

引入虚拟变量的时间序列分解法在卷烟销量预测中的应用^①

罗 彪, 闫维维, 万 亮

(中国科学技术大学 管理学院, 合肥 230026)

摘 要: 时间序列分解法依据时间序列的长期特征和季节性特征对未来进行合理预测, 但处理季节因素时, 在我国会受到传统节日的影响. 以时间序列分解法为基础, 将中国传统节日设定为虚拟变量, 构建基于时间序列分解法和虚拟变量的改进模型. 通过虚拟变量估测传统节日对序列的影响, 对传统方法进行适用性改进. 在对某省卷烟 90 个月总销量预测的算例中, 改进后的预测方法能够提高预测精度, 有利于企业据此合理安排生产销售计划.

关键词: 虚拟变量; 时间序列分解法; 卷烟销量; 预测

Application of Time-series Decomposition with Dummy Variables to Cigarette Sales Forecast

LUO Biao, YAN Wei-Wei, WAN Liang

(School of Management, University of Science & Technology of China, Hefei 230026, China)

Abstract: According to the long term and seasonal trends of time series, time-series decomposition makes a reasonable forecast of future, but when dealing with seasonal factors in China, it'll be influenced by Chinese traditional festivals. Based on time-series decomposition, this paper built a modified model consisting of time-series decomposition and dummy variables which represent Chinese traditional festivals. In an example of the 90 months' cigarette sales forecast in a province, the new model can effectively improve the prediction accuracy, and it's helpful for enterprises to make production and sales plans.

Key words: dummy variables; time-series decomposition; cigarette sales; forecast

我国卷烟行业推行“按订单组织货源”的背景下, 卷烟销量预测的精度直接影响到烟草公司采购计划及库存决策的制定. 影响卷烟销量的因素很多, 如市场规模、经济发展水平、季节波动等, 这些因素互相作用并共同影响卷烟的销量. 同时, 卷烟月度和季度销量具有明显的时间序列二重趋势变化特点, 即呈现整体趋势变动性和季节波动性^[1]. 对二重趋势预测常用的方法有线性回归法、神经网络、时间序列法等. 就我国烟草行业而言, 由于行业计划性极强, 烟民消费需求变化相对稳定, 基本不存在市场化竞争, 可以采用时间序列分解法进行销量预测.

时间序列分解法将序列分解为趋势变动(Trend, T)、季节变动(Seasonal, S)、周期变动(Circle, C)和不规则变动(Irregular, I), 能够很好地对时间序列进行预测^[2]. 但目前的时间序列分解法在分析季节性变动因素时, 主要针对阳历时间进行分析, 无法体现出中国农历节日如中秋、春节的消费特色, 无法准确预测节日性的销量增长^[3], 而这些特定节日本身对卷烟销量产生重要影响. 为解决这一问题, 本文在时间序列分解法的基础上, 引入虚拟变量, 把对卷烟销量有着重要影响的农历传统节日进行虚拟变量化, 以更好地拟合卷烟销量变化趋势来提高预测的精度.

^① 基金项目:国家自然科学基金委青年科学基金(70802058);教育部博士点基金新教师基金(200803581007);国家自然科学基金创新研究群体项目(70821001);安徽省自然科学基金第五批优秀青年科技基金(10040606Y35)

收稿时间:2012-04-23;收到修改稿时间:2012-05-20

1 卷烟月销量预测方法回顾

相比于年销量和季度销量预测, 卷烟月销量预测的难度更大. 年销量和季度销量预测所使用的影响因素的月度数据很难收集, 不便于做多元回归分析. 因此学术界通常使用时间序列法对卷烟月销量进行预测, 并把月销量预测值汇总预测季度或年度销量, 国内学者对卷烟月销量的预测方法可归纳如表 1.

