

# 一种基于在线大数据的高频 CPI 指数的设计及应用<sup>①</sup>

刘涛雄<sup>1,2</sup> 汤珂<sup>1,2</sup> 姜婷凤<sup>3</sup> 仇力<sup>4</sup>

(1. 清华大学社会科学学院经济学研究所; 2. 清华大学创新发展研究院;  
3. 对外经济贸易大学金融学院; 4. 中国社会科学院世界经济与政治研究所)

**研究目标:** 在数字经济时代探索如何运用在线大数据编制实时高频物价指标。  
**研究方法:** 设计了中国第一套基于互联网在线大数据的居民消费价格指数, 从多方面分析指数质量及其应用。研究发现: 在线 iCPI 可实现各层次类别的日、周、月指数无滞后实时更新; 指数数据从采集、清洗到加工和发布均由计算机自动进行, 既节省人力又减少人为干预因素; 指数在代表一般物价变化、精准捕捉典型事件影响、现时预测通货膨胀、实时监测宏观经济形势等方面表现突出。  
**研究创新:** 首次采用在线大数据编制 CPI, 弥补了中国尚无实时高频物价指标的空白。  
**研究价值:** 在线 iCPI 是传统 CPI 的有益补充, 其编制思维和方法可用到更多高频宏观经济指标的构建上。

**关键词** 在线大数据 iCPI 实时高频指标 宏观现时预测

**中图分类号** F222.1; C813 **文献标识码** A

**DOI:**10.13653/j.cnki.jqte.20190906.005

## 引 言

随着人类社会活动的深度互联网化、数字化, 大数据时代已经来临。在这样一个时代, 几乎所有经济活动都留下“数字记载”, 技术发展极大地提升了获取信息的范围和速度。我们认为, 在大数据时代发展实时化、高频化的宏观经济指标已经条件具备, 且意义重大。所谓实时化, 是指尽量缩短数据发布的时滞, 做到基本无滞后, 如本月结束最后一天或第二天便可发布月度数据。所谓高频化, 便是提高数据发布的频率。大多数宏观经济指标频率较低, 周期一般在一个月以上, 本文则将其频率提高到以天为周期。高频化和实时化相辅相成, 实时高频的宏观经济指标在多方面具有重要意义: 一方面, 这有利于提高宏观经济信息获取的时效性, 进而提高宏观经济政策的有效性; 另一方面, 高频宏观数据的涌现必然会为宏观经济学家装上“显微镜”, 为宏观经济学的研究打开一个全新的世界。

然而, 就重要的宏观经济指标而言, 实时化、高频化的指标还基本为空白, 例如, 作为反映通货膨胀的重要指标 CPI, 目前各国仍主要通过定时定点调查实体店的商品和服务价格变化得到, 一般每月发布一次。当下中国正处于互联网经济快速发展的时代, 构建实时高频

<sup>①</sup> 本文获得国家社科基金重大项目 (16ZDA008)、国家杰出青年科学基金项目 (71325007) 和清华大学数据科学研究院的资助。本文通讯作者为姜婷凤。

的通货膨胀指标已经具备很强的必要性和可行性。一方面,中国电子商务发展迅速,网购对居民消费的各方面已经产生深刻影响,线上覆盖的居民生活消费的商品和服务类别越来越广泛,海量的在线价格数据可以实时采集,这为构建实时高频的物价指数提供了丰富的信息源<sup>①</sup>。另一方面,不断发展的大数据算法和计算速度,如机器学习、深度神经网络、并行计算、云计算等,为编制高频物价指数提供了新的工具和方法,具备了在极短的时间内完成从原始数据到指标生成的加工过程。这二者使得构建基于在线大数据的实时高频物价指数成为可能,这将极大地提高通货膨胀预测的时效性与准确性。

据此,本文旨在运用大数据的理念和技术手段,探索利用在线数据构造实时高频的新型居民消费价格指数(iCPI),并实现可视化发布和系统在线运行(见网站 [www.bdecon.com](http://www.bdecon.com)),具有较大的创新和贡献。尽作者所知,iCPI为中国首支基于互联网在线价格数据的、全面涵盖居民生活消费商品和服务篮子的、实时高频的居民消费价格指数,是传统CPI的有益补充,弥补了中国目前尚无在线高频居民消费价格指数的空白,具有重要的理论意义和现实意义。

iCPI有一系列突出的优点:将价格指数高频化、实时化,既包含月指数,又包含日指数、周指数,发布无滞后;数据从采集、清洗到加工和在线发布完全由计算机自动进行,既节省人力又减少人为干预因素。通过采用一系列对比检验及分析,本文发现iCPI既能很好地代表一般物价水平变动,又能体现出高频物价指数在很多方面的独特优势,例如,在线价格调价成本低、反应灵敏,除了能反映传统CPI与现有宏观经济指标的关系,还能及时地反映消费者信心变化、高频宏观经济指标(如利率、汇率等)的影响及其相互作用,为通胀研究增加了新的维度;iCPI周指数和日指数能实时反映假期(如春节)物价指数波动、精准捕捉典型事件(如“双十一”)影响,便于人们更为客观地了解物价指数波动及其反映的宏观形势。构建基于在线大数据的高频物价指标iCPI是新常态下创新宏观经济监测与预测体系的有益尝试,未来可以进一步将该方法应用于构建其他实时高频的宏观经济指标。

## 一、现有价格指数研究

### 1. 传统CPI编制方法

CPI作为三大宏观经济指标之一,对于刻画通货膨胀(紧缩)程度、制定货币政策、分析经济景气状况和经济周期等具有重要参考作用,是最受关注的价格指数之一,其编制准则历来受到各国及国际组织的重视。2004年国际劳工组织、国际货币基金组织等编译的《消费者价格指数手册:理论与实践》及2009年出版的《补充手册:消费价格指数实用编制指南》系统介绍了CPI编制的理论和实践问题,为各国编制CPI提供了准则与指导,但是各国依据国情制定的具体CPI编制方法仍有一定不同。

在中国,国家统计局负责CPI的编制、抽样、计算和发布。CPI编制是以“固定篮子指数”理论为基础,即选定一组具有代表性的商品和服务,以这些商品和服务价格变动代表一般消费商品和服务的价格水平变动,反映消费者购买代表性固定篮子平均支出费用的相对变化。国家统计局采用抽样调查方法,在全国抽选了500个市县作为价格采集点,如超市、服

<sup>①</sup> 国家统计局数据表明,2018年,全年社会消费品零售总额为380987亿元,比上年增长9.0%;全国网上零售额为90065亿元,比上年增长23.9%。其中,实物商品网上零售额为70198亿元,增长25.4%,占社会消费品零售总额的比重为18.4%(具体参见[http://www.stats.gov.cn/tjsj/zxfb/201901/t20190121\\_1645784.html](http://www.stats.gov.cn/tjsj/zxfb/201901/t20190121_1645784.html))。

装店、农贸市场等，对指定的商品篮子价格进行定时、定点采集<sup>①</sup>。中国 CPI 计算参数“五年一大调、一年一小调”，每五年进行一次基期轮换，对调查目录及其权数进行较大的调整，每年小调对权数的调整不大<sup>②</sup>。国家统计局 2016 年 1 月开始使用 2015 年作为新一轮的对比基期，此次 CPI 目录调整参考了联合国制定的《按目的划分的个人消费分类》(COICOP)<sup>③</sup>和国家统计局发布的《居民消费支出分类(2013)》<sup>④</sup>。在现阶段，国家统计局于每月 9 日公布上个月 CPI，遇到节假日往后顺延<sup>⑤</sup>。

陈立双和祝丹(2013)将中国 CPI 编制方法与《消费者价格指数手册：理论与实践》、美国 CPI 编制方法进行对比，表明中国 CPI 编制在样本轮换、权重基期更新的频率、发布工作等方面有较大的提升空间。徐强(2013)从多方面系统比较了 OECD 国家 CPI 的编制方法，包括消费品和服务分类、支出权数来源、指数计算方法和更新频率等方面，建议中国应该使价格数据调查方法更加科学化、加快权数更新的频率并系统引入各种质量调整方法。任栋和王琦(2013)采用 Divisia 指数法解构 CPI，发现消费支出结构变化是影响 CPI 核算的重要因素，而权重与消费支出结构在技术上的错位影响则相对较小。刘向耘和高宏(2016)通过国际对比，发现中国商品篮子确定方式与大多数经济体相同、符合国际标准，但五年更新一次权数的频率低于多数发达经济体，且食品烟酒、衣着类权数偏高，而居住、交通和通信类权数偏低。这些综合表明我国 CPI 编制仍有一定提升空间。

## 2. 在线价格指数编制和应用

随着信息技术及其支持硬件不断发展更新，“大数据”理念和技术应运而生，并且在诸多领域得到运用。目前国内外已有一些机构在探索利用在线数据构造价格指数，影响力较大的有美国麻省理工的“十亿价格项目”和中国的阿里巴巴网购系列指数。

2008 年 3 月，MIT 发布“十亿价格项目”：Billion Price Project (以下简称 BPP)<sup>⑥</sup>，项目旨在运用电商数据计算物价指数，引起了国际上的关注。创始人 Cavallo 和 Rigobon (2016)详细地介绍了 BPP 的数据收集、数据处理和应用，BPP 主要抓取同时在线上线下进行销售的大型零售商的价格数据。但是 BPP 没有服务品价格，该体系认为服务业价格通常很稳定，几乎没有波动源，且它发布的中国价格指数中只包含生鲜食品、超市食品，其他类别没有覆盖。

中国阿里巴巴利用自身数据资源，从 2014 年 6 月开始每月发布网购价格系列指数。该系列指数分为两种：一种是阿里巴巴全网价格指数(简称 aSPI)，它基于生产消费理论，用以反映阿里巴巴系列平台上各类商品价格水平的变化；另一种是根据固定篮子理论编制的阿里巴巴网购核心商品价格指数(简称 aSPI-core)，它用以反映阿里巴巴平台上主流商品的价格变动<sup>⑦</sup>。两种指数虽然原理不同，但是组成结构相似，均由十个大分类、约五百个基本分

① 参见国家统计局新闻办公室、资料中心：《您身边的统计指标——居民消费价格指数》[EB/OL]，2014 年 2 月 26 日，<http://www.stats.gov.cn/tjsz/tjbk/nszbz/201402/P020140226558226322808.pdf>。

