

网络搜索数据与 CPI 的相关性研究^①

张 崇, 吕本富, 彭 赓, 刘 颖

(中国科学院研究生院管理学院, 北京 100090)

摘要: 网络搜索数据蕴含了三亿多市场主体的兴趣与关注, 反映其行为趋势与规律, 为研究宏观经济问题提供了必要的微观数据基础; 本文从商品市场的角度建立概念框架, 以均衡价格理论为基础, 揭示了网络搜索数据与居民消费价格指数(CPI)之间存在一定的相关关系及先行滞后关系; 实证结果表明: 网络搜索数据与 CPI 之间存在协整关系, 模型拟合度达到 0.978, 预测绝对误差为 0.48, 宏观形势搜索指数和供求关系搜索指数相对于 CPI 的先行周期分别为五个月和两个月; 同时模型具有很强的时效性, 比国家统计局的数据发布提前一个月左右; 与传统的预测方法相比, 模型还具备一定的转折点预测能力.

关键词: 网络搜索数据; CPI 预测; 消费价格指数; 协整分析; 转折点预测

中图分类号: F830.91 **文献标识码:** A **文章编号:** 1007-9807(2012)07-0050-10

0 引 言

CPI(居民消费价格指数)是根据与居民生活密切相关的商品和劳务价格统计出来的物价变动指标,反映了居民生活成本及社会通货膨胀水平,是国家制定宏观经济政策、进行经济调控的重要参考依据,因此能够实时监测和提前预测 CPI 具有重要的意义.

我国的 CPI 由国家统计局负责监测,国家统计局每月对食品、烟酒及用品、衣着、家庭设备用品及服务、医疗保健及个人用品、交通和通信、娱乐教育文化用品及服务、居住八大类 263 个基本分类约 700 个规格品种的商品和服务进行价格调查,并根据居民消费支出的比例对不同类别赋以权重,最终计算得出月度 CPI 数据,并于次月中旬发布,因此国家统计局对 CPI 的监测存在两周左右的滞后. 为了能够提前预知 CPI,学术界提出了多种预测方法,这些方法可以分为定性和定量两

大类,定性预测主要是通过对 CPI 影响因素及先行经济指标的分析来判断 CPI 未来的变化方向,但对变化的幅度难以做出精确的判断. 定量预测则多采用时间序列、神经网络、组合预测、灰色预测等方法^[1-4],这些方法的基本原理都是用历史拟合未来,用当前的趋势外推未来的情况,但历史数据难以反映未来的不确定性,无法把握未来的突变情况,因此基于这种原理的预测对结构性变化的预测准确度较低.

CPI 反映的是一组特定商品及服务的价格水平,其本质是价格. 根据均衡价格理论,商品的价格由供求关系决定. 而供求关系不是孤立的,会受多种宏观经济因素影响. 供给与需求由市场主体生产者和消费者产生,无论是生产者的生产决策还是消费者的购买决策,都需要获取大量的信息,如商品及原材料的价格、质量、性能,商家的规模、信誉,宏观政策及经济形势等. 随着互联网的发展,人们获取信息的方式正在由传统渠道向网

① 收稿日期: 2010-05-25; 修订日期: 2012-05-05.

基金项目: 国家自然科学基金资助项目(70972104; 71172199); 中国科学院研究生院院长基金资助项目(Y15101QY00); 博士后科学基金资助项目(2011M500422).

作者简介: 张 崇(1980—),女,河南人,博士生. Email: zczoln@hotmail.com

络渠道转移,网络已超越传统媒体成为三亿多网民获取信息的最重要途径,在众多网络服务中,网民使用率为82.1%^②的搜索引擎成为最重要的信息入口。生产者及消费者通过互联网获取信息的同时,互联网也记载了他们的查询和浏览记录,这些数据能够反映生产者及消费者的关注和兴趣,折射他们在现实中生产及购买行为的变化趋势。因此生产者和消费者的行为在商品市场及互联网中都会有所表现,在商品市场中的表现为交易量及交易价格的变化,在互联网上的表现为网络搜索、浏览等指标的变化,这两种表现是同一事物的反映,因此二者存在一定相关关系。而且由于市场主体的行为对价格产生影响有一定的传导过程,反映在价格上的变化是滞后的,而反映在互联网上的搜索数据变化是即时的,因此价格和网络搜索数据的变化还存在先行滞后关系,网络搜索数据的变化会领先于价格的变化,对CPI的变化具有一定的预测能力。

利用网络搜索数据对社会经济活动的监测及预测研究在国外已经取得了少量的研究成果,这些研究成果大多出现在2008年以后,还尚未形成系统的研究体系,基于Google搜索数据的流感预测^[5-6]、失业率监测^[7]、汽车及房地产等行业销售量预测^[8-9]都取得了较高的准确度,而且网络数据的即时性能够很好的弥补传统的监测方法的滞后,具有更强的时效性。这些研究在实证上取得了一定突破,但是还存在一些不足之处:首先,在内在机理方面挖掘的不够充分,也未形成系统的理论框架;其次,如何从海量搜索数据中提取有预测价值的相关关键词,这一数据预处理方法还有待完善。而且利用网络搜索数据对综合性指数如CPI的预测研究在国内外还尚属空白,因此本文以CPI和网络搜索数据为研究对象,首先在理论上探讨二者的相关关系,揭示其内在机理,建立概念框架;然后对二者之间的关系进行实证分析和检验;在此基础上,又进一步探讨利用搜索数据对CPI进行实时监测及预测的方法;最后对全文进

