

网络搜索数据与消费者信心指数的相关性研究*

刘伟江 樊国虎 李映桥

(吉林大学商学院, 吉林长春, 130012)

摘要: 随着互联网科技的发展, 人们获取信息的方式正在由传统渠道向信息化渠道发展。消费者通过网络搜索引擎获取自身需要信息的同时, 搜索引擎也会将其搜索痕迹记录下来。而消费者在互联网上留下的这些痕迹, 正是消费者购买意图、及消费信心的真实反映。消费者信心指数, 是现代社会经济生活当中的一个重要经济指标, 可以对主要的经济变量的未来走势进行监测。本文利用谷歌趋势获取谷歌网络搜索数据, 采用时差相关分析法确定与消费者信心有关的网络搜索关键词的相关系数和领先阶数, 并合成了消费者预期搜索指数和消费者满意搜索指数, 通过实证结果表明: 网络搜索数据与消费者信心指数之间具有显著相关性, 模型拟合优度 0.954, 预测平均绝对误差 0.49。

关键词: 网络搜索数据 消费者信心指数 相关性

A Research on the correlation between web search query and consumer confidence index

Abstract: With the development of Internet technology, a growing number of people get information from informatizational way instead of traditional one. Search engine record the traces left by users, which reflect consumers' real purchasing willing and consumer confidence. Consumer Confidence Index(CCI) is an crucial economic indicator in modern social economy which can be used to monitor the trends of major economic variables. Owing to the web search data from Google Trends, this paper uses time difference relevance analysis to determine the correlation coefficients and leading periods between CCI and web search keywords, and combines consumer expectations search index and consumer satisfactions search index. The results show that there is a significant correlation between web search query and CCI with model fitting degree of 0.954 and the forecasting mean absolute forecast error of 0.49.

Key words: Web search query CCI Correlation

引 言

搜索引擎、门户网站、社交网络等网络应用的问世, 使得互联网数据迅速膨胀。根据 2014 年 1 月中国互联网络信息中心 (CNNIC) 所公布的《第 33 次中国互联网络发展状况统计报告》, 截至 2013 年 12 月底, 我国网民规模达 6.18 亿, 互联网普及率为 45.8%。互联网

*[基金项目]: 吉林省科技发展规划项目(项目编号: 20120616)

数据中心最新的数字宇宙甚至推测，到人们步入 2020 年的时候，全球的数据存储总量将达到 35ZB。我们正逐步步入海量信息时代，日益增长的海量数据，已经大大超越了人们的获取能力和理解水平。倘若不借助强有力的先进的科学方法和科学工具，那么我们就很难弄清所蕴含海量数据当中极具价值的信息。

经济学家和众多的研究者们一直把对未来发展趋势的预测作为研究的重点内容。传统预测方法的数据，主要来自于政府及相关统计部门的统计报告，除了数据收集的真实性和准确性存在一定质疑外，这些数据的公布往往存在滞后，这极大限制了预测的有效性。因此，如何能够找到真实性和准确性都比较令人满意的高质量的数据资源，成为能否改善预测结果的一大关键因素。目前利用网络搜索数据对相关领域进行研究已成为了许多学者研究创新的热点领域之一。

消费者信心指数已经成为我国经济景气指数体系的有机组成部分，受到国内外的广泛关注。研究消费者信心指数的变化规律，对于我们更好的把握消费者信心的波动及准确测度消费者信心具有重要的现实意义。以便我们更加准确的掌握经济的先行指标，能够更好地为政府部门、企事业单位及个人的经济决策提供服务。本文旨在通过对网络搜索数据与消费者信心指数之间的关系研究，提高对消费者信心指数变化规律的认识和把握，从而为进一步提高我国消费者信心指数的应用发展水平提供了一种新的思路。

1 文献回顾

近年来国内外学者已经把网络搜索数据的研究扩展到健康、社会和经济等多个领域，对很多行为或现象进行相关性的研究和预测。在很多领域已经取得了比较丰硕的科研成果。Polgreen (2008) 利用人们通过搜索引擎对流行性感冒的相关关键词的搜索量来合成搜索指数，发现合成的搜索指数与流感发病率之间存在长期稳定关系，而且可以利用该合成指数预测流感发病情况及流感的死亡率。经研究发现美国劳动部公布的失业率与谷歌趋势上与失业密切相关的互联网搜索关注度存在显著的相关性，并建立了两个互联网关注度变量的对比模型，结果表明网络搜索关键词关注变量的引入，对提高美国失业率预测模型的预测精度有显著的作用。Choi, Varian (2009) 利用部分网络搜索数据对美国汽车及零配件、零售业、房地产和旅行业四个行业的销售量与网络搜索数据之间所存在的相关关系进行了深入的研究，他们引入季节性自回归模型和固定效应模型，并将关键词网络搜索量作为新的影响因素加入到传统的自回归模型中，然后建立了典型行业当期销售量预测模型，通过实证的研究结果发现，把网络搜索数据纳入销售量预测模型后，四个行业的预测精确度都有了显著的提高，其中对汽车及零配件行业的销售量预测改善最大，其预测精度提高了 18%，其次是房地产销量的预测，其预测精度提高了 12%。杨欣,吕本富等 (2013) 以投资者行为视角构建重大突发事件对股价波动理论框架，以“7-23 甬温动车特大交通事故”为研究对象，将网络搜索数据合成市场搜索指数和事件搜索指数，然后构建移动自回归 (ARMA) 模型，通过实证分析，探讨了该事故对中国股市动车概念版的影响强度。实证结果表明网络搜索数据能够及时而又准确的反映该动车特大交通事故对相应板块的股市波动，其模型的拟合优度为 0.888。