表 1 卷烟月销量预测方法总结

作者	数据时间范围	研究主题	预测方法
谢星峰、谢东风、邹平(2007)	1999.01-2004.12	基于 CBP 的卷烟销售二重趋势时间序列预测模型研究与应用	趋势分解的神经网络乘积模型
宋作玲(2007)	2003.01-2006.09	卷烟配送中心订单预测及系统实现	GA-BP 神经网络模型
邹益民(2008)	2007.01-2007.12	基于农历的时间序列预测模型研究及其在卷烟销售预测中的应用	编程将以农历销售数据转换为农历销售数据
关牧玫(2009)	2006.01-2008.12	抚顺市烟草分公司卷烟销量预测研究	季节指数+ARMA
席玮(2009)	2002.01-2008.12	卷烟销量预测模型研究	人工神经网络+灰色模型
罗艳辉、吕永贵、李彬(2009)	2004.01-2007.12	基于 ARMA 的混合卷烟销售预测模型	ARMA+PERT(计划评审技术)
师彦锋(2010)	2006.01-2009.12	基于两种组合模型的安康市卷烟预测思想与运用研究	指数平滑+ARIMA 的组合模型

资料来源: 根据相关文献整理所得.

时间序列分析是根据系统观测得到的时间序列数据, 通过曲线拟合和参数估计来建立数学模型的理论和方法^[4]. 时间序列预测简单实用, 已有的文献对二重趋势时间序列预测提出了多种方法, 其中最常见的是自回归滑动平均(ARMA)模型, 而该模型要求时间序列数据经过差分后具有平稳性^[5], 另外在做多步预测时, 和指数模型和门限回归模型一样容易偏向平均值, 造成误差较大^[6]; BP 神经网络也得到广泛应用, 但它常常会忽略某些巨大噪音或非平稳数据并会忽视序列的整体增长趋势, 使其预测结果普遍低于实际观测值; 尽管 CBP 模型在结构上具有同向性, 对 BP 神经网络进行了部分改进, 但忽视了二重趋势时间序列的周期波动性, 其预测效果也相对较差^[7]; 灰色 G(1, 1)模型仅能较好地拟合时间序列的趋势性部分, 而对于周期波动性, 其预测精度则明显降低^[8]; 而传统的移动平均法和指数平滑法常出现滞后误差^[9].

在对卷烟月销量的预测方法中, 很少考虑农历节

日对阳历销量的影响. 实际预测中通常从定性的角度估计传统节日的影响, 在定量预测的基础上根据经验进行调整^[10], 这种预测方法容易受主观因素影响, 因此只能作为辅助手段. 邹益民(2008)^[11]将阳历销量数据转换为农历数据的预测方法, 不仅需要编程, 还要考虑闰月及各月实际天数等问题, 同时还要频繁地进行阳历和农历的转换, 不便于实际操作, 预测准确性不高, 相对误差最高仍为 132.63%. 本文考虑引入虚拟变量, 将传统节日定量化能够较好地解决这一问题, 同时在操作上更简单易行.

一般在计量经济学的回归模型中, 被解释变量不仅要受到定量因素的影响, 而且还受到定性因素的影响. 在回归模型中, 用人工变量将定性因素定量化, 赋值 0 和 1^[12]. 0 代表变量不具备某种属性, 1 代表变量具备某种属性, 这种人工变量称为虚拟变量(dummy variable). 虚拟变量在计量经济学的回归问题、协整分析等领域中常被用来解决变结构问题, 在金融分析、国民经济预测等方面都有着广泛的应用^[13]. 如果所研究的样本发生了结构性变动, 即回归模型中解释变量的参数发生变化, 则引入虚拟变量会使对问题的描述更简明, 一个方程能达到两个或多个方程的作用, 而且更接近现实^[14], 从而预测更准确. 在本文的预测模型中, 虚拟变量被用来对季节调整后的序列进行回归, 以反映传统的农历节日对卷烟销量的影响.

2 时间序列分解法

时间序列分解法将时间序列的各波动因素分解开, 分别分析趋势变动 T、季节变动 S、周期变动 C 和不规则变动 I, 以揭示时间序列的变化规律. 四种变动因素与原序列的关系被概括为两种模型^[15]:

加法模型(additive model):

$$y_t = T_t + S_t + C_t + I_t \tag{1}$$

乘法模型(multiplicative model):

$$y_t = T_t \cdot S_t \cdot C_t \cdot I_t \tag{2}$$

卷烟销量受以上四种因素的交互影响, 因此使用时间序列分解法中的乘法模型, 具体预测过程如下:

Step 1: 计算长期趋势 \hat{T} . 对原数据进行步长为 12 个月的移动平均, 再进行二次移动平均(即中心化过程), 将季节变动 S 和不规则变动 I 消除, 则所得移动平均的结果只包含长期趋势 T 和周期变动 C. 按时间序列对 TC 进行拟合, 得到卷烟销量的长期变化趋势

\hat{T} .