行总结,并提出未来可扩展的研究方向。

1 理论分析与概念框架

1.1 理论分析:价格影响因素及传导时滞

均衡价格理论认为商品的价格由供求关系决定。而供求关系又受到宏观经济因素的影响,在众多影响价格的宏观经济因素中,货币是研究最多也是最关键的因素,货币政策通过准备金率、利率等工具调控资产的相对价格,改变市场参与者的行为,将政策效果传导至实体经济中,对企业、居民的投融资行为、生产消费行为都产生影响,进而引导商品市场供求变化及其价格体系的变动^[10-13]在经济全球化的背景下,国际汇率波动和国际资本的流动也会影响到本国的货币流通、产出、价格等方面,汇率机制也成为一国货币政策传导的重要途径,一般认为,一国汇率贬值将提高以本币表示的进口替代品和出口品价格,从而带动一般物价水平上升^[14]除货币政策外,财政政策也会影响物价水平,政府税收的增加可以导致价格水平的降低,过高的财政赤字则可能会引起物价水平的上升^[15]除此以外,经济增长、失业率等也是影响价格变化的因素。

由于市场主体对于宏观政策的变化有一个心理预期过程,在政策变动时,首先要分析其经济前景,资金成本、预期收益等情况,最后才决定调整投资行为和购买行为,因此宏观政策的变化,都并不立即伴随商品价格的相应变化,在情势调整到一个新局面出现之前,总有一个间歇期,即任何宏观政策都存在一定的时滞性^[13]货币学派代表人物弗里德曼经过大量的实证研究后认为,从货币供给量的变动到产量的变动,大约需要6-9个月的时间,从产量的变动到物价水平的变动,也大约需要6-9个月的时间;我国学者也对货币政策及货币工具的传导时滞做出很多研究,由于研究方法的差异,得出的结论也不尽相同,总的来说货币政策对产出的传导时滞为4个月到4年,对价格

② 数据来自艾瑞咨询《2009年中国搜索引擎年度数据报告》, <http://news.iresearch.cn/Zt/107468.shtml>.

的传导时滞则从 4 个月到 1 年不等^[16-17]。

根据粘性价格理论,供求关系变化对价格的影响也存在一定时滞。由于商品市场中的信息不对称性,任何一个市场主体都很难掌握其他市场主体的全部信息,因而难以得知某一商品的总需求及总供给量,更无法根据供求关系即时的调整市场行为及商品价格^[18-19];同时由于“菜单成本”的存在,生产者要对调整价格的成本与收益进行比较才能做出价格决策,因此,价格水平不随需求的变动而迅速地改变,存在一定的滞后期。

以上理论分析表明宏观经济及供求关系是影响价格的重要因素,它们对价格的影响都存在一定的时滞。这些影响因素在搜索数据中都可以找到对应的关键词,宏观经济形势和供求关系变化会引起相关关键词搜索量的变化,这种变化是即时的。因此本文认为,搜索数据与价格水平之间存在一定的相关

关系及先行滞后关系,选择相关关键词并建立模型,可以对价格水平的走势作出预测和判断。

1.2 概念框架: 网络搜索数据与 CPI

本节从商品市场的供求关系角度建立概念框架(图 1),分析网络搜索数据与 CPI 之间的关系。概念框架的基本思想是:宏观经济因素影响供求关系,供求关系决定价格;作为产生供给与需求的市场微观主体,生产者及消费者的行为在商品市场和互联网上都有所反映,在商品市场上反映体现为交易量和价格的变化;而在互联网上则体现为网络搜索、浏览等指标的变化;由于市场主体的行为反映在价格上的变化是滞后的,而反映在互联网上的搜索浏览数据变化是即时的,因此二者之间还存在一定的先行—滞后关系;选择互联网上的相关搜索词,利用相关词的搜索量变化可以对商品价格水平的走势做出判断或预测。

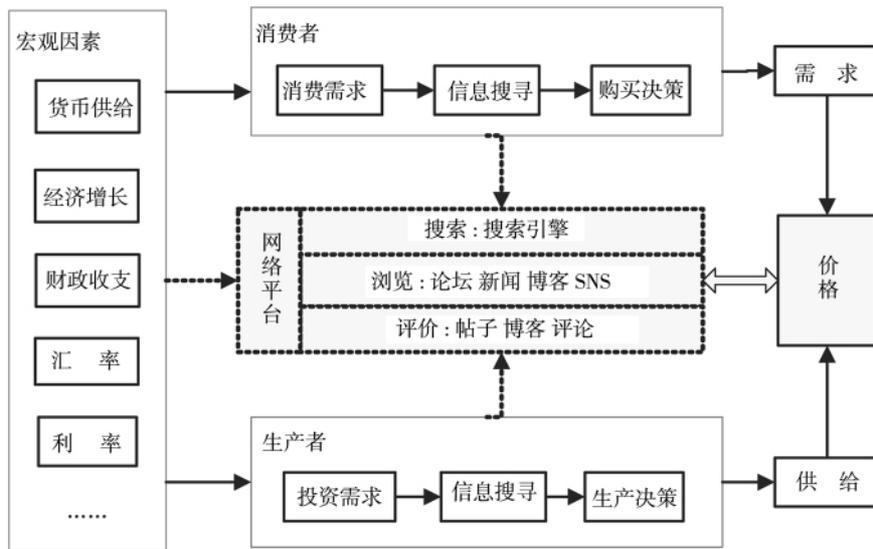


图 1 概念框架: CPI 与网络搜索数据

Fig.1 Concept model: CPI and web search data

根据消费者行为理论,本文将消费者的购买行为划分为消费需求产生、信息搜寻和购买决策三个环节;同样将生产者的行为也划分为投资需求产生、信息搜寻和生产决策三个环节。消费者和生产者的决策过程都需要大量的信息。随着互联网的快速发展,网络成为众多网民获取信息的重要途径,截至 2009 年 6 月 30 日,中国网民规模达到 3.38 亿人,并保持以年 30% 增长率的快速增长;搜索、浏览、论坛是互联网的三大应用,网络

