

论文编号：A0751

# 基于消费数据和微博文本评论数据的电子 消费券经济驱动力研究——以郑州市为例

论文题目：基于消费数据和微博文本评论数据的电子  
消费券经济驱动力研究  
——以郑州市为例

参赛学校：天津财经大学

参赛成员（作者）：王浩红、常梦、陈井妍

指导老师：风丽洲

## 目录

摘要.....	I
一、引言.....	1
(一) 研究背景与意义.....	1
(二) 国内外研究现状.....	3
(三) 报告内容与组织结构.....	6
1. 报告内容.....	6
2. 组织结构.....	6
二、研究对象与数据处理.....	7
(一) 研究对象与数据来源.....	7
(二) 数据预处理.....	8
三、基于线性回归的消费券效用影响分析.....	8
(一) 变量的设计与选择.....	8
(二) 模型的建立与参数估计.....	8
(三) 模型结果分析.....	10
四、基于 ARIMA 模型的消费券效用分析.....	10
(一) 描述性分析.....	10
(二) 模型建立与检验.....	11
(三) 模型结果分析.....	13
五、面向郑州市消费券政策的网民评价分析.....	14
(一) 模型介绍.....	14
(二) 数据准备与算法实现.....	14
(三) 结果实现与分析.....	15
1. 评论词云图实现.....	15
2. 词共现网络主题分析.....	16

3. 文本情感分析.....	18
六、结论与建议.....	19
(一) 结论.....	19
(二) 建议.....	20
参考文献.....	22
致谢.....	24

## 图目录

图 1 郑州市的社会消费品零售总额变化图.....	11
图 2 一阶差分的 acf、pacf 图.....	12
图 3 预测值与真实值对比.....	13
图 4 评论词云图.....	16
图 5 词的部分权重矩阵.....	16
图 6 词共现网络图.....	17
图 7 情感得分图.....	18

## 表目录

表 1 线性回归结果.....	9
表 2 ARIMA 预测模型 .....	12
表 3 社会消费品零售总额的预测值与真实值对比 (亿元) .....	13
表 4 城市消费券的最高门槛和最高优惠值.....	17

## 摘要

消费券是政府促进经济良性循环的重要手段。早在 2008 年经济大危机的时候,我国就已经将这一政策付诸实践,而当年并没有现在这么发达的网络,所以当时政府所发放的纸质消费券在经济上虽然起到了相对的促进作用,但是存在较多的问题,例如纸质消费券不易保存,存在造假现象等。这些原因都直接或间接地对后期政策效果测度产生阻挠作用,同时也无法准确衡量其对经济的拉动作用。现如今,数字消费已经成为人们生活消费的主流,这为通过网络下发消费券提供了契机,政府能通过网络平台直接发放电子消费券并实时记录相关数据。鉴于此,本文选取的研究对象是郑州市,进而分析政府所发放的电子消费券对经济的拉动作用。

首先,通过建立多元线性回归模型,确定消费券对消费的正向影响。其次从郑州市统计局官网上整理出 2018 年 3 月到 2020 年 8 月郑州市社会消费品零售总额的数据,并针对 2018 年 3 月至 2020 年 3 月数据,利用 ARIMA 模型进行建模,预测 2020 年 4 月及其后的总额,与真实的零售总额进行对比分析,根据两者之间的差距判断政策的有效性。最终根据分析发现该政策会刺激经济,能够使经济迅速回暖。另外,从微博上抓取网民对郑州市发放电子消费券这一热点话题的评论,通过对微博评论的文本分析了解民众对于这一政策的看法,发现电子消费券对于大部分民众具有普适性,但也存在一些问题,例如一些网友认为郑州市所发放的消费券门槛过高,抢到了也用不上等结论。最后结合分析结论给政府提出了降低消费券的门槛和注重其存在的公平性与合理性等建议。

关键词：电子消费券 线性回归 ARIMA 预测 文本分析 经济驱动力

# 一、引言

## (一) 研究背景与意义

党的十九届五中全会提出,加快构建以国内大循环为主体、国内国际双循环相互促进的新发展格局以及增强消费对经济发展的重要作用是我国目前的主要任务。在“三驾马车”整体乏力的大背景下,消费成为拉动经济增长的中坚力量。国家统计局发布的数据显示,2020年1、2月,全国社会消费品零售总额与去年同月相比下降20.5%,尤其是,餐饮收入与去年同月相比下降43.1%。2020年3月起国内疫情有所好转,全国开启了“重启键”,但受国际疫情持续蔓延影响,餐饮、文化、旅游、住宿等服务业仍处于低迷状态。政府为了稳定国家经济大环境,出台了一系列宏观调控经济手段的政策,其中,发放电子消费券的目的是用财政补贴的方法刺激消费、释放消费需求。

消费券作为现金替代品,由政府或商家发放,只能用于购买商品或服务,且有一定期限。而在疫情之下,经济不景气,消费券的发放是作为刺激消费的一种手段,可以提高居民的购买欲望和购买能力、刺激消费回暖,甚至可以作为拉动内需提振经济的一种工具。1939年美国的食品券是最早出现的一种消费券。在当时社会大萧条的条件下,美国居民每月只需支付83美元就可购买到价值153美元的食品,这项计划惠及两千万美国民众。1999年日本政府为冲破股市和房市泡沫破灭后经济长期阻滞的状态,决定向特定年龄段(15~65岁)的群众每人发放两万日元的“地域振兴券”。在2008年金融危机时期,消费券在我国出现,主要包括食品消费券、旅游券等。我国金融危机期间发放消费券虽对经济起到一定正向效应,但是存在着许多问题,如需投入大量人力物力,难以发放到流动人口,跨区域使用受到局限,“打假”难度大,资金投入核算不清晰和政策绩效评价滞后等<sup>[2]</sup>。而面对新冠疫情的冲击下,我国政府依靠领先的数字支付技术在网

