

Sales Forecast Based on Markov and ARIMA Models

Jingxi Zhou

School of Information Science and Technology, Hunan Agricultural University, Changsha Hunan
Email: 1119324437@qq.com

Received: Aug. 7th, 2019; accepted: Aug. 21st, 2019; published: Aug. 28th, 2019

Abstract

Many merchants are deeply troubled by the problem of filling orders caused by insufficient stock of goods, and the problem of filling orders mainly has two problems: the occupation of cash flow and the problem of inventory. Therefore, effective and reasonable replenishment strategy is an urgent problem for merchants. Based on the sales data of business over the years, and on the basis of data preprocessing, this paper studies the sales volume of a business from several aspects, such as model hypothesis, model establishment and solution. According to the establishment of ARIMA model, and according to ARIMA model, the sales volume in the next few days is fitted and predicted, and then the results are tested by Markov model, which provides a theoretical basis for the replenishment and inventory of the merchant.

Keywords

ARIMA Model, Sales Forecast, Matlab Software, Markov Model

基于马尔科夫和ARIMA模型的销量预测

周静曦

湖南农业大学信息科学技术学院, 湖南 长沙
Email: 1119324437@qq.com

收稿日期: 2019年8月7日; 录用日期: 2019年8月21日; 发布日期: 2019年8月28日

摘要

许多商家对于因货物存量不足而导致的补单问题存在极大的困扰, 且补单问题主要存在两个问题: 对现金流的占用和库存问题。因此, 有效解决合理的补单策略是商家们急需解决的问题。本文基于某商家的

历年销量数据,在数据预处理的基础上,从模型假设、模型建立与求解等几个方面对该商家的销售量数据进行ARIMA模型的建立,并根据ARIMA模型对未来几天的销售量进行拟合和预测分析,然后用马尔可夫模型进行结果检验,为该商家的补单及存货提供理论依据。

关键词

ARIMA模型,销售预测, Matlab软件, 马尔可夫模型

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

商家在原来已有订单基础上,买家又追加订单,就是补单。有时补单也特指因品质异常而不能满足客户要求导致的补单。商家的补单主要与现金流的占用和库存问题等原因有关。商家的补单数量的多少将直接影响客户需求。过于乐观的补单将造成大量的库存,库存的积压带来的不仅是仓储成本的增加,更为严重的是要考虑清仓问题,清库存的方式无非是加大营销力度,以更低的成本清理货物,这时候不仅资金的流转变慢,而且低价的清仓也会拉低品牌溢价,而过于悲观的补单则会导致货不够卖而造成资源的浪费和利润的流失,这又是企业不愿意看到的,所以若能够将商家的日均销售量预测出来,科学的根据企业自身情况合理制定补单模型和计划也显得尤为重要,这也将极大地减少商家的流动资金占用、库存压力及仓储成本问题。

2. 数据来源与模型假设

本文选取某厂家2018年7月~2019年3月的销售量数据作为研究对象,数据来源于2019年第12届华中地区数学建模邀请赛。为简化建模过程,本文做出以下假设:

假设题目所给未指定日期的销售量都为零;

假设问题所给数据真实可靠,并且筛选掉的数据不会对本文的研究结论做出影响;

假设未考虑的影响因素不会令模型的预测效果产生大幅度偏差。

3. ARIMA 预测模型建立与求解

本文选取基于ARIMA模型的时间序列分析预测[1][2][3]。ARIMA模型的基本思想[4]是首先将非平稳时间序列转化为平稳序列,然后通过将对因变量的滞后项和随机误差项进行回归建立预测模型。它的具体形式可以表达成 $ARIMA(p, d, q)$,其中: p 表示自回归过程的阶数; d 表示差分阶数; q 表示移动平均过程的阶数。

1) 数据预处理及平稳性检验

我们首先随机选取某一个货号作为研究对象,利用SPSS软件将该货号的销售量根据日期进行缺失值的填补,然后做出时序图,如图1所示。

根据时序图可知,原始数据并不平稳,因此,对该非平稳时间序列进行数据预处理,采用差分处理方法使序列平稳。首先采用一阶差分,若效果不好,则采用二阶差分,以此类推并观察动态结果,选择出最优差分解,发现当采用一阶差分时序图再数值0上下波动,符合平稳序列的特点(见图2),因此,