搜索代表了市场主体的关注,网络浏览反映了一种信息的传递,论坛博客则是一种交流和反馈。在众多网络服务中,网民使用率为 82.1% 的搜索引擎成为最重要的信息入口,因此本文将主要探讨网络搜索数据与 CPI 的相关关系。

生产者及消费者在互联网上搜寻的信息可以分为微观和宏观两个层面。微观层面的信息包括特定商品的价格、性能、质量,商家的规模、信誉,原材料的价格等,对这类信息的搜寻产生了如

“笔记本电脑报价”、“数码相机哪个牌子好”、“诺基亚 N95”、“旅游景点”、“水泥价格”、“西本钢铁最新报价”、“大豆价格”等关键词,这些关键词反映了市场主体对某些特定商品及服务的关注与购买意图,其搜索量的变化折射出供需关系的变化。宏观层面的信息包括国家宏观政策,经济走势,市场行情等,对宏观信息的搜寻在互联网上产生了如“外汇牌价”、“存款利率”、“央行加息”、

“经济过热”、“中国物价”等搜索关键词,这些词反映了市场主体对国家宏观政策及经济形势的关注与预期,其搜索量的变化折射出宏观政策及经济形势的变化。图2分别展示了关键词“笔记本电脑报价”与“物价指数”自2005年1月至2009年8月的月搜索量数据与CPI数据的曲线图,可以看出两个词的搜索量变化与CPI走势非常一致。

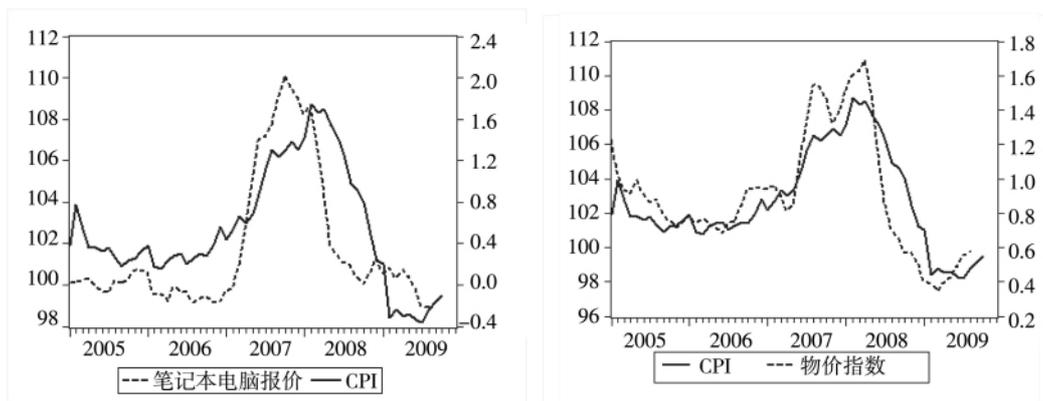


图2 搜索数据与CPI变化曲线比较

Fig.2 Data of CPI and some search keywords

注:搜索量为预处理后的数据,坐标轴横轴为时间,左纵轴为CPI,右纵轴为关键词搜索量

网络搜索中不同关键词代表的含义不同,因此需要采用科学的方法对关键词进行筛选,选择和CPI相关的关键词,并建立适当的模型,才能用于价格波动及CPI变化趋势的判断。

2 实证研究

2.1 搜索词选取与预处理

本文所采用的CPI数据来源于国家统计局网站,采用2004年1月至2009年8月共58个月的CPI同比数据。搜索数据来源于Google搜索解析(<http://www.google.com/insights/search/#>),这是2008年Google发布的一项服务,利用该项服务可以得到一个给定关键词自2004年1月至当前周的周搜索量数据,该搜索量并非是绝对的搜索量,而是一个相对数据,即“在过去的一段时间里,相对于在Google上执行的总搜索量,您所输入字词被搜索了多少次”,因此该数据反映的是在某段时期,某个特定关键词的热门程度和被关注程度。

1) 关键词的选取按照以下三个步骤进行:

- 首先根据CPI的影响因素及构成选取初始关键词:根据CPI的影响因素选取了如“货币政策”、“经济增长”、“汇率”、“加息”等反映宏观经济形势的初始关键词;根据CPI的构成选取了国家统计局在价格调查时采用的衣、食、住、行、医疗保健、休闲娱乐、教育文化等八大类下的部分商品服务名称,如“电脑”、“冰箱”、“汽车”、“猪肉”、“减肥”、“房租”等初始关键词;

- 然后利用Google及百度的关键词推荐功能,逐一输入初始关键词,得到1000多个与初始关键词相关的词,如“人民币汇率”、“今日汇率”、“央行加息”、“降息”、“猪肉价格”、“笔记本电脑报价”等。剔除重复的关键词后,从Google上得到有搜索量的关键词约300个;

- 最后计算每个关键词与CPI的皮尔逊相关系数,该系数表示候选词搜索变化曲线与CPI曲线的相似程度,相关系数越高,二者的相似程度越大,最终选取相关系数大于0.4的关键词共44个。

2) 为了能够更好的拟合CPI,对来自于

Google 的搜索数据做如下预处理:

- 将周数据合并为月数据;
- 使用 3 期移动平均法对搜索数据进行平滑,消除其短期波动,突出其长期变化规律;
- 将平滑后的搜索数据转换为同比数据,与 CPI 数据保持一致.