络平台上发放数字消费券,同时这种数字消费券完全避免了以前纸质消费券所存在的“打假”难度大、使用受到局限、无法评估等一系列问题。而这些问题在当下这个时代都将迎刃而解,大数据时代的来临使得数字消费成为人们生活的一种消费主流,才使得电子消费券能够以一个更快捷便携的方式出现在人们的生活中。同时,政府也能在节省大量物力人力的前提下实施该政策,这正是数据新动能的体现。人人都可以通过网络进行消费券的领取,并且实名制,不会出现造假现象,相对于纸质消费券电子消费券更为方便快捷,这样不但节省了人们的时间,提高了效率,而且更会对经济的推动呈现出乘数效应。数据成为政策实施的一个推手,从而转化为人们新的消费动能,直接刺激消费,带动经济发展。另外,政府发放消费券的根本目的也是想要刺激国民经济回暖,而发放消费券这一政策同时含有经济学的意义:一方面,从微观经济学的角度来看,民众所持有的数字消费券将内化为每个人的当期收入;另一方面,从宏观经济学的角度来看,数字消费券的发放依据的是凯恩斯的乘数原理,当发放的数字消费券转化为民众的当期收入时就意味着这是他们的可支配收入,消费支出一单位的增加将会使国民收入获得成倍增加。消费券通过精准刺激受疫情影响最为严重的服务业,拉动其他行业的景气回升。

因此,为了刺激消费、提振内需,促使经济恢复良性循环,各地政府纷纷采取发放消费券的政策。电子消费券一方面有利于居民消费结构优化升级,另一方面针对制约消费升级和扩大内需的问题,在打通堵点、挖掘潜力上发挥重要功能,有助于促进消费从数量和质量上同时提升、从而重启经济,实现产业和消费“双升级”。

郑州地处中原,不仅作为交通枢纽,在全国经济发展中扮演重要的角色,而且是“中国八大古都”之一和世界历史都邑联盟成员。例如在2020这一年,郑州共创造了12003亿元生产总值,财政一般公共预算收入高达1259.2亿元,这

使得郑州市入围中国城市综合竞争力 20 强 ,同时 13 家超百亿级企业 ,1.2 :39.7 :59 的一二三产业比重 ,也使郑州在全国的作用持续凸显。2020 年 4 月 3 日 ,郑州市政府下发价值 5000 万的普惠型消费券 ,同年 4 月中旬和 6 月又共下发 3.2 亿的社会型消费券。郑州商务局的调查数据显示 ,第一期发放的消费券核销金额达 3925 万 ,合计带动消费 5.5 亿元 ,其杠杆比例达到 1 :14。出于短期的应急目的 ,消费券这个“ 经济药引子 ”的作用持续凸显 ,最终形成长期经济杠杆效果 ,这就是一种数据新动能的表现。

2020 年 4 月 9 日 ,中共中央、国务院印发《关于构建更加完善的要素市场化配置体制机制的意见》,《意见》明确将数据作为一种新型生产要素。这一举措表明数据正式迈入要素市场 ,如何充分发挥数据对其他要素效率的倍增作用 ,促进数据要素市场发展成熟 ,令大数据成为推动经济高质量发展的新动能 ,是当今社会各大领域都关心的话题。而文本数据 ,即以各种自然语言形式存在的数据 ,是一种特殊的大数据 ,经过高效的挖掘与分析 ,以及人工智能技术处理后 ,这些数据可以产生极大价值。在电子消费券发放过程中 ,网民在各大社交媒体上发表评论 ,这些评论数据成为一种新的数据源 ,可被用来研究民众对电子消费券的接受程度 ,在电子消费券经济驱动力研究中提供辅助作用。

因此 ,本文的研究目标是通过了解郑州市发放电子消费券后的消费情况 ,进而分析其是否对经济有带动效用 ,以及民众的态度是否会影响消费券政策的可持续性。

## (二) 国内外研究现状

消费券作为特用券的一种 ,可用于实现经济政策。当经济萧条或突发事件造成居民消费能力急剧下降时 ,消费券是人民未来消费时的支付凭证 ,由政府或者企业发放 ,通过增加民众的购买力与消费欲望以达到增加消费活动、带动生产与



投资增长、加速经济复苏。电子消费券是指一定面值的金额以数字代码的形式呈现，并可凭此代码在消费时抵消一定或全部消费金额。

近年来国内学者对我国消费券的经济效用研究大多停留在理论层面，直至去年，部分学者尝试从实证层面考察消费券发放的效果及其前景。林毅夫等人在《我国消费券发放的现状、效果和展望研究》（2020）一文中总结消费券有三个具体特征：一是反应迅速；二是与民生密切相关、小额面值；三是发放过程兼顾与效率，并创造性地从核销率、消费券消费者用户画像和商家画像三个角度，来初步描述消费券的发放效果，通过双重差分和三重差分方法研究消费券发放对交易笔数和交易金额的影响、消费券发放方式的影响和城市特征的影响。得到如下结论：一、发放消费券可以促进当地的消费，发券地区受支持行业的支付笔数比未发券地区同期同行业高约 26.26%。二、在第三产业占比高的地区，发放消费券明显提高交易量；为低收入人群定向发放消费券可以增加其消费。三、财力受限是部分城市不发放消费券的主要原因，而不是消费券无效<sup>[1]</sup>。周丽凤在《基于结构方程模型的体育消费券效用分析——以江苏省部分城市为例》（2020）一文中运用文献资料法、专家访谈法、问卷调查法和数理统计法，使用结构方程模型构建体育消费券效用模型，并探究影响体育消费券效用的主要因素<sup>[2]</sup>。吴坚平等在《基于旅游电子消费券发放的优化策略评价模型研究》（2020）一文中以杭州市淳安县为例，通过在支付宝平台发放旅游电子消费券的实践中，围绕电子消费券的派发量、领取量、核销量，以及消费属性、消费类型等多维数据，建立结构方程数据评价模型，从旅游消费量和旅游消费券杠杆效应两个方面进行评价，提出旅游消费券效用综合指数，在实践中不断优化迭代，最终起到了倍增效应<sup>[3]</sup>。刘虎等在《基于新浪微博评论数据的消费券政策效果评估分析》（2020）一文中尝试基于新浪微博网民评论数据，对消费券政策进行评估分析，主要方法为基于价值密度的文本分类、灰色关联分析模型，得到影响消费券政策支持度的显著因素。



