

基于 ARMA 模型预测股票价格的实证分析

孟 坤, 李 丽

(安徽财经大学统计与应用数学学院, 安徽 蚌埠 233030)

摘要: 目的 基于 ARMA 模型的基本理论, 建立合理的 ARMA 模型, 以上证综合指数为例对股票价格进行拟合预测和实证分析。方法 首先, 利用 Eviews 软件对原始数据序列进行单位根检验, 判断原始数据序列是否具有平稳性; 若非平稳, 则需对原始序列作一阶差分处理, 再次检验差分后序列的平稳性; 其次, 用自相关图与偏自相关图识别序列的模型形式对已识别的 ARIMA (2, 1, 2) 模型进行参数估计, 包括估计模型的系数及判别模型的阶数; 最后, 运用所建立的模型对上证综合指数日收盘价进行高精度拟合预测。结果 结果表明原始数据序列是非平稳性序列, 但一阶差分后的序列是平稳的; 模型的残差检验显示 ARIMA (2, 1, 2) 模型是有效的, 预测下一个工作日上证综合指数每日收盘价的价格是 3 642.47, 与实际值相差较小, 说明所建立的 ARMA 模型具有一定的准确性。结论 ARMA 模型比较适合于进行短期预测, 同时结合 Eviews 软件可以使得计算过程变得简便、准确。研究上证综合指数日收盘价的短期变动情况对预测股价未来趋势和制定投资策略具有现实意义, 能够为投资者和决策者提供可靠的信息服务及决策指导。

关键词: 时间序列; ARMA 模型; 股票价格预测; Eviews

中图分类号: F 290

文献标识码: A

DOI: 10.3969/j.issn.1673-1492.2016.05.011

Empirical Analysis of Stock Price Based on ARMA Model

MENG Kun, LI Li

(School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu, Anhui 233030, China)

Abstract: Objective For the prediction of Shanghai composite index future short-term closing price, based on the basic theory of ARMA model, a reasonable ARMA model fitting prediction is established and empirically analyzed by taking the Shanghai composite index daily closing price forecasting as an example. **Methods** First, Eviews software of original data sequence of unit root test was used to judge whether the original data sequence was stationary; If it was not stable, the original sequence as a first-order differential needed further processing, and the stability of the sequence difference was re-inspected. Second, self-correlation diagram and partial self-correlation diagram of the recognition sequence model were used to evaluate the parameters of the identified ARIMA (2, 1, 2) model, including the estimated coefficients of the model and discriminant model order. Finally, The established model was used to predict the fitting of the Shanghai composite index daily closing price with high precision. **Results** Results showed that the original data series was non-stationary series, but points after the first-order difference sequence was stationary; examination of the residuals of the model showed that ARIMA (2, 1, 2) model was effective to predict a work. The Shanghai composite index daily closing price was predicted to be 3642.47 and closer to actual value, indicating that the ARMA model was established with a certain degree of accuracy. **Conclusion** ARMA model is more suitable for short-term prediction, and the combination of Eviews software can make the calculation process simple and accurate. Study on the Shanghai composite index daily closing price of short-term changes is of practical significance to forecast the future stock price trends and formulate investment strategy, and it can provide reliable information services and policy guidance for investors

来稿日期: 2016-03-11

基金项目: 国家自然科学基金项目 (11301001); 安徽财经大学教研项目 (acjyzd201429)

作者简介: 孟坤(1994-), 男, 安徽阜阳人, 安徽财经大学在读学生, 研究方向: 经济数学。

通讯作者: 李丽(1980-), 女, 安徽蚌埠人, 讲师, 硕士, 安徽财经大学应用数学系教师, 研究方向: 应用数学与数学建模。

and policy-makers.

Key words: time series; ARMA model; stock price forecasting; Eviews.