2 研究方法

资讯时代里，随着海量数据的堆砌，其在商业上的价值已经成为企业对未来发展的巨大

依托。未来的互联网不再是速度的对决，而是深度的较量。搜索引擎的发展不仅为诸如社会学和经济学等学科的预测研究开辟了新的领域，而且为把握大数据时代经济文化生活的脉络，重塑我们对于世界和生活的认知提供了必要的基础手段，成为我们认识世界，了解世界的窗口。由于信息搜索行为是人们为了满足特定的目标需求，而进行的查找搜索信息的活动，而人们的需求很大程度上又受到外部环境的影响和左右，因此通过关键词的搜索，再利用计算机等数据处理手段，可以发现一些传统研究中所无法得出且非常具有价值的成果及结论。

2.1 理论框架的构建

消费者满意指数和消费者预期指数是构成消费者信心指数的两个重要组成部分。消费者预期指数，主要反映消费者对未来的经济景气程度和自身发展状况的预期。而消费者对未来的预期和评估主要是对宏观经济走势的把握以及对自身成长的预计，无论是宏观经济的走势，还是对个人成长的预计，都主要取决于宏观的经济大环境，例如国家经济景气程度，政府的宏观调控，央行的财政政策和货币政策，就业形势等，这些变动都会对未来经济产生深远的影响，在很大程度上决定了未来的经济走势，因此，对宏观层面的各种要素的考虑，便是通过间接的手段对消费者预期指数的一种有效测度。反映到网络搜索数据上，便是消费者对一些宏观的关键词的搜索。消费者满意指数，是指消费者对现阶段的经济生活的评价。也就是着眼于自己乃至家庭目前生活状态，是一种短期行为。反映的网络生活中，便是消费者对自己日常消费的关注。通过网络搜索数据的宏观和微观两个层面的搜集、分析和处理，便可对消费者预期指数和消费者满意指数进行度量，也就是对整个消费者信心指数进行度量。

消费者的消费信心如何变化，最终会反映到商品市场当中。而商品市场的核心就是供求双方决定价格，而消费者的消费行为很大程度上受到价格的影响。于是本文从商品供求关系的角度入手，提出了网络搜索行为会反映出消费者的信心波动状况，并据此建立了相应的理论框架：供求关系很大程度上受宏观经济因素的影响，而且直接影响价格的波动，企业作为产品或服务的供应主体，消费者作为需求主体，他们的行为不仅在商品交易市场有所表现，而且也会在互联网上有所反映。因此选择互联网上与消费者行为或心理有关的网络搜索关键词，其搜索量的变化反应了消费者关注的热度，在反映商品或服务价格的同时，根据对价格变化的预测可以在一定程度上判断消费者信心的变化，消费主体在整个过程中，没有收到任何的干扰，完全是消费者自发行为的结果。这种方法比起通过发放调查问卷所获得的结果，要更加的真实和准确，也更加的及时。