同时利用潜在狄利克雷分配(LDA)主题聚类模型,对网民评论数据进行探索性分析,探究消费券“套现”问题的深层次原因,最后对优化消费券政策提出相关建议<sup>[4]</sup>。吴翌琳,张旻在《基于PSM-DID模型的消费券效果评估》(2021)一文中通过建立PSM-DID模型,评估发放消费券的效果以及城市异质性对消费券效果的影响,然后建立三重差分模型比较消费券效果的地区差异,发现消费券的发放与否和时间早晚与地区经济、财政、社会状况密切相关;整体而言消费券对社会消费有显著的促进作用:相比而言城镇化水平较低,第三产业比重较低、人均可支配收入较低的地区消费券作用更为显著,受疫情影响较小的地区发放效果更好;消费券的效果存在显著的东西部差异。为此本文建议加大发放力度,同时统筹规划,向低收入群体、低收入地区倾斜<sup>[5]</sup>。姜蒙强,周志太在《新冠疫情下数字消费券刺激消费回补的政策效应》(2021)一文中利用全国省级收入消费面板数据建立个体时点双固定效应模型,分析新冠疫情期间数字消费券的经济学效用原理得出:数字消费券对刺激消费具有很强的正向效应,当居民得到1%的消费券转移性收入时,相应消费支出扩大0.354967%;部分地区存在数字消费券发放规模小、发放策略不够细化、市场乱象、地方政府财政压力等问题<sup>[6]</sup>。

综上所述,国内学者大多利用文本挖掘、消费券发放金额等相关数据,通过双重差分、三重差分等方法,建立结构方程模型,并尝试构建消费券效用评价指标。但不足之处在于:部分研究大多停留在理论分析,无实证分析作出强有力的支撑;其次研究所收集数据仅为静态数据,未能以动态视角考察消费券的效用;另外大多实证分析在较大地区范围内研究消费券的效用如何,故而不能给出针对不同地区的消费券的前景建议。

《NOW IN VOGUE》(2011)这篇文章重点介绍了2009年,上半年美国优惠券使用的增加。它指出了自2008年以来失业率上升与优惠券消费之间的关系。互联网被认为是同期分发优惠券的主要方式,兑换率上升了308%<sup>[13]</sup>。《Coupon

Redemption》1978 阐述了优惠券在广告商中受欢迎的原因有很多，同时介绍了一种计量经济学的优惠券赎回模型，该模型提供了用于评估优惠券赎回的经验估计。

### (三) 报告内容与组织结构

#### 1. 报告内容

主要研究内容：首先建立社会消费品零售总额与消费券政策实施的 OLS 回归，确定发放消费券这一政策对消费的正向影响，然后针对 2018 年 3 月至 2020 年 8 月郑州市的社会消费品零售总额进行描述性分析及 ARIMA 预测模型来分析消费券的效用；同时，通过抓取微博网民对“郑州将发放 4 亿消费券”和“郑州市再发放 1.6 亿元消费券”这两个微博话题的评价词条，进行文本情感分析和词共现网络主题分析。通过分析得到的结论，给郑州市政府未来消费券的发放政策提供一些可行性的建议。

创新之处：本文将采用数据分析与民众对消费券的态度文本分析相结合的方式，其次本文研究了消费券在郑州市这一具体地区的经济效用，为其他地区的研究提供借鉴。

#### 2. 组织结构

第一章，交代选题背景，梳理国内外研究现状，确定本文的研究内容。

第二章，确定研究对象及数据的收集与整理，包括所需的社会消费品零售总额数据、关于郑州市消费券政策文本的网民评价的收集、存储与清洗。

第三章，建立线性回归模型，探究消费券政策的实施对消费的影响方向及影响程度。

第四章，利用 ARIMA 模型对消费券效用进行分析。

第五章，利用情感分析和词共现网络主题分析，基于消费券政策网民评价，

分析民众的态度是否会影响消费券政策的可持续性。第六章,进行全文的总结与展望。

## 二、研究对象与数据处理

### (一) 研究对象与数据来源

影响消费的主要因素有国家的经济发展水平、物价总体水平和收入水平。于是从郑州市统计局官网上整理出了从 2019 年 4 月到 2020 年 6 月的社会消费品零售总额和居民消费物价指数的月度数据,2019 年第二季度到 2020 年第二季度的居民人均可支配收入和地区生产总值。同时建立虚拟变量  $D$ ,2020 年 4 月之前赋值为 0,代表电子消费券政策实施前,2020 年 4 月之后赋值为 1,代表电子消费券政策实施后,建立 OLS 回归,探究电子消费券政策的实施对社会消费的影响。

社会消费品零售总额在一定程度上反应消费力度以及经济的增长情况,而消费券的政策旨在拉动经济增长,这一部分使用的是从郑州市统计局官网上整理出了从 2018 年 3 月到 2020 年 8 月社会消费品零售总额的月度数据,其中基于 2018 年 3 月到 2020 年 3 月的数据建立 ARIMA 模型预测 2020 年 4 月至 8 月的零售总额,再利用实际值与预测值的差距来判断消费券的经济效用。

