

# 农产品市场价格超短期预测研究\*

——基于西红柿日批发价格的现代时间序列法建模

李干琼, 许世卫, 李哲敏, 董晓霞

(中国农业科学院 农业信息研究所、农业部智能化农业预警技术重点开放实验室、  
智能化农业预警技术与系统重点开放实验室, 北京 100081)

**摘要** 为科学分析与预测农产品市场日价格走势, 研究农产品市场日价格波动的随机性特征, 增强价格的预见性和市场的调控性, 选择全国西红柿日批发价格为预测对象, 基于价格序列数据的 ADF 检验和 ARCH 效应检验, 结合 2008—2009 年间 731 天日价格数据分析, 利用 ARIMA、ARCH、GARCH 等现代时间序列法, 分别建立西红柿日批发价格预测模型, 并选取 2010 年 1 月 1—10 日进行样本外区间的评估预测。研究表明, 3 个日价格预测模型的平均绝对百分比误差(MAPE)都在 2% 以内, 其中 GARCH 模型在预测中具有更高的精度; 农产品市场价格超短期预测中, 在没有突发性因素干扰的情况下, 所建立的 3 个模型预测结果的精度比较理想, 但对于突发性事件等引起的价格急剧变化难以量化模拟和预测。

**关键词** 农产品; 市场; 日价格; 短期预测; 模型

**中图分类号:** F323.7   **文献标识码:** A   **文章编号:** 1008-3456(2010)06-0040-06

受生产成本、自然条件、科技、运输方式、储存状况、消费需求、突发性事件、政策等多种因素影响, 农产品市场日价格波动频繁。近年来, 由于农业生产资源紧缺、自然灾害频发, 重大疫情疫病时有发生、农业多功能凸显, 使得农产品市场价格波动发生的频率有所增加, 农业风险与危机发生的危害影响度加大。更为重要的是, 许多不良农情事件出现突然, 迅速波及到市场价格, 产生重大不良结果。为此, 开展农产品市场价格超短期预测研究, 事关我国农业稳定发展、农民增收和人民生活安全, 是新时期发展现代农业迫切需要解决的一个难题。本文拟采用现代时间序列法建模, 对西红柿日批发价格进行短期预测, 为农产品市场日价格走势科学分析与预测提供参考。

## 一、文献回顾

### 1. 农产品市场价格预测的研究进展

从现有文献资料来看, 总体上有关农产品市场价格的分析中长期预测研究的多, 短期预测研究的

少; 定性趋势性分析研究的多, 量化预测研究的少; 宏观分析研究的多, 针对产品品种研究的少。如中长期预测方面, 比较典型的有美国农业部的多国商品联结模型(简称“CCLS”), 它是一个涉及多品种、多国家的部门均衡预测模型, 按年度进行预测, 预测周期一般为 10 年。在量化短期预测研究方面, 大多运用简单的指数预测、趋势外推或者是基于传统供求理论价格形成的预测模型。刘洪宇<sup>[1]</sup>从供给、需求与市场价格三者之间的关系建立了商品市场价格的三时点预测模型, 并运用该模型对美国 1954—1988 年大米市场进行了实证分析; 程贤禄<sup>[2]</sup>采用多因素回归法和马尔可夫链法对北京市农产品批发市场蔬菜价格进行预测, 并结合经验分析对模型预测结果进行修正; 张小栓<sup>[3]</sup>借鉴领域工程思想, 构造了水产品价格预测时序组合模型、水产品价格预测因果总量预测模型与结构模型; 马孝斌等<sup>[4]</sup>建立了基于生猪市场月度价格数据的向量自回归(VAR)模型; 王勇等<sup>[5]</sup>运用马尔可夫模型对小麦期货价格进行预测; 姚霞等<sup>[6]</sup>、傅如南等<sup>[7]</sup>、刘峰等<sup>[8]</sup>

收稿日期: 2010-09-22

\* 国家“十一五”科技支撑计划重点项目“农产品数量安全智能分析与预警的关键技术及平台研究”(2009BADA9B01); 中央公益性科研院所基本科研业务费专项(2010-J-11)。