Step 2:剔除原始数据的长期趋势, 计算季节指数. 采用趋势剔除法分析, $Y/TC = SI$, 剔除了趋势变动 T 和周期变动 C , 新的数列只包括了 SI .

Step 3:标准化季节指数. 为消除不规则变动 I , 对将同月 SI 数据平均得到 $S_i(i=1,2,\dots,12)$, 再对其进行标准化处理, 得到调整后的季节指数为 $S_i / \left(\sum_{i=1}^{12} S_i / 12 \right)$.

Step 4:用季节指数还原预测结果. 将预测结果乘以对应的季节指数, 得到最终预测值:

$$\hat{y}_t = \hat{T}_t \cdot S_i \quad (i = 1, 2, \dots, 12) \quad (3)$$

3 实例分析

3.1 样本数据分析

本文依据某省 2004-2010 年卷烟销量月度数据构建预测模型, 将 2011 年 1-6 月的数据用于预测模型的检验. 在得到真实数据后, 需要对数据进行分析, 从中发现数据本身的特点和联系.

首先, 可将销售数据按时间先后顺序绘制散点图, 见图 1.

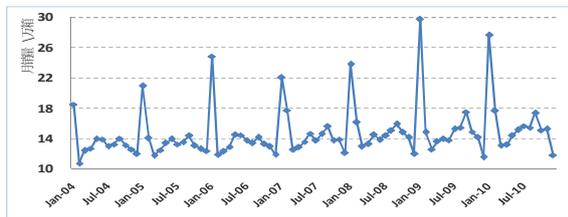


图 1 2004-2010 年卷烟月销量散点图

由图 1 可见卷烟月销量呈明显的周期性波动, 每年的 1、2 月份销量大幅增加, 在 3 月销量大幅回落, 9、10 月也有明显的增加, 在 11 月又大幅回落. 这充分显示了春节、中秋等传统节日对卷烟销量的影响很大, 正是因为这种影响造成的波动较大, 也增加了卷烟月销量预测的可能误差.

其次, 绘制历年各月的销量变化图, 纵坐标为月销量, 见图 2.

由图 2 可以看出, 除了每年 1、2 月份卷烟销量波动较大且呈上涨趋势外, 其余月份均呈平稳上涨趋势. 总体看有同期逐渐递增趋势, 但递增幅度不大.

通过观察图 1 和图 2 可以看出卷烟的月总销量呈

现出显著的季节性周期规律和明显的总体递增趋势.

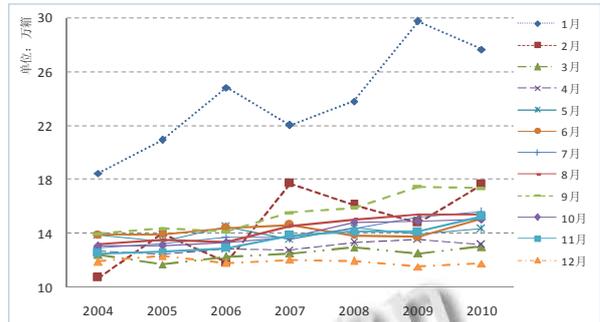


图 2 历年各月份卷烟销量散点图

3.2 时间序列分解法预测

根据第二节中时间序列分解法步骤, 对 2004.01~2010.12 的原始数据进行分析预测.

Step 1:计算长期趋势 \hat{T} . 将经过移动平均和中心化处理后的数据, 绘制成 TC 散点图(如图 3), 进行曲线估计.