② 《国家统计局城市司负责人就 CPI 基期轮换接受中国信息报专访》[N]，《中国信息报》，2016 年 2 月 19 日。

③ 参见联合国，《按目的划分的个人消费分类(COICOP)》[EB/OL]，<https://unstats.un.org/unsd/cr/registry/regist.asp?Cl=5>，2011。

④ 参见国家统计局，《居民消费支出分类》[EB/OL]，2013 年 10 月 21 日，[http://www.stats.gov.cn/tjsj/tjbz/201310/P0201310\\_21349384303616.pdf](http://www.stats.gov.cn/tjsj/tjbz/201310/P0201310_21349384303616.pdf)。

⑤ 国家统计局 2011 年 7 月将 CPI 发布的时间，从原来每月 11 日左右提前到每月 9 日，具体参见《封面故事：基层的统计数据是如何出炉的》[N]，《中国日报》，2012 年 2 月 22 日。

⑥ 详细信息见 BPP 官网 <http://bpp.mit.edu/>，及在 BPP 基础上建立的公司的主页 <http://www.pricestats.com/>。

⑦ 详细信息见阿里研究院官网 <http://www.aliresearch.com/>。

类构成,且所有的原始数据,包括商品价格、销量等都来自于阿里巴巴的在线数据。

随着大数据时代的在线数据可获得性增加,在线价格研究逐渐得到关注。Cavallo (2017) 通过对比 10 个国家 56 家大型多渠道零售商的线上线下价格的差异,发现线上价格和线下价格约 72% 的时间是相同的,二者价格变化虽然不一定同步,但是变化频率及价格调整幅度相似,表明线上价格能较好地代表线下价格,为利用线上价格开展价格研究提供了经验支持。Aparicio 和 Bertolotto (2017) 使用 BPP 价格体系中 2008~2015 年美国、英国、法国、德国以及荷兰的在线价格指数进行通货膨胀预测,发现滞后的在线价格通货膨胀有助于预测未来的线下通货膨胀水平。Hull 等 (2017) 通过搜集瑞典的在线水果和蔬菜价格并将其加入 CPI 预测中,结果显示 CPI 与在线商品价格存在明显的相关性。Gorodnichenko 等 (2018) 使用了美国和英国的线上市场价格数据研究在线价格特点,指出线上价格存在较低的搜索成本和调整成本,发现和线下价格相比,线上价格较为灵活,但仍存在黏性和较大的异质性。

综合上述,国家统计局编制 CPI 所调查的商品价格主要来自于实体销售渠道,发布频率是每月一次;BPP 中的中国价格指数包含的种类十分有限,没有较完整地覆盖 CPI 的商品篮子,且没有在网站上实时发布其价格指数;阿里巴巴网购价格系列指数的数据来源单一,涵盖类别有限,以月为单位并非高频价格指数,且 2016 年 12 月在其官网上停止更新<sup>①</sup>。实际上,作为电子商务平台仅用其平台自身的数据构建物价指数,指标发布和公司商业策略之间存在难以调和的矛盾。因此,目前中国尚无高频在线居民消费价格指数,本文旨在探索编制中国的实时高频在线 CPI,为传统 CPI 补充新的维度,为中国在线价格研究提供新视角;与 BPP 相比,它更切合中国实际、篮子覆盖范围大大拓展;与阿里巴巴网购价格系列指数相比,在数据源、编制与计算方法、发布频率等方面均不同。

## 二、高频通胀指标 iCPI 的编制方法

本文基于在线大数据编制了高频价格指数并已经上线稳定运行<sup>②</sup>,下面我们将这一新型价格指数称为“iCPI”,而将由统计局定期发布的官方 CPI 称为“传统 CPI”。“iCPI”这一指数编制的依据是,电商虽然是在互联网上经营的虚拟商店,但它们的商品来自于实体,与实体门店一样生产、消费,因而两者之间存在紧密联系,线上线下价格变动理论上应是高度相关的。

本文参考了国际劳工组织、国际货币基金组织等编译的《消费者价格指数手册:理论与实践》,借鉴中国国家统计局编制 CPI 的理论方法,来编制高频价格指数 iCPI,基本思想是:先用选定的一篮子商品代表物价总水平,采用计算机每天在固定时间抓取数据,然后算出固定篮子里的代表规格品价格水平的变化率,最后通过逐层加权平均得到相应的指数,该过程全部由计算机系统自动实现。除了 iCPI 月指数和国家统计局 CPI 频率一致,本文还充分挖掘大数据的优势,编制出更高频率的周指数和日指数,能更全面地、更及时地反映居民消费价格水平的变化。iCPI 的主要编制过程分为五步,具体如下:

### 1. 价格调查商品篮子选择

iCPI 商品篮子类别总体上和国家统计局 CPI 保持一致。为了使 iCPI 中一篮子商品和服

<sup>①</sup> 详细信息见阿里研究院官网 <http://www.aliresearch.com/>。

<sup>②</sup> iCPI 具体指数实时发布于网站 [www.bdecon.com](http://www.bdecon.com),并于 2017 年 6 月 1 日在国际著名金融信息服务商 Bloomberg 上线,2019 年 8 月在国内重要金融信息服务商 Wind 上线。

务具有代表性，一方面参考联合国编制的《按目的划分的个人消费分类》、中国国家统计局关于 CPI 的编制说明及其发布的《居民消费支出分类（2013）》，选取涵盖全国城乡居民生活消费的商品和服务<sup>①</sup>，这些商品和服务目录是与居民生活紧密相关的。另一方面考虑到在 2016 年 1 月，国家统计局对 CPI 基期做了一次轮换，本文根据统计局有关解释也做了相应调整，商品和服务篮子共分 8 个大类，46 个中类，262 个子类，具体分类见表 1。

表 1 iCPI 一篮子商品和服务目录

大 类	中 类	子 类
食品烟酒	食品	大米、面粉、其他粮食、粮食制品、薯类、干豆、豆制品、食用植物油、食用动物油、鲜菜、干菜及菜制品、猪肉、牛肉、羊肉、畜肉副产品、其他畜肉副产品、鸡、鸭、其他禽肉及制品、淡水鱼、海水鱼、虾蟹类、其他水产品及其制品、鸡蛋、其他蛋及其制品、鲜奶、酸奶、奶粉、其他奶制品、鲜瓜果、坚果、瓜果制品、食糖、糖果、糕点、其他糖果糕点、食用盐、酱油、食醋、调味酱、味精、其他调味品、方便食品、淀粉及其制品、膨化食品
	茶及饮料	茶叶、固体咖啡、其他固体饮料、饮用水、果汁饮料、其他液体饮料
	烟酒	烟草、白酒、葡萄酒、啤酒、其他酒类
	在外餐饮	正餐、快餐、地方小吃、其他在外餐饮
衣着	服装	男士西服、男士冬衣、男士夹克衫、男士毛线衣、男士运动装、男士衬衫 T 恤、男士裤子、男士内衣、女士外套、女士冬衣、女士毛线衣、女士运动装、女士衬衫 T 恤、女士裤子、女士裙子、女士内衣、婴幼儿服装、儿童上衣、儿童裤子、儿童裙子
	服装材料	服装材料
	其他衣着及配件	袜子、帽子、其他衣着配件
	衣着加工服务费	衣着洗涤保养、衣着加工
	鞋类	男鞋、女鞋、童鞋、鞋类加工服务
居住	租赁房房租	公房房租、私房房租
	住房保养维修及管理	木地板、瓷砖、水泥、涂料、板材、管材、厨卫设备、门窗、其他住房装潢材料、物业管理费、装潢维修费、其他住房费用
	水电燃料	水、电、管道燃气、液化石油气、取暖费、其他燃料
	自有住房	自有住房
生活用品及服务	家具及室内装饰品	柜、床、桌、椅、沙发、其他家具、灯具、其他室内装饰品
	家用器具	洗衣机、电冰箱（柜）、抽油烟机、空调器、热水器、炉具灶具、微波炉、其他大型家用器具、厨房小家电、生活小家电
	家用纺织品	被子、床单被罩、其他床上用品、窗帘门帘、其他家用纺织品
	家庭日用杂品	清洗用品、清洁用具、清洁用纸、厨具、餐具、茶具、家用手工工具、配电附件、雨具、其他日用杂品
	个人护理用品	清洁化妆品、护肤化妆品、彩妆化妆品、化妆器具、清洁类护理用品、护发美发用品、护理器具、其他护理用品
	家庭服务	家政服务、家庭维修服务

<sup>①</sup> 特别地，iCPI 篮子也涵盖服务，虽然和线上商品相比，线上服务价格相对较难获取，但是通过采集有关服务网站和 App 上的数据，仍能获取大部分服务类价格。有一小部分服务价格确实难以获取的，辅助于其他价格调查的方式，其对整个指数的影响几乎可以忽略不计。

(续)

大 类	中 类	子 类
交通和通信	交通	小型汽车、电动自行车、自行车、其他交通工具、汽油、柴油、其他车用能源、停车费、车辆使用费、交通工具零配件、车辆修理与保养、市内公共交通、出租汽车、飞机票、火车票、长途汽车、其他交通费
	通信	固定电话机、移动电话机、通信工具零配件、固定电话费、移动通信费、上网费、其他通信服务、邮政邮递、快递服务
教育文化和娱乐	教育	工具书、教材、参考资料、其他教育用品、学前教育、小学初中教育、高中中职教育、高等教育、课外教育、专业技能培训
	文化娱乐	电视机、照相机、台式计算机、笔记本平板、乐器、音响、其他文娱耐用消费品、书报杂志、纸张文具、体育户外用品、游戏用品和玩具、园艺花卉及用品、宠物及用品、其他文化娱乐用品、电影票、景点门票、有线电视、健身活动、其他文娱服务、旅行社收费、其他旅游
医疗保健	药品及医疗器具	中药材、中成药、抗微生物药、消化系统用药、呼吸系统用药、解热镇痛药、抗肿瘤药、激素及影响内分泌药、心血管系统用药、血液系统用药、治疗精神障碍药、神经系统用药、消毒防腐剂创伤外科用药、泌尿系统用药、维生素矿物质类药、调节水电解质及酸碱平衡药、滋补保健品、医疗卫生器具、保健器具
	医疗服务	一般医疗服务、一般治疗操作、护理、其他综合医疗服务、病理学诊断、实验室诊断、影像学诊断、临床诊断、临床手术治疗、临床非手术治疗、康复医疗、中医治疗、其他医疗服务
其他用品和服务	其他用品类	金饰品、银饰品、铂金饰品、手表、箱包、母婴用品、眼镜
	其他服务类	宾馆住宿、其他住宿、美容、美发、洗浴、养老服务、金融服务、车辆保险、旅行保险、其他保险、中介服务、其他服务