作者简介: 李干琼(1980-), 男, 博士研究生; 研究方向: 农产品市场信息分析。E-mail: lgqxj@caas.net.cn

应用 ARIMA 模型对白菜、肉鸡等农产品价格进行预测;苗开超<sup>[9]</sup>应用指数平滑模法建立了西红柿价格预测模型;王素雅<sup>[10]</sup>以北京新发地批发市场苹果价格为研究对象,利用 GARCH、ARIMA、VAR 3 种模型对日价格、周价格、月度价格进行了模拟与预测。在宏观价格分析方面,比较典型的有美国食品农业政策研究中心(FAPRI)的农产品模型(Meyers, Devadoss 和 Helmar)、国际食物政策研究所(IFPRI)的 IMPACT 模型(Rosegrant, et al)、IIASA 开发的世界农业模型(Parikh t Rabar)、世界银行模型(Mitchell, Ingco 和 Duncan)、中国科学院农业政策研究中心的中国农产品供需贸易政策分析和预测模型(CAPSIM++)、中国农业科学院农业经济与发展研究所的国家农业政策分析平台与决策支持系统等<sup>[11]</sup>。现有的研究表明,有关农产品市场日价格研究的内容极少,而专门针对蔬菜品种的超短期价格预测更为少见。由于决定农产品市场日价格的影响因素众多,农产品价格特别是果蔬产品价格大起大落的现象经常发生,其波动近似随机游走。本研究以现代时间序列法为工具,在提取分析数据特性的基础上,分别建立了 3 个市场日价格预测模型,提高农产品市场短期预测准确度,从而为相关部门提供决策参考的重要依据。

## 2. 常用短期预测方法

根据预测时间长度的不同,农产品市场价格预测可分为中长期预测和短期预测:中长期预测指时间期限为 3 个月以上的预测,短期预测一般为 3 个月以内的预测。短期预测包括日预测、周预测、月度预测、季度预测,其中日预测属于超短期预测。目前,实际预测中常用的短期预测方法主要有:回归分析法、时间序列法、神经网络预测法、灰色预测法、小波分析法、专家系统预测法、组合预测法等<sup>[12-16]</sup>。

# 二、农产品市场日价格短期预测模型的建立及应用

## 1. 数据说明与方法选择

本研究选择西红柿产品为预测对象,采用的日批发价格数据来源于农业部 511 个定点批发市场的监测数据,经加权计算后作为全国市场日价格数据。选取的建模样本数据时间段为 2008—2009 年的全国西红柿日批发价格数据,如图 1 所示。可以看出,

西红柿的日批发价格波动具有季节性、周期性和随机性的特征,一年当中呈现“两头高,中间低”的现象,趋势性不是特别明显,因此趋势外推法、单指数平滑法、双指数平滑法等不适合。

在预测方法选择上,由于相关数据资料收集困难以及考虑到实际应用情况,本研究采用数据驱动(Data-Driven)的时间序列建模方法,根据价格数据的特征拟建立 ARIMA 模型和条件异方差模型<sup>[17-19]</sup>,其他方法的应用研究待数据资料收集完整后将另做比较分析。

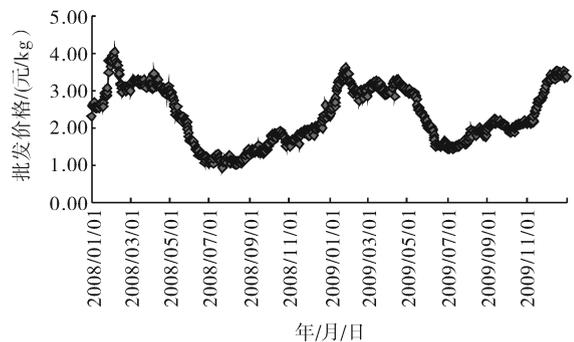


图 1 2008—2009 年全国西红柿日批发价格

## 2. 日价格短期预测模型的建立

(1) 价格序列的平稳性检验。用  $P$  表示为 2008—2009 年全国西红柿日批发价格序列,价格序列  $P$  的单位根检验结果如表 1。由于序列  $P$  增广的迪基—福勒检验(ADF 检验) $t$  值为 0.064,大于 3 个不同水平下(1%、5%、10%)的临界值,故接受原假设,即价格序列  $P$  为不平稳序列,存在单位根。