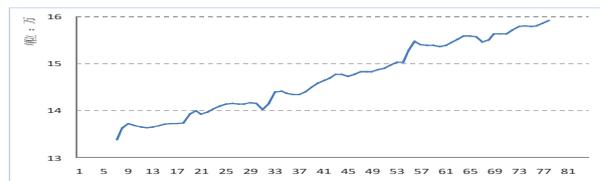


图 3 卷烟月销量 TC 散点图

由图 3 可以看出, 整个图像为一条斜率为正的直线, 因此考虑以时间 t^0 为自变量, 以 TC 为因变量做销量 \hat{T} 的线性拟合, 根据处理得到的卷烟销量数据, 用 Eviews^[16]拟合出模型的回归方程为:

$$\hat{T} = 132022.6 + 351.3205t \quad (4)$$

回归系数在 95% 的置信水平下都是统计显著的; 回归方程的决定系数 $R^2=0.983582$, 说明回归模型的拟合程度较高; F 检验的 P 值接近为 0, 该模型统计意义上显著.

Step 2:计算季节指数, 如表 2 所示.

表 2 卷烟月销量季节指数

月	1	2	3	4	5	6
S	1.6873	1.0407	0.8467	0.8804	0.9463	0.9626
月	7	8	9	10	11	12
S	0.9570	0.9715	1.0395	0.9415	0.9112	0.8153

Step 3:得出预测值. 根据公式(4)计算出 \hat{T} , 再根据公式(3)计算出最终预测值. 月度预测结果中, 相对

误差^②平均值为 4.451%，范围在 0.134%—29.499%间，其中 89.286%的相对误差落入 10%的区间内。月销量与预测结果对比如图 4 所示。

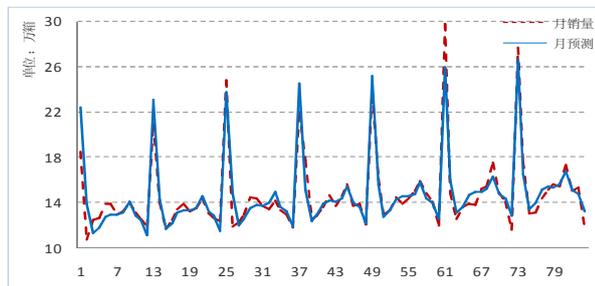


图 4 月销量预测效果对比

由图 4 可以看出，时间序列分解法能够较好地根据历史销量数据实现对卷烟销量的季节性、周期性两大特点的有效模拟，但部分序列点误差偏大。预测结果中，误差大于 5%的序列点有 28 个(如表 3 所示)，其中大于 10%的点基本出现 12、1、2 月份(由下划线标出)，其余分布在 4、5、8、9 月份。

表 3 相对误差大于 5%的序列点汇总

	<u>Jan04</u>	<u>Feb04</u>	Mar04	Apr04	May04	Jun04	Dec04
%	<u>21.08</u>	<u>29.50</u>	9.02	7.00	8.71	6.83	6.52
	<u>Jan05</u>	Dec05	<u>Feb06</u>	May06	Sep06	<u>Jan07</u>	Feb07
%	<u>10.20</u>	6.94	<u>24.50</u>	6.87	5.85	<u>11.11</u>	14.52
	Apr07	Jan08	Jun08	<u>Jan09</u>	Feb09	Mar09	May09
%	0.72	5.85	5.26	<u>12.98</u>	8.42	4.37	5.44
	Jun09	Sep09	<u>Dec09</u>	Feb10	Apr10	May10	<u>Dec10</u>
%	8.93	6.99	<u>11.45</u>	6.55	6.40	5.22	<u>12.07</u>

注：“%”代表相对误差

由表 3 的汇总结果表明即使剔除了季节波动，销量数据中仍包括了某些固定因素的影响，造成预测误差在固定的时期偏大，而这些序列点与中国传统节日相近，因此可以推测，以上几个月份预测误差较大，主要是因为传统的农历节日(如春节、端午、中秋)的影响。传统节日与阳历历法不同，而数据是以阳历为基础汇总的，所以传统节日期间存在异常与此相关，但这个因素干扰了周期性规律，造成预测误差随农历节日的分布出现随机波动，降低了预测的可靠性。