我们初步判定差分次数 d 的值为 1。

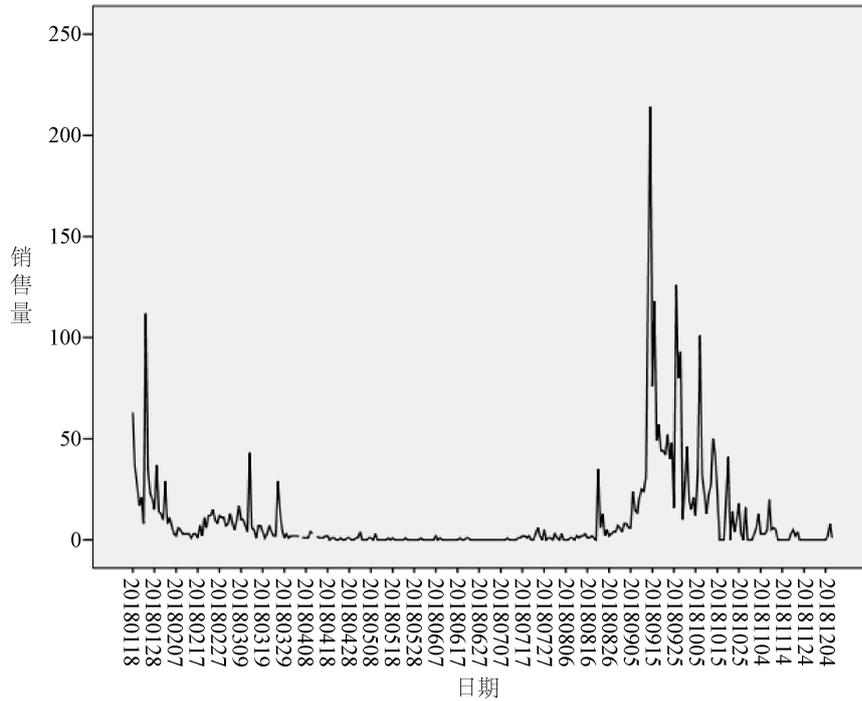


Figure 1. Sales sequence diagram of a selected article number

图 1. 选取某一货号的销售量时序图

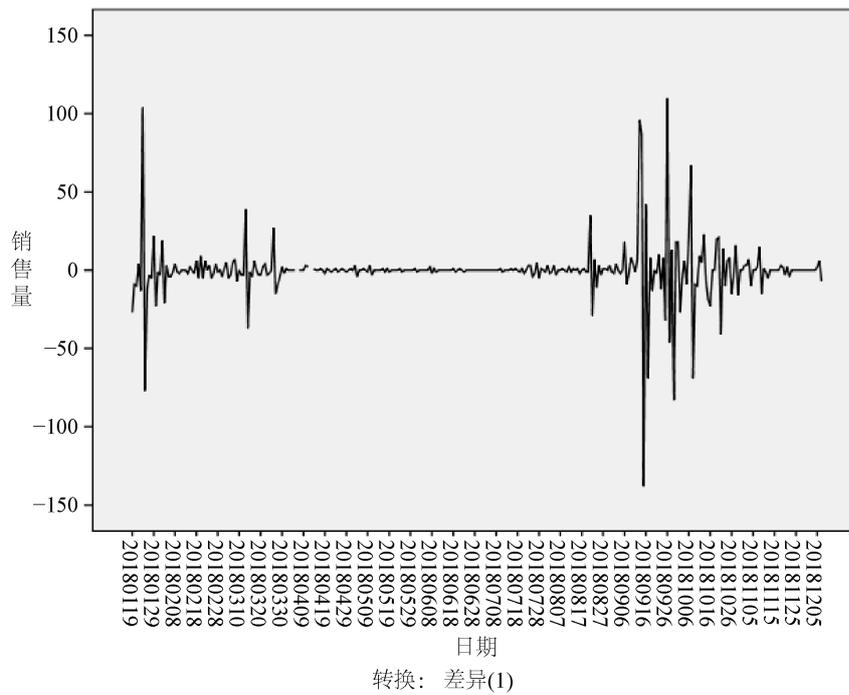


Figure 2. First-order difference processing diagram of sales volume data of a certain article number