注：居民消费商品和服务分类简称为8个大类、46个中类和262个子类，由于分类目录会定期调整，实际的中类和子类数目可能有出入。为了便于表示大类、中类、子类（由于一些类别没有子类，表格中没有列出子类）和子类的关系，用“—”连接各个类别，例如“食品烟酒”大类下的“食品”中类下的“粮食”子类下的“大米”子类可以记为“食品烟酒—食品—粮食—大米”，具体可见iCPI官网的指数查询栏目 <http://www.bdecon.com>。

根据国家统计局制定的价格调查规定“同一规格品的价格必须同质可比”，即需要保证规格品的稳定性、可比性，我们的在线规格品也要遵循类似原则，保证每天持续地采集同一个规格品的价格。由于过少的商品会使指数产生偶然误差，为了使计算的价格具有代表性，每个子类需要选择多种代表商品。商品选取特定品牌、特定规格的标准品，着重选取市场份额大、生活常见、在多个平台销售、一般家庭消费量较大且供给需求相对稳定的商品，每种规格品可以跟踪，另外，行业所占份额及电商平台的销售量是选取规格品的重要参考。

## 2. 价格来源选择及采集方法

(1) 价格来源。价格来源是iCPI和CPI最大的区别，iCPI的价格全部来自线上。为了得到稳定而有代表性的数据，本文力求选择有权威、真实性高的数据来源，对每种商品价格的来源进行了严格筛选，来源主要有两类。一是进行交易的网购平台，网购平台主要分为

B2B (Business to Business)、B2C (Business to Customer)、C2C (Customer to Customer) 等类别, B2B 的买卖双方主要是企业, 和居民生活消费的直接关系相对较弱; C2C 的卖方是小店铺或个人, 规模小, 经营不稳定。因此, iCPI 采价主要从天猫、京东、苏宁易购、亚马逊等大型 B2C 在线购物平台, 且选取这些平台上质量合格、销售量较大、比较稳定的商家。此外, 各类别也会尽可能地结合一些官方网站或者行业网站, 如耐克官网、汽车之家, 以丰富同种类别的价格来源。

二是不进行在线交易但实时更新商品价格数据的一些信息平台, 如“价格行情”网, 它由中国商务部负责运营管理, 每天及时公布全国农产品价格数据库;“搜房网”, 它作为全国最大的房地产家居网络平台, 覆盖全国多个城市的租金价格。这些平台的价格信息种类多, 来源广, 涉及全国各地, 有充分代表性, 能涵盖一些电商平台没有覆盖的类别, 包括一些服务类。

本文尽可能筛选合格的、可靠的在线价格信息来源, 但是由于价格来源的平台数量较多, 难以避免个别信息平台上的价格信息质量不高。然而从总体上考虑, 我们选取的平台总数超过 100 个, 每一个平台的随机因素可以相互抵消, 而且不同类别都有多种价格来源, 不只受单一平台的影响, 大数据以数据量之大弥补了单个商品受噪音扰动的影响, 且可以采用机器学习的方法将一些噪音大的数据清洗掉, 这保证了在线价格来源的可靠性和稳定性。

(2) 价格采集方法。统计局的价格调查方法是每月“定人、定点、定时”, 由采价人员手持数据采集器对规格品进行采价<sup>①</sup>。我们借鉴相似的原则, 只是价格采集通过计算机程序自动完成, 使用开源的数据采集程序, 在指定时间、指定网站、指定位置自动采集所需要的数据。价格每天采集一次, 采集后存入专用数据库。除了基本的商品名称、规格、产地、报价时间、商品价格之外, 为了便于后续的深度分析, 还同时采集了销售量、评价值、发货地、是否促销等多维数据。为了保证价格的代表性、减小系统误差, 同时也为了保证商品价格的稳定性, 本文对每种商品保证多个以上的价格来源。这种通过抓取互联网数据的方法能够极大地提高数据采集频率, 节省人力、物力成本, 减少人为对数据的干预, 使得在线价格指数的实时更新成为可能。

### 3. 确定权重

不同类别的商品和服务在居民日常消费中所占的比重不同, 因而在计算价格指数的时候应该为不同类别设定相应的权数, 通过加权平均得到指数。统计局 CPI 权数是根据居民家庭住户调查资料及相关统计资料整理得出, 必要时会辅以典型调查或专家评估来补充和完善<sup>②</sup>。在实际调查中, 由于各个市县情况不同, 调查统计范围也不同, 所确定的市县级权数也有一定的差别, 因此传统 CPI 的权重与地域有关<sup>③</sup>。iCPI 由逐级分类指数加权平均计算得出, 即子类指数通过加权平均得到中类指数, 中类指数加权平均得到大类指数, 大类指数加权平均得到 iCPI 总类指数, 但权重与地域无关, 只有一套权重。

iCPI 权重确定主要有两种思路: 第一种, 考虑到 iCPI 是基于线上市场的居民消费价格

<sup>①</sup> 参见国家统计局,《流通消费价格调查制度》[EB/OL], 2017 年 1 月 9 日, [http://www.stats.gov.cn/tjsj/tjzd/gjtjzd/201701/t20170109\\_1451375.html](http://www.stats.gov.cn/tjsj/tjzd/gjtjzd/201701/t20170109_1451375.html)。

<sup>②</sup> 参见国家统计局城市司,《走进 CPI——什么是计算 CPI 的权数》[EB/OL], 2011 年 6 月 10 日, [http://www.stats.gov.cn/ztjc/tjzs/zjpci/201106/t20110610\\_71462.html](http://www.stats.gov.cn/ztjc/tjzs/zjpci/201106/t20110610_71462.html)。

<sup>③</sup> 参见国家统计局城市司,《走进 CPI——市县级权重是如何确定的》[EB/OL], 2011 年 6 月 10 日, [http://www.stats.gov.cn/ztjc/tjzs/zjpci/201106/t20110610\\_71482.html](http://www.stats.gov.cn/ztjc/tjzs/zjpci/201106/t20110610_71482.html)。

数据, 权重应该采用居民家庭的线上消费支出结构, 但是目前尚无统一线上市场的、综合各线上平台的居民消费支出统计数据。第二种, 考虑到 iCPI 借鉴了传统 CPI 的编制方法, 线上线下价格高度相关, iCPI 权重确定思路应与传统 CPI 保持一致, 即各级各类权数理论上应依据居民家庭消费支出结构确定。本文当前采用第二种权重确定方式<sup>①</sup>, 由于国家统计局没有直接公布 CPI 权重具体数值<sup>②</sup>, 我们综合多种文献资料以及国际上 CPI 的编制方法推断出不同层次的权重。

本文主要参考各年《中国住户调查年鉴》《中国价格统计年鉴》《中国统计年鉴》及其他各类统计年鉴、Wind 数据库、彭博数据库、各行业研究报告, 特殊类别辅之于问卷调查, 以确定大类、中类和子类的权重, 并保证每年更新一次权重。其中, 《中国住户调查年鉴》是一个重要参考, 它每年更新一次, 是统计局通过对全国 31 个省、自治区、直辖市的 16 万个居民家庭, 开展的住户收支与生活状况调查, 详细记录了家庭收入来源和支出明细。该年鉴得到的居民消费支出结构与统计局 CPI 商品篮子基本一致, 可用于计算不同类别消费品占居民总消费支出的比例, 从而确定不同类别相应的权重。大类和部分中类权重主要从国家统计局有关年鉴中获得, 部分中类和子类权重通过多种参考资料和算法推算得出。考虑到子类权数、中类权数、大类权数都是介于 (0, 1) 的值, 子类权数经过三步加乘计算, 因权数设定的区别而引起的指数偏差被压缩, 影响较小。

#### 4. 指数计算

iCPI 的计算方法依据的是通行的 CPI 的计算方法, 但极大地提高了指数频率, 在月指数的基础上同时计算和公布日指数、周指数。在此, 本文以环比指数的计算过程为例进行具体介绍:

(1) 计算思路。第一步, 计算 iCPI 篮子中各个规格品每日/周/月的价格变化比。第二步, 计算每日/周/月子类价格指数, 由该子类下所有规格品在每个平台上价格变化比的几何平均得到。在这个过程中因为每种商品在子类中的权重难以确定, 从规格品价格变动到子类指数采用几何平均的算法, 这与传统 CPI 的计算方法一致, 也符合国际通行要求和我国国家统计局的做法<sup>③</sup>。各国编制价格指数时在这一层次都采用非加权的平均计算方法, 具体包括 Carli 指数、Dutot 指数及 Jevons 指数, 考虑指数检验功效和商品替代效果等问题, Carli 指数不能通过时间逆检验, Dutot 指数不能通过公度性检验, 只有以几何平均作为计算方法的 Jevons 指数通过了所有的检验, 因而本文采用几何平均的方式计算子类价格指数。第三步, 计算每日/周/月中类、大类及总类价格指数, 采用拉氏指数, 根据所包含的各级分类的基期权数逐级加权平均。编制价格指数通常有两种方法, 一种是拉氏 (Laspeyres) 指数, 它使用基期数量加权、计算当期与基期价格变化比; 另一种是帕氏 (Paasche) 指数, 它是使用当期数量加权、计算当期与基期价格变化比。IMF 等 (2004) 指出典型的消费者价格指数的正统编制方法是基于拉氏价格指数, 通常采用基期支出份额作为权数, 对价格变化或价格比率观察值进行加权平均。而且中国国家统计局也是采用拉氏指数计算 CPI, 因此本文

<sup>①</sup> iCPI 项目组也在积极探索第一种权重确定方法, 未来将对比研究由不同权重得到的 iCPI 各层指数。

<sup>②</sup> 国家统计局表明, 新基期权数主要来源于全国住户收支与生活状况最新调查数据, 同时也适当参考专项调查和行政记录等数据进行相应调整。具体参见《国家统计局城市司负责人就 CPI 基期轮换接受中国信息报专访》[N], 《中国信息报》, 2016 年 2 月 19 日。