2.2 搜索指数合成

根据经济含义的不同,本文将 44 个关键词划

分为两大类并合成为两类指数:宏观形势指数($X1$)和供求关系指数($X2$);宏观形势指数反映了消费者及生产者对宏观政策及经济形势的关注与预期,由与货币政策、经济走势相关的关键词组成;供求关系指数则反映消费者对特定商品及服务的关注及购买意图,大多是查询某种商品服务及报价的关键词.表 1 列出了两类搜索指数的部分关键词.

表 1 两类搜索指数与部分关键词

Table 1 Two search indexes and some keywords

	宏观形势指数($X1$)	相关系数	供求关系指数($X2$)	相关系数
1	货币政策	0.621	粮食	0.477
2	中行外汇牌价	0.830	花生油价格	0.485
3	存款利率	0.604	按摩价格	0.552
4	准备金	0.670	旅游景点	0.499
5	加息	0.698	携程网	0.843
6	物价指数	0.845	减肥药	0.595
7	投资	0.727	冰箱价格	0.613
8	中国经济	0.583	液晶电视价格	0.432

2.3 搜索指数与 CPI 的时滞确定

在前文提到,宏观经济因素及供求关系变化对 CPI 的传导都存在一定的时滞,为了确定搜索指数与 CPI 的时滞阶数,本文分别以第 t 期、 $t-n$ 、 $t+n$ 期的自变量与 CPI 建立回归模型,根据模型 R^2 的高低确定两类搜索指数对 CPI 的时滞阶数(见表 2),结果显

示:与第 t 期 CPI 拟合度最高的分别是第 $t-5$ 、 $t-6$ 期的宏观形势指数和第 $t-1$ 期的供求关系指数, R^2 分别为 0.92 和 0.76. 这说明宏观政策变化及供求关系变化对 CPI 的传导时滞存在差异,宏观政策变化对 CPI 的传导时滞为半年左右,而供需关系变化对 CPI 的传导时滞为一个月左右.

表 2 不同时滞阶数搜索指数与 CPI 的拟合度

Table 2 Fitness of CPI and web search data with different lag period

T	$t+2$	$t+1$	t	$t-1$	$t-2$	$t-3$	$t-4$	$t-5$	$t-6$	$t-7$	$t-8$
宏观形势指数($X1$)	0.16	0.28	0.40	0.54	0.66	0.76	0.86	0.92	0.92	0.85	0.75
供求关系指数($X2$)	0.58	0.68	0.72	0.76	0.72	0.66	0.59	0.47	0.35	0.22	0.16

2.4 搜索指数与 CPI 的协整检验

本文选择 CPI(Y) 为因变量,宏观形势指数($X1$)、供求关系指数($X2$) 为自变量,首先对各变量进行平稳性(单位根)检验,采用扩展的 Dickey-Fuller 检验法(简称 ADF 检验),原假设为序列至少有一个单位根,即不平稳.检验结果如表 3 所示,原序列均为非平稳序列,而在一阶差分下,检验结果均在 1% 的显著性水平上拒绝原假设,说明它们是一阶单整序列.

表 3 变量平稳性检验

Table 3 Checking the smoothness of variables

变量	ADF 检验 t -Stat	MacKinnon 临界值			ADF 检验 结果
		1%	5%	10%	
Y	-2.11	-3.56	-2.92	-2.60	非平稳
$X1$	-1.73	-3.56	-2.92	-2.60	非平稳
$X2$	-1.24	-3.56	-2.92	-2.60	非平稳
ΔY	-6.62	-3.56	-2.92	-2.60	平稳
$\Delta X1$	-4.56	-3.56	-2.92	-2.60	平稳
$\Delta X2$	-5.35	-3.56	-2.92	-2.60	平稳

接下来进一步检验因变量与自变量之间是否具有长期稳定的关系,采用 Engle 和 Granger 提出的两步协整 (co-integration) 关系检验法,第一步是建立回归方程,根据搜索指数与 CPI 的时滞,分别以 $X1_{(t-5)}$ 、 $X1_{(t-6)}$ 和 $X2_{(t-1)}$ 、 $X2_{(t-2)}$ 为自变量,

得到 4 个回归方程;第二步对方程的残差进行单位根检验,自变量和因变量之间若存在协整关系,回归残差序列应具有稳定性. 从表 4 的结果看,4 个回归方程的残差都具有稳定性,因此搜索指数与 CPI 之间存在长期稳定的协整关系.

表 4 回归结果及协整检验

Table 4 Regression results and model testing

	模型 1	模型 2	模型 3	模型 4	模型 5	
截距项	67.890 ***	67.254 ***	69.338 ***	69.440 ***	49.784 ***	
LOG $X1_{(t-5)}$	4.420 ***	4.424 ***			2.559 ***	
LOG $X1_{(t-6)}$			4.728 ***	4.793 ***		
LOG $X2_{(t-1)}$	2.945 ***		2.324 **			
LOG $X2_{(t-2)}$		3.076 ***		2.239 ***	1.102 **	
$Y_{(t-1)}$					0.464 ***	
$Y_{(t-12)}$					-0.116 ***	
修正 R^2	0.938	0.946	0.924	0.925	0.978	
F 值	390.25	471.28	304.21	323.98	491.86	
DW	0.888	1.027	0.787	0.862	2.47	
残差平稳性	ADF 值	-4.70	-5.30	-4.18	-4.49	-8.22
	1% 临界	-3.57	-3.56	-3.57	-3.57	-3.58
	5% 临界	-2.92	-2.92	-2.92	-2.92	-2.93
	10% 临界	-2.60	-2.60	-2.60	-2.60	-2.60
	结论	平稳	平稳	平稳	平稳	平稳
结论	协整	协整	协整	协整	协整	