图 2. 某一货号销售量数据的一阶差分处理图

2) 模型识别与定阶

经过处理差分后的数据已具备平稳序列的要求，接着，再根据 ARIMA 相关性的特征，读取平稳时间序列差分后的自相关图和偏自相关图初步识别 p 、 q 值。图 3 中 ACF 第一阶后呈截尾状，PACF 第四阶后呈拖尾状，因此可初步判断差分后序列适合 ARIMA (3, 1, 3)模型。

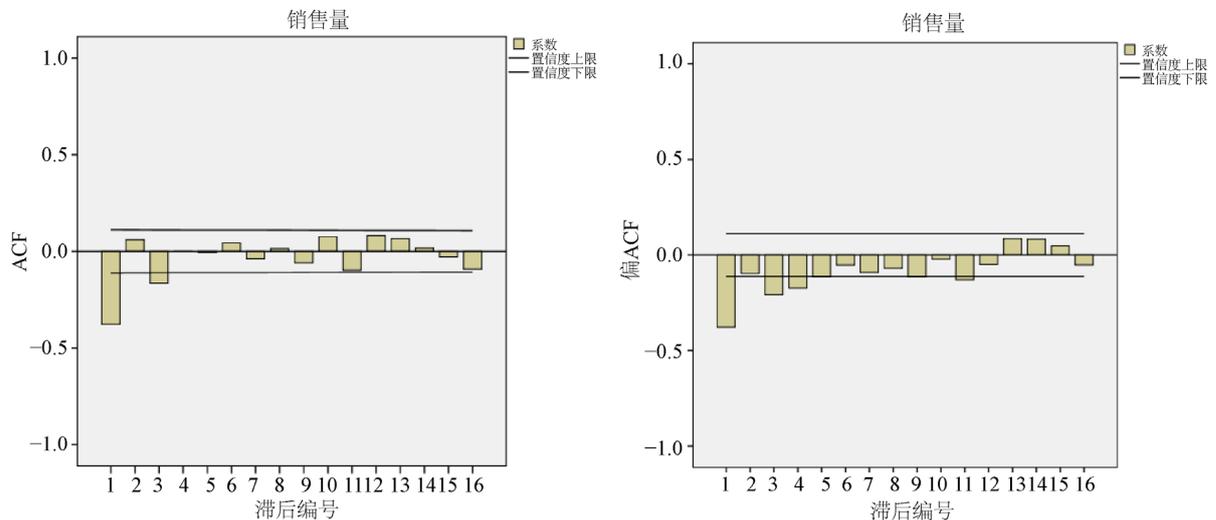


Figure 3. ACF diagram and APCF diagram after sequence difference
图 3. 序列差分后的 ACF 图和 APCF 图

重复拟合 ARIMA (p, d, q)模型中 p 和 q 的各种取值，并计算相应参数对应的显著性 Sig.、 R 的平方来初步判定模型的最佳结束，计算结果见表 1。

Table 1. Test results of fitting model for each parameter in ARIMA model
表 1. ARIMA 模型中各参数的拟合模型检验结果

参数(p, q)	R 的平方	显著性 Sig.
(3, 3)	0.484	0.104
(3, 4)	0.474	0.062
(1, 4)	0.472	0.184
(1, 3)	0.471	0.195
...