为检验价格序列  $P$  是否为  $d$  阶单整序列,对其进行一阶差分,记为  $dp$ 。如表 2,一阶差分后序列增广的迪基—福勒检验(ADF 检验) $t$  值为 -33.46,小于 3 个不同水平下(1%、5%、10%)的临界值,故拒绝原假设,即序列  $P$  一阶差分后是平稳的, $P \sim I(1)$ 。因此,对价格序列  $P$  可建立单整自回归移动平均模型,即  $ARIMA(p, d, q)$ 。

表 1 序列  $P$ 、 $dp$  的平稳性检验结果

序列	检验方法	$t$ 值	1%临界值	5%临界值	10%临界值
$P$	ADF 检验	0.064	-2.568	-1.941	-1.616
$dp$	ADF 检验	-33.463	-2.568	-1.941	-1.616

注:序列  $P$  的 ADF 检验  $t$  值相伴概率为 0.701,序列  $dp$  的 ADF 检验  $t$  值相伴概率为 0.000。

(2) 价格序列的 ARCH 效应检验。首先建立价格随机游走模型,即  $P_t = P_t(-1) + u_t$ 。模型参数

估计结果如表 2。回归方程调整的可决系数( $Adjusted-R^2$ )为 0.987,表明方程拟合效果较好。但对残差做一阶或高阶的序列自相关 LM 检验时,得到的伴随概率为 0.000,故存在自相关。

表 2 价格随机游走模型估计与检验结果

变量	系数	标准误	$t$ 值	伴随概率
$P(-1)$	1.000	0.001	760.760	0.000
可决系数( $R^2$ )	0.987	因变量平均值		2.319
调整的可决系数( $Adjusted-R^2$ )	0.987	因变量标准差		0.759
回归标准差	0.087	赤池信息准则		-2.052
残差平方和	5.471	施瓦茨准则		-2.046
对数似然值	750.328	杜宾-瓦特森统计量		2.395

对残差序列进行 ARCH 效应检验, $q=1$ ,结果

如表 3。 $F$  检验和 LM 检验的相伴概率均为 0,小于显著性水平 0.05,故拒绝原假设,即残差序列存在 ARCH(1)效应。

表 3 ARCH 效应检验结果

$F$ 统计量	28.109	相伴概率( $P$ )	0.000
怀特统计量	27.137	相伴概率( $P$ )	0.000

(3)ARIMA 模型的建立。由于价格序列  $P \sim I(1)$ ,故可建立单整自回归移动平均模型。一阶差分后的序列  $dp$  自相关图和偏相关图如图 2。从中可以看出, $P$  的取值为 1 或 3, $q$  的取值为 1 或 3。经试验比较,组合  $p=1, q=3$  是最优的,故选择建立 ARIMA(1,1,3)模型。

自相关图	偏相关图	滞后阶数	自相关系数	偏相关系数	Q 统计量	相伴概率
** .	** .	1	-0.204	-0.204	30.630	0.000
. .	* .	2	-0.030	-0.075	31.295	0.000
. *	. *	3	0.154	0.138	48.617	0.000
. .	. .	4	-0.020	0.041	48.925	0.000
. .	. *	5	0.058	0.078	51.396	0.000
. .	. .	6	0.037	0.047	52.417	0.000
. *	. **	7	0.189	0.223	78.847	0.000
. .	. .	8	-0.023	0.057	79.225	0.000
. .	. .	9	0.016	0.028	79.407	0.000
. .	. .	10	0.064	0.011	82.410	0.000
. .	. .	11	-0.051	-0.053	84.355	0.000
. .	* .	12	-0.039	-0.111	85.498	0.000
. .	* .	13	0.017	-0.070	85.712	0.000
. *	. *	14	0.116	0.072	95.691	0.000
. .	. .	15	0.001	0.058	95.692	0.000
. .	. .	16	-0.055	-0.034	97.942	0.000