<sup>③</sup> 参见国家统计局城市司, 《走进 CPI——CPI 的个体指数是如何计算的?》[EB/OL], 2011 年 6 月 10 日, [http://www.stats.gov.cn/ztc/tjzs/zjcpi/201106/t20110610\\_71493.html](http://www.stats.gov.cn/ztc/tjzs/zjcpi/201106/t20110610_71493.html)。



也采用拉氏指数计算中类、大类及总类的价格指数。<sup>①</sup>

(2) 计算过程。本部分结合日指数计算公式进一步解释 iCPI 的计算过程。iCPI 有 8 个大类, 分别记为  $D_1, D_2, \dots, D_8$ , 权重分别为  $W_{D1}, W_{D2}, \dots, W_{D8}$ , 且  $W_{D1} + W_{D2} + \dots + W_{D8} = 1$ ; 以  $D_1$  这一大类为例, 假设  $D_1$  有 6 个中类  $Z_1, Z_2, \dots, Z_6$ , 权重分别为  $W_{Z1}, W_{Z2}, \dots, W_{Z6}$ , 且  $W_{Z1} + W_{Z2} + \dots + W_{Z6} = 1$ ; 某一中类  $Z_1$  有 4 个子类  $S_1, S_2, S_3, S_4$ , 权重分别为  $W_{S1}, W_{S2}, W_{S3}, W_{S4}$ , 且  $W_{S1} + W_{S2} + W_{S3} + W_{S4} = 1$ ; 某一子类  $S_1$  有 10 个规格品  $G_1, G_2, \dots, G_{10}$ , 第  $t-1$  天它们的价格分别为  $P_{G1,t-1}, P_{G2,t-1}, \dots, P_{G10,t-1}$ , 它们第  $t$  天的价格分别为  $P_{G1t}, P_{G2t}, \dots, P_{G10t}$ 。计算过程如下:

第一, 计算规格品价格变化, 第  $t$  天规格品  $i$  的环比价格变化为  $H_{Git} = \frac{P_{Git}}{P_{Gi,t-1}}, i = 1, 2, \dots, 10$ ;

第二, “不同规格品→子类价格变化”采用几何平均法, 则第  $t$  天子类  $S_1$  的环比价格变化为  $H_{S1t} = \sqrt[10]{H_{G1t} \times H_{G2t} \times \dots \times H_{G10t}}$ ;

第三, “不同子类→中类价格变化”采用加权算术平均法, 则第  $t$  天中类  $Z_1$  的环比价格变化为  $H_{Z1t} = W_{S1} \times H_{S1t} + W_{S2} \times H_{S2t} + W_{S3} \times H_{S3t} + W_{S4} \times H_{S4t}$ ;

第四, “不同中类→大类价格变化”采用加权算术平均法, 则第  $t$  天大类  $D_1$  的环比价格变化为  $H_{D1t} = W_{Z1} \times H_{Z1t} + W_{Z2} \times H_{Z2t} + \dots + W_{Z6} \times H_{Z6t}$ ;

第五, “不同大类→整个 iCPI”采用加权算术平均法, 则第  $t$  天的环比 iCPI 为  $H_{iCPI t} = W_{D1} \times H_{D1t} + W_{D2} \times H_{D2t} + \dots + W_{D8} \times H_{D8t}$ 。

周指数和月指数计算公式与日指数类似, 只是规格品价格相应地采用周平均价格和月平均价格, 然后算出规格品的周价格变化/月价格变化, 再通过相应的几何平均和加权平均得到子类、中类、大类和总类的周指数/月指数, 本文不一一赘述。

### 5. 系统实现

以上过程已通过编制相关软件, 实现计算机系统自动处理。系统的主要流程包括: 利用爬虫程序不断采集在线价格数据, 实时将数据存入数据仓库, 对数据进行清洗整理, 对各类指数进行计算, 对指数进行可视化实时发布<sup>②</sup>。指数具体包括总类、8 个大类、46 个中类及 262 个子类的日指数、周指数和月指数, 每日凌晨可以查询截至上一天所有类别的日、周、月指数。

特别地, 为了构建实时发布的高频物价指标, 在系统实现过程中, 还有很多实际问题要考虑。例如, 对异常样本的识别和处理、对打折商品过度波动的处理、对规格品下架的及时监测和补充等, 这些都是系统的重要组成部分, 但由于篇幅原因, 此处不一一详述。

## 三、实时高频 iCPI 评测及应用初探

在线的实时高频 iCPI 反映的信息要比一般实体店更为快捷和丰富, 网络调价成本低、相对较为容易、在线价格反应更为灵敏, 实体店调价成本较高、相对更为困难、线下价格调整存在一定时滞, 所以 iCPI 具有独特的研究价值。iCPI 不但包含和传统 CPI 相同频率的

<sup>①</sup> 参见国家统计局城市司, 《走进 CPI——CPI 编制采用什么计算公式?》[EB/OL], 2011 年 6 月 10 日, [http://www.stats.gov.cn/ztc/tjzs/zjpci/201106/t20110610\\_71491.html](http://www.stats.gov.cn/ztc/tjzs/zjpci/201106/t20110610_71491.html)。

<sup>②</sup> 这些流程都由计算机自动完成, 正常情况下不需要人工干预。主要结果在 iCPI 官网 [www.bdecon.com](http://www.bdecon.com) 实时发布。

月指数，而且包含高频的日指数、周指数，不同频率的 iCPI 提供了不同的研究视角，在实际中有广泛的应用前景，对于实时跟踪物价变化、监测宏观经济形势、预测金融市场动态具有重要意义，是传统 CPI 的有益补充<sup>①</sup>。

本部分将结合 2016 年 1 月 1 日~2018 年 12 月 31 日的 iCPI 数据及该期间的有关宏观经济变量，从多方面分析 iCPI 的表现和应用。iCPI 数据均可从网站 (www. bdecon. com) 公开查询，若未加特殊说明，本文中编制的指数都是环比指数，即以“上一期 (上日/上周/上月) =100”，在分析中采用指数环比变化率；除 iCPI 外的其他数据均来自 Wind 数据库；分析采用 Eviews 7. 0。

1. iCPI 与现有宏观经济指标

(1) 月度 iCPI 与传统 CPI 的关系。为考察 iCPI 的代表性，将 iCPI 与传统 CPI 月环比指数进行对比。因价格数据从 2016 年 1 月份开始采集，得到的 iCPI 月环比指数从 2016 年 2 月开始到 2018 年 12 月，共 35 个月，总类 iCPI 和 CPI 的走势如图 1 所示。

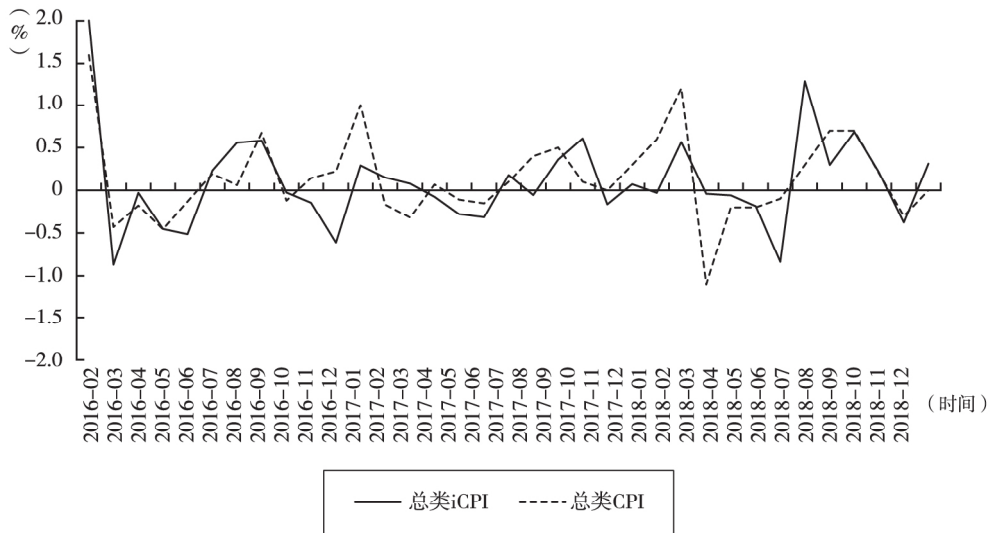


图 1 总类 iCPI 和 CPI 月度环比指数对比

从图 1 可以看出，总类月度 iCPI 与传统 CPI 所代表的价格变化方向 (环比上涨或下降)、幅度均较为接近；进一步计算相关系数，发现总指数 iCPI 与 CPI 的 Pearson 相关系数为 0.66，T 统计量为 5.0333，说明二者在 1% 的显著性水平下显著相关；通过统计 iCPI 和 CPI 的变化方向，发现变化方向一致的月份有 28 个，占比 80%，说明线上线下的总体市场价格变化是趋于一致的，通过基于互联网大数据实时采样的 iCPI 与传统人工采样统计的 CPI 所反映的价格变化总体相似，iCPI 月指数有较好的代表性。

此外，进一步分析实时高频 iCPI 与传统 CPI 变动方向不同的时期，往往是由于有特殊的节假日，例如 2016 年 11 月有“双十一”购物节，iCPI 月环比下降、CPI 环比上升；2017 年 2 月处于春节期间，物价普遍上涨，iCPI 月环比上升、CPI 环比下降。iCPI 和 CPI 的这

<sup>①</sup> iCPI 是传统 CPI 的有益补充，而不是替代，是衡量通货膨胀的一个重要方法。一方面，虽然目前线上零售额占比不是很高，但是线上线下价格高度相关，指数评测表明线上 iCPI 对线下 CPI 有很强的预测作用。另一方面，随着线上市场规模持续高速增长，线上占比将不断提高，线上市场的通胀衡量本身也越来越重要，iCPI 将变得必不可少。

种差异正是由于 iCPI 能更及时地反映节假日的变化,无论是新兴起的线上购物节还是传统的节假日,因此是十分合理的;这种差异从侧面表明新指数 iCPI 能实时地捕捉一些新信息,是传统 CPI 的有益补充,也印证了在这个快速发展的互联网时代编制在线 iCPI 的必要性。