生活中有很多问题都可以看成是时间序列问题,例如银行利率波动、股票收益率变化以及国际汇率变动等问题。所谓的时间序列问题,是指某一统计对象长时间内的数值变化情况。在实际应用中,经常会遇到许多不满足平稳性的时间序列数据,尤其是在经济、金融等领域。因此,能否有效地挖掘非平稳时间序列的有用信息,对于解决一些经济、金融领域的问题显得尤为重要。目前关于预测股票价格的研究文章有很多,这些已有研究大都采用回归分析、组合预测等方法对股票价格未来变动值进行探讨,得出股票价格在未来短期内的变化趋势及预测值,但预测结果并不非常精准,存在较大的误差。ARMA 模型不仅可用于拟合平稳性时间序列问题,而且对非平稳时间序列问题同样具有良好的拟合效果,尤其是在金融和股票领域应用最为广泛。

本文主要针对 2015-01-05 至 2015-12-18 (共计 234 个工作日) 期间上证综合指数每日收盘价的预测问题,建立上证综合指数每日收盘价预测模型,采用 ARMA 模型对上证综合指数每日收盘价进行高精度的拟合预测。研究表明,上证综合指数每日收盘价在短期内将保持平稳上涨,不会有大幅涨跌的情况。研究上证综合指数每日收盘价的短期变动情况了解股票市场变化及制定投资决策具有现实意义,能够为投资者和决策者提供可靠的信息服务及决策指导。

1 ARMA 模型的理论介绍及平稳性检验

1.1 ARMA 模型的理论介绍

1) AR 模型是指利用以前的观察值和当前的干扰值并通过一定的线性组合来进行预测与分析。AR 模型的数学公式为:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t$$

式中: y_t 为一个平稳时间序列, ϕ_i ($i=1, 2, \dots, p$) 表示 AR 模型的待定系数, p 表示 AR 模型的阶数, ϵ_t 为误差。

2) MA 模型是利用以前的干扰值和当前的干扰值并通过一定的线性组合来进行预测。MA 模型的数学公式为:

$$y_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q}$$

式中: y_t 为平稳时间序列, θ_j ($j=1, 2, \dots, q$) 表示 MA 模型的待定系数, q 表示 MA 模型的阶数, ϵ_t 为误差。

3) ARMA 模型: 是由 AR 模型和 MA 模型组合而成, 主要用于描述平稳随机过程, 数学公式为:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q}$$

1.2 ARMA 模型的定阶方法

一般而言, 可以根据序列的自相关系数与偏自相关系数选择合适的模型。(1): 若自相关系数为拖尾, 偏自相关系数为 p 阶截尾, 则选择 AR(p) 模型; (2): 若自相关系数为 q 阶截尾, 偏自相关系数为拖尾, 则选择 MA(q) 模型; (3): 若自相关系数和偏自相关系数都为拖尾, 则选择 ARMA(p, q) 模型^[1]。如果难以确定 ARMA 模型的阶数, 还可以通过 AIC 准则来确定。

1.3 ARMA 模型的平稳性检验——ADF 检验

ADF 检验的基本思想: 首先对

$$\Delta u_t = c + \delta u_{t-1} + \sum_{i=1}^{p-1} \beta_i \Delta u_{t-i} + \epsilon_t$$

作回归, 构造 ADF 检验统计量^[2]:

$$ADF = \frac{\hat{\delta}}{S(\hat{\delta})}$$

其中, $S(\hat{\delta})$ 为参数 δ 的样本标准差。

计算出 ADF 检验统计量后，查 ADF 临界表，看是否可以拒绝原假设 $\delta=0$ 。若 ADF 的绝对值超过 ADF 临界值的绝对值，则不接受原假设 $\delta=0$ ；反之，则选择接受原假设 $\delta=0$ ，说明该序列非平稳。

2 上证综合指数每日收盘价的实证分析与预测

在网易财经网 (http://quotes.money.163.com/trade/lshysj_zhishu_000001.html) 查找上证综合指数每日收盘价的历史数据，时间区间是 2015-01-05 至 2015-12-18，共有 234 个工作日。主要是基于 ARMA 模型的建模理论，建立上证综合指数每日收盘价预测模型，研究上证综合指数每日收盘价的短期变动情况。