注: \* 表示自相关系数或偏相关系数显著不为 0, \*\* 表示自相关系数或偏相关系数非常显著不为 0。

图 2 序列  $dp$  的自相关图和偏相关图

模型 ARIMA(1,1,3)的参数估计与检验结果如表 4-1 和表 4-2。

(4)ARCH 模型的建立。通过 ARCH 效应检验分析,可建立 ARCH(1)模型,方程估计结果如表 5-1 和表 5-2。可知与直接建立价格随机游走模型相比更为合理,对数似然值增加了,赤池信息准则(AIC)和施瓦茨准则(SC)的值均变小了,而可决系数基本没变。此外,ARCH(1)模型的系数为 0.607,小于 1,满足平稳条件。对残差再做条件异方差检验,结果表明已不存在条件异方差。

(5)GARCH 模型的建立。由于价格随机游走模型存在高阶 ARCH 效应,建立 GARCH(2,2)模型,方程参数估计与检验结果如表 6-1 和表 6-2。利用 GARCH(2,2)模型估计得到的结果与 ARCH(1)模型相比,对数似然值增加了,赤池信息准则(AIC)和施瓦茨准则(SC)的值变小了,而可决系数基本没变,可认为模型进一步得到了优化。ARCH 项与 GARCH 项的系数和为 0.993,小于 1,满足参数的约束条件,但由于该值已经接近于 1,说明随机误差项的波动存在强烈冲击,且冲击是持久的,即对

所有未来各期都产生作用<sup>[20]</sup>。再对 GARCH(2,2)模型的随机误差项进行条件异方差检验,结果表明已不存在条件异方差。

表 4-1 ARIMA(1,1,3)模型的参数估计结果

变量	系数	标准误	t 值	相伴概率
AR(1)	-0.215	0.037	-5.842	0.000
MA(2)	-0.071	0.037	-1.882	0.060
MA(3)	0.144	0.037	3.919	0.000

注:因变量为  $D(P,1)$

表 4-2 方程的检验统计指标与量值

检验统计指标	量值	检验统计指标	量值
可决系数( $R^2$ )	0.066	因变量平均值	0.001
调整的可决系数(Adjust- $R^2$ )	0.063	因变量标准差	0.086
回归标准差	0.083	赤池信息准则	-2.129
残差平方和	5.036	施瓦茨准则	-2.110
对数似然值	779.025	杜宾-瓦特森统计量	2.005

表 5-1 ARCH(1)模型的参数估计结果

变量名	系数	标准误	Z 值	伴随概率
$P(-1)$	0.999	0.001	1 190.555	0.000
残差方程				
C	0.004	0.000	15.914	0.000
$RESID(-1)^2$	0.607	0.048	12.561	0.000

注:因变量为  $P$

表 5-2 方程的检验统计指标与量值

检验统计指标	量值	检验统计指标	量值
可决系数( $R^2$ )	0.987	因变量平均值	2.319
调整的可决系数(Adjust- $R^2$ )	0.987	因变量标准差	0.759
回归标准差	0.087	赤池信息准则	-2.200
残差平方和	5.474	施瓦茨准则	-2.181
对数似然值	806.111	杜宾-瓦特森统计量	2.392

表 6-1 GARCH(2,2)模型的参数估计结果

变量名	系数	标准误	Z 值	伴随概率
$P(-1)$	0.999	0.001	932.116	0.000
残差方程				
C	4.05E-05	1.24E-05	3.273	0.001
$RESID(-1)^2$	0.319	0.061	5.233	0.000
$RESID(-2)^2$	-0.308	0.061	-5.091	0.000
GARCH(-1)	1.143	0.089	12.834	0.000
GARCH(-2)	-0.161	0.086	-1.867	0.062

注:因变量为  $P$

表 6-2 方程的检验统计指标与量值

检验统计指标	量值	检验统计指标	量值
可决系数( $R^2$ )	0.987	因变量平均值	2.319
调整的可决系数(Adjust- $R^2$ )	0.987	因变量标准差	0.759
回归标准差	0.087	赤池信息准则	-2.242
残差平方和	5.471	施瓦茨准则	-2.204
对数似然值	824.347	杜宾-瓦特森统计量	2.395

### 3. 日价格短期预测模型的应用

分别应用 ARIMA(1,1,2)模型、ARCH(1)模型、GARCH(2,2)模型对 2010 年 1 月 1 日至 1 月 10 日的西红柿价格进行短期预测,结果见表 7。

