Prediction of US Stocks Based on ARIMA Model

Boyu Xiao(✉)

Guangdong University of Foreign Studies, Guangzhou, China
xby2448330623@163.com

Abstract. Time series analysis method is an important part of statistics. It has practical applications in various fields from economics to engineering. Time series analysis includes analyzing time series data in order to extract meaningful features of data and predict future values. Box-Jenkins method belongs to regression analysis method and is the basic method of time series analysis and prediction. This paper describes the modeling method and implementation process of ARIMA. A time series is a series of data points, usually measured at uniform time intervals. Autoregressive integral moving average (ARIMA) model is a kind of linear model that can represent stationary and non-stationary time series. ARIMA model depends on autocorrelation mode to a large extent. This paper will discuss the application in stock price forecasting, especially the time sampling at different time intervals, to determine whether there are some optimal design frameworks and whether the stock autocorrelation patterns in the same industry are similar.

Keywords: ARIMA · Stock price forecast

1 Introduction

The development of the economy has led to the rapid development of the stock market, and the stock market has become another mirror of the national economy, and more and more people choose to invest idle funds in the low-cost, high-return stock market. Stock price prediction is an operation to find out the law of the stock market, make reasonable trend judgment according to the law, and then guide investment behavior. Therefore, stock price forecasting not only helps the government to carry out macro-control, but also guides investors to make rational choices [1, 2].

2 Modeling and Forecasting of FORECAST

2.1 Stepwise Autoregressive Models in Time Series Analysis

PROC FORECAST can be used to automatically model and predict AMEX closing prices, Jones Industrial average closing prices, and gold spot prices in New York City as data for step-by-step autoregressive and exponential smoothing models.

Before making predictions with PROC FORECAST, the datasets are consolidated and collated. Data set AMEX1, Data set GOLD, Dataset DJsections are shown in Table 1 [3].

Table 2 and Table 3 show the output of a stepwise autoregressive model.

Y. Jiang et al. (Eds.): ICFIED 2023, AEBMR 237, pp. 312–322, 2023.
https://doi.org/10.2991/978-94-6463-142-5_35

Table 1. Data set AMEX1, GOLD and DJ

	AMEX1day	AMEX1close	DJday	DJclose	GOLDday	GOLDclose
1	02AUG93	437.88	03JAN94	1144.88	O3JAH94	393. 7
2	03AUG93	437.29	04JAN94	1123.69	O4JAK94	394.1
3	04AIK93	436,42	05JAN94	1120.66	05JAH94	391.1
4	05AUG93	435.53	06JAN94	1133.16	06JAH94	389.2
5	06AUG93	436.34	07JWI94	1138.76	07JAH94	386.4
6	09AUG93	438.80	10JAN94	1169.91	10JAU94	385
7	10AUG93	439.19	34345	1171.74	UJAK94	388
8	UAUG93	438.87	34346	1153.29	12JAK94	386.3
9	12AUG93	437.58	13JAN94	1145.89	13JMJ94	390
10	I3AUG93	439.08	14JAN94	1149.78	I4JA1R4	389.5

Table 2. The output of a stepwise autoregressive model (1)

	type	day	close	closing price	CGMEX gold Spot Closing Price for Day
1	N	25MAR94	59	59	59
2	WRESID	25MAR94	59	59	59
3	DF	25MAR94	56	56	56
4	SIGhlA	25MAR94	12.43386	2.2308667	31662004
5	CONSTANT	25MAR94	1122.4786	484.24895	386.85501
6	LINEAR	25MAR94	0.8647664	-0.284715	-0.088152
7	ARI	25MAR94	0.880479	0.846846	0.7548211
8	AR2	25MAK94			
9	AR3	25MAR94			
10	AR4	25MAR94			

Table 3. The output of a stepwise autoregressive model (2)

	day	type	dj_close	closing price	CGMEX gold Spot Closing Price for Day
1	28MAR94	FORECAST	1190.8802343	467.9517099	388.6987963
2	29MAR94	FORECAST	1189.8743949	467.5030667	386.7647041
3	30HAR94	FORECAST	1189.0921321	467.7952977	385.3586797

(*continued*)