2.1 原始数据的平稳化处理

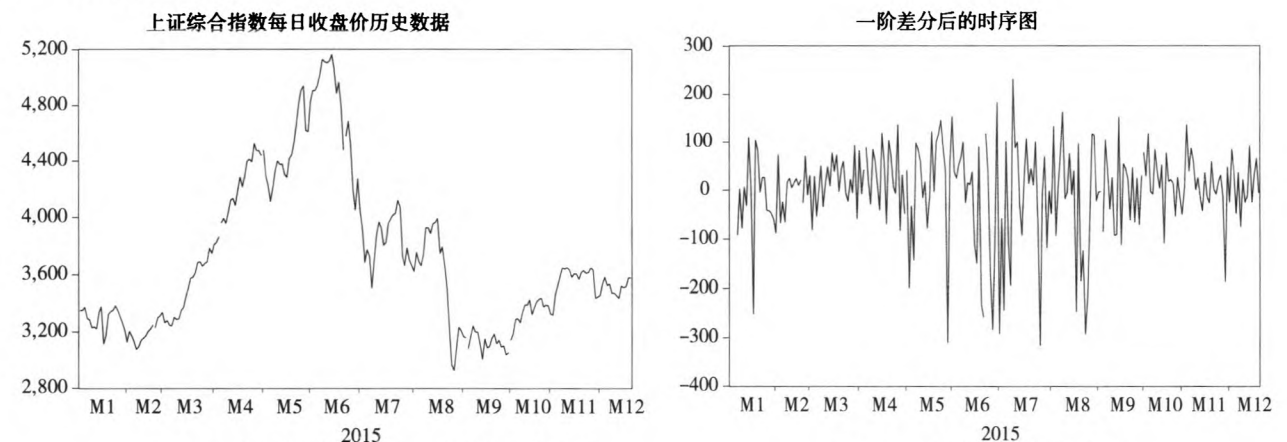


图 1 上证综合指数每日收盘价历史数据

图 2 一阶差分后的序列图

从图 1 可以看出，上证综合指数每日收盘价在 2015-01-05 至 2015-12-18 期间内是非平稳的时间序列，但对原始数据进行一阶差分后的序列却可能是平稳的（图 2）。

首先对原始数据序列进行 ADF 检验，检验结果如图 3。由图 3 可知，在显著性水平为 0.01 条件下存在单位根，故接受原假设，即原始数据序列是非平稳的。

此外，还需要利用 ADF 检验来检验原始数据一阶差分序列的平稳性，若 ADF 检验结果未通过，则该一阶差分序列是非平稳的。ADF 检验结果如图 4 所示，ADF 检验统计量为-13.469 04，其绝对值超过 1% 的显著水平下临界值-3.998 997 的绝对值，因此不接受原假设，一阶差分后的序列是平稳的^[3]。

2.2 利用自相关图与偏自相关图识别序列的模型形式

从图 5 可以看出：股票收盘价的相关系数减弱较慢，说明上证股票收盘价的历史数据序列是非平稳的。由图 6 可知，一阶差分后序列的自相关图与偏自相关图的拖尾性十分显著，所以应该利用 ARMA 模型的建模理论建立模型，最后模型中的 p ， q 值则需要利用 P 值和 AIC 准则加以确定。

