

队伍编号	203295
题号	D

## 基于统计回归模型的新零售企业精准需求预测研究

### 摘要

本文针对精准需求预测问题，运用了人类消费心理学以及经济学方面的相关知识，构建了目标 skc 的销售量模型、目标小类月销售量模型以及目标小类内 skc 的周销售量模型，综合运用 MATLAB 和 SPSS 等软件编程求解，最后结合实际情况进行了模型改进。

针对问题一，题目要求解释在要求时期内产品销售特征，库存信息，节假日折扣等因素对目标 skc 的销售量影响的问题。对此先是运用临近点均值法和标准化处理对所需的数据进行预处理，然后根据各个因素与销售量的都标准化处理后的散点图态势，运用了统计学理论，构建了目标 skc 的销售量模型，最后使用 MATLAB 软件进行求解，得出了销售特征和实际销售价格（折扣与实际销售价格成正比，打折大，价格高）增加时，销售量先增后减的结论，以及库存量增加，销售量先减少后增加的结论。

针对问题二，题目要求预测给定区域内目标小类在 2019 年 10 月 1 日后 3 个月中每个月的销售量，给出每个月预测值的 MAPE。对此先设立了 skc 对应的小编码 tcc 的函数，并运用问题一的方法对数据进行处理，然后结合求得的累计销售额排名前十的小类建立了目标小类月销售量模型，最后用 MATLAB 软件和 EXCEL 进行求解，得到了 10 月 1 日后三个月的目标小类的销售量以及得到了这三个月预测的 MAPE 值，分别为 0.0752、0.0901 和 0.0975。

针对问题三，我们团队在分析实际生活中各种复杂因素影响消费者消费心理的基础上，考虑到销售特征，库存信息，节假日折扣等因素的交互作用也会对销售量产生巨大的影响，对此我们将对问题一的模型进行改进，构建更为精准的需求预测模型——目标小类内 skc 的周销售量模型，然后用 MATLAB 进行求解，得出了 12 周内每周的周销量和每周预测值的 MAPE，为新零售企业在精准需求预测上提供了合理的参考方案和参考依据。

针对问题四，基于问题二和问题三所建立的模型基础上，写一封推荐信给新零售行业的相关负责人，信件内容包括综合分析我们团队所建立的精准需求预测模型相较于其他文献上所采用的预测模型的优势是什么，说明我们模型在新零售企业的预测效果还有适用范围怎么样以及我们模型的优化方向。

最后，对模型的优缺点进行了分析和评价，并且对模型进行了灵敏度分析、误差分析和推广，其中目标小类月销售量模型和目标小类内 skc 的周销售量模型预测的误差范围分别 0-9%和 0-5%之间，表明模型预测的精准性很高。

**关键词：**精准需求预测；统计回归模型；消费心理；交互作用

# 目录

一、问题的重述 .....	1
1.1 问题的基本背景 .....	1
1.2 要求解决的问题 .....	1
二、问题的分析 .....	1
2.1 研究现状综述 .....	1
2.2 本文的研究思路和步骤 .....	2
2.2.1 对具体问题的分析和对策 .....	3
三、基本假设 .....	3
四、符号说明 .....	4
五、模型一的建立与求解 .....	4
5.1 模型准备 .....	4
5.2 数据处理 .....	5
5.2.1 缺失值处理 .....	5
5.2.2 数据标准化处理 .....	5
5.3 模型一的建立与求解 .....	6
5.3.1 目标 skc 的销售量模型的建立 .....	6
5.3.2 目标 skc 的销售量模型的求解 .....	9
5.3.3 模型结果分析 .....	11
5.3.4 模型误差分析和灵敏度分析 .....	11
六、模型二的建立与求解 .....	12
6.1 目标小类的月销售量模型的建立 .....	12
6.2 目标小类的月销售量模型的求解 .....	13
6.3 模型分析 .....	14
6.3.1 结果分析 .....	14
6.3.2 误差分析 .....	15
6.3.3 灵敏度分析 .....	15
七、模型三的建立与求解 .....	16
7.1 目标小类内 skc 的周销售量模型的改进与建立 .....	16
7.2 目标小类内 skc 的周销售量模型的求解 .....	17
7.3 模型分析 .....	18
7.3.1 模型结果分析 .....	18
7.3.2 误差分析 .....	18
7.3.3 灵敏度分析 .....	18
八、模型评价与推广 .....	19
8.1 模型的评价 .....	19
8.1.1 模型的优点 .....	19
8.1.2 模型的缺点 .....	20
8.2 模型的推广 .....	20
九、给企业的一封信 .....	21
十、参考文献 .....	22

## 一、问题的重述

### 1.1 问题的基本背景

随着我国消费市场的不断发展,我国人民消费水平的不断提高以及商品资源的极大丰富,市场上原有的消费模式已由“以物为主”逐步转变为“以客为主”。多元消费观念的趋势下,人们的消费需求也由原来的考虑实用性转向考虑时尚性,特别是将注意力放在了“个性化、时尚、美观”等方面。为了适应这种新的需求,新零售企业应该根据层级复杂,品类繁多的历史销售数据,以区域层级,小类层级乃至门店 skc(单款单色)层级给出精准的需求预测。

### 1.2 要求解决的问题

**问题一:**从产品销售特征、库存信息、节假日折扣等因素,试分析 2018 年国庆节,双十一,双十二和元旦这四个节假日内各种相关因素对销售量时间处于 2018 年 7 月 1 日至 2018 年 10 月 1 日内且累计销售额排名前 50 的 skc 的销售量的影响。

**问题二:**结合问题一的分析结果,预测给定区域内目标小类(历史销售时间处于 2019 年 6 月 1 日至 2019 年 10 月 1 日内且累计销售额排名前 10 的小类)在 2019 年 10 月 1 日后 3 个月中每个月的销售量,给出每个月预测值的 MAPE。

**问题三:**为了满足企业更加精准的营销需求,试着建立相关数学模型,在考虑小类预测结果的同时,预测目标小类内所有 skc 在 2019 年 10 月 1 日后 12 周内每周的周销量,并给出每周预测值的 MAPE(可以考虑 skc 销售曲线与小类销售曲线之间的差异)。

**问题四:**请给企业写一份推荐信,向企业推荐你的预测结果和方法,并说明你们的方案的合理性以及后续的优化方向。

## 二、问题的分析

### 2.1 研究现状综述

王金旺[1]从供应链的角度审视需求预测,他首先对在供应链环境下的需求预测和传统的需求预测作了相应的比较,明确了传统预测的局限性,然后他对供应链环境下的有关需求预测的影响因素从宏观环境、市场环境和业务环境三个方面进行了相应分析,提出供应链环境下的需求预测的体系模型,以谋求供应链整体绩效最优。

张相斌[2]020 模式下连锁零售企业的运营过程进行分析,探讨该模式下影响零售需求量的因素,以及考虑未来随机因素对需求预测的影响,建立概率排列需求预测模型,得出了以 N 个局部平均值构成的预测期望值区间,并以此来对零售需求量进行预测。

张佳慧[3]从木材首饰商品零售企业需求预测方面,以有限时间长度的历史

销售数据分析而实现精度较高的预测的角度，首先运用 SPSS 软件作出时间序列图判断其平稳性，得出了该序列为平稳序列，然后做出了自相关图和偏相关图，分析偏相关函数函数和自相关函数拖尾或截尾，提出了 ARIMA(1, 1, 1) (2, 1, 1) 模型具有良好的拟合优度，最后通过真实值和拟合值的误差分析得出 ARIMA 模型在短期需求预测上具有一定适用性。

我们都知道目前有大量针对需求预测问题进行研究的文献，但是这些文献中尚且还没有涉及到从产品销售特征，库存信息，节假日折扣等因素对销售量需求的精准预测，对此我们团队将在前人的基础上，对次问题进行相关的完善，提出较为合理的的预测模型和预测方法。

## 2.2 本文的研究思路和步骤

本文主要研究新零售目标产品的精准需求预测问题，针对此问题，我们团队将对此分成 4 个问题进行研究。首先我们将运用附件给出的相关数据经处理后拟合出合适的统计回归模型，然后用拟合出的参数分析销售特征、库存信息和折扣分别对销售量的影响。其次，通过设立 skc 与对应的唯一小类编码的 tcc 之间的函数关系和确定要求时间内排名前十的目标小类后建立目标小类的月销售量模型，然后结合 MAPE 计算公式算出该三个月中每个月预测值的 MAPE。接着，当考虑到要增加需求模型预测的精准性的情况时，我们团队在模型一和模型二的基础上进行改进，充分考虑不同因素的交互作用，得到加交互项的目标 skc 的周销售量模型。最后，根据问题一，问题二，问题三的分析结论，给企业提供精准需求预测的方法，并给予后续优化的方向。