当突发事件、社会民生、环境保护等热点问题发生时,民众通过社交媒体进行热烈讨论,导致网络舆情的爆发及传播。同时因为网上社交媒体舆论扩散速度快、数量多及范围广,于是本文选择从微博用户对于发放消费券的评论来看民众对于该政策的态度,在此,以郑州市为代表,在微博客户端抓取了“郑州市将发放 4 亿消费券”以及“郑州再发放 1.6 亿元消费券”这两个热点话题下网民对于该事件的评论。其中,有关郑州市发放消费券的政策话题也登上过新浪微博热搜榜。由此可见其网络舆情的热度之高。

## (二) 数据预处理

1.由于在郑州市统计局官网上所获取的居民人均可支配收入、地区生产总值均为季度数据，于是采取季度数据除以3的方法，粗略转换为月度数据，以进行月度数据的回归。

2.由于在郑州市统计局官网上所获取的月度数据中2019年的1、2月份以及2020年的1、2月份的社会消费品零售总额未发布，因此，本文分别用2019年和2020年的2月份社会消费品零售累计总额求平均得出相应月份的社会消费品零售总额来替代。

3.对所抓取的评论去除话题、一些表情包以及一些比较官方的话，重点提取出网友的评论，同时利用python软件进行结巴分词并去停用词的处理。

## 三、基于线性回归的消费券效用影响分析

### (一) 变量的设计与选择

将社会消费品零售总额作为被解释变量，消费券政策实施与否为解释变量，同时引进地区生产总值、居民人均可支配收入、居民消费价格指数作为解释变量。

对变量作如下定义：sales 指社会消费品零售总额；D 定义为消费券政策实施与否，以2020年4月为时间节点，2020年4月之前，D取0代表消费券政策未实施，2020年4月之后，D取1代表消费券政策已实施；gdp指地区生产总值；income指居民人均可支配收入；price指居民消费价格指数。

### (二) 模型的建立与参数估计

根据上面的变量定义，可以设定一般的计量经济学模型如下：

$$\text{sales} = \beta_0 + \beta_1 D + \beta_2 \text{income} + \beta_3 \text{price} + \beta_4 \text{gdp} + \mu \quad (1)$$

首先进行被解释变量社会消费品零售总额和消费券政策实施与否的一元线性回归，然后逐个加入剩余的解释变量，得到不同模型的回归结果如下：

表 1 线性回归结果

	模型 1	模型 2	模型 3	模型 4
常数	378.097** (43.059)	326.828** (19.172)	118.744 (0.197)	825.826 (1.184)
D	48.608** (4.201)	55.259** (4.360)	55.103** (4.075)	-4.596 (-0.190)
income		0.006** (3.654)	0.006** (3.520)	-0.147** (-2.631)
price			2.069 (0.343)	-3.809 (-0.572)
gdp				0.531** (2.726)
样本量	18	15	15	15
R <sup>2</sup>	0.447	0.726	0.729	0.771
调整 R <sup>2</sup>	0.413	0.680	0.655	0.680
F 值	F(1, 16)=17.645, F(2, 12)=11.461, F(3, 11)=7.282, F(4, 10)=8.831,			
概率	p=0.001	p=0.002	p=0.006	p=0.003

注 \*p<0.05 \*\*p<0.01 括号里面为 t 值

从表 1 中的模型 1 可知，将 D 作为自变量进行回归分析，在 Robust 稳健标准误回归方法下进行研究。在  $\alpha=0.05$  的显著性水平下，模型通过稳健性检验，说明 D 一定会对 sales 产生影响，得到模型： $\text{sales}=378.097 + 48.608 \times D$ 。通过最终的具体分析可知，D 的回归系数值为 48.608，并且呈现出 0.01 水平显著性 ( $t=4.201$ ,  $p=0.000<0.01$ )，说明 D 会对 sales 产生显著的正向影响。

引进 income 解释变量时，由模型 2 结果知，D 会对 sales 产生的正向影响依然显著；引进 income 和 price 解释变量时，由模型 3 结果知，D 会对 sales 产生的正向影响依然显著；但在引进 income、price 和 gdp 解释变量时，由模型 4 结果知，D 会对 sales 产生的正向影响不显著。于是最终确定模型如下：

$$\text{sales} = 326.826 + 55.259 \times D + 0.006 \times \text{income} \quad (2)$$

(19.172) (4.360) (3.654)  $R^2=0.726$

D, income 作为自变量进行回归分析, 并且使用 Robust 稳健标准误回归方法进行研究, 从上表可以看出, 模型的可决系数  $R^2$  为 0.726, 意味着 D、income 可以解释 sales 72.6% 的变化。对模型进行 F 检验时发现模型通过 F 检验 ( $F=11.461$ ,  $p=0.002<0.05$ ), 也即说明 D、income 至少一项会对 sales 产生影响关系。D 的回归系数值为 55.259, 并且在 0.01 的置信水平下统计上显著 ( $t=4.360$ ,  $p=0.000<0.01$ ), 表明 D 会对 sales 产生显著的正向影响。

### (三) 模型结果分析

从模型回归方程来看, 在保持其他变量不变的条件下, 平均而言, 消费券政策的实施会使社会消费品零售总额增加 55.259 亿元。结合上述模型的分析, 联系现实生活, 消费券政策的实施实际上增加了居民的转移性收入, 从而它对社会消费品零售总额的影响程度其实包含在居民可支配收入影响内; 受益于数字经济的迅速发展, 与之前的纸质消费券不同, 正是由于今年许多城市开始以支付宝、微信为平台发放电子消费券, 所以消费券使用效率及对经济的拉动效应均较过去明显提升。