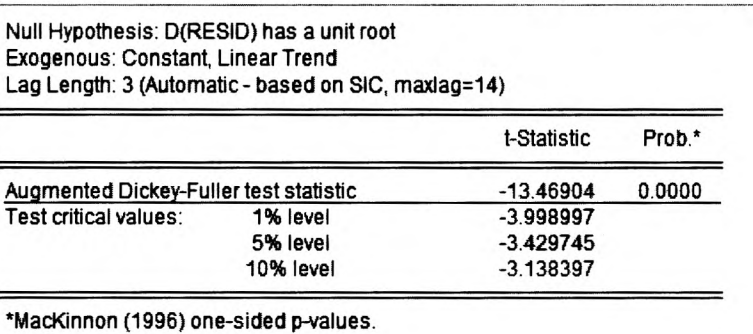


图 3 原始数据序列的 ADF 检验结果

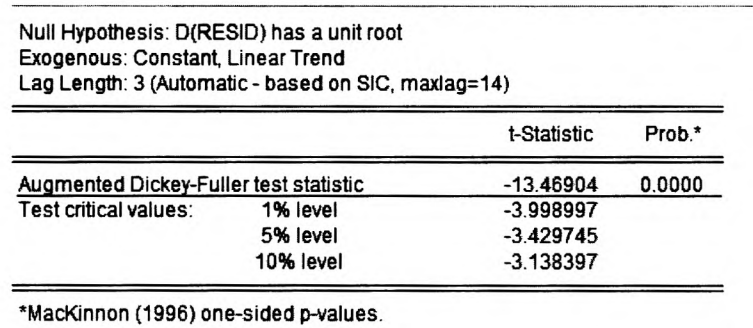


图 4 一阶差分序列的 ADF 检验结果

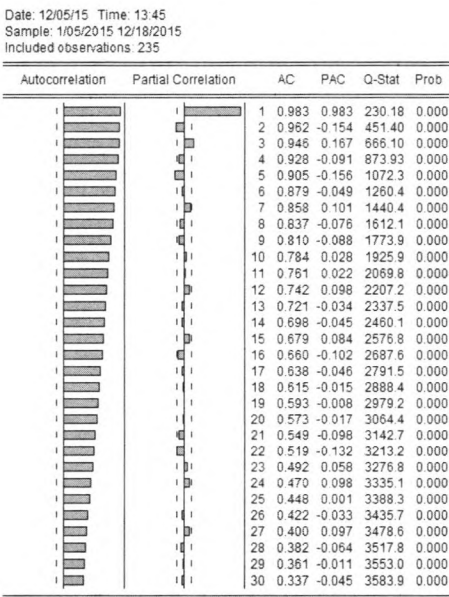


图 5 收盘价的自相关与偏自相关图

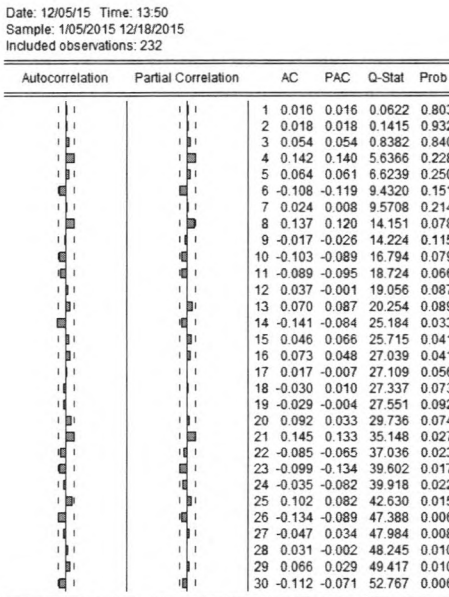


图 6 一阶差分后序列的自相关与偏自相关图

由图 6 可知：由于自相关系数和偏自相关系数均是 2 阶拖尾，因此选取如下 4 种 ARMA 模型进行比较，这 4 种模型分别是 ARIMA (1, 1, 1)、ARIMA (1, 1, 2)、ARIMA (2, 1, 1) 及 ARIMA (2, 1, 2) (图 7)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.972146	7.107146	0.136784	0.8913
AR(1)	-0.352104	0.226573	-1.554043	0.1215
MA(1)	0.564130	0.199596	2.826365	0.0051
R-squared	0.049824	Mean dependent var		0.976472
Adjusted R-squared	0.041561	S.D. dependent var		95.87002
S.E. of regression	93.85664	Akaike info criterion		11.93421
Sum squared resid	2026086	Schwarz criterion		11.97864
Log likelihood	-1387.335	Hannan-Quinn criter.		11.95212
F-statistic	6.030169	Durbin-Watson stat		2.036417
Prob(F-statistic)	0.002802			
Inverted AR Roots	-.35			
Inverted MA Roots	-.56			

ARIMA (1, 1, 1)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.805694	6.271115	0.128477	0.8979
AR(1)	0.010381	0.313128	0.033153	0.9736
AR(2)	-0.168420	0.087672	-1.921024	0.0560
MA(1)	0.182447	0.317283	0.575030	0.5658
R-squared	0.063105	Mean dependent var		0.883664
Adjusted R-squared	0.050777	S.D. dependent var		96.06682
S.E. of regression	93.59604	Akaike info criterion		11.93294
Sum squared resid	1997330	Schwarz criterion		11.99237
Log likelihood	-1380.222	Hannan-Quinn criter.		11.95691
F-statistic	5.119002	Durbin-Watson stat		1.998968
Prob(F-statistic)	0.001909			
Inverted AR Roots	.01+.41i	.01-.41i		
Inverted MA Roots	-.18			