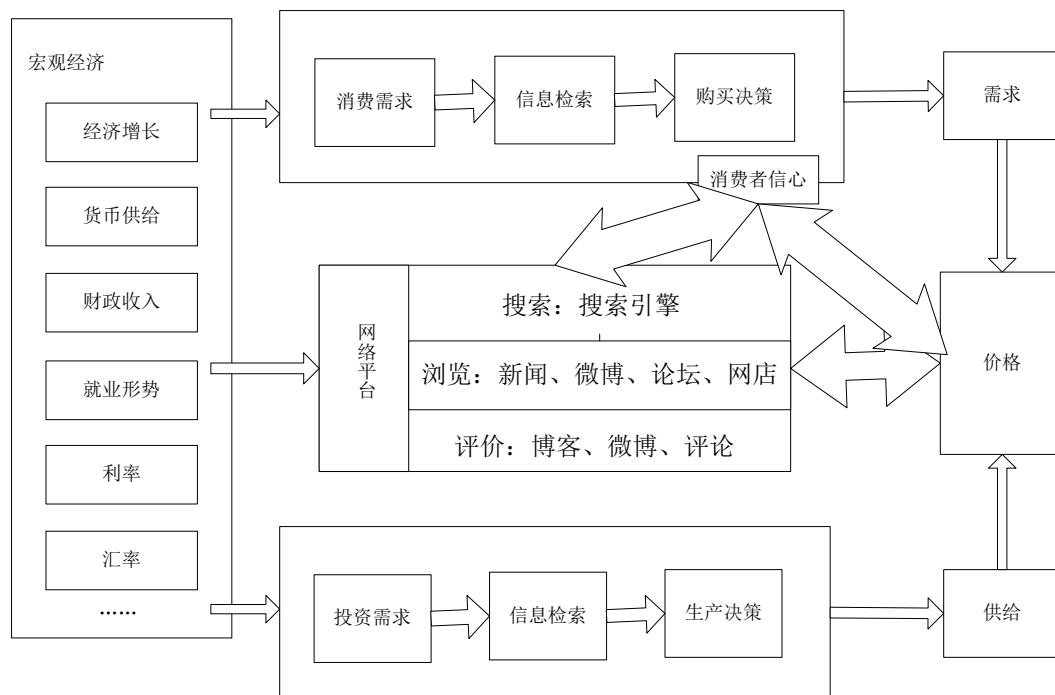


图 2.1 概念框架：消费者信心指数与网络搜索数据

本文将企业行为划分为投资需求产生、网络信息检索及生产决策三个步骤。同样的把消费者的购买行为也大致分成消费需求产生、网络信息检索及购买决策三个步骤。由于企业和消费者在决策过程中，都需要大量的信息。目前网络已经成为消费者获取信息的一个必不可少的途径，而且获取的数据量正日趋增多。企业和消费者尤其是消费者在互联网上搜索的信息主要可以归为宏观和微观两个层面。宏观层面上的信息如经济景气状况、经济运行情况、财政政策、货币政策和就业形势等，在互联网上对宏观信息的搜寻产生了如“存款准备金率”、“就业率”、“物价指数”等关键词，这些词反映了消费者对国家宏观经济景气程度以及我国的宏观经济政策的关注与预期。微观层面上的信息如所需商品或服务的价格、品质、商家信誉和存在问题等，对于此类信息的搜索便收集到了如“伊兰特”、“兰蔻”和“黄金”等这些网络搜索关键词，反映了消费者的购买意图或者是对某些商品或服务的关注，关注热度的变化可以较好的反映供需关系的变化，供需的变化又反映了价格的变化，而价格的变化又会进一步引起消费者信心的变化。

2.2 研究方法及步骤

本文主要对网络搜索数据和消费者信心指数之间的相关性进行了探讨。关键词初选是整个研究的第一步，搜索引擎的推荐、Google correlate 关键词挖掘工具和 SEO 关键字挖掘和 ICTCLAS 汉语分词系统是确定初选网络搜索词的 4 个重要方法，然后基于关键词搜索量的相对值与消费者信心指数的实际变化的相关性强弱，来确定最终用于研究的关键词，还可以在运用时差相关分析法确定相关性大小的过程中，发现其不同关键词的领先或滞后阶数。最后，基于两者间的相关关系，建立预测模型，并对其精度进行分析。