如表 4 所示,前 4 个模型中,以 LOG $X1_{(t-5)}$ 和 LOG $X2_{(t-2)}$ 为自变量的模型 2 拟合度最高,调整 R^2 达到 0.946;该模型 DW 值为 1.027,说明存在序列相关问题,为了消除这一问题,在模型 5 中首先加入了因变量的滞后项 $Y_{(t-1)}$;结果表明该变量是显著的;由于本文采用的是 CPI 当月同比数据,即“第 t 期的 CPI = 第 t 期的物价水平/第 $t-12$ 期的物价水平”,因此当月 CPI 与去年同月即 $t-12$ 期的 CPI 也具有相关性,因此把 $Y_{(t-12)}$ 也加入模型,结果表明该变量是显著的. 增加变量 $Y_{(t-1)}$ 和 $Y_{(t-12)}$ 后,模型 5 的拟合度达到 0.978,且所有变量都在 5% 的显著性水平下通过检验,拟合效果见图 3.

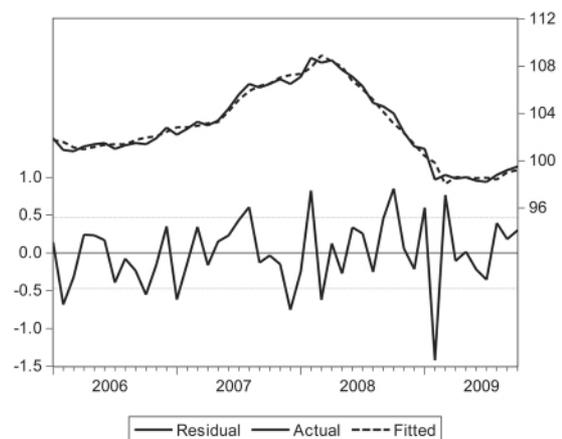


图 3 搜索数据与 CPI 拟合效果图(模型 5)

Fig. 3 Fitting of web search data and CPI(model 5)

2.5 模型解释

模型 5 的回归方程如下,表明 CPI 的变化由宏观经济形势变化,供求关系变化及历史价格水平共同决定.

$$Y_t = 49.784 + 2.559 \text{ Log } X1_{(t-5)} + 1.102 \text{ Log } X2_{(t-2)} + 0.464 Y_{(t-1)} - 0.116 Y_{(t-12)}$$

搜索项:自变量 $\text{Log } X1_{(t-5)}$ 和 $\text{Log } X2_{(t-2)}$ 为搜索项,反映市场主体对宏观经济形势关注度的变化及供求关系的变化. 两个变量的系数均为正,表明两类搜索指数的关注度提高会引起 CPI 的上升.

两个变量滞后阶数的不同,表明了宏观形势及供求关系的变化对价格的传导时滞不同,分别为 5 个月和 2 个月,这一结论与本文第二部分中提到的文献实证结论相吻合,同时也验证了概念框架中“宏观经济政策影响供求关系,供求关系决定价格”这一观点.

两个变量的系数不同,表明了宏观形势及供求关系变化对价格的影响程度不同,两类指数各自变化 1%,对应的价格变化幅度分别为 2.559 个百分点和 1.102 个百分点,这说明宏观政策在“影响企业、居民生产消费行为,进而引导商品市场供求变化及其价格体系的变动”这一过程中,其效力在逐渐减弱,我国宏观政策传导过程存在阻滞.

历史项:模型中的 $Y_{(t-1)}$ 和 $Y_{(t-12)}$ 为历史项,第 t 期的 CPI 与 $t-1$ 期 CPI 显著相关,说明了由于价格粘性的存在,使得价格水平会偏离目标值,通胀具有一定的惯性. 由于本文采用的是 CPI 当月同比数据,即“第 t 期的 CPI = 第 t 期的物价水平 / 第 $t-12$ 期的物价水平”,因此当月 CPI 与去年同期即 $t-12$ 期的 CPI 也显著相关,且系数为负.

2.6 模型预测

为了进一步考察网络搜索数据对 CPI 的预测能力,本文将 2004 年 1 月至 2008 年 12 月的数据作为建模数据,对 2009 年 1 月至 12 月的 CPI 进行预测,结果显示预测的平均绝对误差为 0.48,均方根误差为 0.62,从图 4 中可以看出,预测值与真实非常接近,模型具有良好的预测能力.

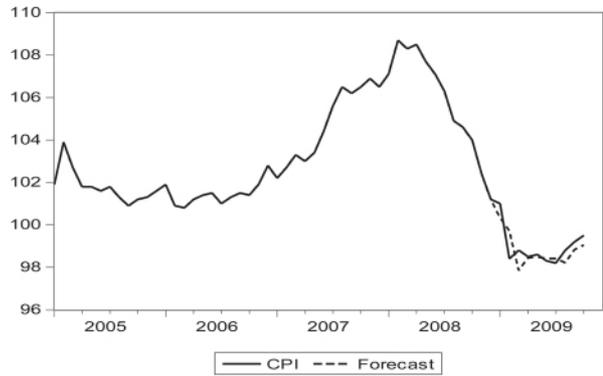


图 4 预测效果图(模型 5)

Fig. 4 Forecasting result(model 5)

由于模型自变量中的搜索项分别是第 $t-5$ 和 $t-2$ 期,历史项为 $t-1$ 和 $t-12$ 期,所以在得知 $Y_{(t-1)}$ 即 $t-1$ 个月的 CPI 数据时便可以预测出第 t 个月的 CPI,能够比国家统计局公布的数据提前 1 个月左右. 因此这一方法克服了传统监测方法的滞后性,具有很强的时效性. 基于这一特点,本文认为搜索指数作为反映价格变化的即时指标,为传统监测方法提供了有效补充,可以用来对 CPI 进行间接的监测.

