

ARIMA 模型在农产品价格预测中的应用

刘峰^{1,2}, 王儒敬¹, 李传席^{1,2}

LIU Feng^{1,2}, WANG Ru-jing¹, LI Chuan-xi^{1,2}

1.中国科学院 合肥智能机械研究所, 合肥 230031

2.中国科学技术大学 自动化系, 合肥 230027

1.Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230031, China

2.Department of Automation, University of Science and Technology of China, Hefei 230027, China

E-mail: fengl@mail.ustc.edu.cn

LIU Feng, WANG Ru-jing, LI Chuan-xi. Application of ARIMA model in forecasting agricultural product price. Computer Engineering and Applications, 2009, 45(25): 238-239.

Abstract: The forecast of future values of a time series based on the current and past values of the prices of agricultural products is helpful to guide agricultural products circulation and production, balance the supply and demand in different regions and provide the basis for agricultural restructuring by the government and farmers. To address this important issue, this paper introduces the non-stationary time series model ARIMA(p, d, q) built on monthly prices of cabbage and attempts to forecast the price in the coming months. The result proves that ARIMA(0, 1, 1) can correctly simulate and forecast the price trend of cabbage, providing an important method for the accurate forecast of agricultural product market information.

Key words: prices of agricultural products; time series; Autoregressive Integrated Moving Average Model (ARIMA); price trend

摘要: 利用农产品价格时间序列的当前值和过去值准确预报未来值, 将有利于正确引导农产品流通和农业生产, 实现农产品区域供求平衡, 并为政府和农户提供结构调整的依据。针对农产品价格这一重要问题, 以白菜月价格数据为例, 构建非平稳时间序列 ARIMA(p, d, q) 模型并预测白菜未来的月价格。结果表明 ARIMA(0, 1, 1) 模型能很好地模拟并预测白菜月价格趋势, 为农产品市场信息的准确预测提供重要方法。

关键词: 农产品价格; 时间序列; 自回归移动平均模型; 价格趋势

DOI: 10.3778/j.issn.1002-8331.2009.25.073 文章编号: 1002-8331(2009)25-0238-02 文献标识码: A 中图分类号: TP311

1 引言

农产品价格数据对于指导农产品生产以及调整农业结构具有重要意义。从而使用农产品的历史价格数据对未来价格趋势进行有效的预测显得尤为重要。一般常用的有时间序列预测和回归预测法, 时间序列预测法一面承认事物发展的延续性, 另一方面又充分考虑到事物发展偶然因素的影响产生的随机性, 为了消除随机波动的影响, 利用历史数据, 进行统计分析, 并用加权平均等方法对数据加以适当的处理, 进行趋势预测。该文将要采用 ARIMA 模型^[1]来预测农产品价格时间序列。

ARIMA 模型全称为自回归移动平均模型^[2](Autoregressive Integrated Moving Average Model, ARIMA), 是由博克思(Box)和詹金斯(Jenkins)于 20 世纪 70 年代初提出的一著名时间序列预测方法, 所以又称为 Box-Jenkins 模型、博克思-詹金斯法。

2 ARIMA 模型的基本思想及数学模型

ARIMA 模型的基本思想是: 将预测对象随时间推移而形成的数据序列视为一个随机序列, 用一定的数学模型来近似描述这个序列。这个模型一旦被识别后就可以从时间序列的过去值及现在值来预测未来值。现代统计方法、计量经济模型在某种程度上已经能够帮助企业对未来进行预测。

数学模型: 指定 3 个参数 p, d, q 用来分析时间序列, 即描述自回归阶数 p 、差分次数 d 和移动平均阶数 q , 通常记为 ARIMA(p, d, q), 表示为:

$$\varphi_p \nabla^d Z_t = \theta_q(B) \alpha_t$$

式中, Z_t 为原序列; α_t 为白噪声序列, 是一列相互之间无关, 其均值为 0, 方差为 σ^2 的随机变量序列; B 为后移算子, 定义为 $BZ_t = Z_{t-1}$, 从而 $B^m Z_t = Z_{t-m}$; φ_p 为自回归算子, $\varphi_p(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots -$

基金项目: 国家自然科学基金(the National Natural Science Foundation of China under Grant No.60774096); 国家高技术研究发展计划(863)(the National High-Tech Research and Development Plan of China under Grant No.2006AA10Z23702); 国家科技支撑计划(the National Scientific and Technical Supporting Programs of China under Grant No.2006BAD10A0502, No.2006BAD10A1410)。

作者简介: 刘峰(1982-), 男, 硕士研究生, 主要研究方向为人工智能、数据挖掘; 王儒敬(1964-), 男, 博士, 研究员, 主要研究方向为智能决策支持系统、数据挖掘; 李传席(1981-), 男, 硕士研究生, 主要研究方向为信息获取、数据挖掘。

收稿日期: 2008-05-07 **修回日期:** 2008-07-25

$\varphi_p B^p$, p 为模型的自回归阶数; θ_q 为移动平均算子, $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$, q 为模型的移动平均阶数; ∇ 为向后差分算子, 它可用 B 表示, 因为, $\nabla Z_t = Z_t - Z_{t-1} = (1-B)Z_t$; θ_0 为常数项, 定义 $\theta_0 = u(1 - \varphi_1 - \varphi_2 - \dots - \varphi_p)$, 其中 u 为平均数。

3 ARIMA 模型预测的基本程序

ARIMA 模型预测的基本程序为^[3]:

(1) 根据时间序列的散点图、自相关函数和偏自相关函数图以 ADF 单位根检验其方差、趋势及其季节性变化规律, 对序列的平稳性进行识别。一般来讲, 农产品价格数据的时间序列都不是平稳序列。