## 四、基于 ARIMA 模型的消费券效用分析

### (一) 描述性分析

根据所整理的郑州市 2018 年 3 月~2020 年 8 月社会消费品零售总额的数据, 利用 R 绘制折线图 (如图 1 所示)。



图 1 郑州市的社会消费品零售总额变化图

根据图 1 观察到社会消费品零售总额整体上市波动上升,但是在 2020 年 1 到 2 月的时候因为疫情的影响致其急剧下降,随后又有所回升上升,而且 2020 年的 4、5 和 6 月份是不断上升的趋势。

根据图 1 的判断以及 2020 年 4 和 6 月份的时候郑州市正好发放消费券来刺激消费以带动经济增长,而且第三部分线性回归模型的结果显示消费券政策的实施对社会消费零售总额有正向影响。因此本文使用了时间序列的 ARIMA 预测模型对 2018 年 3 月至 2020 年 3 月的社会消费品零售总额进行建模,并预测 2020 年 4 月至 2020 年 8 月的社会消费品零售总额。再根据 2018 年 3 月到 2020 年 8 月的真实值和预测值绘制趋势变化图,以此来了解政府发放消费券的政策是否有影响。

## (二) 模型建立与检验

1.使用 R 软件导入所需数据,用 Box.test 检验数据已通过白噪声检验,随后用  $\text{adf.test}()$  检验观察数据是否平稳,根据结果显示在显著性水平 0.05 下,  $P$  值  $=0.226 > 0.05$ ,所以不能拒绝原假设,数据不平稳。

2.进行一阶差分,再次使用  $\text{adf.test}()$  检验,得到  $P$  值  $=0.01 < 0.05$ ,拒绝原假设,可判断一阶差分后的数据平稳。因此,绘制 acf、pacf 图,根据图 2 知差分



后的自相关图与偏自相关图没有超过边界值，可确定数据属于平稳序列。

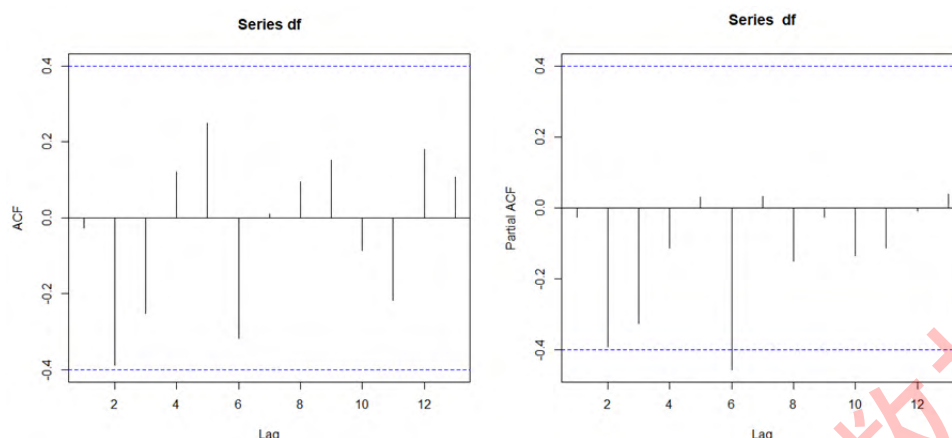


图 2 一阶差分的 acf、pacf 图

3.为了选择合适的 ARIMA 模型，根据 0, 1, 2 从低阶到高阶进行参数抉择，根据 AIC 信息准则来判断最优模型。根据表 2 比较发现 ARIMA(2, 1, 1)的 AIC=224.94 最小而且其参数通过了显著性水平为 0.05 的显著性检验，因此选择此模型进行后期的预测。

表 2 ARIMA 预测模型

模型	aic
ARIMA(1, 1, 0)	229.45
ARIMA(1, 1, 1)	226.56
ARIMA(2, 1, 0)	226.78
ARIMA(2, 1, 1)	224.94

4.对残差序列进行 Box.test()白噪声检验，得出 P 值=0.840> 0.05，不能拒绝原假设，说明残差之间不相关、模型显著成立。因此，ARIMA(2, 1, 1)模型对该时间序列数据拟合成功。

ARIMA(2, 1, 1)模型公式表示为：

$$Y_t = 0.532Y_{t-1} - 0.469Y_{t-2} + e_t - 0.749e_{t-1} \quad (3)$$

从该模型可以看出，前两期的真实（预测）数据对最新估计的数据是负向的影响，上一期的影响则是正向的。

5.使用 ARIMA(2, 1, 1)模型预测 2020 年 4 月至 2020 年 8 月的社会消费品

零售总额，结果如表 3 所示。

表 3 社会消费品零售总额的预测值与真实值对比（亿元）

日期	真实值	预测值	差值
2020 年 4 月	399.86	403.82	-3.96
2020 年 5 月	428.34	389.42	38.92
2020 年 6 月	437.46	372.09	65.37
2020 年 7 月	407.30	369.63	37.67
2020 年 8 月	439.84	376.44	63.40

### （三）模型结果分析

同时为了更好的呈现数据，利用该模型预测了 2018 年 3 月到 2020 年 3 月的社会消费品零售总额（亿元）。考虑到折线图能更为直观地体现数据的变化，接着绘制了预测值和真实值的时序图（如图 3）。