观察表中的观测值，综合比较表中的各个指标，表中参数(3, 3)对应的相关系数 R 的平方最大，且显著性水平 Sig. 大于 0.05，对参数(3, 1, 3)作模型拟合效果图(如图 4 所示)，观察实际拟合效果较好，因此，本实验选中 ARIMA 模型的参数估计为 $p=3, d=1, q=3$ 。

3) ARIMA 模型预测

根据上述得出的 ARIMA 模型，其 R 的平方为 0.484 (如图 5 所示)，达到的拟合程度较好，AR, MA 的系数分别是 0.065 和 0.503。因此，我们认为得到 ARIMA 模型结果为：

$$\Delta x_t = 0.033\Delta x_{t-1} + \varepsilon_t - 0.562\varepsilon_t$$

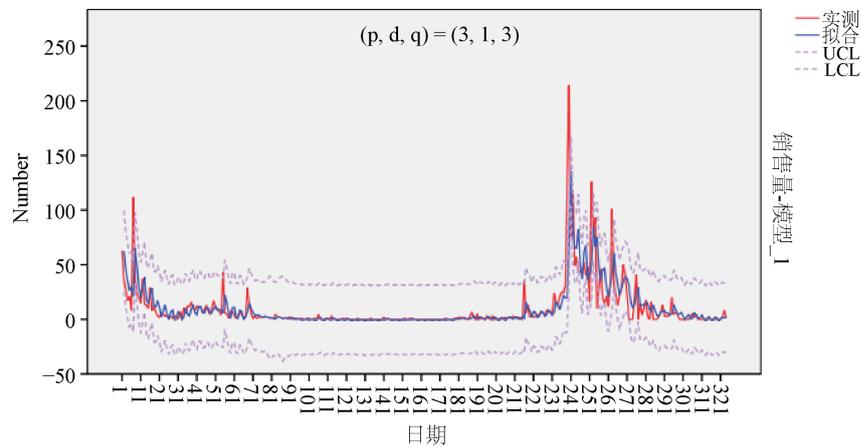


Figure 4. Model fitting diagram corresponding to ARIMA model parameters (3, 1, 3)
图 4. ARIMA 模型参数(3, 1, 3)对应的模型拟合图

模型设计

模型	预测变量数	模型拟合度统计		杨-博克斯Q(18)			离群值数
		平稳R方	R方	统计	DF	显著性	
销售量-模型_1	1	0.260	0.484	18.408	12	0.104	0

ARIMA模型参数

				估算	标准误差	t	显著性		
销售量-模型_1	销售量	不转换	常量	-7965.719	15943.913	-0.500	0.618		
			AR	延迟1	0.033	0.082	0.405	0.686	
				延迟2	-0.735	0.065	-11.228	0.000	
		延迟3		0.238	0.081	2.934	0.004		
		差异	MA	延迟1	0.562	0.054	10.356	0.000	
				延迟2	-0.722	0.045	-15.961	0.000	
				延迟3	0.792	0.052	15.114	0.000	
		日期	不转换	分子	延迟0	0.000	0.001	0.500	0.618

Figure 5. ARIMA (3, 1, 3) model fitting results
图 5. ARIMA (3, 1, 3)模型拟合结果

我们选近 5 天的销售量作为预测目标，首先在 SPSS 中增加 5 个自变量，采用预测建模方程，得到分析结果如图 6 所示，由此，我们可以得出结论：未来 5 天的货物销量分别为：3，3，2，2，3。

4. 马尔可夫预测模型的建立

马尔可夫模型是一种常见的简单随机过程，该过程是研究一个系统的状况及其转移的理论。它通过对不同状态的初始概率以及状态之间的转移概率来研究，来确定状态的变化趋势，从而达到对未来进行预测的目的[5]。

1) 马氏性检验

将某一货号销售量进行等级划分假设。设销售量为 t 。当 $0 < t \leq 10$ 时，认为销量很差；当 $10 < t \leq 20$ 时，认为销量较差；当 $20 < t \leq 30$ 时，认为销量一般；当 $30 < t \leq 40$ 时，认为销量较好；当 $t \geq 40$ 时，认为销量很好。由此可以将销售量划分为 5 个状态等级，如表 2 所示。