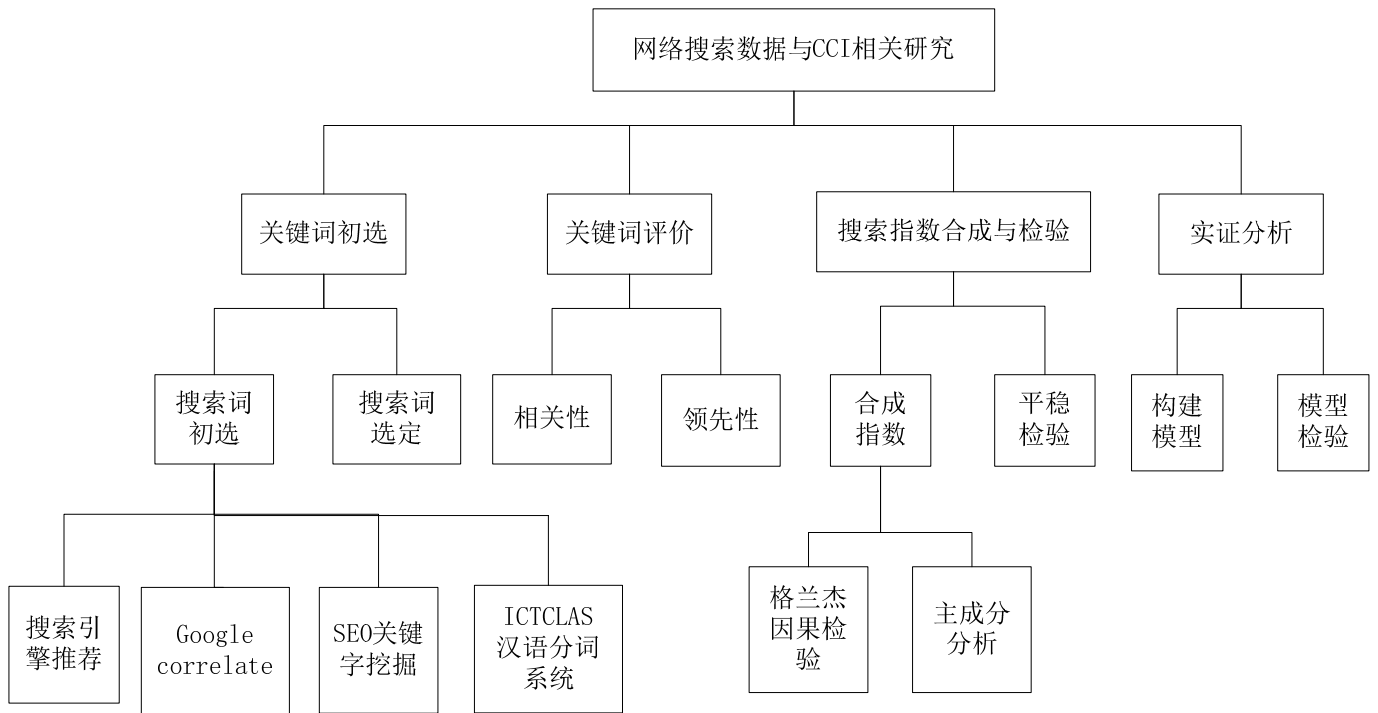


图 2.2 技术路线图

3 实证分析

本文当中使用的网络搜索数据来源于谷歌趋势。消费者信心指数同比数据来自于我国国家统计局官方网站数据库的统计数据。两种数据选取的时段是从 2009 年 11 月到 2013 年 6 月。选取日期是从 2009 年 11 月开始而不是从 2004 年 1 月，主要是因为 2004 年我国无论是网民数量还是互联网的普及率都比较低，这就使得当时的网络搜书数据的代表性不够强，不能很好的反映真实的消费者信心的变化情况，因此选择了从 2009 年 11 月开始至 2013 年 6 月的数据。2005 年至 2013 年我国网民的规模和互联网的普及率见下图。

3.1 关键词初选

本文根据消费者信心指数的编制结构，选定了经济发展形势、家庭收入和就业、物价水平、消费或购买意愿等构成和影响因素，选择了“GDP”、“黄金”和“二手房”等比较原始的网络搜索关键词。除了谷歌相关搜索排行和百度搜索风云榜等比较简单的关键词挖掘方法对与消费者信心比较相关的领域进行了初步的关键词搜集外，还运用了谷歌的数据挖掘工具 Correlate、SEO 关键字挖掘工具及 ICTCLAS（汉语词法分析系统），使得关键词的选取工作更加的有章可循。通过上述的操作，得到 1500 多个与初始关键词相关的词，之后对关键词做了进一步的搜集和整理工作。剔除重复的关键词后，在谷歌趋势搜索字词栏中逐一进行输入，如“公积金买房”、“柴油价格”、“上证指数”、“理财产品”、“居民消费价格指数”等。从谷歌趋势上得到有搜索量并且各个月份数字相对比较完整的关键词约 600 个。

