

基于多变量时间序列的向量自回归周销量预测模型

浙江中烟工业有限责任公司 卢志保 周展民 梁天威

摘要:为解决现有商业周销量预测不够精准的问题,本文提出一种基于多变量时间序列(MTS)及向量自回归(VAR)模型的周销量预测方法,并以ZJ中烟某品牌规格卷烟为研究对象进行了实例应用。核心是通过对订足率的精准预测,实现在可供量既定情况下的销量预测。该方法通过MTS分析对可供量数据进行优选确定参考期,并将参考期内订足率和综合状态指数作为时间序列建立VAR模型,通过机器学习拟合预测订足率。实验结果表明,将向量自回归(VAR)方法应用到订足率预测中是可行的。与传统的定性预测相比,通过定量分析大大提高了预测的客观性和准确性。

关键词:销量;订足率;多变量;向量自回归

DOI: 10.12433/zgkjztz.20240712

工商网配是卷烟货源组织的关键环节之一,开展工商网配是践行行业工作的需求,也是推动企业高质量发展的需求。其中,周销量预测作为整个工商网配流程的起点,对网配准确率有着重要作用。当前,工商网配较普遍的做法是,工业企业依据商业企业1个自然周的销售预测开展卷烟配货。因此,本文解决的问题是,在政策既定即可供量确定的前提下,如何准确的预测周销量,为工商网配量提供决策依据。

一、卷烟周销量预测现状

当前,商业企业预测某一品规周销量 Q_T 的流程及方法如下:

第一步,设定该品规预测周可供量 K_T 。可供量是给零售户的档位分配量,对订足率和销量影响较大,

$K_T = \sum_{i=1}^n X_i Y_i$,其中 X_i 表示第档零售户数量, Y_i 表示第 i 档零售户定量(单户投放条数)。

第二步,根据设定的可供量往前追溯相近可供量的参考周 t 。

第三步,根据参考周的实际订足率 D_t ,结合该品规市场状态调整预期订足率 D_T 。当市场状态趋好时,上调订足率,即 $D_T > D_t$;当市场状态差时,下调订足率,即 $D_T < D_t$;当市场状态平稳时,订足率基本保持不变,即 $D_T \approx D_t$ 。

第四步,根据公式计算预测周销量 $Q_T = K_T \times D_T$ 。

由此可见,预测销量的准确率取决于预测订足率的准确率。但在实际工作中,存在以下难点:第一,可供量完全相同的参考期寻找难度大;第二,参考期与预测期的时间间隔越长,市场状态的波动对订足率的影响越大;第三,根据市场状态对预期订足率的调整受人为主观因素影响较大,缺乏量化指标。

二、文献综述

目前,卷烟销量预测主要使用线性回归方法,例如,王森等利用带季节指数的移动平均和最小平方对卷烟销量进行了预测;王伟民等利用灰色马尔可夫模型对全国卷烟销量进行了年度预测;康静等构建Prophet-VAR组合优化模型,对全国高值卷烟销量进行了年度预测。但这些卷烟销量预测方法存在一些不足:第一,预测期大多以年为单位,与工商网配的实时性不匹配;第二,预测范围大多是省、全国销量为单位,与工商网配的区域性不匹配;第三,预测对象大多以品牌或者地区总量为单位,与工商网配的颗粒度不

匹配。

随着数据挖掘技术的研究和发展,预测技术不断升级。张瑞等提出了一种基于多变量时间序列(MTS)及向量自回归(VAR)机器学习模型的预测方法,并进行了水驱油藏产量的实例应用。但上述预测方法尚未在烟草行业得到应用。

MTS分析是一种重要的数据分析方法,适用于按时间顺序排列的数据集,可以通过建立模型拟合过去的的数据,进而预测未来的趋势和变化。

VAR模型是一种用于处理多变量时间序列数据的机器学习模型,可以揭示多个时间序列数据之间的相互线性依赖关系。模型中每个变量都可以通过自身的滞后值、其他相关变量、常数项和误差项表示,并通过迭代实现对未来值的预测。

机器学习是一种模拟人类学习过程的方法,一般可以分为学习过程和决策过程两个主要阶段。在学习过程中,通过计算机对数据进行分析和处理,得出一定的模型或数字规律;在决策过程中,利用建立的模型或规律预测未来,提供合理的决策参考意见。