类似的结果及推测在其他学者的研究中也有提及。例如，谢星峰等(2005)^[17]在采用 CBP 方法对卷烟月销量进行预测时，年初年末的预测数据误差偏大，

他提出主要的原因是农历的春节假期与公历时间表存在偏差；师彦锋(2010)^[18]用指数平滑和 ARIMA 拟合出的数据在 1-2 月份与实际数据有较大的出入，相对误差范围也在 0.80%—34.62%，他也将原因解释为春节因素的影响。

以上学者只是指出由于春节的影响造成年初和年末的预测误差较大，并没有提出合理的处理方法来改进预测效果，因此，本文引入虚拟变量，希望能通过量化传统节日的影响来改进预测效果。

3.3 引入虚拟变量的时间序列分解法预测

传统的时间序列分解法利用季节指数和长期趋势的乘积预测卷烟销量，对季节类因素的处理方法仅限于移动平均，忽略了传统节日的影响，难以保证预测值与原始数据的拟合度，导致卷烟销量预测值与实际值有较大偏差。因此，本节引入虚拟变量来模拟传统节日的影响。

3.3.1 设定虚拟变量

我国农历节日对卷烟销售有极强的拉动作用，节前卷烟销量会形成峰值。阳历节日的日期固定，它们的影响已经包含在季节指数中，但传统的农历节日(如春节、端午、中秋)与公历变动不同步，因此需要对传统的农历节日进行针对性调整。

此外，春节跨度较长，按照中国传统习俗，春节从正月初一开始到正月十五结束。反映在阳历日期上，正月初一和正月十五可能出现在相邻的两个月份中，为更好地估计春节对卷烟月销量的影响，有必要将正月初一和正月十五分开设定。

因此，我们将新年(正月初一)、元宵、端午、中秋等四个传统节日设定为虚拟变量，具体如下：

$$D_{xn} = \begin{cases} 1, & \text{此月包含正月初一} \\ 0, & \text{此月不含正月初一} \end{cases} \quad D_{yx} = \begin{cases} 1, & \text{此月包含正月十五} \\ 0, & \text{此月不含正月十五} \end{cases}$$

$$D_{dw} = \begin{cases} 1, & \text{此月包含五月初五} \\ 0, & \text{此月不含五月初五} \end{cases} \quad D_{zq} = \begin{cases} 1, & \text{此月包含八月十五} \\ 0, & \text{此月不含八月十五} \end{cases}$$

3.3.2 调整时间序列分解法

因传统时间序列分解法的移动平均不只消除了不规则因素的影响，也将农历的传统节日当作不规则因素消除，所以我们需要调整原时间序列分解法，调整后的时间序列分解法的步骤如下：

Step 1:进行移动平均和中心化得到长期趋势和周期变动的乘积 TC。

Step 2:采用趋势剔除法计算季节指数. $Y/TC = SI$, 再做标准化既得季节指数.

Step 3:用剔除法 $Y/S_i = y_{tci}$, 得到包括了长期趋势、周期循环和不规则变动因素的序列 y_{tci} , 在此基础上将虚拟变量作为解释变量, 来测度传统节日对卷烟销量的影响. 回归模型如下:

$$\hat{y}_{tci} = \alpha + \beta_0 t + \beta_1 D_{xn} + \beta_2 D_{yx} + \beta_3 D_{dw} + \beta_4 D_{zq} + \beta_5 D_{xn} D_{yx} \quad (5)$$

以模型(5)为基础, 用 Eviews 进行回归, 并将统计上不显著的虚拟变量去除, 得到的回归模型如下:

$$\hat{y}_{tci} = 133640.1 + 314.3173t - 23825.75 D_{yx} + 33389.49 D_{xn} D_{yx} \quad (6)$$

回归系数在 95% 的置信水平下都是统计显著的; 回归方程的决定系数 $R^2=0.614984$, 说明回归模型的拟合效果较好^③; F 检验的 P 值接近为 0, 该模型统计显著; Durbin-Watson 检验值为 1.589842, 位于区间[1.5, 2.5]之间, 可判定在 95% 的置信水平下该模型残差项不存在正相关.

由公式(6)可知, 传统农历节日中, 端午和中秋节不会造成卷烟销量的大幅不规则变化, 而中国人最重要的新年和元宵节对卷烟销量的影响很大. 根据上式, 当新年和元宵节处于同一阳历月份时, 将会比只有新年的月份增加 9563.74^④箱卷烟销量, 而当新年和元宵节处于相邻月份时, 元宵所在月份的销量将会比正常情况减少 23825.75 箱.

