与理论的动态分析[J]. *经济学(季刊)*, 2019(1):23-50.
- [15] HUANG Y, LUK P. Measuring economic policy uncertainty in China[J]. *China Economic Review*, 2020, 59:101367.
- [16] 章上峰, 李荣丽, 王玉颖. 中国宏观经济不确定性的统计测度研究[J]. *统计与信息论坛*, 2015(6):14-19.
- [17] 马丹, 何雅兴, 翁作义. 大维不可观测变量的中国宏观经济不确定性测度研究[J]. *统计研究*, 2018(10):44-57.
- [18] 杜龙波. 经济不确定衡量及对股票市场的影响: 基于网络信息检索技术[J]. *技术经济与管理研究*, 2014(3):78-83.
- [19] 滕飞, 郑超美, 李文. 基于长短期记忆多维主题情感倾向性分析模型[J]. *计算机应用*, 2016(8):2252-2256.
- [20] 於雯, 周武能. 基于 LSTM 的商品评论情感分析[J]. *计算机系统应用*, 2018(8):159-163.
- [21] 鲁晓东, 刘京军. 不确定性与中国出口增长[J]. *经济研究*, 2017(9):39-54.

## Measuring economic uncertainty of China based on Weibo data and its impact on macroeconomy

HE Jing, ZOU Xiao

(School of Statistics, Southwestern University of Finance and Economics, Chengdu 611130, P. R. China)

**Abstract:** Economic uncertainty is an important economic index, which measures the degree of inconsistency between the economic conditions and the expectations of economic actors, and is closely related to the macroeconomy performance and policy making. The increase of economic uncertainty may have negative effects on economic activities. Therefore, it is important to study economic uncertainty for understanding fluctuations in the economy, formulating macro-control policies, and stimulating market vitality. Since the economic uncertainty is not directly observable, it is necessary to develop a method to quantify the uncertainty and assess its macroeconomic effects. We propose to construct an economic uncertainty index based on the Weibo data, which involves more people's views regarding the economy. This method takes the public opinions into account and can show the timely changes of the economic uncertainty. First, we use a long and short-term memory (LSTM) model to classify the sentiments of the comments posted on Weibo on the economic issues. This method saves a lot of time and cost to classify the sentiments of the Weibo comments manually and is helpful for the real-time measurement of the economic uncertainty index. Then, we construct the Economic Uncertainty Index(EUI) by measuring the inconsistency of the public opinions. The empirical analysis results show that: 1) Weibo data has a good reflection on the important economic issues, and EUI is consistent with the economic development trends and shows spikes around the major events; 2) EUI is consistent with other economic uncertainty indices which are commonly used in literatures. For example, EUI shows similar patterns with EPU proposed by Baker et al. (2016) and provides timely information for important economic issues. Furthermore, EUI is positively correlated with the market volatility index and the stock return variance and negatively correlated with the Consumer Confidence Index and Fund Manager Confidence Index; 3) Using the vector autoregressive (VAR) model to assess the impact of EUI on the macroeconomy, the results show that EUI has a significant negative impact on imports, exports and consumption.

**Key words:** economic uncertainty; Weibo; big data; LSTM model; VAR model

(责任编辑 傅旭东)