根据本文的研究思路，先给出整体的思路流程图，如图 2-1 所示

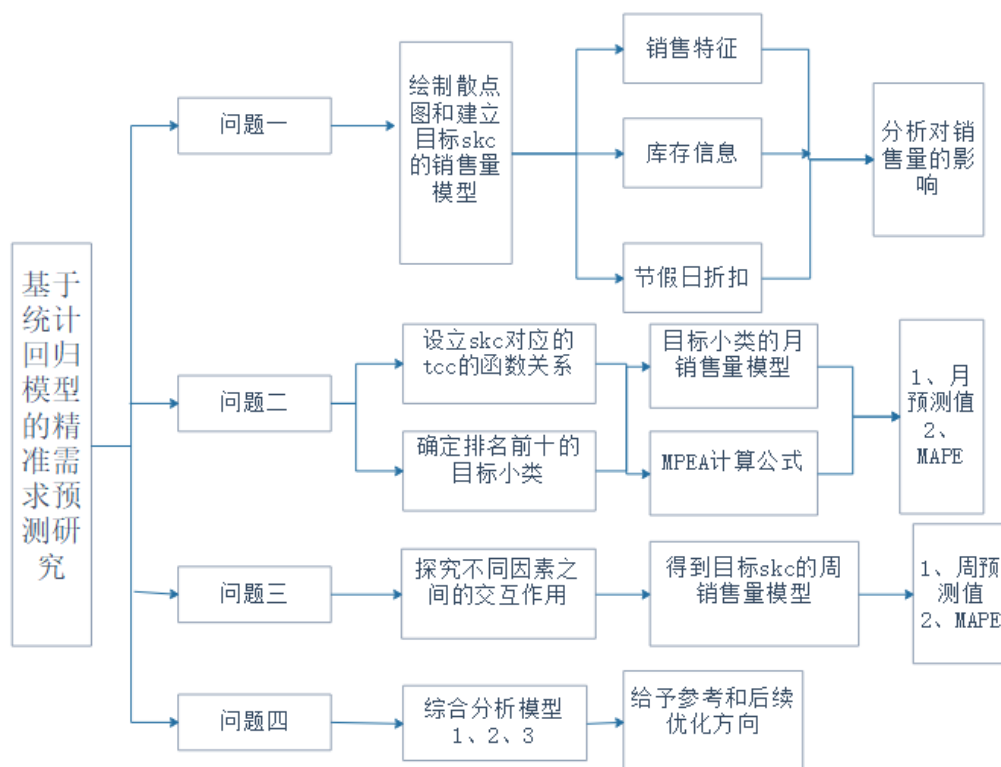


图 2-1：整体思路流程图

## 2.2.1 对具体问题的分析和对策

### (1) 对问题一的分析

问题一要求我们考虑产品销售特征，库存信息（即库存数量），节假日折扣等因素下分析 2018 年国庆节，双十一，双十二和元旦这四个节假日内各种相关因素对销售量时间处于 2018 年 7 月 1 日至 2018 年 10 月 1 日内且累计销售额排名前 50 的 skc(即目标 skc)的销售量的影响。对此问题，首先我们先运用 MATLAB 软件确定目标 skc，并分别对 2018 年的该目标 skc 的销售量与各影响因素之间分别进行统计回归分析。即对销售量分别与销售特征、库存信息和折扣进行统计回归分析。运用 MATLAB 软件分别进行初步拟合并绘图观察分析，结合起来构建出最为合适的统计回归模型，然后运用附件中所给的相关数据进行拟合，并对拟合结果和所得参数进行分析。

### (2) 对问题二的分析

问题二要求我们结合问题一的分析结果，预测给定区域内目标小类（历史销售时间处于 2019 年 6 月 1 日至 2019 年 10 月 1 日内且累计销售额排名前 10 的小类）在 2019 年 10 月 1 日后 3 个月中每个月的销售量，给出每个月预测值的 MAPE。对此，我们首先设立 skc 对应的小编码 tcc 的函数，然后结合问题一分析并在要求时间内按累计销售额大小确定排名前十的目标小类，确定了目标小类后建立目标小类的月销售量模型，最后结合 MAPE 计算公式算出该三个月中每个月预测值的 MAPE。

### (3) 对问题三的分析

问题三要求我们试着建立相关数学模型，在考虑小类预测结果的同时，预测目标小类内所有 skc 在 2019 年 10 月 1 日后 12 周内每周的周销量，并给出每周预测值的 MPEA(可以考虑 skc 销售曲线与小类销售曲线之间的差异)。对于此问题，我们主要通过探究不同因素之间是否具有交互作用，充分模拟现实生活中不同因素相互作用对消费者消费心理的影响，从而反映到对销售量的影响上。所以若有交互作用，则在模型一和模型二的基础上对模型进行改进，增加需求预测的精准性，从而建立起目标 skc 的周销售量模型，然后结合 MAPE 计算公式算出每周预测值的 MAPE。若无交互作用，则重新建立新的精准需求预测模型进行预测。

### (4) 对问题四的分析

问题四要求我们向企业推荐我们的预测结果和方法，并说明我们的方案的合理性以及后续的优化方向。对于此问题，我们主要根据问题一得到的统计回归模型、问题二得到的目标小类的月销售量模型和问题三得到的目标 skc 的周销售量模型进行综合分析，给企业提供精准需求预测的参考，并给出后续优化的方向。

## 三、基本假设

- (1) 假设打折扣前的实际花费即按标签价格销售
- (2) 对同一 skc，假设其销售量内的每一件的实际销售价格都一样

- (3) 假设时间是离散的，这样有利于计算简便
- (4) 假设计算的误差在合理范围内，对数据结果的影响可以忽略
- (5) 假设除主要影响因素的其他影响因素对销售量的影响忽略不计
- (6) 假设随机变量  $\varepsilon$  满足均值为 0 的正太分布

## 四、符号说明

序号	符号	符号说明
1	$z_{skc}$	skc 在 $T_1$ 内的累计销售额
2	$z'_{tcc}$	tcc 在 $T_2$ 内的累计销售额
3	$s_{skc,t}$	$t$ 时, skc 的日销售量
4	$s_{tcc,t}$	$t$ 时, tcc 的日销售量
5	$n_{skc}$	skc 的销售量模型公式的最高次数
6	$rp(sk, t)$	$t$ 时, skc 的日实际销售价格
7	$rc(sk, t)$	$t$ 时, skc 的日实际花费
8	$sc(sk, t)$	$t$ 时, skc 的日销售特征
9	$ie(sk, t)$	$t$ 时, skc 的日库存数量
10	$dc(sk, t)$	$t$ 时, skc 的日折扣
11	$tp(sk, t)$	$t$ 时, skc 的日标签价
12	$S_{tcc,T}$	$T$ 时, tcc 的月销售量
13	$S_{skc,\tilde{T}}$	$\tilde{T}$ 时, skc 的周销售量
14	$S_{tcc,\tilde{T}}$	$\tilde{T}$ 时, tcc 的周销售量

注：剩余部分未标注的符号将在文中进行进一步说明。

## 五、模型一的建立与求解

### 5.1 模型准备

**统计模型的概念：**“统计模型”是描述随机变量随其他随机变量或随机过程或一般变量变化而变化的依赖关系的表达式或“方程式”或“模型”。[4]

统计回归模型与一般回归模型的不同之处在于统计模型具有更强的灵活性，形式更容易确定，而且可以考虑不同自变量之间是否存在交互作用，这更加能体现现实生活中问题的复杂性。而对于黑箱系统目标 skc 的销售量与各影响因素之间的大致公式我们是不知道的，只能通过回归近似拟合估计，然而黑箱问题中，

统计回归模型也比一般回归模型更加精准且灵活，因此在这里，我们团队决定选择运用统计回归模型对销售量的需求进行预测。

## 5.2 数据处理

### 5.2.1 缺失值处理

运用 SPSS 软件，对附件的数据进行分析，并且对销售量、实际花费和库存量缺失的数据进行缺失值处理。其中在 SPSS 中现有的方法中对缺失值的替换主要 5 种：

#### (1) 序列均值

序列均值就是使用该样本中有效观察到的所有观察值的平均数来代替缺失部分的数值，从而使得进一步计算得以展开。具体用法是如果缺失值是数值型的，就根据该变量在其他所有对象的取值的平均值来填充该缺失的变量值；如果缺失值是非数值型的，就根据统计学中的众数原理，用该变量在其他所有对象的取值次数最多的值来补齐该缺失的变量值。[5]

#### (2) 临近点均值

临近点均值就是利用该缺失值前后的  $n$  个观察值做平均数，以代替该缺失的数值，这种方法一般适用于缺失值较少以及数据同质性较大，变化不太大的数据中

#### (3) 临近点的中位数

临近点的中位数就是利用有效观察到的  $n$  个数值的的中位数来代替缺失的数值，这种方法首先要对数据进行排序，与位置排列有关。

#### (4) 线性插值法

线性插值法就是首先利用已知数据拟合出相应的函数关系式，简单来说就是用过 A 点和 B 点的直线来近似表示原函数，然后通过回代得到缺失值。

#### (5) 点处的线性趋势

点处的线性趋势就是以编号为自变量，用线性回归的方法，先拟合出函数表达式，然后将缺失表编号带入回归方程中得到预测值来代替缺失值。

基于上述几种数据缺失值的处理方法，我们团队主要采用的是临近点均值这种替代方法，因为附件中销售量，实际花费和库存量的缺失值不多，用五种方法替代的缺失值无差别，因此选哪一种都可以。

### 5.2.2 数据标准化处理

在实际问题中，不同变量的测量单位往往是不一样的，为消除变量的量纲效应，常对数据进行标准化处理。在此次我们团队对销售量、销售特征，库存信息（即库存数量），节假日折扣都进行了标准化处理，为的是消除数据在接下来的模型建立中不同的单位对模型的准确性造成影响。

不同于假设已有长度为  $m$  的原始数据序列  $\{X_k\}$ ，标准化处理如下

$$X_k^* = \frac{X_k - \frac{1}{m} \sum_{i=1}^m X_i}{\sqrt{\frac{1}{m-1} \sum_{j=1}^m (X_j - \frac{1}{m} \sum_{i=1}^m X_i)^2}}, k=1,2,\dots,m \quad (5-1)$$