(2) 对非平稳序列进行平稳化处理。如果数据序列是非平稳的, 并存在一定的增长或下降趋势, 则需要对数据进行差分处理, 如果数据存在异方差, 则需对数据进行技术处理, 直到处理后的数据的自相关函数值和偏相关函数值无显著地异于零。

(3) 根据时间序列模型的识别规则, 建立相应的模型。若平稳序列的偏相关函数是截尾的, 而自相关函数是拖尾的, 可断定序列适合 AR 模型; 若平稳序列的偏相关函数是拖尾的, 而自相关函数是截尾的, 则可断定序列适合 MA 模型; 若平稳序列的偏相关函数和自相关函数均是拖尾的, 则序列适合 ARI-MA 模型。

- (4) 进行参数估计, 检验是否具有统计意义。
- (5) 进行假设检验, 诊断残差序列是否为白噪声。
- (6) 利用已通过检验的模型进行预测分析。

4 实验

从 www.sounong.net 上抽取 2004 年 1 月到 2007 年 9 月某市白菜的月价格数据, 建立一个简单合理预测准确的模型来预测 2007 年 10 月至 2008 年 2 月该市的白菜月价格, 并与实际价格相比较, 如模型合理则预测比较准确, 若不合理则优化参数直至达到预期的结果。

4.1 数据稳定性检验

从图 1 时间序列图可以看出, 2004 年 1 月~2007 年 9 月该市白菜价格呈上升趋势, 增长幅度不同, 这说明此数列既存在上升趋势又存在方差不齐, 需进行平稳化处理, 对原序列数据进行一阶差分转换并对转换得到的新序列作自相关和偏相关图, 分别如图 2 和图 3 所示。自相关图显示自相关系数很快落入置信区间, 说明新的时间序列具有平稳性。

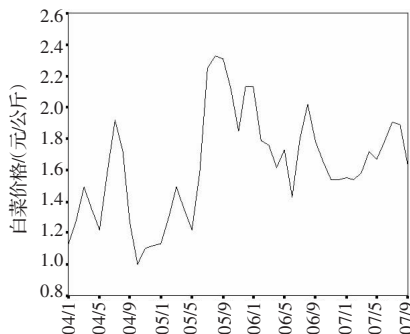


图 1 2004 年 1 月至 2007 年 9 月白菜价格时序图

4.2 参数估计与模型选择

利用 SPSS11.5 软件对不同模型进行拟合, 得到其中的参数如表 1 所示, 其中 AIC-Akaike^[4]信息准则, AIC 值较小的模型

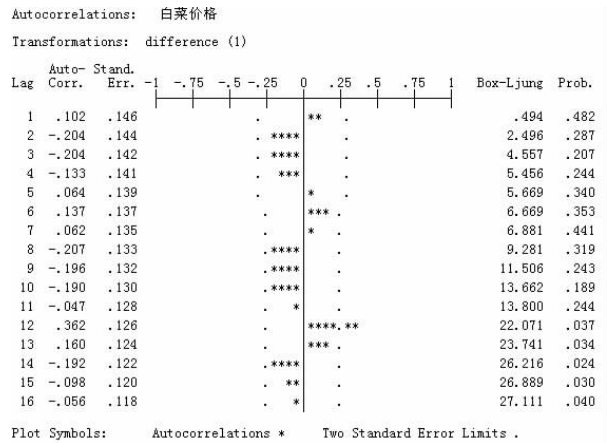


图 2 原时间序列一阶自相关图

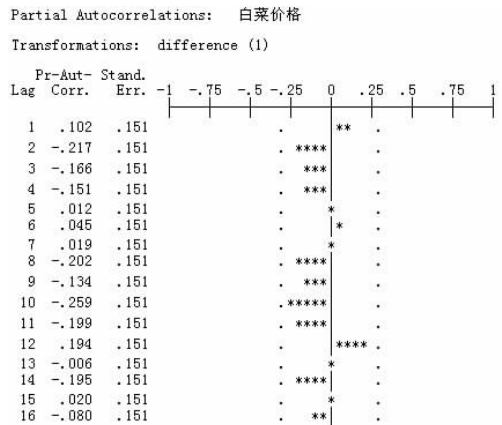


图 3 原时间序列一阶偏相关图

表 1 统计量参数值对比

ARIMA	(0,1,0)	(0,1,1)	(1,1,0)	(1,1,1)
Standard error	0.224 557 35	0.216 524 63	0.223 162 44	0.215 797 29
Log likelihood	3.786 168 2	5.801 234 2	4.543 198 9	6.431 301 7
AIC	-5.572 336 3	-7.602 468 4	-5.086 397 9	-6.862 603 3
SBC	-3.788 146 7	-4.034 089 1	-1.518 018 6	-1.510 034 4

相对较好; SBC-Schwarz 贝叶斯准则, 比 AIC 多考虑了残差个数的影响, SBC 较小的模型较好。故选择 AIC 和 SBC 均最小的 ARIMA(0,1,1)模型进行拟合。

4.3 时间序列的预测与检验

通过分析 2004 年 1 月~2007 年 9 月白菜的价格数据, 得到 ARIMA 最优模型 ARIMA(0,1,1), 利用该模型对 2007 年 10 月~2008 年 2 月的价格进行预测并作图, 如图 4 所示。在表 2 中列出实际数据和预测数据, 并分析相对误差。

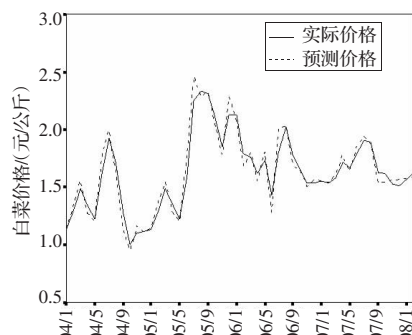


图 4 预测效果图