(2) iCPI 与其他宏观经济指标。相比于传统 CPI,新指数 iCPI 对于宏观经济是否具有更好的解释力呢?我们进一步分析 iCPI、CPI 与其他宏观经济指标的关系:一方面,采用和 CPI 同频的月度宏观经济指标,便于进行对比分析;另一方面,采用更高频的周度 iCPI,分析其与周度宏观经济指标的关系。根据经典的货币数量论、费雪效应、汇率传递效应、理性预期等理论,我们选取的宏观变量主要包括货币增速、利率、汇率、PPI、PMI 等变量。特别地,我们还选取了消费者信心指数,用以对比考察消费者信心强弱和线上线下物价变化的关系;采用不同期限的中国国债到期收益率之差衡量利率期限结构,因为它包含有关未来利率和预期通胀变动的因素,用以考察通胀预期和物价变化的关系。具体的变量含义及描述性统计见表 2。

表 2 变量含义及描述性统计

变量	含义	观测量	均值	最大值	最小值	标准差
<i>iCPI</i> __monthly	在线 iCPI 月环比变化率 (%)	34	0.09	2.00	-0.87	0.5609
CPI	传统 CPI 月环比变化率 (%)	34	0.15	1.60	-1.10	0.5177
M2	广义货币月环比增长率 (%)	34	0.73	2.63	-0.34	0.7017
Shibor __monthly	上海同业拆借利率月环比变化率 (%)	34	0.43	4.67	-5.68	2.4152
YTMdiff __monthly	中国国债到期收益率息差月环比变化幅度 (1 个月减 0 年) (BP)	34	-0.20	160.27	-157.03	43.9466
CFETS __monthly	CFETS 人民币汇率月环比变化 (%) (人民币对一篮子货币的汇率指数, 正向变化表示人民币升值)	34	-0.24	1.76	-3.40	1.0003
PPI	生产者物价指数月环比变化率 (%)	34	0.37	1.60	-0.40	0.4877
PMI	采购经理人指数月环比变化率 (%)	34	0.04	2.45	-1.95	1.0118
Consumer Confidence	消费者信心指数月环比变化率 (%)	34	0.49	4.47	-4.21	2.0014
<i>iCPI</i> __weekly	在线 iCPI 周环比变化率 (%)	154	0.02	1.59	-0.89	0.3738
Shibor __weekly	上海同业拆借利率周环比变化率 (%)	154	0.10	8.79	-6.91	2.0835
YTMdiff __weekly	利率期限结构息差周环比变化幅度 (BP)	154	0.39	61.98	-80.96	18.7072
CFETS __weekly	CFETS 人民币汇率周环比变化 (%)	154	-0.05	1.26	-1.77	0.4414

注:考虑到有些变量截至 2018 年 12 月 31 日尚未公布,分析中统一采用 2016 年 2 月~2018 年 11 月的月度变量。表中采用的“中国国债到期收益率息差”等于“1 个月的中国国债到期收益率减去 0 年的中国国债到期收益率”,“息差月环比变化”等于“这个月的息差减去上个月相同期间的息差”,“息差周环比变化”等于“这个周的息差减去上周相同期间的息差”。

第一,平稳性检验。对这些变量进行单位根检验,采用 ADF 检验法,结果如表 3 所示。可以看出,大部分变量都在 1% 的显著性水平下拒绝“存在单位根”的原假设,个别变量在 5% 的显著性水平下拒绝原假设,说明这些变量都是平稳的,不存在单位根。

表3 变量单位根检验

变 量	检验形式 (C, T, L)	ADF 检验 t 值	P 值	是否平稳
<i>iCPI</i> _monthly	(C, T, 3)	-6.1874***	0.0000	是
<i>CPI</i>	(0, 0, 0)	-6.2297***	0.0000	是
<i>M2</i>	(C, T, 3)	-6.1226***	0.0000	是
<i>Shibor</i> _monthly	(0, 0, 0)	-5.4615***	0.0000	是
<i>YTMdiff</i> _monthly	(0, 0, 4)	-5.6237***	0.0000	是
<i>CFETS</i> _monthly	(0, 0, 0)	-5.1471***	0.0000	是
<i>PPI</i>	(C, T, 5)	-4.0166**	0.0199	是
<i>PMI</i>	(C, T, 0)	-8.0855***	0.0000	是
<i>Consumer Confidence</i>	(0, 0, 0)	-6.1177***	0.0000	是
<i>iCPI</i> _weekly	(0, 0, 1)	-10.41166***	0.0000	是
<i>Shibor</i> _weekly	(0, 0, 2)	-11.4643***	0.0000	是
<i>YTMdiff</i> _weekly	(0, 0, 0)	-11.0423***	0.0000	是
<i>CFETS</i> _weekly	(0, 0, 1)	-10.6301***	0.0000	是

注：\*\*\*、\*\*、\* 分别表示在 1%、5%、10% 水平下显著；检验形式中 C、T、L 分别代表是否加入常数项、趋势项及滞后阶数（根据 SIC 准则判断）。

第二，相关性分析。进一步计算 *iCPI*、*CPI* 与同期宏观经济指标的 Pearson 相关系数，结果如表 4 所示。根据月度指标的相关系数，*iCPI* 和 *CPI* 有共同之处：二者自身在 1% 显著性水平下显著正相关，相关系数为 0.66；均与 *PMI* 在 10% 显著性水平下显著负相关（*CPI* 是在 1% 显著性水平下），相关系数在 -0.3 ~ -0.5 之间；均与 *Shibor* \_monthly、*YTMdiff* \_monthly、*CFETS* \_monthly 等同期不显著相关，后面将进一步检验二者与这些变量滞后项的关系。这说明传统 *CPI* 能反映的信息，*iCPI* 往往也能反映。

然而，*iCPI* 和 *CPI* 也有不同之处：*iCPI* \_monthly 和 *M2* 在 5% 显著性水平下显著负相关，相关系数为 -0.34，表明当期通胀压力增加时，货币当局可能减少同期货币供应，其影响可能到后期才能显现，后面将进一步研究二者滞后期的相互影响；*iCPI* \_monthly 和 *Consumer Confidence* 在 10% 显著性水平下显著正相关，相关系数为 0.29，表明消费者信心指数越高，消费者的消费意愿越强，并且越快地通过在线市场表现出来。但是 *CPI* 与这两个变量均不相关，说明在线 *iCPI* 反应较为灵敏，低成本的网络调价能捕捉一些传统 *CPI* 无法及时反映的信息。

此外，根据更高频的周度相关系数，*iCPI* \_weekly 与 *Shibor* \_weekly 在 1% 显著性水平下显著正相关，这表明在通胀压力上升时，货币当局可能减少货币供应，利率可能上升；与 *YTMdiff* \_weekly 在 5% 显著性水平下显著负相关，表明当收益率曲线斜率变得更陡峭时（息差加大），市场参与者预期未来存在较大的通胀压力，而当前的通胀压力则相对较小；与 *CFETS* \_weekly 在 5% 显著性水平下显著负相关，表明人民币对外升值（指数上升），国内通胀程度会有所缓解。这些结果表明高频通胀指标与高频宏观指标有紧密的联系，它能较好地捕捉一些低频通胀指标所无法反映的信息，为通胀研究增加了高频维度。

表 4 变量间的同期相关性

	<i>CPI</i>	<i>M2</i>	<i>Shibor _monthly</i>	<i>YTMdiff _monthly</i>
<i>iCPI _monthly</i>	0.6646***	-0.3419**	0.0874	-0.0612
	5.0309	-2.0584	0.4966	-0.3471
	0.0000	0.0478	0.6229	0.7308
<i>CPI</i>	1.0000	0.0206	-0.0256	-0.0464
	—	0.1166	-0.1451	-0.2625
	—	0.9079	0.8855	0.7946
	<i>CFETS _monthly</i>	<i>PPI</i>	<i>PMI</i>	<i>Consumer Confidence</i>
<i>iCPI _monthly</i>	-0.1007	-0.1511	-0.3245*	0.2924*
	-0.5725	-0.8648	-1.9404	1.7298
	0.5710	0.3936	0.0612	0.0933
<i>CPI</i>	0.1275	0.1081	-0.5147***	0.1641
	0.7274	0.6153	-3.3956	0.9413
	0.4723	0.5427	0.0018	0.3536
	<i>Shibor _weekly</i>	<i>YTMdiff _weekly</i>	<i>CFETS _weekly</i>	
<i>iCPI _weekly</i>	0.2337***	-0.1854**	-0.2049**	
	2.9629	-2.3265	-2.5803	
	0.0035	0.0213	0.0108	

注：\*\*\*、\*\*、\* 分别表示在 1%、5%、10% 水平下显著；表中三个数值由上到下分别表示相关系数、T 统计量和 P 值。

第三，格兰杰因果检验。通货膨胀与其他宏观经济变量间往往会相互影响，且存在滞后作用，我们进一步进行格兰杰因果检验，结果如表 5 所示。从月度变量来看，*iCPI* 和 *CPI* 有一些共同之处：在 5% 显著性水平下，*M2* 均是 *iCPI* 和 *CPI* 的格兰杰原因，其中 *iCPI* 的 F 统计量 (6.1110) 略高于 *CPI* (5.0578)，表明货币供应量变化会逐渐影响线上线下的通胀水平；在 10% 显著性水平下，*YTMdiff \_monthly* 均是 *iCPI* 和 *CPI* 的格兰杰原因，其中 *CPI* 的 F 统计量 (9.3862) 明显高于 *iCPI* (2.4035)，表明基于利率期限结构的通胀预期对于线上线下的通胀均有预测作用。

同样地，*iCPI* 和 *CPI* 也有不同之处：在 10% 显著性水平下，*iCPI* 也是 *M2* 的格兰杰原因，即 *iCPI* 和 *M2* 互为格兰杰因果，说明通胀水平也会影响到货币供应量增速，货币当局观察到通胀率较高时，可能会减少后续货币供应；在 1% 显著性水平下，*Shibor \_monthly* 是 *iCPI* 的格兰杰原因，表明利率变化会影响资金成本，进而抑制过热的经济或者刺激低迷的经济，并逐渐传导到物价上；在 5% 显著性水平下，*iCPI* 是 *CFETS \_monthly* 的格兰杰原因，这表明国内的通胀水平（人民币对内价值）也会逐渐影响人民币汇率（人民币对外价值）。这说明相比于传统 *CPI*，在线 *iCPI* 与利率、汇率等宏观变量的传导关系更为顺畅，能较好地反映宏观经济变化。