表 7 不同模型的预测结果

日期	实际价格/ (元/kg)	ARIMA(1,1,3)		ARCH(1)		GARCH(2,2)	
		预测值/(元/kg)	误差/%	预测值/(元/kg)	误差/%	预测值/(元/kg)	误差/%
2010-01-01	3.27	3.38	3.36	3.39	3.67	3.39	3.67
2010-01-02	3.35	3.37	0.60	3.39	1.19	3.39	1.19
2010-01-03	3.38	3.38	0	3.38	0	3.39	0.30
2010-01-04	3.41	3.37	-1.17	3.38	-0.88	3.39	-0.59
2010-01-05	3.47	3.37	-2.88	3.38	-2.59	3.39	-2.31
2010-01-06	3.47	3.37	-2.88	3.38	-2.59	3.39	-2.31
2010-01-07	3.45	3.37	-2.32	3.37	-2.32	3.39	-1.74
2010-01-08	3.43	3.37	-1.75	3.37	-1.75	3.39	-1.17
2010-01-09	3.30	3.37	2.12	3.37	2.12	3.39	2.73
2010-01-10	3.34	3.37	0.90	3.37	0.90	3.39	1.50

为比较分析不同模型的预测精度,通常采用平均绝对百分比误差(MAPE)来评价<sup>[21]</sup>,其计算公式如下:

$$MAPE = \frac{1}{n} \sum_{i=1}^n |p_i|$$

式中, $n$ 为样本数据个数, $p_i$ 为相对百分比误差。

3 个模型预测的评价结果如表 8。总体上看,3 个模型的预测精度都非常高,非常适合于农产品市场日价格短期预测。其中 GARCH 模型的预测精度最高,平均绝对百分比误差仅为 1.75%。

表 8 不同预测模型的精度比较

模型	ARIMA(1,1,3)	ARCH(1)	GARCH(2,2)
误差/%	1.80	1.80	1.75

## 参 考 文 献

## 三、结论与讨论

## 1. 结 论

针对农产品市场价格影响因素多样,建模困难的特点,以及日价格波动的随机性特征,提出了基于时间序列法的日价格短期预测建模方法。通过对 2008—2009 年西红柿日批发价格的平稳性测试和 ARCH 效应检验,分别建立了 3 个不同的日价格短期预测模型,并对 2010 年 1 月 1—10 日的价格进行了预测,得出结论:

(1)西红柿日批发价格的波动既有随机性,同时又存在季节效应,使得一年中的价格呈现“两头高、中间低”的态势。

(2)基于数据驱动的现代时间序列建模方法突破了传统因果模型的限制,并且取得了较好的预测效果。所建立的 ARIMA、ARCH、GARCH 三个模型的日预测准确率都在 95%以上,个别预测日甚至达到 100%。较高的预测精度有效保证了这种超短期预测的实际应用价值。

(3)从预测精度平稳性(MAPE)看,GARCH 模型在预测中具有更高的精度,高达 98.25%。

## 2. 讨 论

由于农产品日价格波动频繁,因此任何预测方法都无法产生绝对准确的结果。这不仅是方法本身的问题,而更重要的是一些未来的因素常常不能被准确地预见或加以有效地估计。在实际应用中,短期价格预测模型应根据不同农产品价格数据特征选择最优参数。需要进一步研究的问题:

(1)如何提高模型的解释性。基于数据驱动的建模方法难以对价格变化原因作出解释,而通常情况下短期价格的波动更多地来自突发性事件的影响,如何提前预知可能的事件及量化研究其影响是短期预测需要突破的重大难题。

(2)模型的选择问题。不同模型均有一定的适用条件,且对信息的提取程度不尽相同,这是否意味着组合模型就比单一模型更有优势?或许组合模型是未来一个趋势,但如何组合的问题仍需进一步研究。