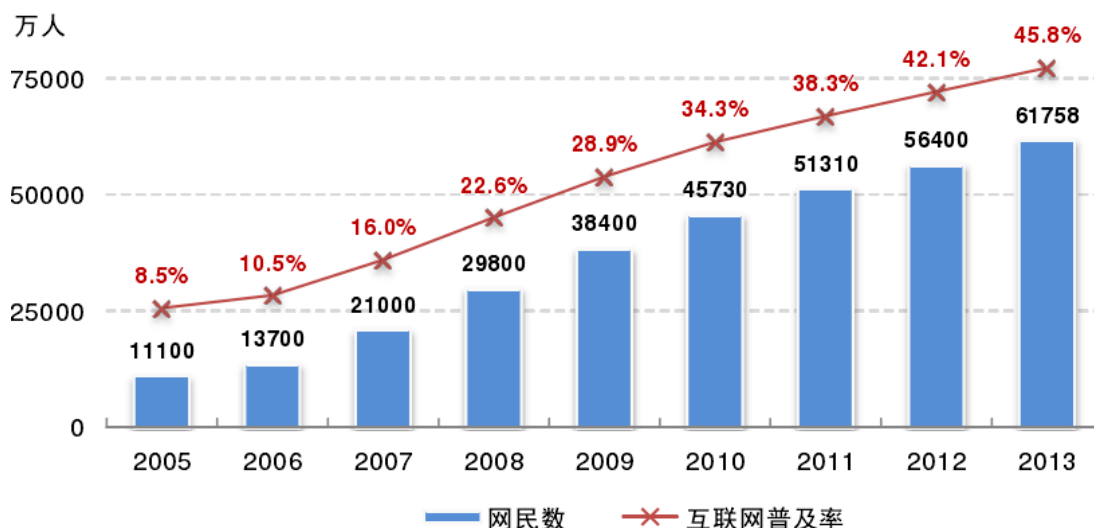


图 3.1 中国网民规模和互联网普及率

3.2 关键词评价

关键词的选取作为网络搜索方面研究的核心环节，必须对选出的关键词在引入最终的回归模型前，进行比较科学合理的评价，也只有这样才能保证整个研究的准确性和客观性。因为不同的网络搜索关键词与消费者信心指数之间所表现出不同的时差关系。一个关键词只能从一个侧面体现消费者信心的一种行为特征，而消费者信心指数的走势绝不会是由一个关键词就可以刻画出来的，必需由若干个关键词共同刻画，因此，需要有必要对每一个初选的关键词的时差进行进一步的判断。本文中采用了时差相关分析法分别计算各个初选网络搜索关键词的领先阶数和相关性。时差相关分析法是用来验证时间序列问题当中，变量与变量间的时差和相关系数大小的一种比较常用的方法。时差相关分析具体表达式见（3.1）

$$r_p = \frac{\sum_{p=1}^n (x_{t-p} - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{p=1}^n (x_{t-p} - \bar{x})^2 \sum_{p=1}^n (y_t - \bar{y})^2}} \quad (3.1)$$

$$p = 0, \pm 1, \pm 2, \dots, \pm n$$

式中， p 为某一网络搜索关键词相对与消费者信心指数的领先阶数， r_p 表示当时差为 p 的两个变量之间的相关系数， x 表示网络搜索关键词搜索热度的月度变化率， \bar{x} 为网络搜索关键词搜索热度的月度变化率均值， y 表示消费者信心指数， \bar{y} 为消费者信心指数的均值，而这个最大值即为关键词与消费者信心指数两者之间的相关系数。本文在关键词评价过程中，符合要求的关键词其 p 值必须大于 0。同时为网络搜索数据与消费者信心指数的相关系数设定一个临界值标准（本文中的临界值为 0.5），通过以上两方面的综合评价，对关键词进行进一步筛选。

3.3 搜索指数合成与检验

按照网络搜索数据与消费者信心指数的理论框架的分析结论，将经过关键词评价步骤，

筛选出来的既符合领先性要求,又符合相关性要求的那些网络搜索关键词根据宏观层面和微观层面也就是从消费者预期和消费者满意两个方面来进行分类。表 3.1 列出了满足与消费者信心时差相关分析领先性和相关系数大于临界值要求的部分关键词,以及其与消费者信心指数之间的皮尔森相关系数。

表 3.1 部分优秀关键词相关系数表

	消费者预期方面	相关系数	消费者满意方面	相关系数
1	国内生产总值	0.663	黄金价格	-0.674
2	收入	0.546	中药	-0.708
3	物价指数	0.630	坐车	-0.701
4	货币供应	0.599	房价	0.647
5	投资	0.533	醋	-0.598
6	利息	0.660	男装品牌	0.565
7	就业情况	0.750	节能汽车	-0.661

为了检验通过关键词评价所剩下的这些关键词是否是引起消费者信心变化的原因,本文采用格兰杰因果检验法对网络搜索关键词逐一进行格兰杰因果检验(置信水平为 0.05),表 3.1 列出了通过格兰杰因果检验的部分关键词。

表 3.2 格兰杰因果检验筛选出的部分关键词

编号	关键词名称	原假设 (H_0)	F 值	P 值
01	中药	01 不是引起 CCI 变化的格兰杰原因	4.718	0.015
02	VERO MODA	02 不是引起 CCI 变化的格兰杰原因	5.112	0.011
03	实木家具	03 不是引起 CCI 变化的格兰杰原因	4.585	0.017
04	银行存款	04 不是引起 CCI 变化的格兰杰原因	3.476	0.042
04	黄金价格	05 不是引起 CCI 变化的格兰杰原因	3.398	0.046
06	营业税	06 不是引起 CCI 变化的格兰杰原因	3.535	0.039
07	公积金	07 不是引起 CCI 变化的格兰杰原因	4.951	0.013