此外，从周度变量来看，在 1% 显著性水平下，*CFETS \_weekly* 是 *iCPI* 的格兰杰原

因，表明人民币升贬值的影响在后期会逐渐传导到物价上，人民币外部价值变化对内部价值影响较为迅速；在10%显著性水平下，*iCPI*是*Shibor \_weekly*的格兰杰原因，这符合前面的结论，货币当局观察到通胀压力上升时，会减少货币供应，进而影响资金价格。这说明高频通胀指标*iCPI*能较好地反映高频宏观变量的影响，甚至是一些宏观变量的先行指标，它能够揭示一些低频通胀指标所无法反映的关系。

综合以上分析，我们发现高频*iCPI*不但能反映传统CPI与现有主要宏观经济变量的关系，还能捕捉一些传统CPI无法反映的信息，这既包括月度*iCPI*能及时地反映同期消费者信心变化、与月度利率/汇率间有较顺畅的传导关系，又包括周度*iCPI*能灵敏地反映高频宏观经济指标的影响及其相互作用，为通胀研究增加了新的维度。

表5 变量间的格兰杰因果检验

原假设	F 统计量	P 值	原假设	F 统计量	P 值
iCPI 组			CPI 组		
<i>M2</i> 不是 <i>iCPI</i> 的 Granger 原因	6.1110***	0.0065	<i>M2</i> 不是 <i>CPI</i> 的 Granger 原因	5.0578**	0.0136
<i>iCPI</i> 不是 <i>M2</i> 的 Granger 原因	2.9285*	0.0706	<i>CPI</i> 不是 <i>M2</i> 的 Granger 原因	0.4867	0.6200
<i>Shibor _monthly</i> 不是 <i>iCPI</i> 的 Granger 原因	5.2811***	0.0037	<i>Shibor _monthly</i> 不是 <i>CPI</i> 的 Granger 原因	0.9445	0.4014
<i>iCPI</i> 不是 <i>Shibor _monthly</i> 的 Granger 原因	0.5337	0.7481	<i>CPI</i> 不是 <i>Shibor _monthly</i> 的 Granger 原因	1.9892	0.1564
<i>YTMdiff _monthly</i> 不是 <i>iCPI</i> 的 Granger 原因	2.4035*	0.0776	<i>YTMdiff _monthly</i> 不是 <i>CPI</i> 的 Granger 原因	9.3862***	0.0008
<i>iCPI</i> 不是 <i>YTMdiff _monthly</i> 的 Granger 原因	0.4314	0.8208	<i>CPI</i> 不是 <i>YTMdiff _monthly</i> 的 Granger 原因	1.5407	0.2325
<i>CFETS _monthly</i> 不是 <i>iCPI</i> 的 Granger 原因	0.3057	0.7391	<i>CFETS _monthly</i> 不是 <i>CPI</i> 的 Granger 原因	0.0407	0.9602
<i>iCPI</i> 不是 <i>CFETS _monthly</i> 的 Granger 原因	3.9561**	0.0311	<i>CPI</i> 不是 <i>CFETS _monthly</i> 的 Granger 原因	0.5449	0.5861
<i>PPI</i> 不是 <i>iCPI</i> 的 Granger 原因	0.2663	0.7682	<i>PPI</i> 不是 <i>CPI</i> 的 Granger 原因	1.4889	0.2436
<i>iCPI</i> 不是 <i>PPI</i> 的 Granger 原因	0.3161	0.7317	<i>CPI</i> 不是 <i>PPI</i> 的 Granger 原因	0.4307	0.6544
<i>PMI</i> 不是 <i>iCPI</i> 的 Granger 原因	0.2765	0.7606	<i>PMI</i> 不是 <i>CPI</i> 的 Granger 原因	0.6578	0.5261
<i>iCPI</i> 不是 <i>PMI</i> 的 Granger 原因	0.2778	0.7596	<i>CPI</i> 不是 <i>PMI</i> 的 Granger 原因	0.0674	0.9350
<i>Consumer Confidence</i> 不是 <i>iCPI</i> 的 Granger 原因	0.6447	0.5327	<i>Consumer Confidence</i> 不是 <i>CPI</i> 的 Granger 原因	1.0582	0.3610
<i>iCPI</i> 不是 <i>Consumer Confidence</i> 的 Granger 原因	0.4678	0.6313	<i>CPI</i> 不是 <i>Consumer Confidence</i> 的 Granger 原因	0.2425	0.7864
<i>Shibor _weekly</i> 不是 <i>iCPI _weekly</i> 的 Granger 原因	0.2322	0.7931			

(续)

原假设	F 统计量	P 值	原假设	F 统计量	P 值
iCPI 组			CPI 组		
<i>iCPI</i> <i>_weekly</i> 不是 <i>Shibor</i> <i>_weekly</i> 的 Granger 原因	2.7996*	0.0641			
<i>YTMdiff</i> <i>_weekly</i> 不是 <i>iCPI</i> <i>_weekly</i> 的 Granger 原因	0.1290	0.8791			
<i>iCPI</i> <i>_weekly</i> 不是 <i>YTMdiff</i> <i>_weekly</i> 的 Granger 原因	0.3864	0.6802			
<i>CFETS</i> <i>_weekly</i> 不是 <i>iCPI</i> <i>_weekly</i> 的 Granger 原因	5.9884***	0.0032			
<i>iCPI</i> <i>_weekly</i> 不是 <i>CFETS</i> <i>_weekly</i> 的 Granger 原因	2.0075	0.1380			

注:\*\*\*、\*\*、\* 分别表示 1%、5%、10% 显著性水平。

(3) 通货膨胀 (CPI) 预测。根据上面的分析, 我们可以探索利用实时高频的在线 *iCPI* 优化传统的通胀预测模型。考虑到 *iCPI* 于每月 1 日发布上一个月的月指数, 早于统计局发布 CPI 的时间 (每月 9 日左右), 且 *iCPI* 和传统 CPI 有显著的相关性, 因此可用于帮助预测同期传统 CPI。这部分将对加入和没有加入 *iCPI* 的通胀预测模型的效果<sup>①</sup>。

根据前面的格兰杰因果检验, *M2* 的滞后项可帮助预测通货膨胀 (CPI), 本文将 *M2* 增长率作为解释变量, 构造传统的通胀预测模型:

$$CPI_t = \alpha + \sum_{i=1}^p \beta_i CPI_{t-i} + \sum_{k=1}^m \varphi_k M2_{t-k} + \varepsilon_t \quad (1)$$

把在线 *iCPI* 作为解释变量, 作为新的通胀预测模型, 设定如式 (2) 所示。特别地, 考虑到 *iCPI* 早于传统 CPI 公布, 式 (2) 加入了第 *t* 期的 *iCPI*<sub>*t*</sub> (即 *h*=0), 表示为:

$$CPI_t = \alpha + \sum_{i=1}^p \beta_i CPI_{t-i} + \sum_{h=0}^n \gamma_h iCPI_{t-h} + \varepsilon_t \quad (2)$$

进一步综合传统变量和新的变量, 在式 (1) 中同时加入在线 *iCPI*<sub>*t*</sub>, 表示为:

$$CPI_t = \alpha + \sum_{i=1}^p \beta_i CPI_{t-i} + \sum_{k=1}^m \varphi_k M2_{t-k} + \sum_{h=0}^n \gamma_h iCPI_{t-h} + \varepsilon_t \quad (3)$$

本文结合 AIC、SC 等准则得出相应模型的最优结果, 结果如表 6 所示。

<sup>①</sup> 本文主要利用同频的月度 *iCPI* 预测统计局通货膨胀, 未来可以进一步拓展到用混频 *iCPI* (*iCPI* 日指数、周指数和月指数) 预测统计局通货膨胀率。目前宏观经济的混频研究主要是把股票指数、同业拆借利率、大宗商品价格等金融市场高频数据加入模型中, 高频 *iCPI* 为宏观经济的混频预测提供了新的指标, 将有利于提高宏观经济预测的准确性和及时性。

模型	模型估计结果	拟合优度
(1) $p=1, m=2$	$CPI_t = 0.10 + 0.17CPI_{t-1} + 0.22M2_{t-1} - 0.15M2_{t-2} + \epsilon_t$ (0.61)(0.90) (2.01) (-1.41)	$R^2 = 0.28$ , 调整 $R^2 = 0.20$ AIC=1.17, SC=1.36
(2) $p=n=0$	$CPI_t = 0.10 + 0.61iCPI_t + \epsilon_t$ (1.40)(5.03)	$R^2 = 0.44$ , 调整 $R^2 = 0.42$ AIC=1.03, SC=1.12
(3) $p=n=0, m=2$	$CPI_t = 0.28 - 0.26M2_{t-2} + 0.55iCPI_t + \epsilon_t$ (2.96)(-2.72) (3.65)	$R^2 = 0.40$ , 调整 $R^2 = 0.36$ AIC=0.90, SC=1.04

注：模型拟合结果系数下面括号中的数值是对应系数的t值。

可以看出，式(2)的调整  $R^2$  明显高于式(1)，且AIC、SC值均小于式(1)，说明采用  $iCPI$  预测通货膨胀率的式(2)优于采用货币增速预测通胀的式(1)。同时加入两种变量的式(3)的表现明显优于只加入货币的式(1)，AIC、SC均低于式(2)，但是调整  $R^2$  低于式(1)，这可能是由于  $iCPI$  和  $M2$  货币增速存在一定的多重共线性(前面相关系数显示二者显著负相关)，所以在一定程度上削弱式(3)的表现。式(1)~式(3)综合表明，无滞后发布的总体  $iCPI$  在实时预测总体通货膨胀(通货紧缩)方面表现突出，这为宏观经济分析提供了新的工具。

### 2 $iCPI$ 与物价实时监测

(1)  $iCPI$  周环比指数——实时反映假期物价指数波动。周指数的时间周期介于日指数和月指数之间，不会像日指数因频率过高而变化太小，太小的价格波动难以捕捉和分析，同时因计算精度限制和扰动而容易造成误差；也不会像月指数频率太低，时滞严重，一些商品的价格有可能在一个月内就产生了大波动，而从月指数中不能有效反映出，周指数更适合捕捉“假日经济”带来的价格变动，总类  $iCPI$  周环比指数如图2所示。<sup>①</sup>