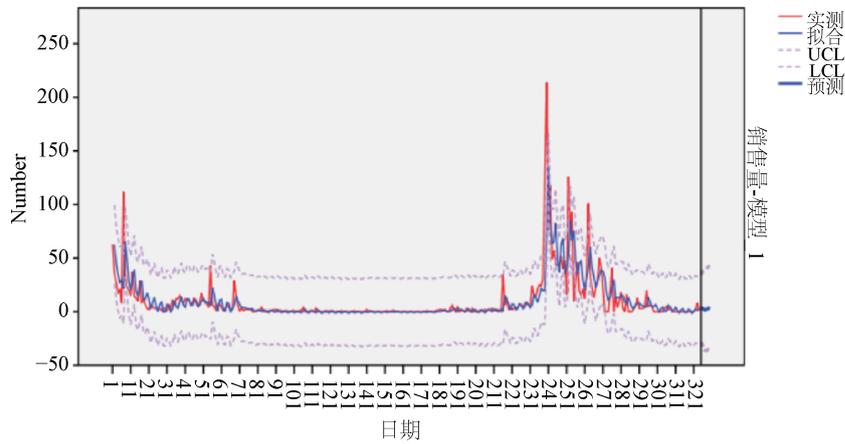


Figure 6. Sales volume forecast by ARIMA model for the next 5 days
图 6. ARIMA 模型预测未来 5 天的销售量图

Table 2. Sales grade status division
表 2. 销售量等级状态划分

状态	销量	等级
1	$0 < t \leq 10$	很差
2	$10 < t \leq 20$	较差
3	$20 < t \leq 30$	一般
4	$30 < t \leq 40$	较好
5	$t \geq 40$	很好

由此计算出初始概率，可得：

$$f_1 = \begin{pmatrix} 190 & 7 & 4 & 1 & 2 \\ 9 & 8 & 4 & 2 & 1 \\ 3 & 6 & 3 & 1 & 2 \\ 1 & 1 & 3 & 0 & 3 \\ 2 & 2 & 1 & 4 & 12 \end{pmatrix}$$

系统在时刻 t 从状态 m 出发，经过 i 步后处于状态 n 的概率为状态转移概率，其中一步转移后形成的矩阵称为一步转移概率矩阵。各步转移概率为：

$$p_1 = \begin{pmatrix} 0.9314 & 0.0343 & 0.0196 & 0.0049 & 0.0098 \\ 0.3750 & 0.3333 & 0.1667 & 0.0833 & 0.0417 \\ 0.2000 & 0.4000 & 0.2000 & 0.0667 & 0.1333 \\ 0.1250 & 0.1250 & 0.3750 & 0 & 0.3750 \\ 0.0952 & 0.0952 & 0.0476 & 0.1905 & 0.5714 \end{pmatrix}$$

$$p_2 = \begin{pmatrix} 0.9310 & 0.0394 & 0.0099 & 0.0099 & 0.0099 \\ 0.4583 & 0.2083 & 0.2083 & 0.0417 & 0.0833 \\ 0.1333 & 0.2667 & 0.2000 & 0.1333 & 0.2667 \\ 0 & 0.7500 & 0 & 0.1250 & 0.1250 \\ 0.1429 & 0.0476 & 0.2381 & 0.0476 & 0.5283 \end{pmatrix}$$

$$p_3 = \begin{pmatrix} 0.9163 & 0.0591 & 0.0148 & 0.0049 & 0.0049 \\ 0.4167 & 0.2500 & 0.1250 & 0.0833 & 0.1250 \\ 0.2143 & 0.0714 & 0.2143 & 0.2143 & 0.2857 \\ 0.2500 & 0.1250 & 0.3700 & 0 & 0.2500 \\ 0.1905 & 0.1905 & 0.0952 & 0.0476 & 0.4762 \end{pmatrix}$$

$$p_4 = \begin{pmatrix} 0.9261 & 0.0493 & 0.0148 & 0.0049 & 0.0049 \\ 0.4167 & 0.2917 & 0.0833 & 0.1250 & 0.0833 \\ 0.2143 & 0.0714 & 0.2857 & 0.0714 & 0.3751 \\ 0.2500 & 0.1250 & 0.2500 & 0.1250 & 0.2500 \\ 0.1000 & 0.2000 & 0.1500 & 0.0500 & 0.5000 \end{pmatrix}$$

$$p_5 = \begin{pmatrix} 0.9212 & 0.0493 & 0.0246 & 0.0049 & 0 \\ 0.3750 & 0.2500 & 0.1667 & 0.0833 & 0.1250 \\ 0.2857 & 0.2143 & 0.1429 & 0.0714 & 0.2857 \\ 0.2500 & 0.1250 & 0.1250 & 0 & 0.5000 \\ 0.1579 & 0.1579 & 0.0526 & 0.1579 & 0.4737 \end{pmatrix}$$