3 扩展性应用

3.1 转折点预测

在 CPI 及其它经济指标的预测中,众多经济学家及相关部门最关心的问题往往并非准确数值,而是如何判断预测经济指标的结构性变化,即转折点预测. 传统的预测方法对转折点都难以做出准确判断,因为这些方法的基本思想都是用历史去拟合未来,用历史的趋势去外推未来,而历史数据很难把握未来的不确定性及突发变化,网络搜索数据能够即时的捕捉市场主体的关注与兴趣,映射其行为意图和趋势,能够即时甚至提前反映经济现象的变化趋势,为转折点预测提供了一种新的方法和思路.

将宏观形势指数和供求关系指数合并为总搜索指数,图 5 展示了 CPI 及总搜索指数的变化曲线,可以看出自 2005 年 1 月至 2009 年 8 月,CPI 及搜索数据都经历了三次大的转折:2006 年初由平

稳到急剧上升的第一次转折; 2007 年末 2008 年初由急剧上升到急剧下降的第二次转折; 2009 年初由急剧下降到平稳的第三次转折. 从图中也可以明显看到, 搜索指数的三次转折点都提前于 CPI 的转折点, 针对三次峰谷值, 提前的时间平均约为 4 个月, 因此本文认为搜索数据对 CPI 转折点具有一定的预测能力, 可以通过搜索指数的转折点预测 CPI 的转折点.

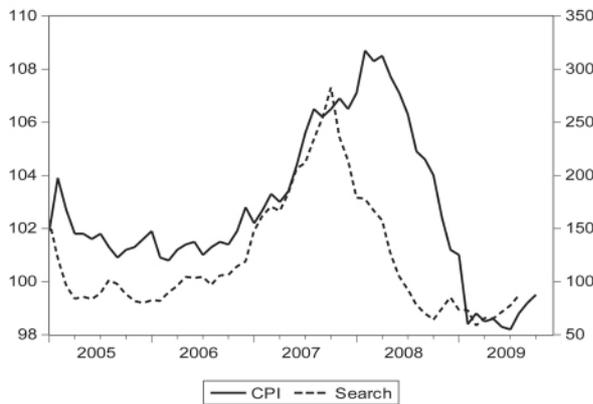


图5 CPI及搜索指数变化曲线图(左纵坐标为CPI, 右纵坐标为搜索指数)

Fig. 5 Comparing web search data and the total search index

3.2 对通胀程度的预警监测

CPI 是国际上衡量通货膨胀最常用的指标, 通胀程度根据 CPI 的高低通常被划分为几个级别: CPI 小于 1% 称为通货紧缩, 1%—3% 之间称为温和通胀, 3%—5% 之间称为轻度通胀, 大于 5% 称为严重通胀. 由于网络搜索数据和 CPI 之间存在着先行一滞后的相关关系, 可以将搜索指数作为通胀程度的领先监测指标, 通过搜索指数的变化对通胀的程度进行预警监测, 根据总搜索指数与 CPI 的回归方程, 可以计算得到通胀程度临界值 1%、3%、5% 所对应的总搜索指数临界值分别为 89.2、129.2、169.2, 当搜索指数的变化接近或跨越临界值时, 说明 CPI 也即将接近或跨越临界值, 通胀的程度即将发生变化. 图 6 展示了不同通胀程度所对应的搜索指数区间, 可以根据搜索指数的大小提前预知通胀程度的高低, 并采取相应的政策进行宏观调控.

1) 当搜索指数小于 89.2 时, CPI 低于或即将低于 1%, 处于通货紧缩期.

2) 当搜索指数大于 89.2 小于 129.2 时, CPI 将大于 1% 小于 3%, 处于温和通胀期.

3) 当搜索指数大于 129.2 小于 169.2 时, CPI 将大于 3% 小于 5%, 处于轻度通胀期.

4) 当搜索指数大于 169.2 时, CPI 将大于 5%, 处于严重通胀期.

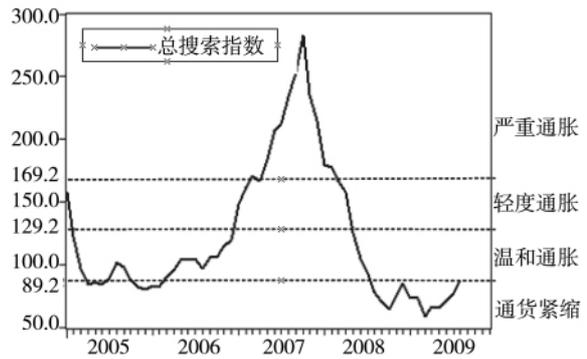


图6 搜索指数监测通胀程度

Fig. 6 Monitor the inflation level by search index

4 结束语

本文首先从理论上分析网络搜索数据与 CPI 之间的相关关系, 揭示了基于网络搜索数据监测或预测 CPI 的内在机理, 进而通过实证研究对其相关关系进行检验, 并建立模型对 CPI 进行预测, 主要结论及创新点如下:

1) 理论框架及内在机理的探讨. 基于网络搜索的社会经济行为预测是一个新的思想及尝试, 相关文献大多出现于 2008 年以后, 这些文献多注重于实证检验而对其内在机理挖掘不够充分, 也未形成系统的理论框架. 本文对内在机理给出较为全面的解释: 消费者与生产者的行为在互联网和实际商品市场中均有所表现, 在互联网上的表现为搜索及浏览指标的变化, 在实际商品市场中的表现为商品交易价格的变化, 两个表现是同一事物的反映, 因此二者存在一定相关性; 并且由于反映时滞的差异两种表现存在先行一滞后关系, 因此采用科学的方法筛选相关关键词, 并利用这些关键词的搜索数据可以实现对价格走势的判断和预测. 本文提出的理论框架对其它在互联网上具有普遍意义的社会经济行为也具有普

适性.