这个变化规律也与实际情况相符. 烟草公司一般规定每周订货一次, 且不同类别的零售客户有不同的订货限量, 再加上烟草公司春节假期不接受订货, 因此卷烟零售商们为了避免货源节日期间供应不足的问题, 在正月初一之前就开始抓紧备货, 扩充库存. 所以当新年和元宵节在一个阳历月份时, 正月初一前的提前备货再加上元宵节期间的销量将会造成此月的销售量猛增; 而当新年和元宵节在相邻的两个阳历月份时, 提前备货已经在正月初一所在月完成, 新年与元宵节间只相差十五天, 上月新年前的备货在当月还未完全消化, 所以元宵节所在月的销量就会相对减少.

Step 4:得出预测值. 根据公式(6)计算出 \hat{y}_{tci} , 再与季节指数相乘计算出最终预测值. 月度预测结果中, 相对误差平均值为 3.843%, 范围在 0.112%—21.539%

间, 其中 92.857% 的相对误差落入 10% 的区间内.

3.4 预测效果对比

3.4.1 相对误差对比

表 4 是对两个模型相对误差的比较, 可以看出在 12、1、2 月份, 模型二对模型一均有较大程度改进, 模型一中误差较大的 2004 年 2 月、2006 年 6 月、2007 年 2 月、2010 年 2 月改进幅度很明显, 分别为 22.24%、20.44%、5.75%、4.99%. 其他除个别月份没有改进外, 都有 1% 左右的改进幅度.

表 4 两模型相对误差对比

	Jan04	Feb04	Mar04	Apr04	May04	Jun04	Dec04
1	21.08	29.50	9.02	7.00	8.71	6.83	6.52
2	21.54	7.76	7.99	5.97	7.73	5.86	5.72
	Jan05	Dec05	Feb06	May06	Sep06	Jan07	Feb07
1	10.18	6.93	24.50	6.87	5.85	11.11	14.52
2	11.09	6.45	4.06	6.51	6.14	11.30	8.77
	Apr07	Jan08	Jun08	Jan09	Feb09	Mar09	May09
1	0.72	5.85	5.26	12.98	8.42	4.37	5.44
2	0.82	5.71	4.99	13.34	8.85	3.89	4.91
	Jun09	Sep09	Dec09	Feb10	Apr10	May10	Dec10
1	8.93	6.99	11.45	6.55	6.40	5.22	12.07
2	8.36	7.54	10.71	1.56	5.60	4.40	11.04

注: 1、2 分别表示传统的时间序列分解法和改进后的时间序列分解法, 下同.

对两个模型的相对误差做散点图, 如图 5 所示, 效果更直观, 模型二的相对误差值普遍小于模型一的相对误差值.

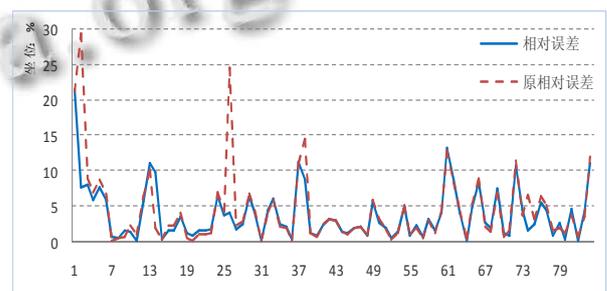


图 5 两模型相对误差对比图

通过对两个模型相对误差的比较, 可以看出在 12、1、2 月份, 模型二对模型一均有较大程度改进, 模型一中误差较大的 2004 年 2 月、2006 年 6 月、2007 年 2 月、2010 年 2 月改进幅度很明显, 分别为 22.24%、20.44%、5.75%、4.99%. 其他除个别月份没有改进外, 都有 1% 左右的改进幅度.