即得到标准化处理后长度也为  $m$  的新序列  $\{X_k^*\}$ ，为方便符号说明，以下变量若带“\*”即表示作了标准化处理。

## 5.3 模型一的建立与求解

### 5.3.1 目标 skc 的销售量模型的建立

我们仅考虑题意中影响目标 skc 的销售量的主要因素为销售特征、库存信息（即库存数量）和折扣，而目标 skc 为销售时间处于 2018 年 7 月 1 日至 2018 年 10 月 1 日内且累计销售额排名前 50 的 skc，skc 即单款单色产品编码，其中该累计销售额即

$$z_{skc} = \sum_{t \in T_1} r c(sk, t) \quad (5-2)$$

其中  $T_1$  表示 2018 年 7 月 1 日至 2018 年 10 月 1 日内这段的时间集合，接着，即对附件 1 相应数据整理并借助 EXCEL 软件即可完成上述排序，整理得如下表

表 5-1：目标 skc

r	skc(r)	$z_{skc(r)}$	r	skc(r)	$z_{skc(r)}$
1	796573650638	211860.00	26	390572118052	55587.50
2	396573870660	175782.50	27	696572117667	55572.50
3	690572114917	143528.75	28	402573650792	53596.25
4	996573870572	120558.75	29	502573650759	53073.75
5	996572333025	114653.75	30	608572774763	52331.25
6	802573650495	97361.25	31	296572551749	50805.00
7	496572333586	85876.25	32	102572118283	49938.75
8	302572775577	84007.50	33	690572118833	49305.00
9	408572660891	83631.25	34	496573321265	49287.50
10	499572333179	80830.00	35	596572118162	48653.75
11	196573870759	75267.50	36	802573870715	48628.75
12	100572118316	70478.75	37	496572118514	48467.50
13	102573650748	67987.50	38	702573650594	48291.25
14	902573870176	66261.25	39	696572115379	47846.25
15	490572114587	64887.50	40	102573321111	47628.75
16	902572118613	64520.00	41	396573870484	47262.50
17	896572118954	63412.50	42	296572114433	46665.00
18	496572117700	62620.00	43	596573650847	46583.75
19	596572118723	62505.00	44	702573870715	44631.25
20	902573320275	62043.75	45	702572225797	44328.75
21	433572114917	61551.25	46	696573650781	44182.50



22	896573870660	59062.50	47	396572118745	44033.75
23	602572551672	58691.25	48	802573650803	43652.50
24	391572223674	57077.50	49	702573870858	43081.25
25	890572114477	55922.50	50	790572118833	42422.50

不失一般性，随机取样取排名为 $r=20$ 的  $skc$  的销售量与各影响因素之间分别进行统计回归分析。对销售量与销售特征。销售特征即以排名为 $r=20$ 的  $skc$  的销售量占排名前 50 的  $skc$  的总销售量的比例为代表近似估计，即有

$$sc(sk(r),t) \approx \frac{s_{sk(r),t}}{\sum_{i=1}^{50} s_{sk(i),t}} \quad (5-3)$$

即对附件 1 相应数据整理并借助 MATLAB 软件作如上式的数据处理与标准化处理。由于该问题属于黑箱问题，因此，这里我们借助 MATLAB 软件的 cftool 工具箱以 Polynomial 类函数对标准化后的销售量与销售特征进行适宜的拟合，得出销售量与销售特征的线性拟合函数，其对应的关系散点图及拟合曲线如下

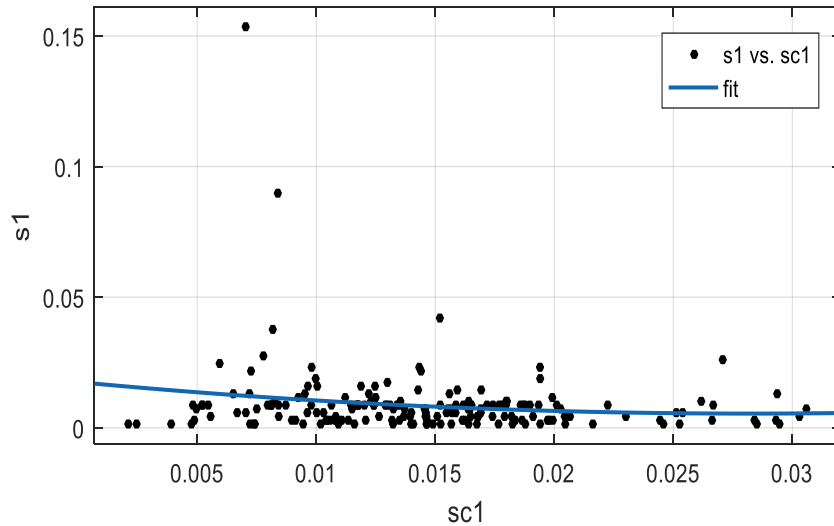


图 5-1:  $skc(20)$  的  $s_{sk(20),t}^*$  与  $sc^*(sk(20),t)$  的关系散点图（附拟合曲线）

从图中，不难发现都标准化处理后的销售量与销售特征是呈开口向上的抛物线趋势，即当销售特征不断增大时，销售量先是随着销售特征增大而单调递减，当销售特征达到一定值时，销售量达到最小值，然后随着销售特征再次增加，销售量开始呈单调递增趋势。因此该拟合曲线方程形式如下

$$s_{sk(20),t}^* = \alpha_0 + \alpha_1 \cdot sc^*(sk(r),t) + \alpha_2 \cdot [sc^*(sk(r),t)]^2 + \varepsilon \quad (5-4)$$

其中  $\varepsilon$  为随机误差， $\alpha$  即参数。同理，对销售量与其他两种主要因素，对应的，都标准化处理后的销售量与库存信息的关系散点图（附拟合曲线）如下

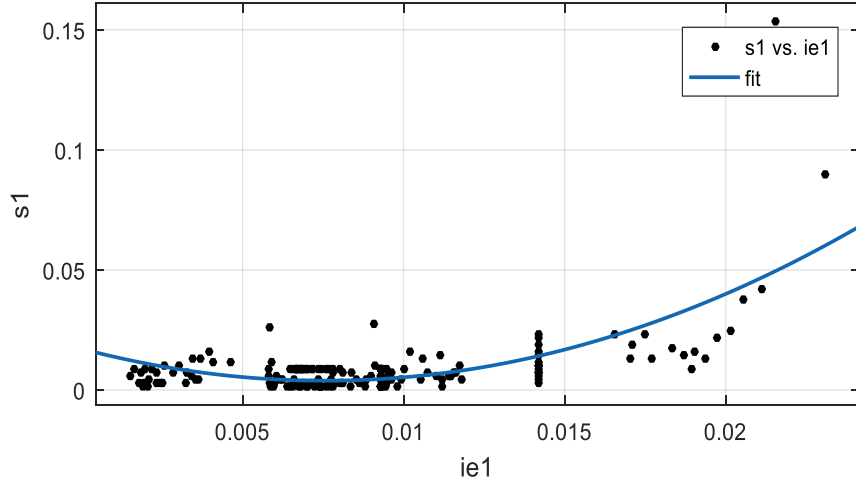


图 5-2:  $skc(20)$  的  $s_{skc(20),t}^*$  与  $ie^*(skc(20),t)$  的关系散点图（附拟合曲线）

显然，从图中不难发现都标准化处理后的销售量和库存信息的散点图是呈现出抛物线的态势的。当库存信息增加时，销售量先是呈现出单调递减的趋势；当库存信息达到一定值时，销售量达到最小值；当库存信息继续增加时，销售量呈现出单调递增的趋势。因此都标准化处理后的销售量与库存信息对应的拟合曲线方程形式如下

$$s_{skc(20),t}^* = \alpha_0 + \alpha_1 \cdot ie^*(skc(r),t) + \alpha_2 \cdot [ie^*(skc(r),t)]^2 + \varepsilon \quad (5-5)$$

由于附件 2 的标签价对应的时间单位是年，因此，对问题一要求的 2018 年的研究分析，即可对标签价视为与小于年的时间单位无关的常数，即折扣与实际销售价格成正比

$$dc(sk(r),t) = \frac{10}{tp(sk(r),t)} \cdot rp(sk(r),t) \quad (5-6)$$

其中实际销售价格与实际花费有如下关系

$$rp(sk(r),t) = \frac{rc(sk(r),t)}{s_{sk(r),t}} \quad (5-7)$$

因此，对销售量与折扣的统计回归分析等价于对销售量与实际销售价格的统计回归分析，即可作都标准化处理后的销售量与实际销售价格的关系散点图（附拟合曲线）如下

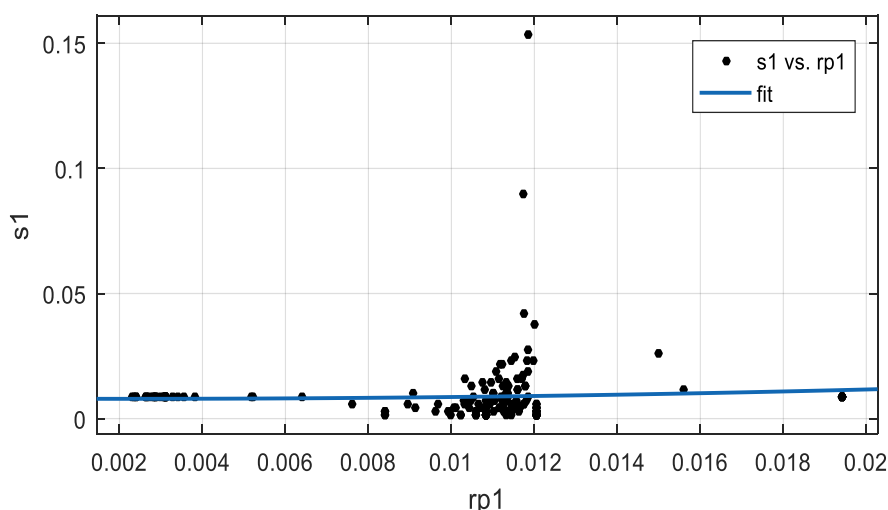


图 5-3:  $skc(20)$  的  $s_{skc(20),t}^*$  与  $rp^*(skc(20),t)$  的关系散点图（附拟合曲线）

从图中不难发现从都标准化处理后的销售量和实际销售价格的散点图是呈现出抛物线的态势的。当实际销售价格增加时，销售量先是呈现出增加的趋势；实际销售价格达到一定值时，销售量达到最大值；实际销售价格继续增加时，销售量呈现出减少的趋势。因此其对应的拟合曲线方程形式如下

$$s_{skc(20),t}^* = \alpha_0 + \alpha_1 \cdot rp^*(skc(r),t) + \alpha_2 \cdot [rp^*(skc(r),t)]^2 + \varepsilon \quad (5-8)$$