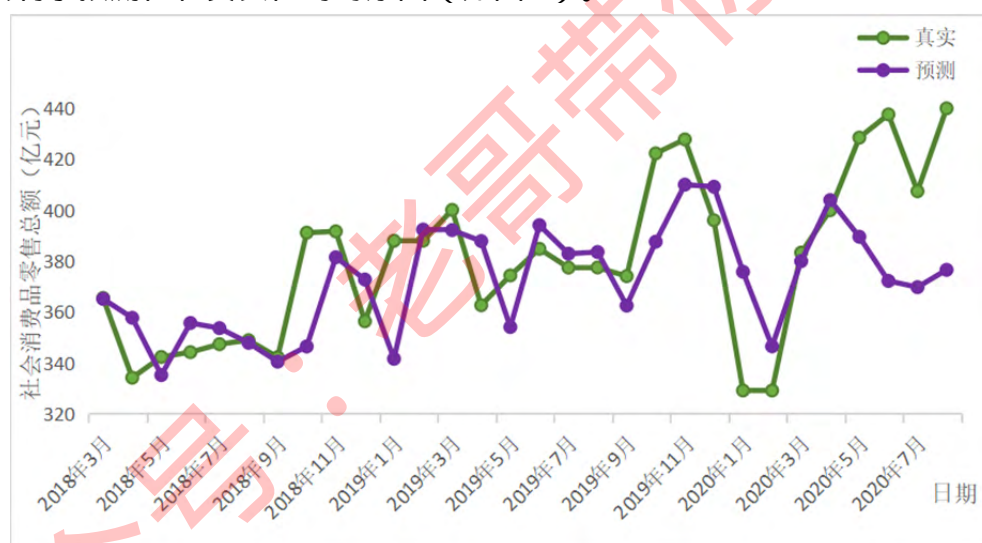


图 3 预测值与真实值对比

从图 3 看出，2018 年 3 月至 2020 年 3 月的预测值普遍比真实值高，即使低于真实值时也没有连续几期差距特别大的。然而在 2020 年 4 月份以及后期的预测值普遍比实际值低很多，且与 2020 年 4 月（政策实施前）的预测数据相比，发放消费券的政策实施后预测数据与实际数据差距特别大。由于消费券是采用支付宝等线上平台发放的，可以理解为是电子消费券的影响，同时其使用是有有效期的，所以其对经济的影响是需要一定时间缓冲的，在此电子消费券的新动能通

过 2020 年 5、6、7 月的预测值与真实值的巨大差距可以解释。因此消费券的这个政策对经济有一定推动力的,这也与第三部分所做的线性回归模型的分析结果是吻合的。

## 五、面向郑州市消费券政策的网民评价分析

基于上文的分析发现,发放消费券的确会拉动消费、促进经济增长,因此,利用文本数据挖掘技术,针对抓取到的电子消费券政策评论数据进行分析与建模。网民通过各平台和媒体对社会上各种突发事件、民生问题、环境保护等方面的热点事件的讨论,会导致事件热度迅速扩散,最终会使许多热点话题的传播呈现出指数爆炸的形式,不论是传播范围还是传播频率都非常高。网络上的各种评论信息表现了民众的内心情感色彩和情感倾向性。基于此,通过浏览带有主观色彩的评论来了解大众舆论对于政策的看法,并以此为依据进行分析。

### (一) 模型介绍

1.文本情感分类模型,主要是对采集到的文本数据进行情感分析,该模型可以对包含主观信息的文本按照情感倾向进行极性分类。即文本所包含的情感是“好”还是“坏”,在这里把“好”和“坏”形象地看作是褒义和贬义两个极性。

2.词共现网络主题模型,主要是对分词后的文本进行向量化表示,然后根据权重的大小确定其重要性,选取权重大的部分词以词共现网络的形式体现,通过研究可以发现复杂网络中普遍存在着“社区内部联系紧密,社区间联系稀疏”的特点,可以根据这一特性进行主题提取,从而确定人们所关注的重点。

### (二) 数据准备与算法实现

1.首先,对经过去停用词和分词处理后的文本评论数据绘制词云图,初步了

解一下民众对于政府所实行的消费券政策的看法及态度。

2.利用 TF-IDF (词频-逆向文件频率) 向量空间模型进行文本向量化表示, 将纯文本表示为机器所能识别和处理的向量形式。

3.利用词共现网络主题分析将文本数据以网络形式呈现, 从而转化为复杂网络中的社区发现问题。基于 TF-IDF 向量空间模型可以得到所有词的重要性排序, 从而得到每个词的权重, 从中选取权重最大的 30 个词, 以词作为网络中的节点, 并认为两个节点之间存在一条边, 通过文本向量计算两个词在文本中的相似性并将其作为边的权重。由此, 可以将文本的关键词以复杂网络的形式体现, 从而得知文本的主题。

4.使用文本情感分析模型针对民众基于郑州市发放消费券这个事件的评论进行褒贬极性分析。

### (三) 结果实现与分析

#### 1.评论词云图实现

为了了解大众对这个话题的主要观点及看法, 利用抓取的评论绘制了词云图, 将复杂无序的文本可视化, 以便更为直观抓住网民基于郑州市发放消费券的关注点。



图 4 评论词云图

根据图 4 发现民众微博评论的关键词主要有“刺激消费”、“感谢政府”、“商家”、“抢到”、“经济”、“买”、“错过”等；可以看出人们对于政府发放数字消费券来刺激消费的普适性，人们普遍认为发放消费券可以拉动经济，从词云图中看到“支付宝”、“错过”、“羡慕”等词，可以看出民众对于消费券的接受度比较高，所以此政策在一定程度上是可行的。

## 2.词共现网络主题分析

将进行分词和去停用词处理的评论进行文本向量化表示,并得到相应的词频

权重矩阵（部分如下）：

[illegible]

图 5 词的部分权重矩阵

从上面算出的词频权重中选取权重最大的前 30 个词进行词共现网络的实

现，得到相应的结果如下：



图 6 词共现网络图

从图 6 可知，“政府”、“刺激”、“消费”、“感谢”之间联系最为紧密，这也正是消费券政策的要点所在；其次，“开心”、“哈哈”、“终于”的联系较为紧密，这一部分体现出抢到消费券的民众的心理特征；另外，“错过”、“门槛”、“花钱”等词的联系比较紧密，依据消费券的面值及满减情况可以推测出还是有一部分民众对于消费券政策抱有负面情绪，认为使用消费券的门槛过高。