(表 3.2 原假设中的数字编号代表对应行关键词的名称, C C I 表示消费者信心指数)

为了合成最终的消费者预期搜索指数和消费者满意搜索指数,需要对通过格兰杰因果检验剩下的关键词进行进一步的合并,本文采用的是主成分分析法,将宏观层面通过格兰杰因果检验的网络搜索关键词合并成消费者预期搜索指数 (X_1),其领先于消费者信心指数 2 个月,将微观层面通过格兰杰因果检验的网络搜索关键词合并成消费者满意搜索指数 (X_2),其领先于消费者信心指数 1 个月。

3.4 模型构建

本文选择消费者信心指数 (Y) 为被解释变量,消费者预期搜索指数 (X_1) 和消费者满意搜索指数 (X_2) 为解释变量,因为所有的变量都是月度数据,是典型的时间序列数据,所以最终建立的一系列模型也是时间序列的模型,为了确定各个研究变量是否平稳,需要分

别对各个变量进行单位根检验，本文采用的是扩展的 Dickey-Fuller 检验法（ADF 检验法）。各个变量的检验结果如表 3.3 所示。

表 3.3 变量平稳性检验

变量	ADF 检验 t-Stat	MacKinnon 临界值			ADF 检验结果
		1%	5%	10%	
Y	-1.771	-3.634	-2.952	-2.610	非平稳
X1	-1.267	-3.634	-2.952	-2.610	非平稳
X2	-1.669	-3.634	-2.952	-2.610	非平稳
ΔY	-3.326	-3.634	-2.952	-2.610	平稳
$\Delta X1$	-4.751	-3.634	-2.952	-2.610	平稳
$\Delta X2$	-5.627	-3.634	-2.952	-2.610	平稳

通过单位根检验的结果，可以看出，所有变量的原序列都不是平稳的时间序列，因此对各个变量又进行了一阶差分下的单位根检验，检验结果为所有变量在 5% 的置信水平都已平稳。

为了对网络搜索数据与消费者信心指数之间的相关关系进行定量分析，就需要建立两者相互关系的模型，过去的网络搜索的研究，很多指数可以利用自回归模型对指数本身进行预测，通过对消费者信心指数自相关和偏自相关的图的分析，建立消费者信心指数的自回归模型（3.2），并将此作为基础模型，进行进一步的分析。

$$Y_t = c + \beta_1 Y_{t-1} + \varepsilon \quad (3.2)$$

接着把合成的与消费者信心指数有重要影响的消费者预期搜索指数（X₁）和消费者满意搜索指数（X₂）和消费者信心指数（Y）构建了回归方程（3.3）

$$Y_t = c_1 + \beta_1 X1_{t-2} + \beta_2 X2_{t-1} + \varepsilon_1 \quad (3.3)$$

国内外很多学者都把网络搜索数据某个关键词数据搜索指数或者合成的搜索指数引入到相应的模型中，可以提高模型的拟合效果和预测效果，于是本文将消费者预期搜索指数消（X₁）和消费者满意搜索指数（X₂）引入到了（3.3）模型中得到了方程（3.4）

$$Y_t = c_2 + \beta_1 X1_{t-2} + \beta_2 X2_{t-1} + \beta_3 Y_{t-1} + \varepsilon_2 \quad (3.4)$$

为了确定模型被解释变量与解释变量确实具有长期稳定的相互关系，本文用到了 Engle 和 Granger 的两步协整关系检验法来检验其长期趋势，首先要建立消费者信心指数与网络搜索数据的搜索指数之间的回归方程，接着对回归方程的残差再进行单位根检验，如果被解释变量和解释变量之间存在协整关系，那么所构建的网络搜索指数与消费者信心指数的回归残差序列就应该是平稳的。从表 3.4 的结果显示，残差项都是稳定性，因此网络搜索指数与消费者信心指数之间具有长期稳定的协整关系。

表 3.4 回归结果及协整检验

参数	模型 1	模型 2	模型 3
截距项	43.502	113.846	-10.079
β_1	0.574	-5.734	-1.081
β_2	-	1.863	1.062
β_3	-	-	0.077
修正 R ²	0.350	0.625	0.954

F 值		21.58	33.26	289.65
残	ADF 值	-7.153	-5.348	-3.115
差	1%临界	-3.641	-3.634	-3.641
平	5%临界	-2.955	-2.952	-2.955
稳	10%临界	-2.611	-2.610	-2.611
性	结论	协整	协整	协整