综上所述，在假设的前提下建立目标 skc 的销售量模型如下

$$s_{skc(r),t}^* = \alpha_0 + \alpha_1 \cdot sc^*(skc(r),t) + \alpha_2 \cdot ie^*(skc(r),t) + \alpha_3 \cdot rp^*(skc(r),t) + \alpha_4 \cdot [sc^*(skc(r),t)]^2 + \alpha_5 \cdot [ie^*(skc(r),t)]^2 + \alpha_6 \cdot [rp^*(skc(r),t)]^2 + \varepsilon \quad (5-9)$$

### 5.3.2 目标 skc 的销售量模型的求解

这里，我们借助 MATLAB 软件结合附件以 2018 年的相应数据即可求解式 (5-9) 的参数估计值，并代入式 (5-9) 得排名为  $r=20$  的 skc 的销售量模型公式

$$s_{skc(r),t}^* = 0.0332 - 0.7519sc^*(skc(r),t) - 3.6880ie^*(skc(r),t) - 2.0361rp^*(skc(r),t) + 21.2869\alpha_4 \cdot [sc^*(skc(r),t)]^2 + 245.4926[ie^*(skc(r),t)]^2 + 91.0263[rp^*(skc(r),t)]^2 + \varepsilon \quad (5-10)$$

同理也可求得排名前 50 的其他排名 skc 的销售量模型公式，整理成模型参数估计值表如下

表 5-2:  $skc(r \in [1,50])$  的销售量模型参数估计值表

r	1	2	3	4	5	6	7	8	9	10
$\hat{\alpha}_0$	0.0116	0.0122	-0.0083	0.0099	0.0483	0.0132	-0.0084	0.1138	-0.0285	0.0128
$\hat{\alpha}_1$	0.1512	0.1352	0.1646	1.2695	0.0318	0.6710	0.6778	0.0566	0.1197	0.1979
$\hat{\alpha}_2$	-1.1347	-2.7540	0.2577	-7.2061	-5.4850	0.0424	4.1052	-0.8868	-0.1478	0.0447
$\hat{\alpha}_3$	-0.1738	7.2558	5.4021	6.5099	-1.3270	-2.6049	0.2325	-33.1299	20.5581	-2.4030
$\hat{\alpha}_4$	-8.2760	-3.2143	-9.9315	-98.3490	-7.5415	-41.5211	-45.6135	4.8797	-8.4189	-13.5967

$\hat{\alpha}_5$	105.5591	113.1617	9.1266	282.7649	223.7769	60.9865	-292.2315	67.4588	22.5657	85.1034
$\hat{\alpha}_6$	-4.3399	-995.3011	-527.2024	-213.8575	59.5359	141.0767	-23.7922	2536.9903	-2773.8435	127.2261
<b>r</b>	<b>11</b>	<b>12</b>	<b>13</b>	<b>14</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>18</b>	<b>19</b>	<b>20</b>
$\hat{\alpha}_0$	-0.0133	-0.0089	0.0074	0.0088	0.0105	0.0035	-0.0355	-0.0290	-0.0419	0.0332
$\hat{\alpha}_1$	0.1950	0.1687	0.4199	0.8889	-0.1209	0.8883	0.0174	0.9885	-0.5392	-0.7519
$\hat{\alpha}_2$	4.6213	-1.1209	-0.2014	-1.8632	1.7955	1.9264	0.8066	3.6149	4.8321	-3.6880
$\hat{\alpha}_3$	-1.4122	7.0282	-0.3896	1.1680	-3.1707	-1.1840	13.3149	5.3554	-2.8075	-2.0361
$\hat{\alpha}_4$	0.7885	-4.5022	-10.3857	-47.3636	0.4534	-50.8955	1.9391	-49.1882	27.3465	21.2869
$\hat{\alpha}_5$	-172.1934	53.8480	19.1019	80.6007	-38.5847	-65.9527	-19.9914	-246.2596	-55.3934	245.4926
$\hat{\alpha}_6$	47.5409	-498.8723	3.5290	-45.3075	232.6243	32.6505	-1133.6269	-303.5564	148.0407	91.0263
<b>r</b>	<b>21</b>	<b>22</b>	<b>23</b>	<b>24</b>	<b>25</b>	<b>26</b>	<b>27</b>	<b>28</b>	<b>29</b>	<b>30</b>
$\hat{\alpha}_0$	0.0342	-0.0727	-0.0798	0.0133	0.0608	-0.1718	-0.0065	0.0287	-0.0107	0.0736
$\hat{\alpha}_1$	0.2063	0.7143	0.2681	-0.3371	0.9483	0.3446	-0.1510	0.1339	-0.0867	0.6034
$\hat{\alpha}_2$	-2.4766	-0.9021	3.0687	-0.8802	-6.0840	-0.5802	5.5961	-4.1743	-3.1831	-7.2359
$\hat{\alpha}_3$	-3.8893	25.6000	22.2077	0.2224	0.4111	94.6175	-3.5416	-1.9957	7.4914	-7.6301
$\hat{\alpha}_4$	-14.8948	-29.6376	-7.5992	-2.5904	-71.6507	-23.6603	-3.5747	-9.7970	1.6403	-19.0676
$\hat{\alpha}_5$	196.4971	30.9266	-90.6103	78.2677	158.5675	15.0792	-264.8839	254.1903	210.6894	265.6146
$\hat{\alpha}_6$	186.7609	-1899.5583	-1836.7584	-29.0712	-15.8844	-11645.1883	282.4693	151.4615	-470.3137	599.0481
<b>r</b>	<b>31</b>	<b>32</b>	<b>33</b>	<b>34</b>	<b>35</b>	<b>36</b>	<b>37</b>	<b>38</b>	<b>39</b>	<b>40</b>
$\hat{\alpha}_0$	0.0062	0.0228	0.1948	0.0197	-0.0437	0.0043	0.0316	0.0022	0.0091	-0.0236
$\hat{\alpha}_1$	0.2943	-0.0604	-1.2070	0.3912	-0.1147	0.3540	0.1113	0.4147	0.3162	0.3779
$\hat{\alpha}_2$	-3.5846	-2.2824	0.4531	-4.6355	1.7913	-2.5334	-5.9527	2.3792	0.0273	2.3784
$\hat{\alpha}_3$	1.6184	-3.0058	-23.4822	-0.0163	12.1862	4.8052	-0.3879	-2.0865	-1.0214	6.9842
$\hat{\alpha}_4$	-14.4675	-5.0279	51.5040	-21.3885	10.9375	-8.0562	-6.0792	-26.9180	-13.7409	-33.1016
$\hat{\alpha}_5$	398.3257	141.8917	-4.2320	415.3916	-35.3790	168.9041	412.9743	-68.6236	69.1194	-112.7410
$\hat{\alpha}_6$	-108.3516	320.9470	775.5676	-28.2876	-1034.8470	-466.7825	-1.4563	105.5325	22.2032	-628.1740
<b>r</b>	<b>41</b>	<b>42</b>	<b>43</b>	<b>44</b>	<b>45</b>	<b>46</b>	<b>47</b>	<b>48</b>	<b>49</b>	<b>50</b>
$\hat{\alpha}_0$	0.0120	0.0007	0.0213	0.0044	0.0120	0.0091	-0.0415	0.0110	-0.0011	-0.0117
$\hat{\alpha}_1$	0.4218	0.2694	0.1128	-0.2538	0.2956	0.4842	0.0986	0.8802	0.4777	0.7562
$\hat{\alpha}_2$	0.4717	-3.0875	-3.1240	0.0030	-0.9853	0.3407	2.3098	-0.4689	1.8130	0.9267
$\hat{\alpha}_3$	-4.2218	4.2316	-0.6896	0.9643	-0.6954	-1.2621	15.3571	-2.6673	0.5223	4.1124
$\hat{\alpha}_4$	-18.0922	-11.4697	-9.7452	4.9263	-25.3439	-24.9369	-2.5400	-41.8292	-25.5340	-36.0780
$\hat{\alpha}_5$	-18.5122	241.2828	281.2514	48.7572	118.7331	7.4333	-86.0837	29.7672	-164.9932	-69.9898
$\hat{\alpha}_6$	409.6168	-261.3718	14.8348	-112.5020	19.6909	55.2402	-1407.5692	218.7751	-34.3851	-273.4064

5.3.3 模型结果分析

从表 5-2 可以看出， $sc^*(skc(20),t)$  的二次系数百分之八十左右都是负的，这表明单独且都标准化处理后的销售特征与销售量拟合出来的函数应该是一条开口向下的抛物线，即销售特征增加时，销售量是呈现出先增加后减少的趋势，这就要求零售企业在在这四个时期内生产产品时应该适当注重产品的时尚、个性化等方面的销售特征，但是消费者时尚的定义会发生变化，消费者偏好也会发生改变，因此零售企业也应该要提高对时尚的风向标的预知能力。