经 Matlab 检验, 给定显著水平 α , $\chi^2 = 2 \sum_{i=1}^m \sum_{j=1}^m f_{ij} \left| \ln \frac{p_{ij}}{p_j} \right| > \chi_\alpha^2((m-1)2)$, 可认为序列符合马氏性, 因此, 可对该序列用马尔科夫模型。接下来可利用马尔科夫模型进行预测工作。(Matlab 程序见附录三)

2) 马尔可夫模型预测及分析

首先计算各阶段自相关系数及权重。采用自相关系数作为相依关系强弱的度量, 第 s 步的自相关系数为:

$$r_s = \frac{\sum_{t=1}^{n/s} (X_t - \bar{X})}{\sum_{t=1}^{n/s} (X_t - \bar{X})^2}$$

由此求得序列各阶自相关系数 $r = (0.7091, 0.6558, 0.6131, 0.6331, 0.6448)$ 。

将自相关系数规范化可得各步马尔科夫的权重:

$$\omega_s = \frac{|r_s|}{\sum_{t=1}^t |r_s|}$$

计算出的权重为 $\omega = (0.2178, 0.2014, 0.1883, 0.1944, 0.1980)$

因此, 根据以上数据可预测得到未来 5 天的销售状态等级:

Table 3. 2018/12/08 sales grade forecast

表 3. 2018/12/08 销售等级预测

初始日期	初始状态	滞后阶数	状态 1	状态 2	状态 3	状态 4	状态 5
2018/12/03	1	5	0.9212	0.0493	0.0246	0.0049	0
2018/12/04	1	4	0.9261	0.0493	0.0148	0.0049	0.0049
2018/12/05	1	3	0.9163	0.0591	0.0148	0.0049	0.0049
2018/12/06	1	2	0.9310	0.0394	0.0099	0.0099	0.0099
2018/12/07	1	1	0.9314	0.0343	0.0196	0.0049	0.0098
Pi			0.9254	0.0459	0.0168	0.0059	0.0060

由表 3 可知, 2018/12/08 销售量为状态 1 的概率最大, 故符合我们 ARIMA 模型预测结果。经过同样计算验证, 得出 2018/12/09 至 2018/12/12 销售状态等级也为 1, 符合 ARIMA 模型估计。

5. 模型检验

对 ARIMA 模型的参数 $p = 3, d = 1, q = 3$ 进行残差分析检验, 图 7 中可以看出残差图基本上都是平稳的, 且预测出的观测值基本都在预测值的预测区间之内(见图 8), 因此, 我们认为 ARIMA (3, 1, 3)模型是合理的, 且判定其拟合效果较好。