另外,改进后的相对误差的平均值和波动范围更小,因此从误差的角度可以认为模型二的预测精度更高。

3.4.2 拟合度对比

分别计算两种模型预测值和实测值的拟合度,拟合度^[19]=1-残差平方和/原始序列平方和。结果表明,模型一预测值与实测值的拟合度为 89.627%,模型二的拟合度为 91.931%。

由拟合度对比可以看出,模型二拟合度更理想。

3.4.3 预测能力对比

因为有些模型虽然可以实现对历史数据的较好拟合,但预测能力较弱^[20]。为了对两种模型的预测能力进行比较,我们用两种预测方法分别对 2011 年的前六个月进行预测,将预测期内的数据作为检验数据,以样本标准误差作为预测精度的评价指标,其定义为:

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (7)$$

其中 n 为预测样本个数, \hat{y}_i 为预测销量, y_i 为实际销量。结果表明,模型一的 σ 为 34923.320,模型二的 σ 为 30301.900,可见改进后的时间序列分解法预测能力更好。

为了更清晰地比较这两种模型的预测能力,将两种模型的预测量与实际销量进行比较,如表 5 所示。

表 5 两模型对 2011 年上半年的预测结果对比

月份	实际销量	模型 1		模型 2		优化 %
		预测值	相对误差	预测值	相对误差	
1	333631.00	273152.58	18.127	273574.73	18.001	0.13
2	122880.83	168842.17	37.403	142417.78	15.899	21.50
3	131680.21	137660.85	4.542	136304.64	3.512	1.03
4	132363.62	143446.35	8.373	142003.62	7.283	1.09
5	142791.38	154524.85	8.217	152939.02	7.107	1.11
6	151875.52	157517.97	3.715	155869.27	2.630	1.09

由表 5 可以看出,模型二预测的整体效果优于传统的时间序列分解法,尤其是 2 月份,优化幅度高达 21.504%,其余月份均有 0.127%—1.111%不同程度的优化。

因此,从对未来时间点的预测能力来看,引入虚拟变量的时间序列分解法预测效果更好,更符合实际。

4 结语

时间序列分解法以阳历时间为基础,未考虑我国

农历节日的影响,由于农历和阳历不同步,农历节日就成为浮动的影响因素,如果未对其进行分析和修正,则会在较大程度上影响预测精度。

本文通过对某省 2004 年至 2010 年卷烟月销量样本的检验,引入虚拟变量的时间序列分解模型更加贴切地拟合卷烟月销量变化的规律和趋势,并能具体测量出传统节日对卷烟销量的影响程度。此外,误差、拟合度和预测能力三方面的对比均表明,改进后的模型能够显著地提高预测的准确性,能够很好地帮助烟草公司设置安全库存以及减少资金占用。同时,模型中的虚拟变量可以根据实际预测需要进行设定,具有高度的灵活性和可行性,能够对其他方面的预测工作起到较好的借鉴作用。

参考文献

- 1 罗艳辉,吕永贵,李彬.基于 ARMA 的混合卷烟销售预测模型.计算机应用研究,2009,26(7):2664-2668.
- 2 Armstrong SJ, Collopy FJ, Yokum T. Decomposition by causal forces: a procedure for forecasting complex time series. International Journal of Forecasting, 2005,21:25-36.
- 3 席玮.卷烟销量预测模型研究.合肥:合肥工业大学,2009.
- 4 Sun ZL, Choi TM, Au KF. Sales forecasting using extreme learning machine with application in fashion retailing. Decision Support Systems, 2008,46:411-419.
- 5 Guillenb RA, Marqueza L, Pasadasa M. Hybridization of intelligent techniques and ARIMA models for time series prediction. Fuzzy Sets and System, 2008,159:821-845.
- 6 张晓峒,王健,杜勇宏.季节虚拟回归中参数及统计量的分布特征.系统工程学报,2006,24(4):361-365.
- 7 宋仙磊,刘业政,陈思凤,许波.二重趋势时间序列的灰色组合预测模型.计算机工程与应用,2011,47(8):115-117.
- 8 方匡南,朱建平,谢邦昌.有序聚类虚拟变量法及其应用.数理统计与管理,2012,31(2):333-340.
- 9 郭秀英,尹兴国,张艳云.季节性波动时间序列预测的分解季节指数法.数理统计与管理,2000,19(6):60-64.
- 10 宋作玲.卷烟配送中心订单预测及系统实现.济南:山东大学,2007.
- 11 邹益民.基于农历的时间序列预测模型研究及其在卷烟销售预测中的应用.呼和浩特:内蒙古大学,2008.