$ie^*(skc(20),t)$  的二次系数有大概百分之九十左右是正的，这表明单独且都标准化处理后的库存量与销售量拟合出来的函数应该是一条开口向上的抛物线，即库存量增加时，销售量呈现出先减少后增加的趋势。所以库存和销售量存在着单调两面性，但是基于这四个全民狂欢的节假日，零售企业可以适当增加库存量。

$rp^*(skc(20),t)$  的二次系数总体上看是负数，这表明单独且都标准化处理后的实际销售价格与销售量拟合出来的函数应该是一条开口向下的抛物线，即实际销售价格增加时，销售量是呈现出先增加后减少的趋势。这是因为实际销售价格增加到某一特定值时，销售量达到最大，因此要降低价格才能使销售量增加。这就要求零售企业在题目要求的四个节假日时期内，加大对商品的优惠力度，形成薄利多销的格局。

又因从表格中可以看出  $rp^*(skc(20),t)$  二次系数的绝对值是最大的所以零售企业在分析 2018 年国庆节，双十一，双十二和元旦这四个节假日内目标 skc 的销售量受折扣的影响因素最大，因此零售企业应该在协调好销售特征、库存信息对销售量影响的同时应该加大优惠力度，形成薄利多销的格局。

5.3.4 模型误差分析和灵敏度分析

不失一般性，对排名为  $r=20$  的 skc 的销售量模型的参系数估计值进行误差与灵敏度分析，对此，我们借助 MATLAB 软件在置信水平即犯错误的可能性上限为 5%的前提下求解各参系数的置信区间与及该模型的拟合度如下表

表 5-3: 参系数的统计量表

参系数	参系数估计值	置信区间	拟合度
$\alpha_0$	0.0332	[0.0217, 0.0447]	$R^2 = 0.5004 > 0.5$
$\alpha_1$	-0.7519	[-1.7633, 0.2594]	
$\alpha_2$	-3.6880	[-4.8964, -2.4795]	
$\alpha_3$	-2.0361	[-3.3266, -0.7456]	
$\alpha_4$	21.2869	[-9.0609, 51.6347]	
$\alpha_5$	245.4926	[191.4604, 299.5247]	
$\alpha_6$	91.0263	[26.0084, 156.0443]	

可见拟合度稍微较好，即该类 skc 的销售量仅有超过 50%可由模型决定，其对应的拟合曲线对比图（附上下限置信区间）如下

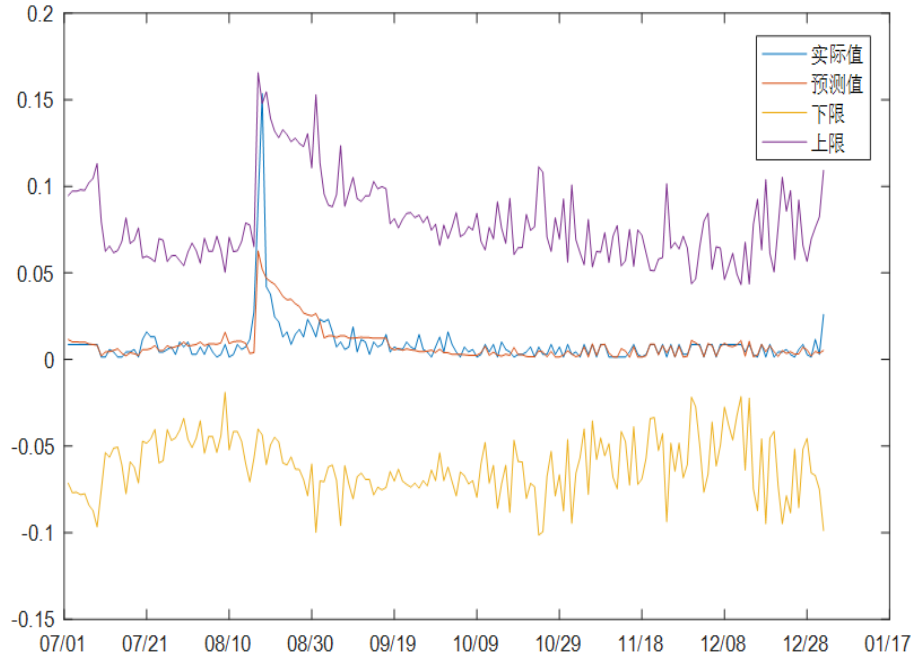


图 5-4:  $skc(20)$  的  $s_{skc(20),t}^*$  拟合曲线对比图 (附上下限置信区间)

由图可见, 初期拟合效果并不佳, 而后期的拟合效果可观。同时, 对参数的置信区间, 即参数估计值的大小都在置信区间范围内, 在 95% 置信度的保证下可认为该模型是稳定的, 但又存在特别是销售特征因素对应的参数置信区间都包含 0 点, 显然, 销售特征因素可有可无, 这与实际并不相符, 可见模型还不够精准, 还需有待改进, 对此, 在接下来的问题三中会对该模型改进进一步讨论。

## 六、模型二的建立与求解

### 6.1 目标小类的月销售量模型的建立

首先设函数

$$tcc = f(skc) \quad (6-11)$$

即表示  $skc$  对应得唯一的小类编码  $tcc$ , 即  $skc$  到  $tcc$  的映射。

结合问题一的分析, 目标小类即历史销售时间处于 2019 年 6 月 1 日至 2019 年 10 月 1 日内且累计销售额排名前 10 的小类, 其中该累计销售额为

$$z'_{tcc} = \sum_{skc \in \{f^{-1}(tcc)\}} z_{skc} = \sum_{skc \in \{f^{-1}(tcc)\}} \sum_{t \in T_2} rc(skc, t) \quad (6-12)$$

其中  $\{f^{-1}(tcc)\}$  表示满足  $tcc = f(skc)$  的所有  $skc$  组成的集合,  $T_2$  表示 2019 年 6 月 1 日至 2019 年 10 月 1 日内这段时间的集合。也可借助 EXCEL 软件完成上述排序, 整理得如下表

表 6-1：目标小类

r	tcc(r)	$z'_{tcc(r)}$
1	27050401	3400685
2	27217089	1181793.8
3	27196225	919157.5
4	27164944	877360
5	27060804	873528.75
6	27206656	391203.75
7	27112849	328790
8	27092025	303085
9	27071209	224920
10	27102436	165153.75

即确定了目标小类，同理，即可建立目标小类的日销售量模型如下

$$S_{tcc(r),t} = \sum_{skc \in \{f^{-1}(tcc(r))\}} S_{skc,t} \quad (6-3)$$

其中  $S_{skc,t}$  与目标 skc 的销售量模型原理类似，即

$$\left\{ \begin{array}{l} S_{skc,t}^* = \mathbf{a} \cdot \mathbf{F}(sc^*(skc,t), ie^*(skc,t), rp^*(skc,t)) \\ sc(skc,t) \approx \frac{S_{skc,t}}{\sum_{i \in \{f^{-1}(tcc(r \in [1,10]))\}} S_{i,t}} \\ rp(skc,t) = \frac{rc(skc,t)}{S_{skc,t}} \end{array} \right. \quad (6-4)$$

其中  $\mathbf{a}$  即参数组成的向量， $\mathbf{F}(sc^*(skc,t), ie^*(skc,t), rp^*(skc,t))$  即由参数组成的向量，该分量的最高次数不超过  $n_{skc}$ 。因此，即可建立能预测给定区域内目标小类在 2019 年 10 月 1 日后 3 个月中每个月的销售量的目标小类的月销售量模型

$$S_{tcc(r),T} = \sum_{t \in \{T\}} S_{tcc(r),t} \quad (6-5)$$

其中  $\{T\}$  表示  $T$  这个月内的时间集合。接着，结合 MAPE 计算公式有如下

$$MAPE(T) = \sum_{r \in [1,10]} \frac{APE(tcc(r),T)}{10} = \sum_{r \in [1,10]} \frac{|S_{tcc(r),T} - \hat{S}_{tcc(r),T}|}{10S_{tcc(r),T}} \quad (6-6)$$

其中  $\hat{S}_{tcc(r),T}$  为  $S_{tcc(r),T}$  的估计值。也即可得该 3 个月中每个月预测值的 MAPE。

## 6.2 目标小类的月销售量模型的求解

目标小类的月销售量模型的求解重点在于求解式 (6-4) 的参数组估计值，与目标 skc 的销售量模型求解原理类似，结合附件的相应数据借助 MATLAB 软件即

可求解出参数估计值，也即可预测给定区域内在 2019 年 10 月 1 日后 3 个月中每个月目标小类对应的所有 skc 的总月内累计销售量，即目标小类的月销售量，整理得如下表

表 6-2：2019 年 10 月 1 日后 3 个月的目标小类的月销售量预测值表

$r$	$\hat{S}_{tcc(r),2019.10}$	$\hat{S}_{tcc(r),2019.11}$	$\hat{S}_{tcc(r),2019.12}$
1	10273.6087	10276.2114	8246.2283
2	1322.3977	1167.9652	1086.0969
3	1392.7575	996.3443	993.9275
4	2559.2138	2367.7201	2685.2136
5	3129.3537	3024.3662	3307.1700
6	275.3421	162.4939	186.6564
7	1383.9809	1816.6332	1892.5086
8	340.6589	243.4683	365.1413
9	298.6411	177.8755	193.2468
10	268.4177	143.4536	179.9067