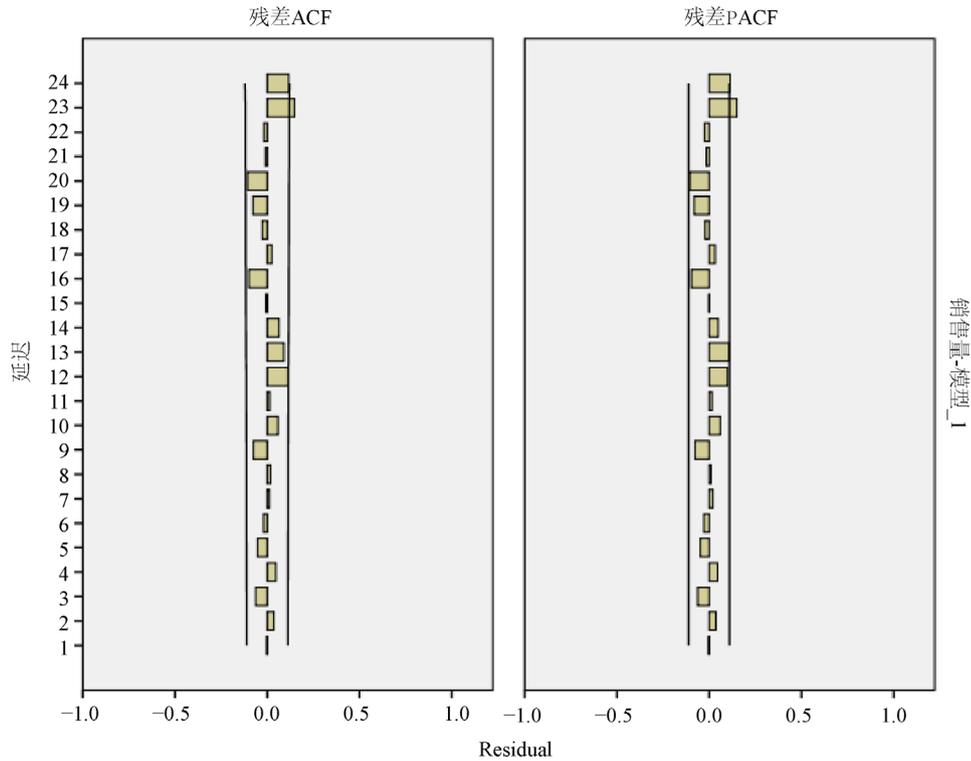


Figure 7. ACF and PACF diagrams of model residuals of ARIMA (3, 1, 3)
图 7. ARIMA (3, 1, 3)模型残差的 ACF 和 PACF 图

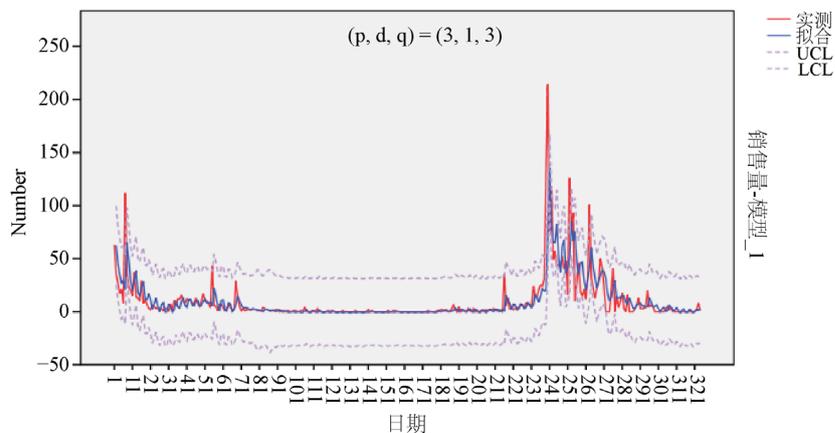


Figure 8. Prediction fitting effect of ARIMA (3, 1, 3) model
图 8. ARIMA (3, 1, 3)模型的预测拟合效果图

6. 结论

本文通过运用 ARIMA 模型对某厂家的销售量进行预测,同时采用马尔科夫模型进行检验,两种模型相结合,结果基本相一致,且所达到的预测结果较好。该文提出的预测结果也将有利于相关厂家指定相关的补单和存货方案来应对不同的订单问题所带来的后果,做到防患于未然。但本文目前所提出的两种销量预测模型仍有一些地方需要进一步改进,以提高预测的准确度与稳定性。

参考文献

- [1] 葛娜, 孙连英, 赵平, 万莹. 基于 ARIMA 时间序列模型的销售量预测分析[J]. 北京联合大学学报, 2018(8): 28-33.
- [2] 魏杰, 张璇, 王柳, 汪丽萍. 基于灰色预测和 ARIMA 模型的疾病分析与预测[J]. 数理医药学杂志, 2018, 31(12): 1739-1742.
- [3] 徐梦茹, 王学明. 马尔可夫与 ARIMA 组合模型对地区降雨量的预测研究[J]. 计算机应用与软件, 2019, 36(3): 34-37.
- [4] 董大勇, 刘珂言. 基于 ARIMA 模型的工业品出厂价格指数预测与分析[J]. 统计与决策, 2016(1): 179-181.
- [5] 数学建模之马尔可夫预测[J/OL]. <https://wenku.baidu.com/view/4bc708f66e1aff00bed5b9f3f90f76c660374c53.html>, 2018-02-04.