(下转第 148 页)

表 2 不同阈值函数去噪后的 SNR 和 MSE

处 理 方 法	评 价 依 据	阈值函数				
		固定 阈值	stein 无偏似 然估计阈值	启发式 阈值	极大极 小阈值	改进 阈值
硬 阈 值	S N R	72.0181	82.5166	82.5166	74.1395	83.0985
	M S E	11.2671	11.3193	11.3193	11.3036	8.8748
软 阈 值	S N R	79.8636	73.5837	73.5837	88.5789	89.8234
	M S E	9.3826	11.2646	11.2646	10.1489	6.3935

由图 1~4 和表 2 可知, 采用改进阈值函数软阈值去噪算法去噪效果较好, 有效地去除了含噪信号中混有的基线漂移、工频干扰、肌电干扰, 并且心电波形畸变较少、光滑性较好, 较好地保留了心电信号中有用成分。

3 结语

本文给出了一种基于小波阈值的心电信号去噪算法, 并对不同的阈值函数分别进行了硬阈值和软阈值的去噪处理, 提出了改进的阈值函数, 从 Matlab 仿真结

果可以清楚的看出, 改进小波阈值的软阈值去噪算法去噪效果较好, 能够较好地去除心电信号中混杂的基线漂移、工频干扰和肌电干扰, 并较好地保留了心电信号中有用信息, 有利于进一步的波形识别和特征提取。

参考文献

- 1 赵晴,赵捷,魏珑.基于小波变换的心电信号去噪算法.现代生物医学进展,2007,7(10):1566-1568.
- 2 王佳文,凡友华.心电信号去噪中的小波方法.数理医药学杂志,2009,22(1):85-89.
- 3 马文.基于小波变换的心电信号去噪研究[硕士学位论文].兰州:兰州大学,2009.
- 4 潘泉,孟晋丽,张磊,程咏梅,张洪才.小波滤波方法及应用.电子与信息学报,2007,29(1):236-242.
- 5 文莉,刘正士,葛运建.小波去噪的几种方法.合肥工业大学学报(自然科学版),2002,25(2):167-172.
- 6 潘磊,张军,邹采荣.基于小波变换的一种心电信号去噪算法.科技资讯,2006(36):53-54.
- 7 宋喜国,邓亲恺.MIT-BIH 心率失常数据库的识读及应用.中国医学物理学杂志,2004,21(4):230-232.
- 8 Daubechies I. The wavelet transform, time-frequency localization and signal analysis. IEEE Transactions on Information Theory,1990:961-1005.
- 9 Alfaouri M, Daqrouq K. ECG Signal Denoising By Wavelet Transform Thresholding. American Journal of Applicable Sciences, 2008, 5(3):276-281.

(上接第 220 页)

- 12 李道清,任玉珑,胡大江,朱丽娜.基于虚拟变量法的我国交通能源消费量预测.科技管理研究,2005,11:139-141.
- 13 刘中学,张毅.包含虚拟变量的澳门酒店入住率预测模型及实证.统计与决策,2007,(20):98-100.
- 14 据亮,贺昌政,何跃.四川省六大城市资产负债率预测研究.经济研究,2010,11:37-38.
- 15 Harvey AC. The Econometric Analysis of Time Series. The MIT Press, 1991:79-95.
- 16 易丹辉.数据分析与 Eviews 应用.北京:中国人民大学出版社,2008.101-120.
- 17 谢星峰,谢东风,邹平.基于 CBP 的卷烟销售二重趋势时间序列预测模型研究与应用.控制理论与应用,2007,24(6):1015-1020.
- 18 师彦锋.基于两种组合模型的安全市卷烟预测思想与运用研究.经济师,2010,9:231-233.
- 19 尚红云.税收收入模型预测精度的比较.统计与决策,2008,3:40-43.
- 20 庞晓波,孙叶萌,王晨.基于 ANN 方法对汇率波动非线性的检验与预测比较.吉林大学社会科学学报,2008,48(1):76-81.