2) 实证结果表明: 网络搜索数据与 CPI 具有长期稳定的协整关系. 基于网络搜索数据的 CPI 预测模型拟合度达到 0.978, 预测绝对误差为 0.48, 搜索指数的提高会引起 CPI 的上升, 其中宏观形势指数和供求关系指数各变化 1%, 对应的 CPI 变化幅度分别为 2.559 个百分点和 1.102 个百分点; 两类搜索指数相对于 CPI 的提前周期分别约为 5 个月和 2 个月.

3) 与传统的监测方法相比, 基于网络搜索数据的 CPI 监测具有更强的时效性. 这一方法弥补了传统监测方法长达两周的信息发布滞后, 比国家统计局的数据发布提前一个月左右, 实现真正意义上的 CPI 实时监测, 因此网络搜索数据可以作为 CPI 的间接监测指标, 为传统监测方法提供有效的补充.

4) 与传统的预测方法相比, 基于网络搜索数据的 CPI 预测具有一定的转折点预测能力. 这一预测方法将影响价格变化的即时因素带入了模型, 克服了传统时间序列、神经网络等方法仅用历史数据预测未来的局限性; 搜索数据的转折点均提前于 CPI 的转折点, 平均提前周期为 4 个月, 因此对 CPI 的转折点具有一定的预测能力.

对于这种新的预测方法, 本文认为难点和重点在于关键词的选择及模型的完善. 相关关键词选择的是否准确与全面, 将直接关系到预测的效果, 如何建立一套科学的选词方法, 建立词表的“进入”和“退出”机制对词表进行动态维护, 使其能够快速准确的反映社会经济最新现象, 成为下一步的研究重点之一. 同时这一方法内在机理的深入挖掘、拐点判断及该预测方法在其它社会经济领域的适用性, 也是本文关注的重点.

参 考 文 献:

- [1] 王 宇, 李旭东, 等. 基于 BP 神经网络的我国 CPI 预测与对策 [J]. 计算机科学, 2009, (10): 256-264.
Wang Yu, Li Xudong, et al. Prediction and countermeasure of Chinese CPI based on BP neural network [J]. Computer Science, 2009, (10): 256-264. (in Chinese)
- [2] 贺 勇, 诸克军, 黄毓芝. 基于 GEP 的经济时间序列组合预测方法研究 [J]. 科技管理研究, 2008, (5): 180-182.
He Yong, Zhu Kejun, Huang Yuzhi. Economic time series forecasting method research based on GEP [J]. Science and Technology Management Research, 2008(5): 180-182. (in Chinese)
- [3] 董景荣, 杨秀苔. 经济时间序列的非线性组合建模与预测方法研究 [J]. 中国管理科学, 2000, (3): 27-33.
Dong Jingrong, Yang Xiutai. Research on nonlinear combination modeling and forecasting method of economic time series [J]. Chinese Journal of Management Science, 2000, (3): 27-33. (in Chinese)
- [4] 张鸣芳, 项燕霞, 齐东军. 居民消费价格指数季节调整实证研究 [J]. 财经研究, 2004, (3): 133-144.
Zhang Mingfang, Xiang Yanxia, Qi Dongjun. Empirical study on the seasonal adjustment of consumer price index [J]. Journal of Finance and Economics, 2004, (3): 133-144. (in Chinese)
- [5] Ginsberg J, Mohebbi M H, Patel R S, et al. Detecting influenza epidemics using search engine Query data [J]. Nature, 2009, 457, 1012-1014.
- [6] Doornik J A. Improving the Timeliness Of Data On Influenza-Like Illnesses Using Google Search Data [R]. Technical Report, University of Oxford, 2009.
- [7] Askitas N, Zimmermann K F. Google econometrics and unemployment forecasting [J]. Applied Economics Quarterly, 2009, (2): 107-120.
- [8] Wu Lynn, Erik Brynjolfsson. The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales [C]. Proceedings of the 2009 International Conference on Information Systems, 147.
- [9] Cho H i, Varian H. Predicting the Present with Google Trends [C]. Technical Report, 2009, Google Inc.
- [10] Samuelson P A, Solow R M. The problem of achieving and maintaining a stable price level: Analytical aspects of anti-infla-