ARIMA (2, 1, 1)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.994027	6.496030	0.153021	0.8785
AR(1)	0.038075	0.506212	0.075216	0.9401
MA(1)	0.154350	0.501426	0.307821	0.7585
MA(2)	-0.137452	0.125778	-1.092813	0.2756
R-squared	0.056210	Mean dependent var		0.976472
Adjusted R-squared	0.043846	S.D. dependent var		95.87002
S.E. of regression	93.74472	Akaike info criterion		11.93605
Sum squared resid	2012468	Schwarz criterion		11.99529
Log likelihood	-1386.549	Hannan-Quinn criter.		11.95994
F-statistic	4.546229	Durbin-Watson stat		1.998077
Prob(F-statistic)	0.004075			
Inverted AR Roots	.04			
Inverted MA Roots	.30	-.46		

ARIMA (1, 1, 2)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.827988	5.669860	0.146033	0.8840
AR(2)	-0.920095	0.055802	-16.48859	0.0000
MA(2)	0.803387	0.085327	9.415388	0.0000
R-squared	0.089923	Mean dependent var		0.883664
Adjusted R-squared	0.081975	S.D. dependent var		96.06682
S.E. of regression	92.04509	Akaike info criterion		11.89528
Sum squared resid	1940156	Schwarz criterion		11.93985
Log likelihood	-1376.853	Hannan-Quinn criter.		11.91326
F-statistic	11.31359	Durbin-Watson stat		1.615404
Prob(F-statistic)	0.000021			

ARIMA (2, 1, 2)

图 7 四种 ARMA 模型检验结果

根据上面 4 个模型的检验结果并结合 AIC 准则和 P 值这两项检验结果可得：在 ARIMA (2, 1, 2) 模型中，由于 AR (2) 和 MA (2) 的系数比较显著且它们所对应的 P 值都为 0，而且该模型的 AID 值比较小，因此可用 ARIMA (2, 1, 2) 模型对原始数据序列进行建模及预测分析^[4]。

2.3 上证综合指数收盘价序列模型的建立与参数估计

通过前面对模型的识别与选择，认为建立 ARIMA (2, 1, 2) 模型是最佳的选择。如图 7 中的 ARIMA (2, 1, 2) 所示，是 ARIMA (2, 1, 2) 模型的参数估计结果及模型的检验结果，从图中可以看出

在 ARIMA (2, 1, 2) 模型的参数估计中, AR (2) 和 MA (2) 所对应的 P 值都趋于 0, 但是常数项 C 对应的 P 值为 0.884 0, 并不趋于 0。因此需要去掉常数项 C , 重新对 ARIMA (2, 1, 2) 模型进行估计和残差检验^[5], 其结果如图 8 所示:

由图 8 可知: 除去常数 C 后, 重新对模型进行参数估计和残差检验, 结果表明模型的系数比较显著, 且 AIC 值变得更小了, 说明去掉常数项 C 后的 ARIMA (2, 1, 2) 模型更加准确。下面是该模型所对应的数学表达式:

$$X_t = -0.9201X_{t-1} + \epsilon_t + 0.8035\epsilon_{t-1}$$

式中 ϵ_t 为残差序列。

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(2)	-0.920128	0.055670	-16.52813	0.0000
MA(2)	0.803450	0.085126	9.438413	0.0000
R-squared	0.089839	Mean dependent var	0.883664	
Adjusted R-squared	0.085881	S.D. dependent var	96.06682	
S.E. of regression	91.84905	Akaike info criterion	11.88675	
Sum squared resid	1940337.	Schwarz criterion	11.91647	
Log likelihood	-1376.863	Hannan-Quinn criter.	11.89874	
Durbin-Watson stat	1.615305			

2.4 收盘价序列模型的残差检验

如果一个残差序列是白噪声, 则该残差序列中的有用信息已被提取完; 反之, 则说明还存在部分未被提取的有用信息, 还需要对模型做进一步的改进; 若残差序列的自相关系数均在随机区间内, 则该残差序列是白噪声。残差序列^[6]的检验结果如图 9 所示:

图 8 ARIMA (2, 1, 2) 建模结果

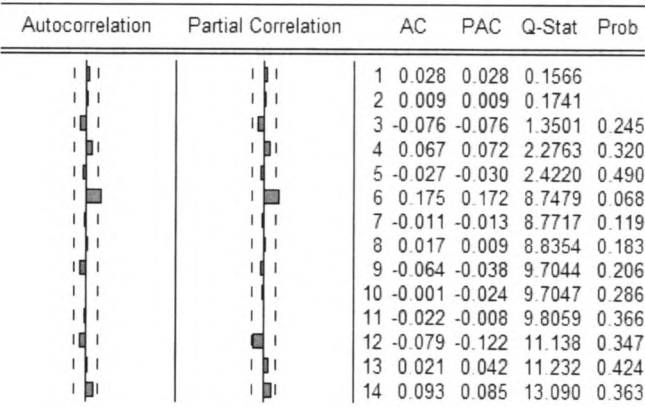


图 9 ARIMA (2, 1, 2) 模型残差相关图

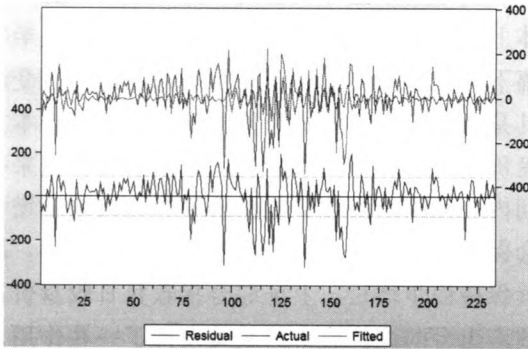


图 10 ARIMA (2, 1, 2) 模型拟合图

由图 9 可知: 由于 P 值都大于 0.05, 说明所有 Q 值都小于 0.05 的检验水平下的卡方分布临界值, 该模型的残差序列为白噪声, 因此可以确定该模型的拟合效果较好。

2.5 收盘价序列模型的预测分析

用已建立的 ARMA 模型对上证综合指数未来几日的收盘价进行短期预测, 预测未来 1 个工作日内上证综合指数每日收盘价。

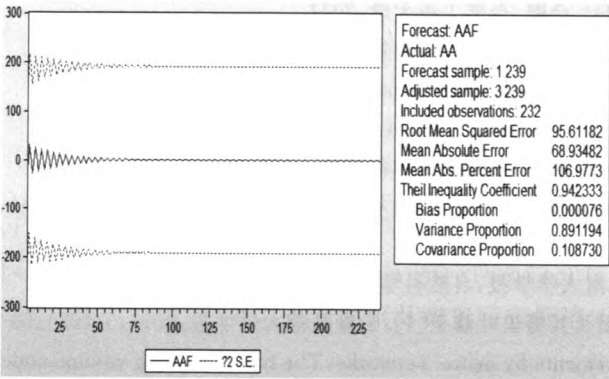


图 11 序列动态预测图

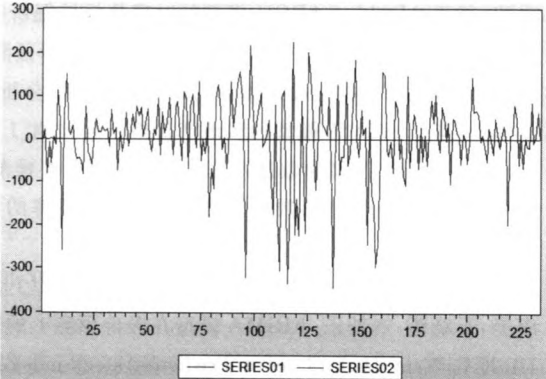


图 12 动态预测效果图

如图 11、12 所示，由于动态预测值大致为直线，说明动态预测效果并不理想^[7]。

下面开始进行静态预测，如图 13 所示，预测值仍在 AFF 中^[8]，所以本文所建模型的静态预测效果比较理想。

经过向前 1 步预测，可以得到未来 1 个工作日的预测值为 63.51，考虑上 1 个工作日（2015-12-18）的收盘价为 3 578.96，可以得出未来 1 个工作日（2015-12-21）上证综合指数日收盘价为 3 642.47，而 2015-12-21 实际的上证综合指数日收盘价是 3 631.2，预测值与实际值相差较小，说明所建立的 ARMA 模型具有一定的准确性。

