第 33 卷 第 5 期 电子科技大学学报 Vol.33 No.5
2004 年 10 月 Journal of UEST of China Oct. 2004

绿色农产品定价方法的ARIMA状态空间研究

魏 来[1]，陈良均[2]

(1. 四川农业大学都江堰分校 四川都江堰 611830; 2. 电子科技大学应用数学学院 成都 610054)

【摘要】根据当前绿色农产品定价研究的现状，从网络环境提供的充分竞争性、信息公开性以及海量数据出发，提出了引入ARIMA过程的状态空间模型的原因。通过ARIMA过程的状态空间表示和用于预报的Kalman滤波关系式，讨论了在状态空间模型时的似然函数和新息方法求解，探讨了ARIMA过程的状态空间模型的精确预报方法。其结果对大力发展绿色农产品，改善农民的收入结构和国家的农产品更好地打破出口的“绿色壁垒”有积极的理论和现实意义。

关 键 词 绿色农产品; 网络环境; 状态空间; 似然函数

中图分类号 O211.61; F304.2 文献标识码 A

Research of Green Produce Price Method by ARIMA Space Model

Wei Lai[1]，Chen Liangjun[2]

(1. Dujiangyan Campus of Sichuan Agriculture Univesity，Sichuan Dujiangyan 611830;

2. School of Applied Mathematics UEST of China Chengdu 610054)

Abstract Base on the actuality of green produce make a price research, start form sufficient competition and information opening and mass data offered by the network. Put forward the reason of introduce the state space model of the ARIMA process, through the express of its state space model and Kalman filter relation for forecast, discuss the likelihood function and fresh interest means which based on state space model, probe into exact forecast means of the state space model of the ARIMA process.To develop green farm produce energetically, improve the income structure of farms, the produce of our country break the “green bulwark” of export, has active theoretic and realism meaning.

Key words green produce; network environment; state space; likelihood function

绿色农产品普遍具有生命周期较短的特点，随着互联网络的迅速发展，人们在很短的时间内就可以通过电视、互联网络知道全国乃至世界范围内某种产品的报价，及时的调整可以获取更大的利益，从而连带效应的结果也能使广大的农民知道应该种植什么样的绿色农产品,所以首先面临的是如何处理海量的价格数据。在预报具体的经济活动时，许多时间序列适合于用非平稳模型来描述，因为它们都不具有在时间上不变的均值水平，所对应的随机模型是自回归求和滑动平均过程(ARIMA过程)，这类过程提供了较广的模型范围：平稳模型；非平稳模型。该预测方法比较简明，ARIMA预测模型不直接考虑其他相互因素的变动，只要求掌握必要的计算手段，就可以作指标数量不大，但预测频度较高的预测工作。然而一般的预测，往往从单一的品种入手，由于经济生活中要考虑的因素较多，平常所用的ARIMA方法无法满足现实的需要，于是开始将目光转移到从多个品种、多个变量，利用其状态空间模型从多维的角度研究。

收稿日期：2004– 06 – 04

作者简介：魏 来(1969 –)，男，在职硕士生，讲师，主要从事信息管理与电子商务技术方面的研究；陈良均(1944 –)，男，教授，主要从事随机过程研究.

1　ARIMA过程的状态空间模型及求解

1.1　ARIMA过程的状态空间模型表示

状态空间模型用于精确预报，为了预报也为了模型的设定和参数的极大似然估计，用状态空间形式来表示ARIMA模型是发展的趋势[1](大写字母除特别说明外，均表示向量)。

对于$ARIMA(p,d,q)$过程，$f(B)z_t = q(B)a_t$，其中，$f(B)$为自回归算子是p阶，$q(B)$为滑动平均算子是q阶，差分次数取为d，a_t表示白噪声过程。定义预报值如下

$$\hat{z}_t(j) = E_t[z_{t+j}] \qquad j = 0,1,\cdots,r-1 \tag{1}$$

其中　$r = \max(p+d, q+1)$，及$\hat{z}_t(0) = z_t$，

$$\hat{z}_t(j-1) = \hat{z}_{t-1}(j) + y_{j-1} + a_t \qquad j = 1,2,\cdots,r-1 \tag{2}$$

此外，当 j 等于r，且$r>q$时，有

$$\hat{z}_t(j-1) = \hat{z}_{t-1}(j) + y_{j-1}a_t = \sum_{i=1}^{p+d} j_i \hat{z}_{t-1}(j-i) + y_{j-1}a_t \tag{3}$$