三、构建基于MTS及VAR模型的预测方法

基于以上分析,本文提出一种基于多变量时间序列(MTS)及向量自回归(VAR)机器学习模型的短期销量预测方法,通过MTS分析对可供量数据进行优选确定参考期,并将参考期内订足率和综合状态指数作为彼此相关的时间序列建立VAR模型,通过模型拟合预测周订足率,根据预测的订足率得到预测周销量,进而为工商网配提供参考。

(一)多变量时间序列分析

销量历史数据中包含了时间序列观测数据,例如订足率。同时,某一规格的订足率变化还受到市场状态的变化的影响,因此需要进行多变量时间序列(MTS)分析。

MTS分析的关键是相关性分析。相关系数是衡量两个变量之间相关程度的指标,通常用Pearson相关系数来衡量两个变量之间的线性相关程度。设有n个关于a和b两个变量的数据对 (a_i, b_i) , $i=1, 2, \dots, n$, 则其计算公式为:

$$r = \frac{\sum_{i=1}^n (a_i - \frac{1}{n} \sum_{i=1}^n a_i) (b_i - \frac{1}{n} \sum_{i=1}^n b_i)}{\sqrt{\sum_{i=1}^n (a_i - \frac{1}{n} \sum_{i=1}^n a_i)^2} \sqrt{\sum_{i=1}^n (b_i - \frac{1}{n} \sum_{i=1}^n b_i)^2}} \quad (1)$$

其中, r 的取值范围为 $[-1, 1]$; r 的符号表示变量之间的相关方向, $r>0$ 表示线性正相关, $r<0$ 表示线性负相关; $|r|$ 的大小可用来定量分析变量之间的线性相关程度。

(二)基于综合状态指数的向量自回归模型

1. 参数假设

预测期:要预测销量的那一周,记为 T 。

最优参考期:对某一个品规来说,在 $\{1, 2, \dots, T-1\}$ 这些周,寻找和第期的可供量最相似的那一周,记为 $t \in \{1, 2, \dots, T-1\}$ 。

$$\text{最优参考期的向量: } \mathbf{t} = \begin{bmatrix} t_1 \\ t_2 \\ \dots \\ t_n \end{bmatrix}。$$

订足率:记为 d ,其中第 T 周的订足率记为 d_T 。

$$\text{参考期的订足率向量:记为 } D_t, \text{ 即 } \mathbf{D}_t = \begin{bmatrix} d_{t1} \\ d_{t2} \\ \dots \\ d_{tn} \end{bmatrix}。$$

综合状态指数:记为 Z ,其中第 T 周的状态指数记为 Z_T 。

$$\text{参考期的综合状态指数向量:记为 } Z_t, \text{ 即 } \mathbf{Z}_t = \begin{bmatrix} Z_{t1} \\ Z_{t2} \\ \dots \\ Z_{tn} \end{bmatrix}。$$

说明:综合状态指数是ZJ中烟根据市场状态综合评价模型,集合市场价指数、零售价指数、社会库存可销天数、进货面、订足率、订足面等指标,按照一定规则和权重进行赋分,是对品规市场状态的综合评价得分。

2. 模型构建及核心步骤

第一,分析各个品规的可供量和历史状态。

第二,基于规则识别最优参考期 t 。

第三,自回归模型构建(预测未来 T 这一周的需求)

$$d_T = a + \alpha \mathbf{D}_t + \beta (\mathbf{Z}_{T-1} - \mathbf{Z}_t) \quad (2)$$

其中, a 为截距项, α 和 β 均为参数向量, $\alpha = \begin{bmatrix} \alpha_{t1} \\ \alpha_{t2} \\ \dots \\ \alpha_{tn} \end{bmatrix}$,