附录

马尔可夫预测 matlab 程序:

```
clc
clear
a=xlsread('code.xlsx');
E=a(:,4)';
fori=1:5
forj=1:5
a1=0;
a2=0;
a3=0;
a4=0;
a5=0;
form=1:272
ifE(n)==i
ifE(n+1)==j
a1=a1+1;
end
end
f1(i,j)=a1;
end
form=1:271
ifE(n)==i
ifE(n+2)==j
a2=a2+1;
end
end
f2(i,j)=a2;
end
form=1:270
ifE(n)==i
ifE(n+3)==j
a3=a3+1;
end
end
f3(i,j)=a3;
end
form=1:269
ifE(n)==i
```

```
ifE(n+4)==j
a4=a4+1;
end
end
f4(i,j)=a4;
end
for n=1:268
ifE(n)==i
ifE(n+5)==j
a5=a5+1;
end
end
f5(i,j)=a5;
end
end
end
f1
for i=1:5
for j=1:5
p1(i,j)=f1(i,j)/sum(f1(i,:));
p2(i,j)=f2(i,j)/sum(f2(i,:));
p3(i,j)=f3(i,j)/sum(f3(i,:));
p4(i,j)=f4(i,j)/sum(f4(i,:));
p5(i,j)=f5(i,j)/sum(f5(i,:));
end
end
p1,p2,p3,p4,p5
for n=1:5
a1=0;
a2=0;
for i=1:5
a1=a1+f1(i,n);
for j=1:5
a2=a2+f1(i,j);
end
end
S(n)=a1/a2;
end
S
```

```
fori=1:5
forj=1:5
x_tjl(i,j)=2*f1(i,j)*abs(log(p1(i,j))-log(S(j)));
end
end
x_tjl(4,4)=0
x_tjl=sum(x_tjl(:));
x_bzl=32;
ifx_tjl>x_bzl
disp('该序列符合马氏性')
elsex_tjl<x_bzl
disp('不能对该序列用马尔科夫模型')
end
E_jun=sum(E(1,:))/length(E);
form=1:5
a1=0;
a2=0;
fort=1:(length(E)-n)
a1=a1+(E(t)-E_jun)*(E(t+n)-E_jun);
a2=a2+(E(t)-E_jun)^2;
end
r(n)=a1/a2;
end
r
form=1:5
a3=0;
fort=1:5
a3=a3+r(t);
end
w(n)=r(n)/a3;
end
w
k=325
K(1,:)=p5(E(k-5),:);
K(2,:)=p4(E(k-4),:);
K(3,:)=p3(E(k-3),:);
K(4,:)=p2(E(k-2),:);
K(5,:)=p1(E(k-1),:);
K
```

```
for i=1:5
    a1=0;
    for j=1:5
        a1=a1+K(j,i)*w(6-j);
    end
    P(i)=a1;
end
P
```

知网检索的两种方式:

1. 打开知网首页: <http://cnki.net/>, 点击页面中“外文资源总库 CNKI SCHOLAR”, 跳转至: <http://scholar.cnki.net/new>, 搜索框内直接输入文章标题, 即可查询;
或点击“高级检索”, 下拉列表框选择: [ISSN], 输入期刊 ISSN: 2325-2251, 即可查询。
2. 通过知网首页 <http://cnki.net/> 顶部“旧版入口”进入知网旧版: <http://www.cnki.net/old/>, 左侧选择“国际文献总库”进入, 搜索框直接输入文章标题, 即可查询。

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: sa@hanspub.org