从而可以将k时刻的“状态”向量$\boldsymbol{Y}_t$定义为具有r个分量的形式

$$\boldsymbol{Y}_t = (z_t, \hat{z}_t(1), \cdots, \hat{z}_t(r-1))' \tag{4}$$

$\boldsymbol{Y}_t$满足方程

$$\boldsymbol{Y}_t = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ j_r & j_{r-1} & j_{r-2} & \cdots & j_1 \end{bmatrix} \boldsymbol{Y}_{t-1} + \begin{bmatrix} 1 \\ y_1 \\ \\ \vdots \\ y_{r-1} \end{bmatrix} a_t \tag{5}$$

当$i > p+d$时，$j_i = 0$，有

$$\boldsymbol{Y}_t = \boldsymbol{F}\,\boldsymbol{Y}_{t-1} + \boldsymbol{Y}\,a_t \tag{6}$$

及量测方程：

$$\boldsymbol{Z}_t = z_t + \boldsymbol{N}_t = [1,0,\cdots,0]\boldsymbol{Y}_t + \boldsymbol{N}_t = \boldsymbol{H}\,\boldsymbol{Y}_t + \boldsymbol{N}_t \tag{7}$$

假设过程z_t是在附加白噪声的条件下而被观测到的，则上式的附加噪声$\boldsymbol{N}_t$存在，否则便认为$\boldsymbol{Z}_t = \boldsymbol{H}\,\boldsymbol{Y}_t$，以上两式构成了模型状态空间表示。由于$\boldsymbol{Y}_t$的结构不是唯一的，所以ARIMA 模型的状态空间形式也不是唯一的。状态空间模型的稍微一般形式的状态方程

$$\boldsymbol{Y}_t = \boldsymbol{F}\,\boldsymbol{Y}_{t-1} + a_t \tag{8}$$

量测方程

$$\begin{cases} \boldsymbol{F}_t = \boldsymbol{F} \\ \boldsymbol{H}_t = \boldsymbol{H} \\ Z_t = \boldsymbol{H}\,\boldsymbol{Y}_t + \boldsymbol{N}_t \end{cases} \tag{9}$$

式中　a_t和$\boldsymbol{N}_t$是相互独立的白噪声过程，而a_t是协方差阵为Σ_a的向量白噪声过程，$\boldsymbol{N}_t$是方差为s_N^2的白噪声。状态向量(非观测的)$\boldsymbol{Y}_t$概括了动态系统在t时刻的状态，状态方程刻画了动态系统随时间的演变，而量测方程则表明，观测值z_t是受附加白噪声干扰的状态变量的线性组合所生成的。

在式(8)中的矩阵$\boldsymbol{F}_t$是$r\times r$的状态转移矩阵，式(9)中的$\boldsymbol{H}_t$是$r\times r$的向量，它们均可随时间而变化，通常在应用中都是作为常数矩阵，$\boldsymbol{F}_t = \boldsymbol{F}, \boldsymbol{H}_t = \boldsymbol{H}$，即对于所有$t>0$，这些矩阵都不依赖于$t$。在这种情况下，系统或模型称为非时变的。在状态空间模型中状态向量$\boldsymbol{Y}_t$的最小维数必须足够大，以使系统的动态特征能用形如式(8)的简单的一阶Markov结构描述。为了减少地方性的消费偏好等造成的误差，尽可能的拟合现实情况，在对于绿色农产品网络环境下的定价问题，本文采用农业部发布价格信息常用的19种蔬菜：大白菜，

西红柿，黄瓜，青椒，芹菜，土豆，白萝卜，茄子，豆角，胡萝卜，菜花，韭菜，蒜薹，大葱，葱头，油菜，菠菜，洋白菜，莴笋。

1.2 用于预报的Kalman滤波关系式及新息方法求解求解

对于ARIMA过程的状态空间形式，可利用Kalman滤波方程：从某些适当的初值$\boldsymbol{Y}_0=\hat{\boldsymbol{Y}}_{0|0}$及$\boldsymbol{V}_0=\hat{\boldsymbol{V}}_{0|0}$，最优滤波估计$\hat{Y}_{t|t}$可以用以下递推关系给出

$$\hat{\boldsymbol{Y}}_{t|t}=\hat{\boldsymbol{Y}}_{t|t-1}+\boldsymbol{K}_t(\boldsymbol{Z}_t-\boldsymbol{H}_t\hat{\boldsymbol{Y}}_{t|t-1}) \tag{10}$$

其中

$$\boldsymbol{K}_t=\boldsymbol{V}_{t|t-1}\boldsymbol{H}_t'[\boldsymbol{H}_t\boldsymbol{V}_{t|t-1}\boldsymbol{H}_t'+s_N^2]^{-1} \tag{11}$$