与及这 3 个月的 MAPE，整理得如下表

表 6-3：2019 年 10 月 1 日后 3 个月的预测值的 MAPE 表

$T$	2019.10	2019.11	2019.12
MAPE	0.0752	0.0901	0.0975

### 6.3 模型分析

#### 6.3.1 结果分析

由表 6-2 的 2019 年 10 月 1 日后 3 个月的目标小类的月销售量预测值可以看出来，结果并非我们认为想象——这十种小类的月销售量会逐月增加，而结果恰恰是销售量逐月减少，但是差距不太大。造成这种局面的原因有两个：

(1) 因为这三个节日靠的比较近，很多商家在节日前期就开始局部降价，进行节日预热活动，而在这三个节日前是很少有优惠活动的，甚至商家还会为了节假日热卖效果进行涨价，所以在这三个节假日之首——国庆节，消费者就开始报复性消费了，这导致国庆时期销量指数性上涨。虽然大部分消费者在国庆节期间的消费欲得到满足，但是双十一优惠力度实在太具有诱惑力，这导致大部分消费者还是会冲着价格便宜进行消费，不可否认的是确实会存在一部分理性人抵抗了价格便宜的诱惑，这就是为什么双十一总体销量会比国庆有所下降。双十二与双十一比较也同理可得。

(2) 当然，造成这种局面的最重要因素是相比于其他两个节日而言，国庆节是我国普天同庆的日子，不管是线上还是线下的商家都开展极大的优惠活动。而大部分国民都有七天小长假，在这七天里，他们有足够的进行消费，这种消费不局限于线上消费，而是线上线下同步进行，新零售企业在这期间就能充分体现线上线下一体化的极大优势，实现销售量的最大化。所以这就是国庆节比其双十一双十二销售量高一点的主要原因。



### 6.3.2 误差分析

由表 6-3 可以看出来,这三个月的 MAPE 值分别是 0.0752、0.0901 和 0.0975,三个值都不超过 0.1,而由出题组给的数据来看,MAPE 的计算过程实际上就是这三个月销售量预测值与真实值的平均绝对百分比误差,所以可以用这三个值作为误差。从数值大小上看,这三个数都不超过 10%,这就说明真实值与预测值之间的误差很小,也说明了模型在需求预测上的精准性很高。

### 6.3.3 灵敏度分析

与问题一的灵敏度分析同理,不失一般性,随机取样取排名为  $r=1$  的目标小类对应的  $skc=102572121792$ ,对此,我们借助 MATLAB 软件在置信水平即犯错误的可能性上限为 5%的前提下求解各参数的置信区间与及该模型的拟合度如下表

表 6-4: 参数的统计量表

参数	参数估计值	置信区间	拟合度
$\alpha_0$	-1.5667	$[-2.9344, -0.1989]$	$R^2 = 0.8673 > 0.5$
$\alpha_1$	0.0694	$[0.0166, 0.1223]$	
$\alpha_2$	0.0223	$[0.0181, 0.0264]$	
$\alpha_3$	-0.0127	$[-0.0184, -0.0070]$	
$\alpha_4$	-0.000652	$[-0.0011261, -0.0001782]$	
$\alpha_5$	-0.000015	$[-0.0000212, -0.0000090]$	
$\alpha_6$	-0.000006	$[-0.0000185, -0.0000070]$	

其他数据见支撑材料。由上表可见拟合度非常良好,即该类  $skc$  的销售量仅有超过 86.7%可由模型决定,其对应的拟合曲线对比图(附上下限置信区间)如下

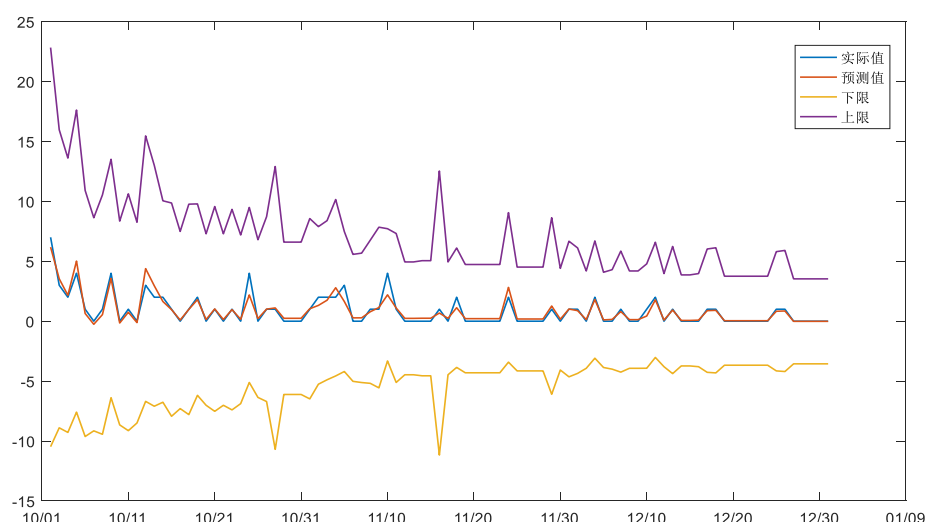


图 6-1:  $skc=102572121792$  的  $s_{skc,t}$  拟合曲线对比图

由图可见拟合效果可佳。对参数的置信区间,即参数估计值的大小都在置信区间范围内,在 95%置信度的保证下可认为该模型是稳定的。

## 七、模型三的建立与求解

### 7.1 目标小类内 skc 的周销售量模型的改进与建立

首先,因问题一的目标 skc 的销售量模型并未考虑到主要因素间对销售量的影响有无交互作用,仅仅是在这 3 种影响因素对销售量的影响都是互相独立的前提下研究问题,并没有体现出各影响因素之间是否存在协同作用,显然,与实际不相符,也并不满足企业更加精准的营销需求。因此,本段就有无交互作用对问题一的目标 skc 的销售量模型进行改进。

若目标 skc 的销售量模型  $\exists(r, n_{skc(r)}) \in [1, 50] \times [2, +\infty)$ , 则可在目标 skc 的销售量模型的基础上添加交互项, 即改进后的目标 skc 的销售量模型

$$s_{skc(r),t}^* = \sum_{i+j+k \leq n_{skc(r)}} \alpha_{(ijk)} \cdot [sc^*(skc(r),t)]^i \cdot [ie^*(skc(r),t)]^j \cdot [rp^*(skc(r),t)]^k + \varepsilon \quad (7-1)$$

其中  $i, j, k \in \mathbb{N}$ 。最后结合问题二的分析, 建立更加精准的目标小类内所有 skc 的周销售量模型

$$\left\{ \begin{array}{l} S_{skc, \tilde{T}} = \sum_{t \in \{\tilde{T}\}} s_{skc,t} \\ \left\{ \begin{array}{l} s_{skc,t}^* = \sum_{i+j+k \leq n_{skc}} \alpha_{(ijk)} \cdot [sc^*(skc,t)]^i \cdot [ie^*(skc,t)]^j \cdot [rp^*(skc,t)]^k + \varepsilon, \quad n_{skc} \in [2, +\infty) \\ s_{skc,t}^* = \alpha_0 + \alpha_1 \cdot sc^*(skc,t) + \alpha_2 \cdot ie^*(skc,t) + \alpha_3 \cdot rp^*(skc,t) + \varepsilon, \quad n_{skc} \notin [2, +\infty) \end{array} \right. \\ sc(sk, t) \approx \frac{s_{skc,t}}{\sum_{i \in \{f^{-1}(tcc(r \in [1, 10]))\}}} s_{i,t} \\ rp(sk, t) = \frac{rc(sk, t)}{S_{skc,t}} \end{array} \right. \quad (7-2)$$

其中  $\{\tilde{T}\}$  表示  $\tilde{T}$  这个周内的时间集合,  $skc \in \{f^{-1}(tcc(r \in [1, 10]))\}$ 。同时目标小类的周销售量有

$$S_{tcc(r), \tilde{T}} = \sum_{skc \in \{f^{-1}(tcc(r))\}} S_{skc, \tilde{T}} \quad (7-3)$$

在考虑小类预测结果的同时, 即可预测目标小类内所有 skc 在 2019 年 10 月 1 日后 12 周内每周的周销量。接着, 再结合 MAPE 计算公式有如下

$$\left\{ \begin{array}{l} MAPE1(\tilde{T}) = \sum_{skc \in \{f^{-1}(tcc(r \in [1, 10]))\}} \frac{APE1(sk, \tilde{T})}{10} = \sum_{skc \in \{f^{-1}(tcc(r \in [1, 10]))\}} \frac{|S_{skc, \tilde{T}} - \hat{S}_{skc, \tilde{T}}|}{10S_{skc, \tilde{T}}} \\ MAPE2(\tilde{T}) = \sum_{r \in [1, 10]} \frac{APE2(tcc(r), \tilde{T})}{10} = \sum_{r \in [1, 10]} \frac{|S_{tcc(r), \tilde{T}} - \hat{S}_{tcc(r), \tilde{T}}|}{10S_{tcc(r), \tilde{T}}} \end{array} \right. \quad (7-4)$$

其中  $\hat{S}_{skc, \tilde{T}}$  为  $S_{skc, \tilde{T}}$  的估计值,  $\hat{S}_{icc(r), \tilde{T}}$  为  $S_{icc(r), \tilde{T}}$  的估计值。也即可得每周预测值的 MAPE1 和 MAPE2。

### 7.2 目标小类内 skc 的周销售量模型的求解

同理，结合上述和附件相应数据借助 MATLAB 软件即可求解出参系数估计值，也即可预测目标小类内所有 skc 在 2019 年 10 月 1 日后 12 周内每周的周销量，整理得周销售量预测表见支撑材料。对此，由于数据量过大，因此，这里，不失一般性，随机取样 10 种 skc 数据，整理得周销售量预测表如下