模型 3 的回归方程 (3.4) 拟合效果是最优的, 该方程量化了网络搜索数据与消费者信心指数之间的关系, 表明消费者信心指数的变化是由消费者未来生活的预期的变化和消费者对现有生活的满意情况的变化以及历史的消费者信心水平来共同决定。

$$Y_t = -10.079 - 1.081X1_{t-2} + 1.062X2_{t-1} + 0.077Y_{t-1} \quad (3.5)$$

模型当中解释变量 $X1_{t-2}$ 和 $X2_{t-1}$ 代表的是消费者通过网络搜索相关内容对消费者信心指数刻画的贡献, 从宏观和微观两个层面, 以消费者预期和消费者满意为代表, 反映了在网络生活中, 消费者预期和消费者满意网络关注度和热度的变化, 其中 $X1_{t-2}$ 的系数为负值, 表示消费者预期搜索指数关注度的提高会带来消费者信心指数的下降, 而 $X2_{t-1}$ 的系数为正值, 表明消费者满意搜索指数的关注度的提高会引起消费者信心指数的上升。两个综合的搜索指数变量相对与消费者信心指数领先阶数是不同的, 这表明消费者预期和消费者满意对消费者信心的影响时差是不同的, 其中消费者预期搜索指数是领先消费者信心指数 2 个月, 消费者满意搜索指数领先消费者信心指数 1 个月。两个搜索指数变量的系数不同, 这说明消费者预期和消费者满意两个影响因素的变化对消费者信心的影响程度是不同的。

为了直观的了解模型 3 的拟合效果, 把消费者信心指数的真实值和模型 3 对消费者信心指数的拟合值的曲线图绘制在了一起, 便于进行比较, 具体见图 3.2

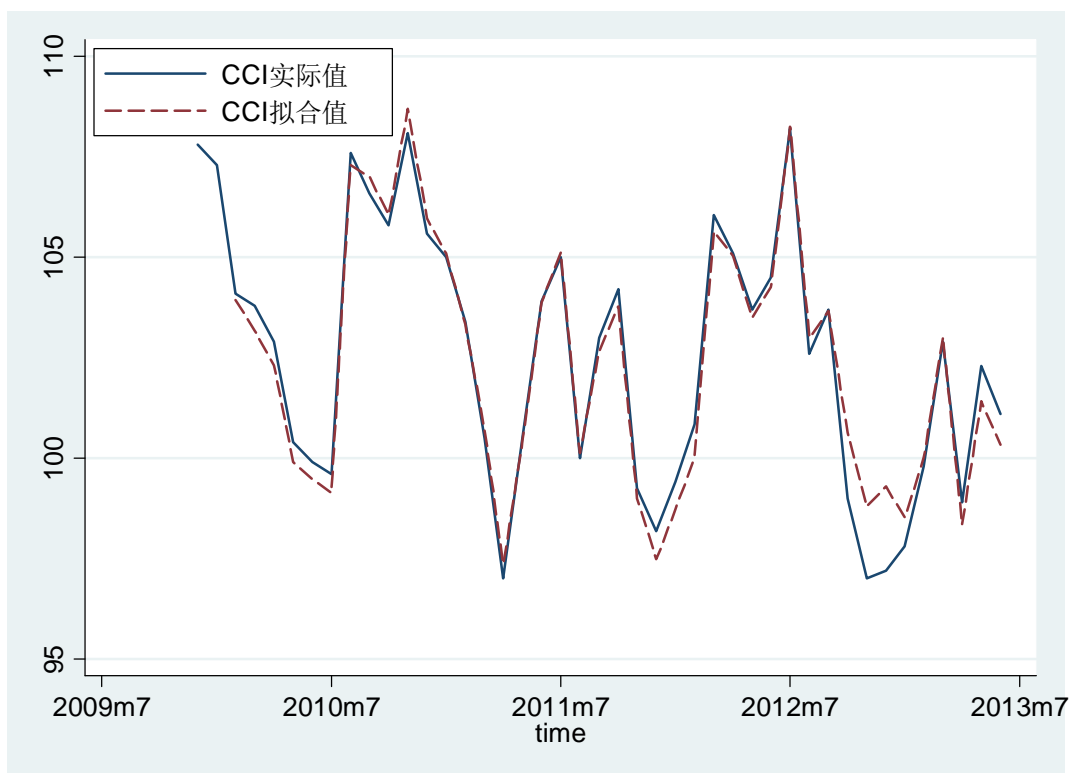


图 3.2 拟合效果图

可以看出模型 3 对于消费者信心指数的拟合效果基本上令人满意,两者的走势基本上是同步的,有待之后进一步的探讨和改进。为了检验模型的预测效果如何,本文选取了从 2013 年 7 月至 2013 年 12 月 6 个月的数据进行预测,结果显示预测的均方根误差为 0.69,平均绝对误差为 0.49。从整体上看模型 3 无论是对样本内数据的拟合还是样本外的预测,都有一定的效果,也就是说网络搜索数据可以在一定程度上对消费者信心指数进行预测,但是在某些时段,消费者信心指数的模型拟合值和消费者信心指数的真实值之间还是存在较大的偏差,模型进行预测的误差也有待进一步的降低。