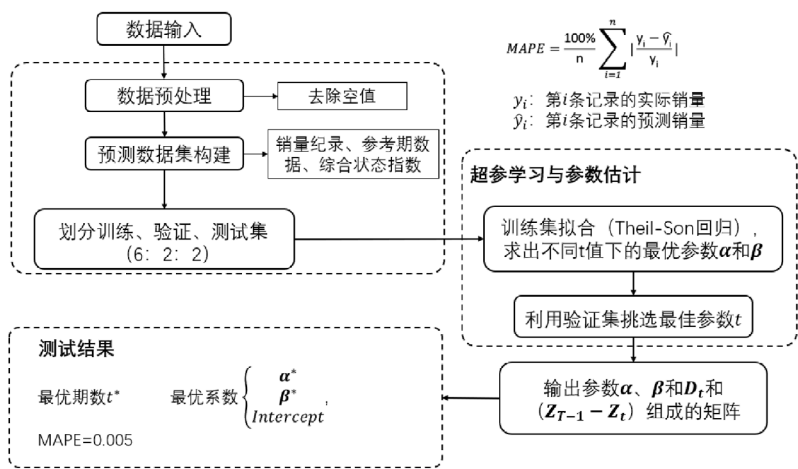
$$\beta = \begin{bmatrix} \beta_{t1} \\ \beta_{t2} \\ \dots \\ \beta_{tn} \end{bmatrix}, \text{ 需要通过数据和算法求解。}$$

注:加粗符号表示向量,不加粗符号为标量。

四、实验与结果

(一)实验设计

以某中烟某产品2022年第10周至2023年第15周数据为基础,进行订足率预测实验。在实验中,按6:2:2的比例将预测数据集划分为训练集、验证集和测试集,通过机器学习,得出最优参数 α 和 β ,以及最优参考期数 t 。利用MTS及VAR机器学习模型可以对订足率进行预测,具体实验流程如图1所示。



(二)结果与分析

在不同的*t*值下建立拟合函数 $d_t=a+aD_t+\beta(Z_{T-1}-Z_t)$,通过验证集数据预测订足率,并与实际订足率进行误差检验。去除空值的测试结果如表1所示。

表1 去除空值的测试结果

	<i>t</i> =1	<i>t</i> =2	<i>t</i> =3	<i>t</i> =4	<i>t</i> =5	<i>t</i> =6	<i>t</i> =7	<i>t</i> *=5
验证集 MAE	1.6823	0.8391	1.0146	1.0098	0.2841	0.2926	0.3061	
验证集 MAPE	0.0170	0.0085	0.0102	0.0102	0.0029	0.0029	0.0031	
测试集 MAE								0.5462
测试集 MAPE								0.0054

实验结果显示,当*t*=5时,得到的平均绝对百分比误差(MAPE)和平均绝对误差(MPE)最小,此时得到最优解。

$t^*=5$

$$\alpha^* = \begin{bmatrix} 0.0006 \\ -0.0011 \\ -0.0004 \\ -0.0002 \\ -0.0001 \end{bmatrix}$$
$$\beta^* = \begin{bmatrix} 0.0269 \\ 0.0294 \\ -0.0021 \\ 0.023 \\ 0.0409 \end{bmatrix}$$

为了进一步检验VAR模型预测的准确性,将其与传统经验预测的误差结果作对比,具体如表2所示。结果表明,相比传统经验预测方法,基于VAR模型的预测结果平均绝对百分比误差(MAPE)缩小了一半,准确率明显提升。

表2 不同方法预测结果对比

方法	误差 (MAPE)
VAR模型预测	0.54%
传统经验预测	1.18%

五、结语

针对商业周销量预测问题,本文提出了一种基于多变量时间序列(MTS)及向量自回归(VAR)机器学习模型的周销量预测方法,核心是在政策既定即可供量确定的前提下,对订足率的预测。实验结果表明,将该方法应用到订足率预测中是可行的。本文方法采取定量分析,无需依靠专家经验即可进行预测和评价,可以提高预测的客观性和准确性,有效规避供过于求、供非所求、供不应求等问题,提升响应市场、满足市场的能力,进一步提高货源管理效益,推动货源组织高质量发展。

参考文献:

[1]王森,吴春明.烟草系统卷烟需求预测模型的探讨[J].现代机械,2007(02):4-6+9.
[2]王伟民,汪沅,张国安.基于灰色马尔科夫模型的全国卷烟需求预测研究[J].中国烟草学报,2009,15(06):66-69.
[3]康静,姚春玲. Prophet-VAR组合优化模型在高值卷烟销量预测中的应用[J].中国烟草学报,2023,29(01):127-134.
[4]张瑞,贾虎.基于多变量时间序列及向量自回归机器学习模型的水驱油藏产量预测方法[J].石油勘探与开发,2021,48(01):175-184.
[5]杨海民,潘志松,白玮. 时间序列预测方法综述[J].计算机科学,2019,46(01):21-28.