表 7-1：10 种 skc 的 2019 年 10 月 1 日后 12 周的预测值

skc	1	2	3	4	5	6	7	8	9	10	11	12
102572121792	17.49	9.80	4.53	7.29	4.93	7.97	2.29	2.36	3.37	1.03	1.64	1.30
102572881892	21.92	15.37	9.10	12.22	17.21	12.24	29.17	24.14	24.20	27.07	54.22	43.14
502572121891	12.63	34.24	21.47	29.87	29.19	31.93	37.54	26.50	29.68	23.49	15.89	33.23
902572441760	16.48	10.83	16.84	24.67	32.10	37.71	43.50	37.19	43.87	36.93	24.84	30.10
908572115687	36.40	19.48	24.96	24.32	24.73	27.19	16.50	15.23	17.64	16.31	19.08	7.49
996572122111	41.33	14.67	36.04	22.21	24.32	24.44	23.86	14.49	18.72	9.34	11.50	14.25
996572227238	43.80	16.56	26.04	22.10	14.81	13.75	10.58	12.42	10.45	3.07	3.08	2.07
996572776919	26.40	14.56	15.90	18.95	18.63	23.75	27.68	24.72	17.43	15.11	18.14	7.04
996573321309	59.36	24.84	19.71	20.90	28.22	38.07	26.69	22.26	20.85	28.01	26.13	14.25
996573321474	17.10	23.32	19.90	23.10	35.23	26.16	40.38	37.89	18.90	20.28	14.88	8.28

也求得这 12 周的 MAPE1 和 MAPE2，整理得如下表

表 7-2：2019 年 10 月 1 日后 12 周的预测值的 MAPE 表

$\tilde{T}$	MAPE1	MAPE2
1	0.0013	0.0012
2	0.0106	0.0073
3	0.0149	0.0156
4	0.0057	0.0063
5	0.0285	0.0281
6	0.0011	0.0017
7	0.0164	0.0171
8	0.0200	0.0207
9	0.0141	0.0148

10	0.0050	0.0056
11	0.0444	0.0441
12	0.0322	0.0330

## 7.3 模型分析

### 7.3.1 模型结果分析

基于预测目标小类内所有 skc 在 2019 年 10 月 1 日后 12 周内每周的周销量所得预测数据十分庞大，我们团队只截取了一部分放在论文中，即表 7-1 内的数值，从表中可知总的来说节假日是国庆节、双十一和双十二坐在周（标红）的目标 skc 的销售量会相对大一点，其相邻周的销售量会相对小一点。

### 7.3.2 误差分析

由表 7-2 可以看出来，2019 年 10 月 1 日后 12 周中 skc 和 tcc 的预测值的 MAPE 分别为 MAPE1 和 MAPE2 对应列的值，同模型二一样，可将它们看成是这 12 周销售量预测值与真实值的平均绝对百分比误差。从数值上看，最大的数值不超过 0.05，所以可以看出来模型的预测精度高达百分之九十五，符合题目要求的精准需求预测，基于时间精力有限，日后再对模型进行改进，从而提高预测精度。

### 7.3.3 灵敏度分析

在问题二的灵敏度分析和问题三的模型改进的基础上，不失一般性，同对 skc=102572121792，对此，我们借助 MATLAB 软件在置信水平即犯错误的可能性上限为 5%的前提下求解各参系数的置信区间与及该模型的拟合度如下表

表 7-3: 参系数的统计量表

参系数	参系数估计值	置信区间	拟合度
$\alpha_{(000)}$	-0.5639	$[-1.3756, 0.2477]$	$R^2 = 0.9707 > 0.5$
$\alpha_{(100)}$	-0.0383	$[-0.0463, -0.0304]$	
$\alpha_{(010)}$	0.0484	$[0.0422, 0.0546]$	
$\alpha_{(001)}$	0.0244	$[-0.0068, 0.0556]$	
$\alpha_{(200)}$	0.000279	$[0.000243, 0.000315]$	
$\alpha_{(020)}$	-0.0000036	$[-0.0000089, 0.0000017]$	
$\alpha_{(002)}$	-0.000227	$[-0.000509, 0.0000553]$	
$\alpha_{(110)}$	-0.000298	$[-0.000334, -0.000261]$	
$\alpha_{(101)}$	0.0000745	$[-0.000046, 0.000195]$	
$\alpha_{(011)}$	-0.000061	$[-0.000155, 0.000033]$	

其他数据见支撑材料。由上表可见拟合度非常好，即该类 skc 的销售量仅有超过 97%可由模型决定，其对应的拟合曲线对比图（附上下限置信区间）如下

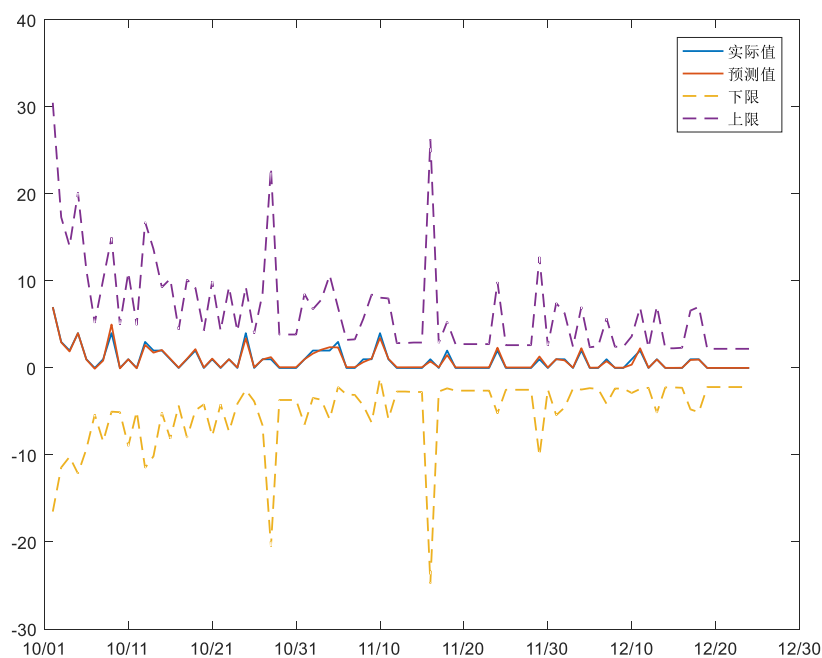


图 7-1:  $skc=102572121792$  的  $s_{skc,t}$  拟合曲线对比图

由图可见拟合效果非常好。同时，对参数的置信区间，即参数估计值的大小都在置信区间范围内，在 95%置信度的保证下可认为该模型是稳定的，同时发现，交互项销售特征与库存信息的置信区间不包含 0，显然，销售特征与库存信息之间必存在交互作用。

## 八、模型评价与推广

### 8.1 模型的评价

本文所建立的精准需求预测模型是基于统计回归模型上的，模型建立的主要思路是通过研究用 MATLAB 拟合出合适的销售量分别与销售特征、库存信息和节假日折扣的散点图，分型散点图的性态，结合起来构建出最为合适的统计回归模型，然后运用附件所给的信息进行拟合，求出每个自变量的参数，通过分析参数对模型的影响来表示销售特征、库存信息和节假日折扣对销售量的影响。最后对于精准需求的预测，主要是通过改进原有的数学模型，综合考虑各种自变量之间的交互作用来了解自变量之间复杂的关系，这能有效的模拟现实生活中各种复杂因素的综合作用使得消费者的消费心理发生改变，从而使得商品销售量发生改变。当然，数学模型只是从实际问题中抽象出来的符号表达，它除了具有能解决实际问题的有点之外，它也有很多局限的的方，下面我们将具体叙述我们所建立的精准需求预测模型的优缺点。

#### 8.1.1 模型的优点

(1) 模型一中，对于对销售量分别与销售特征、库存信息和节假日折扣作逐步的统计回归，可以清楚知道每个自变量对因变量是否存在影响，然后通过自变量

的交互作用了解实际生活中各种相关因素的相互作用对消费者消费心理的影响，从而反映到商品的销售量上来，更贴近实际。

（2）相对于参考文献中的其他预测需求的方法，我们的模型更加的简单，灵活性更加强，更加容易确定表达式的形式，并且预测的效果不输于其他模型。

（3）模型三在模型一和模型二的基础上进行改进，较好的实现了问题之间的连接性，在提高需求预测的精准性的同时也很好地体现了文章的层次性。

（4）建立的模型能与实际紧密联系，结合情况对问题进行求解，使得模型具有很好的通用性和推广性。

### **8.1.2 模型的缺点**

（1）由于时间和精力有限，在建立模型过程中忽略了某些对销售量有一定影响的因素，这使得模型预测的精准性会有所下降。

（2）假设过多，也会降低模型的精准性。

## **8.2 模型的推广**

（1）本文主要建立了精准需求预测模型，该模型同样适用于各种经济活动的预测，例如投资额预测以及人员薪金管理方面。

（2）本文建立的模型在未来对库存管理上有一定适用性，可以求出一定库存量相对应的大致销售量从而有助于生产经营决策，能在一定程度上储存生产成本。

## 九、给企业的一封信

尊敬的 XXX 领导：

你好！我们是 XXX 学校的学生。我们都知道随着我国消费市场的不断发展，我国人民消费水平的不断提高以及商品资源的极大丰富，市场上原有的消费模式已由“以物为主”逐步转变为“以客为主”。多元消费观念的趋势下，人们的消费需求也由原来的考虑实用性转向考虑时尚性，特别是将注意力放在了“个性化、时尚、美观”等方面。在这种消费观念的驱使下，谁能精准预测到消费者在这些种类商品的需求量，谁就能在这种消费模式下获得最大的利益。在得知贵公司无法精准预测消费者的需求后，我们团队通过共同协作，建立了销售特征、库存量以及节假日折扣对销售量共同影响的数学模型，该模型将有助于贵公司在生产经营活动中做出正确的决策，下面就让我们来介绍一下我们的模型。