4 结论与展望

本文先从理论上分析网络搜索数据与消费者信心指数之间的相关关系,然后进一步分析了网络搜索数据能够反映消费者信心变化的原理,进而通过实证研究本文得出了以下结论:

(1) 通过实证分析的表明:网络搜索数据与消费者信心指数之间是协整的。消费者预期搜索指数的提高会引起消费者信心指数的下降;消费者满意搜索指数的提高会引起消费者信心指数的上升。消费者预期搜索指数相对与消费者信心指数的提前期为 2 个月,而消费者满意搜索指数的提前期为 1 个月。

(2) 经过理论分析和实证分析表明,网络搜索数据中的部分数据和消费者信心之间具有显著的相关性,这也证实了理论部分的分析,消费者信心是一种心理感受,难以直接度量,但消费者信心会通过消费者的行为加以反映。消费者互联网上的行为便是消费者行为的客观反映。因此部分网络搜索数据和消费者信心指数之间存在较强的相关性是有说服力的。

(3) 利用网络搜索数据合成的消费者预期搜索指数和消费者满意搜索指数与消费者信心指数之间存在不同时间的时差。其中消费者预期搜索指数领先消费者信心指数两个月,消费者满意搜索指数领先消费者信心指数一个月。这是由于消费者预期指数主要反映的是消费者受宏观经济等大环境影响消费者对自身及家庭的收入和发展的评估和预判,本身就具有超前性。而消费者满意搜索只是消费者对现阶段经济生活的评价,主要受到微观因素的影响,也就是消费者日常的消费行为的影响较大,因此消费者满意搜索指数的领先阶数比消费者预期搜索指数要少一个月。

与传统数据相比网络搜索数据具有数据量大、可瞬时产生、获取成本低、准确和真实等优点,已经成为了国内外学者争相研究的热点之一,通过网络搜索可以对互联网影响较大的诸多方面进行研究,为传统数据很难获取且与互联网关系密切的领域提供了研究的可能。目前的研究领域已经涉及到了医药健康、社会民生和经济政治等领域,相信随着互联网硬件和软件的进一步发展,网络服务的进一步完善,网络搜索数据的研究领域将会得到进一步拓展,研究的结果将更加准确,产生的经济和社会价值将难以估计。

参考文献

- 国敏. 2012. 基于网络搜索技术的游客量预测方法研究. 北京: 首都师范大学.
- 郭洪伟, 2012. 消费者信心指数的研究. 北京: 首都经济贸易大学.
- 刘颖, 吕本富, 彭廉. 2011. 网络搜索对股票市场的预测能力:理论分析与实证检验. 经济管理, (1): 172~179.
- 王曙. 2013. 基于百度指数的投资者关注度与股票指数表现的相关性研究——来自概念股的相关性研究. 南京: 南京大学.

- 张崇, 吕本富, 彭赓, 刘颖. 2012. 网络搜索数据与 CPI 的相关性研究. 管理科学学报, (7): 50~58.
- 中国互联网络信息中心. 2014. 《第 33 次中国互联网络发展状况统计报告》.
- 杨欣, 吕本富等. 2013. 基于网络搜索数据的突发事件对股票市场影响分析. 数学的实践与认识, (23): 17~28.
- Eysenbach G. 2006. Infodemiology: Tracking Flu-related Searches on the Web for Syndromic Surveillance. AMIA Annual Symposium Proceedings. American Medical Informatics Association. 2006: 244~248.
- Polgreen P M, Chen Y, Pennock D M, Nelson F D. 2008. Using Internet Searches for Influenza Surveillance. Clinical Infectious Diseases, 47(11): 1443~1448.
- Askatas N, Zimmermann K F. 2009. Google econometrics and unemployment forecasting. Applied Economics Quarterly,(92): 107~120.
- Cho H, Varian H. 2012. Predicting the Present with Google Trends. Economic Record, 88(s1): 2~9.
- Chris H, Guy J. 2012. Searching for the picture: forecasting UK cinema admissions using Google Trends data. Applied Economics Letters. 19 (11): 1051~1055.
- D'Amuri F. 2009. Predicting Unemployment in Short Samples with Internet Job Search Query Data. Germany. University Library of Munich.
- Vosen S, Schmidt T. 2011. Forecasting private consumption: survey - based indicators vs. Google trends. Journal of Forecasting, 30(6): 565~578.
- Wu L, Erik B. 2013. The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales. Economics of Digitization. University of Chicago Press.