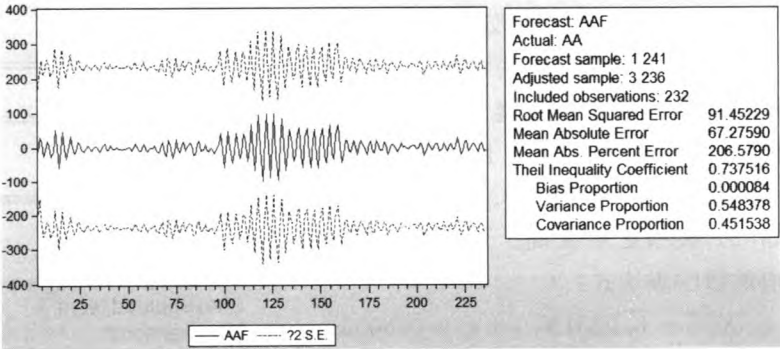


图 13 静态预测图

3 结 论

本文基于 ARMA 模型的建模理论^[9-12]，采用 ARMA 模型对上证综合指数日收盘价进行高精度的拟合预测和实证分析，研究上证综合指数日收盘价的短期变动情况。

从原始数据可以看出，2015-01-05 至 2015-12-18（共有 234 个工作日）期间上证综合指数日收盘价总体上呈先上升后急剧下降的趋势。其中上半年上证综合指数日收盘价显著上升，但 6 月份至 12 月份却大幅下降。这主要是由于今年国内股市剧烈变动所致，可见国内的股票市场仍存在非常大的起伏和波动，尤其是短时间大范围的股价剧烈变动。基于本文所建立的 ARIMA（2，1，2）模型并对上证综合指数日收盘价进行拟合预测和分析，结果表明在未来一段时期内上证综合指数日收盘价仍会在低位有小幅变化，短期内上证综合指数日收盘价不会有大幅度增长。但从未来长期的整体趋势来看，上证综合指数日收盘价将会保持持续上涨。这可能的原因有：第一，政府出台一系列针对股市的宏观调控政策；第二，国民经济持续保持稳步增长。上证综合指数每日收盘价在短期内将保持平稳上涨，不会有大幅涨跌的情况，这是由于政府出台的调控股市的政策发挥了一些作用。

总之，ARMA 模型对预测非平稳时间序列数据具有很好的效果，同时结合 Eviews 软件进行 ARMA 模型的建立与求解，能够为投资者和决策者提供可靠的信息服务及决策指导。

参考文献：

[1] 郭雪,王彦波. 基于 ARMA 模型对沪市股票指数的预测[J]. Economic & Trade Update,2006,48(04):58-59.

[2] 邓军,杨宣,王玮,等. 运用 ARMA 模型对股价预测的实证研究[J]. 企业导报,2010,(06):266-267.

[3] 贾澎涛,何华灿,刘丽. 时间序列数据挖掘综述[J]. 计算机应用研究,2007,24(11):15-18.

[4] 李琼. 基于 ARMA-GARCH 模型的时序数据挖掘研究[D]. 合肥:合肥工业大学,2011.

[5] 黄俊,周猛,王俊海. ARMA 模型在我国能源消费预测中的应用[J]. 统计与决策,2004,180(12):49-50.

[6] 阳光宏义,陈平,朱梅. 股票指数的时间序列模型分析[J]. 数学实践与认识,2006,36(08):8-9.

[7] 欧廷皓. 基于 ARMA 模型的房地产价格指数预测[J]. 统计与决策,2007,(07):92-93.

[8] 何永沛. ARMA 模型参数估计算法改进及在股票预测中的应用[J]. 重庆工学院学报,2009,23(02):109-112.

[9] 何树红,陈浩. 基于非参数模型对上海股市收益率的实证分析[J]. 云南民族大学学报:自然科学版,2010,19(03):162-165.

[10] 范玉妹,玄婧. ARMA 算法在 GDP 预测中的应用[J]. 江南大学学报:自然科学版,2010,(06):736-740.

[11] 李美,干晓蓉,刘新乐. ARIMA 模型在股价预测上的应用及其傅里叶修正[J]. 云南师范大学学报,2011,31(05):50-55.

[12] Hu X T, Wang H, Xu L, et al. Predicting stock index increments by neural networks: The role of trading volume under different horizons[J]. Expert Syst Appl,2008,34(04):3043-3054.

【责任编辑：刘守义 英文编辑：刘彦哲】