- [1] 刘洪宇. 商品市场价格的三时点预测模型及其应用[J]. 预测, 2001(3):51-54.
- [2] 程贤禄. 北京农产品批发市场蔬菜价格预测预报体系研究[J]. 北京农业科学,2002(2):1-10.
- [3] 张小栓. 水产品价格预测支持系统研究[D]. 北京:中国农业大学经济管理学院,2003:1-13.
- [4] 马孝斌,王婷,董霞,等. 向量自回归法在生猪价格预测中的应用[J]. 中国畜牧杂志,2007(23):4-6.
- [5] 王勇,张浩. 小麦期货价格预测的马尔科夫模型[J]. 安徽农业科学,2008,36(5):1721.
- [6] 姚霞,彭汉良,朱艳,等. 时鲜农产品价格预测的 ARIMA 时序模型构建与运用[J]. 农业系统科学与综合研究,2007(1):89-94.
- [7] 傅如南,林丕源,严尚维,等. 基于 ARIMA 的肉鸡价格预测建模与应用[J]. 中国畜牧杂志,2008(20):17-21.
- [8] 刘峰,王儒敬,李传席. ARIMA 模型在农产品价格预测中的应用[J]. 计算机工程与应用,2009,45(25):238-239.
- [9] 苗开超. 基于指数平滑模型的农产品价格预测研究[D]. 安徽:合肥工业大学计算机学院,2009.
- [10] 王素雅. 农产品短期价格分析及预测方法选择——以北京新发地批发市场苹果为例[D]. 北京:中国农业科学院研究生院,2009.
- [11] 廖永松,黄季焜. 21 世纪我国粮食安全保障与灌溉水需求预测[J]. 中国水利,2004(1):36-38.
- [12] 陈思杰,周浩. 电力市场电价预测综述[J]. 继电器,2006(11):54-60.
- [13] 沈巍. 股票价格预测模型研究[J]. 财经问题研究,2009(7):89-93.
- [14] 王强,秦泰增. 上证指数的短期预测模型[J]. 电脑编程技巧与维护,2008(11):14-15.
- [15] 杨奎河. 短期电力负荷的智能化预测方法研究[D]. 陕西:西安电子科技大学计算机学院,2004:7-16.
- [16] 邓聚龙. 灰色系统理论教程[M]. 武汉:华中理工大学出版社,1990:25-30.
- [17] 东梅,赵俊杰. 市场分形理论与 2009 年中国小麦价格预测[J]. 乡镇经济,2009(9):16-18.
- [18] 刘春海,佟凤奇,崔涛. 粮食价格预测方法探讨[J]. 农业发展与金融,2007(6):84.
- [19] 徐枫. 股票价格预测的 GARCH 模型[J]. 统计与决策,2006(18):107-109.
- [20] 李国柱,刘德智,等. 计量经济学试验教程[M]. 北京:中国经济出版社,2009:280-287.
- [21] 蒋白懿,代进,高金良. 城市日用水量预测模型比较研究[J]. 沈阳建筑大学学报:自然科学版,2008(2):278-281.

# Study on Super Short-term Forecasting for Market Price of Agro-products

——Based on Modern Times Series Modeling of Daily Wholesale Price of Tomatoes

LI Gan-qiong, XU Shi-wei, LI Zhe-min, DONG Xiao-xia

*(Agricultural Information Institute, the Chinese Academy of Agricultural Sciences/  
Key Lab of Digital Agricultural Early Warning Technology, Ministry of Agriculture/  
Key Lab of Digital Agricultural Early Warning Technology and System,  
the Chinese Academy of Agricultural Sciences, Beijing, 100081)*

**Abstract** Many factors may cause the fluctuation of agro-product market prices, so the agro-product price often experience ups and downs, whose fluctuations are similar to the random walk. In order to scientifically analyze and predict the trend of daily price of agro-product market, this paper selected the daily wholesale price of tomatoes in China as object to model, and the data used in the modeling are between 2008 and 2009 with daily prices for 731 days. The ultimate goal is to provide technical support for price forecasting and market regulation. According to the random features of daily price fluctuation of agro-product market as well as ADF test and ARCH effect test based on price series data, this paper employed the modern time series methods of ARIMA, ARCH and GARCH to establish daily wholesale price forecasting models of tomatoes respectively, and applied the models to forecast the tomato price from January 1, 2010 to January 10, 2010 as evaluation. The result shows that mean absolute percentage error (MAPE) of the three daily price forecasting models is less than 2%, among which the highest accuracy in forecasting is GARCH model. Accuracy of the three models forecasting is ideal if unexpected incidents don't occur in super short-term agro-product market price forecasting. But it is hard to simulate and forecast quantitatively for emergencies causing dramatic changes.

**Key words** agro-products; market; daily price; short-term forecasting; model

(责任编辑:陈万红)