表 4 城市消费券的最高门槛和最高优惠值

城市	最高门槛（元）	最高优惠（元）	平均门槛（元）	平均优惠（元）	总金额（亿元）	城镇常住人口（万人）	人均金额（元）
成都	40	20	40	20	2	1233.8	16.210
福州	90	30	60	20	1.50	544	27.574
杭州	255	45	132.50	27.50	16.80	813.3	206.566
嘉兴	800	200	350	82	2.5	323.5	77.280
南京	50	100	75	50	3.18	707.2	44.966
秦皇岛	0	100	0	60	0.05	191	2.618
沈阳	50	50	50	50	0.30	674.1	4.450
温州	130	38	48	13.60	18	655.6	274.558
武汉	50	50	33.33	26.66	23	902.5	254.848
郑州	2500	500	408	90.70	4	772.1	51.807

注：表格信息来自于文献<sup>[4]</sup>作者的整合



经过对文本的一系列分析,发现网民对于郑州市消费券的一个关注点为门槛过高,因此列举出了 10 个城市消费券的最高门槛和最高优惠值,可以看出郑州市无论是最高门槛还是平均门槛都位于 10 个城市的首位,而网友抱怨说郑州市部分消费券门槛过高,抢到了也用不着不无道理;所以政府可以根据这一点对政策进行相应的调整。

### 3.文本情感分析

为更直观的了解到民众的主观情绪的极性分布情况,在上文分析的基础上,进行文本情感倾向性分析。首先,提取经过预处理的评论文本中所能体现积极情绪的词作为判断词的极性的正向词典、提取体现消极情绪的词作为判断词的极性的负向词典;利用 SnowNLP 来训练模型并生成新的模型,进而绘制情感分的直方图如下:

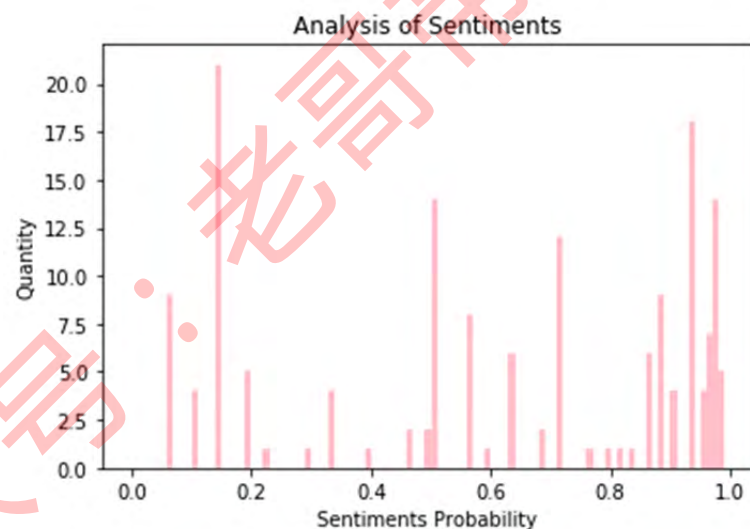


图 7 情感得分图

从图 7 可以看出民众对于消费券政策的态度更偏向积极情绪,因为得分在 0.8-1.0 的评论更为密集且多,而得分在 0-0.3 和 0.3-0.8 之间的评论则比较稀疏并且相对于 0.8-1.0 的评论来说较少,而这也与评论相呼应,就民众对郑州市发放消费券的评论而言,主要体现了三个观点:一、大部分民众认为消费券体现了积



极作用并愿意使用消费券进行消费；二、少部分民众认为对自己影响不大，可能平时消费用不到，持中立态度；三、较少部分民众则认为发放消费券最终还是得自己花钱，而且消费券门槛太高，达不到某些消费数额就不能使用，意义不大。

综上所述，从词云图到词共现网络再到文本情感分析，对民众基于消费券政策的评论进行全面地分析。俗话说：“知政失者在草野。”，从某种程度上反映出网民的评论是检验一个政策好坏的“晴雨表”，他们站在不同的位置，以不同的视角思考问题，从而可以得到针对政策的更为全面的评价。如果网民的评价多数为正面，则说明这个政策可行性比较强，反之，则说明该政策的执行成本比较高。另外，网民的评论是各抒己见的，所以，也可以从他们的评论中找到政策制定时的漏洞和不足，进而做出相应的调整 and 改革。就网民对郑州市发放消费券的评论来看，主要矛盾集中于部分消费券面额与实际消费情况不符，例如在奶茶店消费满 300 才能满减，这个的确不太符合常情，不过根据情感分析可以看出还是有相当一大部分网民支持该政策的实施并积极响应。

## 六、结论与建议

### （一）结论

从建立的线性回归模型结果来看，2020 年 4 月消费券政策的实施使郑州市社会消费品零售总额明显提升，在其他外部条件保持不变的情况下，平均而言，消费券政策的实施会使社会消费品零售总额提高约 55.259 亿元。

根据建立的 ARIMA 模型预测郑州市的社会消费零售总额的结果显示，在没有政府发放消费券政策的支持下，经济回温的过程会比较缓慢，预测模型的结果显示在 2020 年 5、6、7 和 8 月份社会消费品零售总额会呈现先降后缓慢回升的情况。然而，真实值是迅速上升后下降又回升的情形同时也远高于预测值，呈现一种“V”型结构，说明当有政策刺激时就会加大民众们的消费动机，促进消费，