且有

$$\hat{\boldsymbol{Y}}_{t|t-1}=\boldsymbol{F}_t\hat{\boldsymbol{Y}}_{t-1|t-1},\boldsymbol{V}_{t|t-1}=\boldsymbol{F}_t\boldsymbol{V}_{t-1|t-1}\boldsymbol{F}_t'+\Sigma_a \tag{12}$$

$$\boldsymbol{V}_{t|t}=[\boldsymbol{I}-\boldsymbol{K}_t\boldsymbol{H}_t]\boldsymbol{V}_{t|t-1}=\boldsymbol{V}_{t|t-1}-\boldsymbol{V}_{t|t-1}\boldsymbol{H}_t'[\boldsymbol{H}_t\boldsymbol{V}_{t|t-1}\boldsymbol{H}_t'+s_N^2]^{-1}\boldsymbol{H}_t\boldsymbol{V}_{t|t-1} \qquad t=1,2,\cdots \tag{13}$$

从而得出精确(有限样本)的提前一期预报$Z_{t-1}(l)=E[\boldsymbol{Z}_t\,|\,\boldsymbol{Z}_{t-1},\cdots,\boldsymbol{Z}_1]$及其误差方差$\boldsymbol{V}_t=\boldsymbol{H}\boldsymbol{V}_{t|t-1}\boldsymbol{H}'$。对于计算ARIMA过程$n$个观察值$\boldsymbol{Z}_t,\cdots,\boldsymbol{Z}_n$的似然函数非常有用，并可用于模型参数的极大释然估计[2]。

在讨论利用Kalman滤波方程求解时，首先假定来自于$ARMA(p,q)$模型的n各给定的观察值为$w_1,w_2,\cdots,w_n$，Kalman滤波方程(14~17)已经给出了状态向量Y_t的一步外推预报值：$\hat{Y}_{t|t-1}=E[\boldsymbol{Y}_t\,|\,w_{t-1},\cdots,w_1]$，以及误差协方差阵$\boldsymbol{V}_{t|t-1}=E[(\boldsymbol{Y}_t-\hat{\boldsymbol{Y}}_{t|t-1})(\boldsymbol{Y}_t-\hat{\boldsymbol{Y}}_{t|t-1})']$。对于$ARMA(p,q)$模型的状态空间形式，$t$=1,2,⋯,$n$, 递推方程为

$$\hat{\boldsymbol{Y}}_{t|t}=\hat{\boldsymbol{Y}}_{t|t-1}+\boldsymbol{K}_t(w_t-\hat{w}_{t|t-1}) \tag{14}$$

其中

$$\boldsymbol{K}_t=\boldsymbol{V}_{t|t-1}\boldsymbol{H}'[\boldsymbol{H}\,\boldsymbol{V}_{t|t-1}\boldsymbol{V}']^{-1}$$

$$\hat{w}_{t|t-1}=\boldsymbol{H}\hat{\boldsymbol{Y}}_{t|t-1}$$

且有

$$\begin{cases}\hat{\boldsymbol{Y}}_{t|t-1}=\boldsymbol{F}\,\hat{\boldsymbol{Y}}_{t-1|t-1}\\ \boldsymbol{V}_{t|t-1}=\boldsymbol{F}\,\boldsymbol{V}_{t-1|t-1}\boldsymbol{F}'+s_a^2\boldsymbol{Y}\boldsymbol{Y}'\end{cases} \tag{15}$$

其中

$$\boldsymbol{V}_{t|t}=[\boldsymbol{I}-\boldsymbol{K}_t\boldsymbol{H}]\,\boldsymbol{V}_{t|t-1} \tag{16}$$

预报向量的第一分量是$\hat{w}_{t|t-1}=\boldsymbol{H}\hat{\boldsymbol{Y}}_{t|t-1}=E[w_t\,|\,w_{t-1},\cdots,w_1]$，元素$s_a^2v_t=\boldsymbol{H}\boldsymbol{V}_{t|t-1}\boldsymbol{H}'=E[(w_t-\hat{w}_{t|t-1})^2]$是预报误差方差。

为了得到n个观察值的向量$w'=(w_1,w_2,\cdots,w_n)$的精确似然函数，w的联合分布被因子分解为

$$p(w\,|\,f,q,s_a)=\prod_{t=1}^{n}p(w_1\,|\,w_{t-1},\cdots,w_1,f,q,s_a) \tag{17}$$