首先，我们在统计回归模型基础上建立了目标 skc 的销售量模型、目标小类月销售量模型以及目标小类内 skc 的周销售量模型，其中目标小类内 skc 的周销售量模型不但能够单独反映销售特征、库存量以及节假日折扣各自对销售量的影响，而且它还能反映不同因素之间的交互作用对销售量产生的协同影响，这能够充分反映现实生活中的复杂情景，因此我们的模型是紧密联系现实生活的。

其次，我们的模型相对于现在已有的对需求的预测方法而言更具有优势，这体现在我们的模型所用到的算法以及原理更加简单，并且在预测效果上也非常良好。以目标小类月销售量模型为例，该模型对于月销售量的预测值精准度达到百分之九十，模型的拟合度也达到了百分之八十六，已经实现了精准需求预测，想必这将有助于贵公司实现库存管理成本最低，与此同时，我们的模型还能够清晰的反映哪种因素对销售量的影响最大，它们如何交互影响消费者的消费心理，这将有助于贵公司在生产经营决策中做出正确的方针。

最后，我们的模型应用的领域非常广泛，它在经济生产经营活动中，不仅仅可以在需求预测上应用，它还可以应用于员工的薪金管理、还有投资风险预测等方面，这是我们团队后期模型优化的方向，所以希望贵公司采纳我们的模型。

敬祝：  
生意兴隆

建模小组 203295 号  
2020 年 5 月 24 日

## 十、参考文献

- [1]王金旺. 供应链环境下市场需求预测体系研究[D]. 北京交通大学, 2008
- [2]张相斌, 罗玲桃. 020 模式下连锁零售网点需求预测及资源调度[J]. 系统管理学报, 2020, 29(01):168-174.
- [3]张佳慧, 马继东. 木材首饰商品零售企业的 ARIMA 需求预测研究[J]. 森林工程, 2016, 32(04):75-78+82.
- [4]胡良平. 回归建模的基础与要领( I )——统计模型种类的划分方法[J]. 四川精神卫生, 2018, 31(06):489-492.
- [5]张朝雄, 沈旻, 张慧. 缺失数据插补方法比较研究[J]. 市场研究, 2007(09):33-35



## 附录

### 问题一

```
1. clear all
2. close all
3. clc
4. DATA=dlmread('t1skr.txt','\t');%读取数据
5. %s 销量, rc 实际花费, ie 库存, cp 售价, sc 销售特征
6. X=DATA;
7. [m,n]=size(X);
8. S=DATA(:,1);
9. IE=DATA(:,3);
10. RP=DATA(:,5);
11. SC=DATA(:,6);
12. a=0;
13. for j=1:n
14.     for i=1:m
15.         a=a+X(i,j)^2;
16.     end
17.     A(1,j)=sqrt(a);
18.     a=0;
19. end
20. A= repmat(A,m,1);
21. data=X./A;
22. s=data(:,1);
23. rc=data(:,2);
24. ie=data(:,3);
25. rp=data(:,5);
26. sc=data(:,6);
27. alpha=0.05;%置信水平
28. i=20;
29. S1=S(i:50:9200);SC1=SC(i:50:9200);IE1=IE(i:50:9200);RP1=RP(i:50:9200);
30. o=ones(184,1);
31. s1=s(i:50:9200);sc1=sc(i:50:9200);ie1=ie(i:50:9200);rp1=rp(i:50:9200);
32. x22=[o,sc1,ie1,rp1,sc1.^2,ie1.^2,rp1.^2];
33. X22=[o,SC1,IE1,RP1,SC1.^2,IE1.^2,RP1.^2];
34. [b22,bint22,r22,rint22,stats22]=regress(s1,x22,0.05);
35. [B22,Bint22,r22,rint22,Stats22]=regress(S1,X22,0.05);
36. xx=x22;
37. yhat=x22*b22;%预测值
38. yhat1=x22*bint22(:,1);
39. yhat2=x22*bint22(:,2);
40. x=1:184;
41. plot(x,s1,x,yhat,x,yhat1,x,yhat2);%绘图
42. dateaxis('x',6,'1-JULY');%设置时间为横坐标
43. XX=X22;
44. Yhat=x22*B22;%预测值
45. Yhat1=x22*Bint22(:,1);
46. Yhat2=x22*Bint22(:,2);
47. x=1:184;
48. figure
49. plot(x,S1,x,Yhat,x,Yhat1,x,Yhat2);
50. dateaxis('x',6,'1-JULY');
51. %stats: R^2,F 统计量, 显著性 P 值, 模型误差,R 平方越大拟合程度越高。
52. for i=1:50
53. s1=s(i:50:9200);sc1=sc(i:50:9200);ie1=ie(i:50:9200);rp1=rp(i:50:9200);o=ones(184,1);
54. x22=[o,sc1,ie1,rp1,sc1.^2,ie1.^2,rp1.^2];
55. [b22,bint22,r22,rint22,stats22]=regress(s1,x22,0.05);
56. bbb(:,i)=b22;
57. end
```

## 问题二

### 程序一

```
1. clear all
2. close all
3. clc
4. %计算具体 skc 预测值
5. data=dlmread('t2skc.txt','\t');%读取数据
6. %s 销量, rc 实际花费, ie 库存, rp 售价, sc 销售特征
7. s=data(:,1);
8. rc=data(:,2);
9. ie=data(:,3);
10. sc=data(:,4);
11. alpha=0.05;%置信水平
12. [n,m]=size(data);
13. o=ones(n,1);
14. x=[o,rc,ie,sc,rc.^2,ie.^2,sc.^2];
15. [b,bint,r,rint,stats]=regress(s,x,0.05);
16. xx=x;
17. yhat=x*b;%预测值
18. yhat1=x*bint(:,1);
19. yhat2=x*bint(:,2);
20. x=1:91;
21. plot(x,s,x,yhat,x,yhat1,x,yhat2);
22. dateaxis('x',6,'1-OCT');
```

### 程序二

```
1. clear all
2. close all
3. clc
4. data=dlmread('t2tcc.txt','\t');%读取数据
5. %s 销量, rc 实际花费, ie 库存, rp 售价, sc 销售特征
6. s=data(:,1);
7. rc=data(:,2);
8. ie=data(:,3);
9. sc=data(:,4);
10. alpha=0.05;%置信水平
11. [n,m]=size(data);
12. o=ones(n,1);
13. x=[o,rc,ie,sc,rc.^2,ie.^2,sc.^2];
14. [b,bint,r,rint,stats]=regress(s,x,0.05);
15. xx=x;
16. yhat=x*b;%预测值
17. %计算 MAPE
18. mape1=0;
19. for i=1:10
20. mape1=abs(s(i)-yhat(i))/(10*s(i))+mape1;
21. end
22. mape2=0;
23. for i=11:20
24. mape2=abs(s(i)-yhat(i))/(10*s(i))+mape2;
25. end
26. mape3=0;
27. for i=21:30
28. mape3=abs(s(i)-yhat(i))/(10*s(i))+mape3;
29. end
```

### 问题三

#### 程序一

```
1. clear all
2. close all
3. clc
4. data=dlmread('t3skc1.txt','\t');%读取数据
5. %s 销量, rc 实际花费, ie 库存, rp 售价, sc 销售特征
6. s=data(:,1);
7. rp=data(:,2);
8. ie=data(:,3);
9. sc=data(:,4);
10. alpha=0.05;%置信水平
11. [n,m]=size(data);
12. o=ones(n,1);
13. x=[0,sc,ie,rp,sc.^2,ie.^2,rp.^2,sc.*ie,sc.*rp,ie.*rp];
14. [b,bint,r,rint,stats]=regress(s,x,0.05);
15. xx=x;
16. yhat=x*b;%预测值
17. yhat1=x*bint(:,1);
18. yhat2=x*bint(:,2);
19. x=1:84;
20. plot(x,s,x,yhat,x,yhat1,x,yhat2);
21. dateaxis('x',6,'1-OCT');
```

#### 程序二

```
1. clear all
2. close all
3. clc
4. %skc MAPE
5. x=[
6. 16825.0107 16600.8174
7. 8412.5357 9306.5750
8. 3738.8643 4296.9781
9. 6543.0428 6916.4554
10. 6543.0428 4678.8966
11. 7477.7893 7561.8770
12. 1869.4321 2176.8626
13. 1869.4321 2243.0877
14. 2804.1786 3200.3142
15. 934.7464 981.2863
16. 2804.1786 1557.8702
17. 934.7464 1235.9189
18. ];
19. for i=1:12
20. m(i)=abs(x(i,1)-x(i,2))/(10*x(i,1));
21. end
22. %TCC MAPE
23. x1=[
24. 17229.5916 17029.5188
25. 9068.2253 9727.9248
26. 3627.2537 4192.1722
27. 6347.7698 6747.7939
28. 6347.7698 4564.7949
29. 7254.5681 7377.4187
30. 1813.6572 2123.7609
31. 1813.6572 2188.4064
32. 2720.4554 3122.2415
33. 906.7982 957.3480
34. 2720.4554 1519.8364
35. 906.7982 1205.7834
```

```
36. ];  
37. for i=1:12  
38. m1(i)=abs(x1(i,1)-x1(i,2))/(10*x1(i,1));  
39. end
```