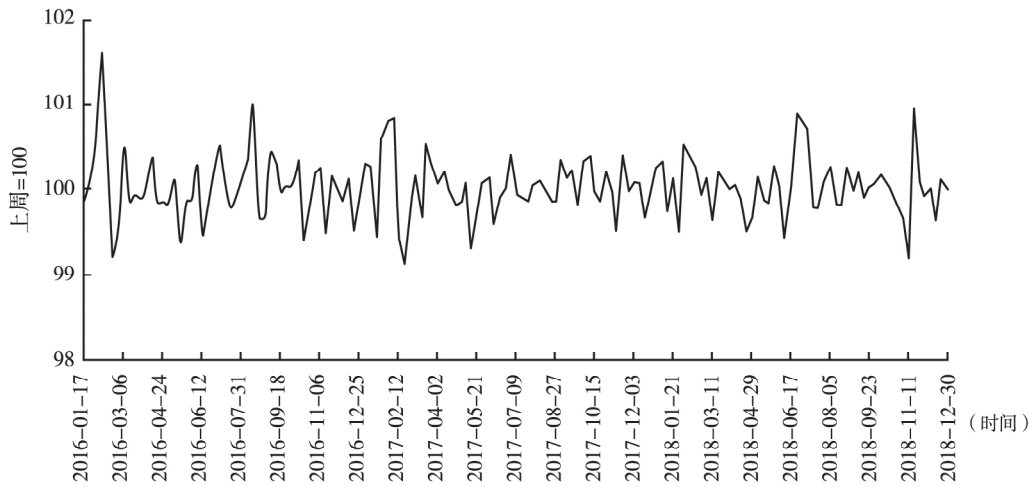


图2 总类  $iCPI$  环比周指数变化

资料来源： $iCPI$  官网 [www.bdecon.com](http://www.bdecon.com)<sup>①</sup>。

<sup>①</sup> 在  $iCPI$  官网 [www.bdecon.com](http://www.bdecon.com) 上的“指数查询”栏目，可以查询总类、8个大类、46个中类、262个子类的日、周、月指数，并下载相关指数及其变化趋势图。



众所周知，春节前后通常是一年中物价波动幅度较大的时期，2016年2月8日、2017年1月28日、2018年2月16日分别为2016年、2017年、2018年的正月初一。可以看出，这几年iCPI周指数均呈现节前和节中涨幅较大、节后下跌明显的特点。例如，2017年1月23日~2017年1月29日、2017年1月30日~2017年2月5日、2017年2月6日~2017年2月12日、2017年2月13日~2017年2月19日iCPI周指数分别环比上涨0.79%、上涨0.85%、下降0.39%、下降0.89%。在此基础上，还可以进一步结合各大类、中类和子类的周指数，分析具体哪些类别波动较大，iCPI周指数为洞察春节前后的物价变化提供了新工具。

(2) iCPI日环比指数——精准捕捉典型事件影响。日指数具有频率高、实时更新的优点，总类iCPI日环比指数如图3所示。可以看出，绝大多数日期里iCPI小幅度波动，也有个别日期的日波动较大，大幅波动往往反映了一些特殊事件，如“双十一狂欢节”“京东618购物节”。此外，一些极端天气，如严重的雾霾天气、强降雨天气，可能导致特定类别波动剧烈。iCPI提供了更详细的大类、中类和子类日指数，通过借助日指数的“显微镜”，观察到的特定波动状况，能反映传统CPI所无法反映的一些特点，为物价分析提供更多维度，我们后文将具体分析。

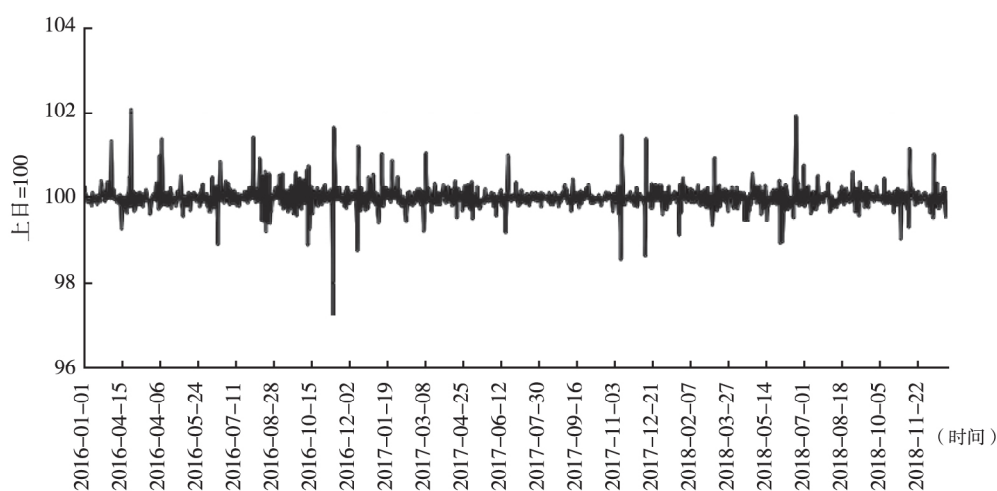


图3 总类iCPI环比日指数变化

资料来源：iCPI官网 [www.bdecon.com](http://www.bdecon.com)。

本文进一步主要以高频iCPI日指数为例，对比分析2016年、2017年和2018年“双十一”购物节期间物价变化特点，如图4和图5，具体异同点如下：

第一，总类变化之趋势。从图4可以看出，iCPI总指数在三年的“双十一”前后变化趋势很相似，但变化幅度有一定差异。具体来看，在“双十一”当天，iCPI总水平呈下降趋势，2016~2018年分别下降2.81%、1.46%、0.70%，下降幅度依次减小；2016年“双十一”之前出现小幅上涨，2017年、2018年节前上涨、下跌趋势交错；“双十一”过后，物价明显回升，总类iCPI在三年的11月12日分别上涨1.62%、1.44%、1.14%，但上涨幅度也逐年减小。此外，通过计算iCPI总类在2016~2018年的11月日指数的标准差和两两相关系数，发现三年标准差分别为0.64、0.42、0.33，说明购物节所在月份指数波动下降明显；2016年/2017年、2017年/2018年、2016年/2018年相关系数均在1%显著性水

平下显著为正,分别为0.90、0.78、0.64,表明总类指数在相邻两年波动较为相似,跨年后相似性有所下降。

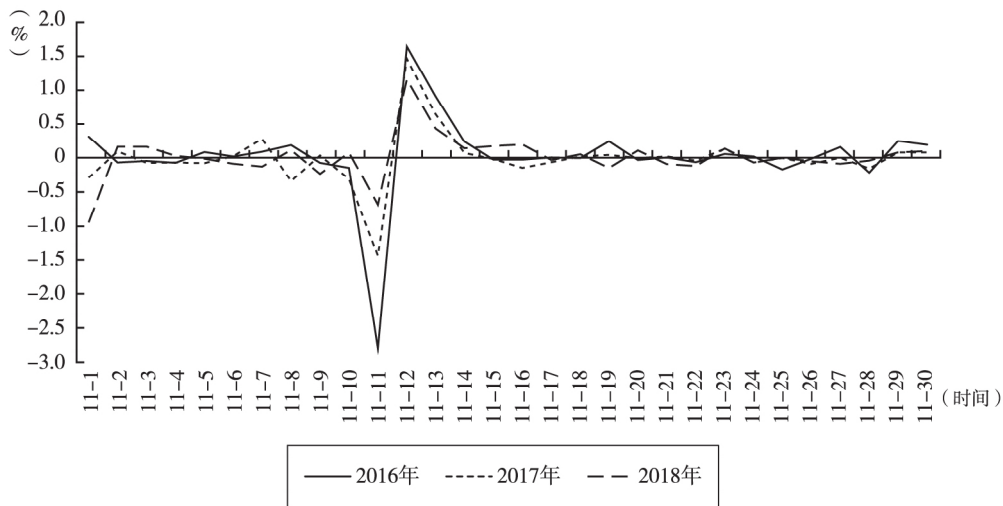


图4 2016~2018年“双十一”期间总类iCPI日指数对比

第二,大类变化之差异。从图5中可以看出,2016~2018年“双十一”iCPI八大类价格指数波动差异十分明显,除了“居住”和“教育文化和娱乐”,其他六大类连续三年指数都下降。其中,“衣着”连续三年在“双十一”当天的下降幅度最大,分别下降14.49%、3.48%、3.30%,但是打折幅度也在逐年递减。在全民购物狂欢潮中,女性朋友往往是抢购的“主力军”,其中“衣着”消费额占其总消费额较大比例,是商家每年打折力度较大的类别。此外,通过计算iCPI八大类在2016~2018年的11月11日指数的标准差和两两相关系数,发现三年标准差分别为4.52、1.10、1.50,说明八大类差异先下降后上升;2016年/2017年、2017年/2018年、2016年/2018年相关系数分别为0.83、0.80、0.75,表明八大类在这三年间“双十一”的波动关系较为相似。

第三,价格变化之时点。2016年“双十一”前后总类和八大类价格指数波动呈“V”形曲线,主要集中在11月11日当天打折,而2017年、2018年“双十一”总类和八大类价格指数波动呈“W”形曲线,部分商家在11月11日之前已经开始打折,“医疗保健”“衣着”前期波动明显。表明随着购物节不断发展,2017年“双十一”打折比往年来得更早,而且这种趋势延续到2018年,一些电商抢在11月11日前开始打折,分散了“双十一”当天全网物价下跌压力,所以2017年、2018年的11月11日当日物价下降幅度不及2016年11月11日,但2017年、2018年的11月日指数波动均显著下降(2016~2018年的11月日指数的标准差分别为0.64、0.42、0.33)。

总之,通过三年“双十一”购物节对比可以看出,在电商平台兴起的购物促销在一定程度上也传导到其他部门,“互联网+”渗透的范围越来越广。虽然各大类影响程度有所不同,但也反映了不同类别线上线下融合程度的差异,特别是随着时间推移,这种差异也在发生变化。从最开始由某一电商平台发起的电商节,逐渐发展到多家电商平台的大战、再到全民参与狂欢,这反映出中国消费者的消费意识、消费方式和消费结构正在发生重大转变。