式中 $p(w_t\,|\,w_{t-1},\cdots,w_1,f,q,s_1)$表示给定$w_{t-1},\cdots,w_1$时$w_t$的条件分布。在正态假定下，该条件分布是具有(条件)均值$\hat{w}_{t|t-1}=E[w_t\,|\,w_{t-1},\cdots,w_1]$和(条件)方差$s_a^2v_t=E[(w_t-\hat{w}_{t|t-1})^2]$的正态。因此，$w$的联合分布可以表示为：

$$p(w\,|\,f,q,s_a)=\prod_{t=1}^{n}(2\pi s_a^2v_t)^{-\frac{1}{2}}\exp[-\frac{1}{2s_a^2}\sum_{t=1}^{n}\frac{(w_t-\hat{w}_{t|t-1})^2}{v_t}] \tag{18}$$

式中 $\hat{w}_{t|t-1}$和s_a^2由Kalman滤波方程确定。Kalman滤波所需的初值由r维零值向量$\hat{\boldsymbol{Y}}_{0|0}=0$和$\boldsymbol{V}=\mathrm{cov}[\boldsymbol{Y}_0]$给出。$g_k$利用关系式$w_{t+j}=\hat{w}_t(j)+\sum_{k=0}^{j-1}y_k a_{t+j-k}$得出，$\boldsymbol{V}_{0|0}$的元素很容易确定为$w_t$的$ARMA(p,q)$过程的自协方差

和权 y_k 的函数。

例如，在 w_t 的 $ARMA(1,1)$ 模型的情况，有 $\boldsymbol{Y}_t' = (w_t, \hat{w}_t(1))$，则：

$$\boldsymbol{V}_{0|0} = \mathrm{cov}[\boldsymbol{Y}_0] = \begin{bmatrix} g_0 & g_1 \\ g_1 & g_0 - s_a^2 \end{bmatrix} = s_a^2 \begin{bmatrix} s_a^{-2} g_0 & s_a^{-2} g_1 \\ s_a^{-1} g_1 & s_a^{-1} g_0 - 1 \end{bmatrix}$$

通常情况下，一期预报值 $\hat{w}_{t|t-1}$ 和相应的误差方差 $s_a^2 v_t$ 相当迅速地接近于稳态，Kalman滤波在某些阶段(如在时间 t_0 之外)，可以化为简单形式：$\hat{w}_{t|t-1} = \sum_{i=1}^{p} f_i w_{t-1} - \sum_{i=1}^{q} q_i a_{t-i|t-i-1}$ 以及 $s_a^2 v_t = \mathrm{var}[a_{t|t-1}] = s_a^2$，对于 $t > t_0, a_{t|t-1} = w_t - \hat{w}_{t|t-1}$。

用(18)形式所表示似然函数通常称为新息形式，量 $a_{t|t-1} = w_t - \hat{w}_{t|t-1}, t = 1, \cdots, n$ 是(有限样本)新息。似然函数的新息形式也可以不直接用状态空间模型式得到，而是通过使用新息算法得到[3]。这种方法主要涉及到一个导出的 $MA(q)$ 过程 $w_t' = w_t - f_1 w_{t-1} - \cdots - f_p w_{t-p} = a_t - q_1 a_{t-1} - \cdots - q_q a_{t-q}$，要对该过程的 $n \times n$ 带形协方差阵做Cholesky分解。当序列 z_t 的某些值未被观测到，即在 z_t 的顺序值中存在丢失值的情形，有关确定似然函数的“新息”状态空间方法完全可以用于这种情况下ARMA模型的参数估计问题[4]。

2　结束语

本文着重在研究方法方面的探讨。针对网络环境下绿色农产品定价的实际问题，引入ARIMA过程的状态空间模型表示方法。在处理ARIMA过程的状态空间模型的求解过程中，一方面可利用Kalman滤波方程作为计算精确释然函数的简便方法，另一方面，也可以采用Cholesky分解的新息算法，后者尤其适用于序列的某些值未被观察到的情况。

本文得到四川农业大学科技基金资助，在此表示感谢。

参　考　文　献

[1] George E P，Box Gwilym M Jenkins，Gregory C Reinsel著. 时间序列分析预测与控制[M] (第三版). 顾岚主译. 北京：中国统计出版社，1999. 112

[2] Gardner G, Harver A C，Phillips G D A . Algorithm As 154: an algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman fitering [J]. Appl Statist，1980，29: 311-322

[3] Brockwell P J，Davis R A. Time series: theory and methods [M]. New York: Springer-Verlag, 1987

[4] Wincek M A，Reinsel G C . An exact maximum likelihood estimation procedure for regression-ARIMA time series models with possibly nonconsecutive data [J]. J Roy Statist Soc(B), 1986, 48: 303-313

编　辑　徐安玉