- tion policy [J]. *American Economic Review*, 1960, (2): 177–194.
- [11] Friedman. *Inflation: Causes And Consequences* [M]. New York: Asia Publishing House, 1963.
- [12] 艾慧. 中国当代通货膨胀理论研究 [D]. 上海: 复旦大学, 2004.
- Ai Hui. *Theoretical research on contemporary Chinese Inflation* [D]. Shanghai: FuDan University, 2004. (in Chinese)
- [13] 张成思. 中国通胀惯性特征与货币政策启示 [J]. *经济研究*, 2008, (2): 33–43.
- Zhang Chengsi. *The nature of inflation inertia in China and its implications on monetary policy* [J]. *Economic Research Journal*, 2008, (2): 33–43. (in Chinese)
- [14] 张帅, 赵昕, 王茂林. 人民币汇率波动与通货膨胀率之间关系的实证分析 [J]. *上海金融*, 2008, (4): 46–50.
- Zhang Shuai, Zhao Xin, Wang Maolin. *Empirical analysis of the relationship between the RMB exchange rate fluctuations and inflation rate* [J]. *Shanghai Finance*, 2008, (4): 46–50. (in Chinese)
- [15] 龚六堂, 邹恒甫. 财政政策与价格水平的决定 [J]. *经济研究*, 2002, (2): 10–16.
- Gong Liutang, Zou Hengfu. *Financial policy and price level decision* [J]. *Economic Research Journal*, 2002, (2): 10–16. (in Chinese)
- [16] 麦元勋, 田晔. 我国货币政策效应时滞的实证研究 [J]. *商业研究*, 2004, (5): 90–94.
- Mai Yuanxun, Tian Ye. *An empirical research on time lag of monetary policy in China* [J]. *Commercial Research*, 2004, (5): 90–94. (in Chinese)
- [17] 谭旭东. 中国货币政策的有效性——基于政策时间不一致性的分析 [J]. *经济研究*, 2008, (9): 46–57.
- Tan Xudong. *The effectiveness of China's monetary policy: Analyzing from the viewpoint of time inconsistency* [J]. *Economic Research Journal*, 2008, (9): 46–57. (in Chinese)
- [18] 刘金全, 邵欣炜. 流动性约束与消费行为关系的实证研究 [J]. *管理科学学报*, 2004, (4): 90–94.
- Liu Jinquan, Shao Xinwei. *Empirical analysis of relationship between liquidity constraints and consumer behaviors* [J]. *Journal of Management Sciences in China*, 2004, (4): 90–94. (in Chinese)
- [19] 邹小芄, 黄峰, 杨朝军. 流动性风险、投资者流动性需求与资产定价 [J]. *管理科学学报*, 2009, 12(6): 139–149.
- Zou Xiaopeng, Huang Feng, Yang Chaojun. *Liquidity risk, liquidity demand of investors and asset pricing* [J]. *Journal of Management Sciences in China*, 2009, 12(6): 139–149. (in Chinese)

A study on correlation between web search data and CPI

ZHANG Chong, LV Ben-fu, PENG Geng, LIU Ying

GS School of Management, Chinese Academy of Sciences, Beijing 100090, China

Abstract: The web search data, which recorded hundreds of millions of searchers' concerns and interests, reflected the trends of their behaviors and provided an essential data basis for the study of macro-economic issues. This paper established a concept framework based on the commodity market and equilibrium price theory, and revealed there is a certain correlation and lead-lag relationship between web search data and consumer price index (CPI). Empirical results indicated there is a co-integration relationship between web search data and CPI. The model was able to obtain a good fit with CPI. The model fitting is 0.978 and the absolute forecast error is 0.48. Simultaneously, the model has very strong time effectiveness compared with traditional CPI monitor method which had a 2-week lag, the model's forecast result can be obtained one month ahead of the

(下转第70页)

[15] 郎茂祥, 胡思继. 车辆路径问题的禁忌搜索算法研究 [J]. 管理工程学报, 2004, 18(1): 81 - 84.

Lang Maoxiang, Hu Siji. Study on the tabu search algorithm for vehicle routing problem [J]. Journal of Industrial Engineering and Engineering Management, 2004, 18(1): 81 - 84. (in Chinese)

Fuzzy dynamic location-routing problem in post-earthquake delivery of relief materials

DAI Ying¹, MA Zu-jun¹, ZHU Dao-li², FANG Tao¹

- 1. Institute for Logistics and Emergency Management, School of Economics and Management, Southwest Jiaotong University, Chengdu 610031, China;
- 2. Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200052, China

Abstract: Optimizing the delivery system for post-earthquake relief materials is an important way to enhance its operational performance. From the view point of integrated optimization, the joint decision problem of locating distribution centers and scheduling routes of delivery vehicles was studied. A chance-constrained programming model for the fuzzy dynamic location-routing problem was developed by considering the following characteristics: fuzzy and dynamic demand of relief materials, time window constraints, dynamic rehabilitation of affected road networks, heterogeneous capacitated delivery vehicles, and split deliveries. The goal is to minimize the total time in delivering relief materials to all affected areas. A two-phase heuristic algorithm was proposed to solve the model. Finally, the feasibility and validity of the model and algorithm was demonstrated by a numerical example.

Key words: earthquake disasters; relief materials; location-routing problem; fuzzy demand; dynamic decision; heuristic algorithm

~~~~~  
(上接第 59 页)

State Statistical Bureau's report. Compared with the traditional forecasting methods, the model also has the predictive power for a certain turning point.

**Key words:** web search data; CPI forecasting; consumer price index; co-integration analysis; turning point forecasting

附录: 关键词列表

|             |             |          |            |            |
|-------------|-------------|----------|------------|------------|
| 1. 通胀       | 11. CPI     | 21. 央行   | 31. 粮食     | 41. 旅游景点   |
| 2. 物价       | 12. CPI 是什么 | 22. 中国投资 | 32. 中国粮食   | 42. 携程网    |
| 3. 货币政策     | 13. CPI 指数  | 23. 央行加息 | 33. 猪肉     | 43. 瘦身减肥   |
| 4. 美元汇率     | 14. 物价上涨    | 24. 股票   | 34. 花生油价格  | 44. 中国物价上涨 |
| 5. 物价指数     | 15. 准备金     | 25. 基金   | 35. 花生价格   |            |
| 6. 中国银行外汇牌价 | 16. 央行准备金   | 26. 证券投资 | 36. 液晶电视价格 |            |
| 7. 中行外汇牌价   | 17. 存款      | 27. 证券股票 | 37. 冰箱价格   |            |
| 8. 外汇牌价表    | 18. 上证指数    | 28. 中国股市 | 38. 按摩床    |            |
| 9. 通货膨胀     | 19. 加息      | 29. 货币流动 | 39. 按摩价格   |            |
| 10. 存款利率    | 20. 股票行情    | 30. 利息   | 40. 按摩     |            |