使得经济能在短时间内迅速回升。从中可以看出适当地发放消费券来短期刺激经济是有效的，也是可取的，可以为经济的增长奠定基础。

同时，从微博网民评论分析中可以发现广大网民对于政府发放消费券这一举措持有不同意见，当然绝大多数人还是表现出支持的态度。就所抓取的网友们对郑州市发放消费券这一热点话题的讨论来看，有大量评论是其他地区网友呼吁当地政府“抄作业”。由此可见，政策的可行性较强，民众对这个政策的认可度也比较高，毕竟这是一个惠民的举措。从政府的角度来说，电子消费券的使用会使得国民经济迅速回升；从消费者的角度来看，在消费的时候使用消费券抵消部分金额可以为自己节省一部分支出。短期来看这一政策的实施的确会促进我国国民经济快速回暖，拉动经济增长，且如今数字经济的大环境更是成为政策顺利执行的推手，民众可以从各网络平台轻松获取到各式各样的信息，从而做出决定。当电子消费券成为民众手中的可支配收入时，就像凯恩斯的乘数效应一样，借助大数据时代的网络平台，一单位的消费券带来的将是数倍的消费增长。在大数据的社会背景下，数字消费转化为消费的动能，对消费的拉动效果可见一斑。

当然，网友们所指出的问题也是政府需要着重考虑的，毕竟每项政策的实施对象都是广大的人民群众，大家好才是真的好，因此，本文将根据分析得到的结果给出一些建设性的建议。

## （二）建议

1.消费券政策可以推动消费，所以当经济不景气或者需要振兴经济的时候，政府可以适当实施消费券政策，利用人们的心理特征来带动消费，同时发放的次数不能过于频繁，门槛也不能过高。发放次数频繁会让民众产生依赖心理，门槛过高可能会消磨民众的耐心，容易产生抵触心理，所以“适度”很重要。

2.显然，消费券是一种短期纾困政策，所以政府就更需要根据广大民众所反

馈出的意见和问题对该政策做出及时的调整，这样政策才能更好地实施，经济发展才会以预期的效果得到提升。

3. 提高消费券政策的公平性。首先，可以选取多个支付平台来发放消费券，从而扩大受众群体；其次，可以注重商家参与的丰富性，这样更能方便居民；另外，关注群众中的少数群体，如不会使用智能手机的老年人及一些低收入群体。

4. 注重部分消费券存在的合理性。根据网友们的评论可以看出部分网友有提到奶茶店消费够 300 才能使用消费券，这就存在不合理之处，单人一次在奶茶店消费满 300 的确不太容易，除非团购。那么这种类型的消费券就失去了它存在的意义，所以政府可以针对这些消费券做出相应地调整。根据分析发现，基于郑州市消费券政策的实施，网民主要认为消费券门槛过高，普遍体现出部分消费券抢到手也不会用的情况，而从表 4 中也了解到郑州市消费券的最高门槛、平均门槛均为 10 个城市中最高的，所以政府应该更加科学地设置消费券的面值和满减金额。一方面，过高的门槛可能会使消费者望而却步，另一方面，门槛过高有可能会使居民认为政府在“套路”消费者，反而会适得其反，违背政府当初发放消费券的宗旨。另外，政府可以根据当地居民的收入水平，制定合适的面值，而不是单纯地复制“作业”。

## 参考文献

- [1] 北京大学国家发展研究院课题组, 林毅夫, 沈艳, 孙昂.我国消费券发放的现状效果和展望研究[J].中国经济报告, 2020(04):21-33.
- [2] 周丽凤.基于结构方程模型的体育消费券效用分析——以江苏省部分城市为例[J].体育科学研究, 2020, 24(03):17-24.
- [3] 吴坚平, 张旻, 姜明.基于旅游电子消费券发放的优化策略评价模型研究[J].中国旅游评论, 2020(02):93-101.
- [4] 刘虎, 王艺奇, 许蓉蓉. 基于新浪微博评论数据的消费券政策效果评估分析 [A]中国统计教育学会.2020 年(第七届)全国大学生统计建模大赛优秀论文集[C].中国统计教育学会:中国统计教育学会, 2020:84.
- [5] 吴翌琳, 张旻.基于 PSM-DID 模型的消费券效果评估[J].调研世界, 2021(01):14-24.
- [6] 姜蒙强, 周志太.新冠疫情下数字消费券刺激消费回补的政策效应[J/OL].宜宾学院报:1-7[2021-05-15].<http://kns.cnki.net/kcms/detail/51.1630.Z.20210423.1322.003.html>.
- [7] 林毅夫, 沈艳, 孙昂.中国政府消费券政策的经济效应[J].经济研究, 2020, 55(07):4-20.
- [8] 刘祥. 江苏省体育消费券政策效用评价体系构建与实证研究[D].南京体育学院, 2019.
- [9] 汪侠, 甄峰, 吴小根.旅游消费券效用差异及其成因分析[J].财经问题研究, 2012(08):123-129.
- [10] 孙杰. 消费券对刺激消费的实证研究[D].浙江大学, 2012.
- [11] 李梅志.旅游消费券助推郑州经济发展[J].中国商贸, 2010(19):156-157.
- [12] 郭方, 郭荣朝.旅游消费券效用分析与应用对策——以郑州市为例[J].黑龙江

对外经贸, 2010(09):113-114.

- [13] How Will Consumption Coupons Stimulate Economic Recovery?[J].Beijing  
Review, 2020, 63(25):46-47.

公众号·老哥带你学数模

## 致谢

在这里,首先感谢本次统计建模大赛的主办方提供这次宝贵的机会,让我们挑战自己,将理论知识应用于实际问题。其次由衷地感谢我们队的指导老师在选择与论文写作的过程中给予我们许多意见,不厌其烦地陪我们修改论文,给予我们很多帮助。同时也感谢我们自己能够坚持下来,并顺利完成了报告。最后再次向各位老师表达真诚的感谢!