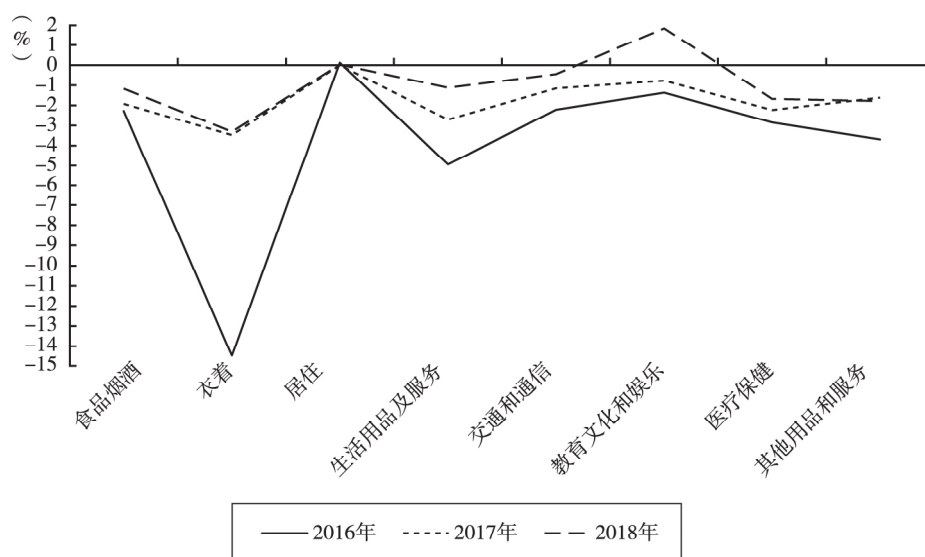


图5 2016~2018年“双十一”当天iCPI八大类日指数对比

### 3. 指数不足之处

上述研究表明基于在线大数据的iCPI确实存在一些独特的优势。但是该指数也有不足之处，这是iCPI未来有待完善的方向。一方面，iCPI高频日指数易受多种因素干扰，噪声成分相对较高。高频iCPI日指数和周指数的优点是能实时监测物价变化、捕捉一些特殊事件的影响，但也容易受到异常扰动的影响。虽然线上市场规模在不断扩大，在社会零售额中的占比在逐渐增大，但是仍有较大部分的线下市场不能覆盖，这也是iCPI目前不能替代CPI的重要原因。未来我们会进一步扩大商品来源和数量，加强数据质量监测，提高样本代表性；综合运用多种算法提高对有效信息和噪声的判断准确性，并增强日指数的稳定性和抗扰动性。

另一方面，iCPI尚无地域概念，无法观测地区的价格变化情况。传统CPI是通过在全国抽选500个市县、确定价格采集点<sup>①</sup>，包括城市和农村，但是iCPI中的重要价格来源之一电商平台，是全网统一价格。监测各地区的价格走势有利于为地域经济发展提供参考、分析总体价格波动来源，未来iCPI会进一步拓展到地区，例如按照商家所在地划分地区，采用新的方法构造地区iCPI。

## 四、结 论

近年来，国内外社会经济的不确定性日益增加，经济金融的短期波动变得更加频繁，如何对经济、金融进行实时高频的监测与预测是一个挑战性问题。本文探索了如何将物价指数实时化、高频化，编制了中国首套基于在线大数据的实时高频居民消费价格指数iCPI，具有重要的理论意义和现实意义，主要结论具体如下：

第一，本文通过运用大数据的理念和方法，设计出一整套基于互联网在线数据的居民消费价格指数(iCPI)，将物价指数实时化、高频化，并实现总类、8个大类、46个中类与262个

<sup>①</sup> 参见国家统计局新闻办公室、资料中心：《您身边的统计指标——居民消费价格指数》[EB/OL]，2014年2月26日，<http://www.stats.gov.cn/tjzs/tjbk/nszb/201402/P020140226558226322808.pdf>。

子类的日指数、周指数和月指数的无滞后更新。本文充分证明了利用在线大数据构建实时高频通货膨胀指标的科学性和可行性，弥补了国内物价指数实时化和高频化方面的研究空白。

第二，高频物价指标 iCPI 的有效性得到了多方面的验证，为通胀研究增加了新的维度。一方面，高频 iCPI 不但能反映传统 CPI 与现有主要宏观经济变量的关系，还能捕捉一些传统 CPI 无法反映的信息，这既包括月度 iCPI 能及时地反映同期消费者信心变化、与月度利率/汇率间有较顺畅的传导关系，又包括周度 iCPI 能灵敏地反映高频宏观经济指标的影响。另一方面，高频 iCPI 可用于实时监测物价变化，iCPI 日指数在精准捕捉典型事件（如“双十一”）影响、iCPI 周指数在捕捉“节假日效应”（如春节效应）等方面均表现突出。

第三，基于在线大数据编制的物价指数 iCPI 也显示出了独特的优越性。一方面，基础数据全部来自于互联网，编制过程经过初次配置后，之后的大部分工作都交由计算机程序自动实现，降低了人力、物力、财力成本，减少人为干预因素；另一方面，基于在线大数据的方法大幅度提高了数据采集和处理的效率、指数发布的频率，8 个大类、46 个中类与 262 个子类的日/周/月指数均可在线实时查询，因此高频 iCPI 有助于更及时地进行分析决策。另外，iCPI 的数据来自互联网全网，在数据选择时有意扩大数据来源，平台总数量超过 100 个，避免过于依赖某些平台而造成系统性误差，以保证价格指数的独立性和代表性。这些结果足以证明高频 iCPI 对价格的代表性以及应用该指数对宏观、微观经济问题研究的有效性。

第四，实时高频在线价格指数 iCPI 有广阔的应用前景。高频的 iCPI 包含更快捷更细致的价格信息，网络调价成本低、相对较为容易，能快速地反映供给和需求的变化，对于及时地分析宏观经济问题、微观经营策略具有十分重要的参考价值。iCPI 可用于实时监测宏观经济形势和物价变化，特别是网络购物节、春节等价格的短期剧烈波动都可以显现出来，通过横向和纵向对比能反映出居民消费结构和消费方式的变化；iCPI 日指数在金融市场预测方面有很大的潜力。

总之，在当今互联网经济快速发展的时代，构建实时高频物价指标的任务变得愈加迫切，本文设计的基于互联网在线数据的居民消费价格指数（iCPI）做出了较大的创新和贡献，在实时高频物价指标研究方面迈出了第一步，是从 0 到 1 的突破。本文为利用在线大数据构建实时高频宏观经济指标提供了一个范式，对其他问题的研究有重要推广价值。今后我们将进一步扩大商品数量、研究核心通货膨胀、拓展不同区域 iCPI、深入挖掘不同类别的在线价格波动规律；加强数据质量监测，提高样本代表性、增强日指数的稳定性和抗扰动性；运用 iCPI 对经济、金融进行实时高频的监测与预测，深入研究高频 iCPI 对股票、大宗商品等金融产品的预测作用。

#### 参 考 文 献

[1] Aparicio D., Bertolotto M. I., 2017, *Forecasting Inflation with Online Prices* [EB/OL], <http://dx.doi.org/10.2139/ssrn.2740600>.

[2] Cavallo A. F., Rigobon R., 2016, *The Billion Prices Project: Using Online Prices for Measurement and Research* [J], *Journal of Economic Perspectives*, 30 (2), 151~178.

[3] Cavallo A. F., 2017, *Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers* [J], *American Economic Review*, 107 (1), 283~303.

[4] Gorodnichenko Y., Sheremirov S., Talavera O., 2018, *Price Setting in Online Markets: Does It Click?* [J], *Journal of European Economic Association*, 16 (3), 1~48.

[5] Hull I., Lof M., Tibblin M., 2017, *Price Information Collected Online and Short-Term Inflation*

*Forecasts* [R], IFC-Bank Indonesia Satellite Seminar on “Big Data” at the ISI Regional Statistics Conference.

[6] 陈立双、祝丹:《中国 CPI 编制方法与国际〈CPI 手册〉及美国之比较分析》[J],《统计研究》2013 年第 11 期。

[7] 国际劳工组织、国际货币基金组织、经济合作与发展组织等:《消费者价格指数手册:理论与实践》[M],国家统计局国际合作司译,中国财政经济出版社,2008。

[8] 国际劳工组织、国际货币基金组织、经济合作与发展组织等:《补充手册:消费价格指数实用编制指南》[M/OL],[http://www.unece.org/fileadmin/DAM/stats/publications/Practical\\_\\_Guide\\_\\_to\\_\\_Producing\\_\\_CPI.pdf](http://www.unece.org/fileadmin/DAM/stats/publications/Practical__Guide__to__Producing__CPI.pdf), 2009。

[9] 国家统计局城市社会经济调查司:《解读 CPI100 问》[M],中国统计出版社,2009。

[10] 刘向耘、高宏:《CPI 编制的国际比较》[R],中国金融论坛工作论文,2016 年第 59 期。

[11] 任栋、王琦:《中国 CPI 权重调整问题探析》[J],《数量经济技术经济研究》2013 年第 12 期。

[12] 徐强:《OECD 国家 CPI 编制的国际比较及借鉴》[J],《统计研究》2013 年第 6 期。

## Design and Application of Novel CPI Based on Online Big Data

Liu Taoxiong<sup>1,2</sup> Tang Ke<sup>1,2</sup> Jiang Tingfeng<sup>3</sup> Zhang Li<sup>4</sup>

(1. Institute of Economics, School of Social Sciences, Tsinghua University;

2. Institute for Innovation and Development, Tsinghua University;

3. School of Banking & Finance, University of International Business and Economics;

4. Institute of World Economics and Politics, Chinese Academy of Social Sciences)

**Research Objectives:** The paper aims to explore the design and application of real-time high-frequency price indicators with online big data in the era of digital economy. **Research Methods:** We design the first set of Internet-based Consumer Price Indices (iCPI) in China, and analyze the index quality and its applications from multiple aspects. **Research Findings:** Firstly, iCPI can realize real-time updating of daily, weekly, and monthly indices of various levels of products and service (see official website [www.bdecon.com](http://www.bdecon.com)). Secondly, iCPI is automatically generated in the computation procedure from the data collection, cleaning, processing and final online publishing, which effectively saves time and avoids human intervention. Thirdly, high-frequency iCPI performs well in representing general price changes, accurately capturing the effects of special events, nowcasting the inflation and real-time monitoring the macroeconomic situation. **Research Innovations:** We firstly employ online big data to design CPI in China, which makes up for the research gap of real-time high-frequency price indicators. **Research Value:** Online iCPI is the beneficial complement of official CPI, and the proposed realization roadmap can be applied to construct different high-frequency macroeconomic indicators in the future.

**Key Words:** Online Big Data; iCPI; Real-time High-Frequency Indicators; Macroeconomic Nowcasting

**JEL Classification:** E31; C82

(责任编辑:焦云霞)