Table 3. (*continued*)

	day	type	dj_close	closing price	CGMEX gold Spot Closing Price for Day
4	31KAR94	FORECAST	1188.5067238	466.6772541	384.2757695
5	01APR94	FORECAST	1188.0946419	466.2929833	383.4367631
6	04APR94	FORECAST	1187.8361702	466.9239600	382.7818327
7	05APR94	FORECAST	1187.7100685	465.5678488	382.2658719
8	06APR94	FORECAST	1187.7032769	465.2226724	381.8548007
9	07APR94	FORECAST	1187.8006547	464.8867559	381.5229024
10	08APR94	FORECAST	1187.9897516	464.5586812	381.2507655

Table 4. Exponential smoothing model data

	type	day	dj_close	closing price	COMEX gold Spot Closing Price for Day
1	N	25-Mar-94	59	59	59
2	KRISID	25-Mar-94	59	59	59
3	DF	25-Mar-94	56	56	56
4	WEIGHT	25-Mar-94	0.2	02	0.2
5	SI	25-Mar-94	11914582	47010675	38767696
6	S2	25-Mar-94	11800942	469.71516	38493688
7	S3	25-Mar-94	1169.0996	46939076	38293243
8	SICMA	25-Mar-94	17901338	26944957	38799429
9	COKSTAITT	25-Mar-94	1203.1918	04705655	39115267
10	LIBEAR	25-Mar-94	3.0372982	01335819	10758252
…	…	…	…	…	…

2.2 Exponential Smoothing Model in Time Series Analysis (as Table 4)

2.3 Empirical and Comparative of Stepwise Autoregressive Models and Exponential Smoothing Models

For any time series, you must choose the model that best reflects the trend of the data, and the goodness-of-fit statistic is a common criterion for selecting models. Table 5 compares the variable CLOSE using a stepwise autoregressive model and an exponential smoothing model goodness-of-fit statistics.

It can be observed from the table that in the statistics (SSE, MSE, PMSE, MAPE, MPE, MAE, ME), the values corresponding to the stepwise autoregressive model are all smaller than the exponential smoothing model, such as SSE 293.28, which is smaller than

Table 5. Stepwise autoregressive models and exponential smoothing models goodness-of-fit comparison table

Extrapolate the time series model AMEX index closing price		
Statistics	AMEX Index Closing Price Model	
	Gradual self-regression	Exponential smoothing
SSE	293.28332	376.95883
MSE	5.2372021	6.7314077
RMSE	2.2884934	2.5944957
MAPE	0.3464997	0.3906927
MPE	0.0001961	0.0467643
MAE	1.6465708	1.8553361
ME	0.0108127	0.22273171
RSQUARE	0.8798778	0.8456062
ADJRSQ	0.8755877	0.8400922
RW_RSQ	-0.04846	-0.347592
ARSQ	0.8670076	0.829064
APC	5.5035505	7.0736827
AIC	100.6125	115.42131
SBC	106.84511	121.65393

SSE376.96; in addition, The value of the stepwise autoregressive model is closer to 1 than the exponential smoothing model, for example, RSQUARE (stepwise autoregressive model) 0.88 is greater than RSQUARE (exponential smoothing model) 0.85. This shows that the stepwise autoregressive model fits the past values of the AMEX index closing price series better.

2.4 Forecasting for Extrapolated Time Series Models

From this, the model predictions of the stepwise autoregressive and exponential smoothing models are obtained (as Fig. 1 and Fig. 2).

Interpretation of the results: The predictions of the autoregressive model and the exponential smoothing model predict different expectations. The autoregressive model predicts an uptrend for the DJIA and a downtrend for the AMEX index closing price and gold spot price, while the exponential smoothing model predicts an uptrend for the DJIA and an uptrend for the AMEX index and gold.

Forecast Stepwise Autoregressive Model						
DJIA, AMEX, and Gpld						
Obs	day	_TYPE_	_LEAD_	dj_close	close	gold
1	28MAR94	FORECAST	1	1190.88	467.952	388.599
2	29MAR94	FORECAST	2	1189.87	467.503	386.765
3	30MAR94	FORECAST	3	1189.09	467.080	385.359
4	31MAR94	FORECAST	4	1188.51	466.477	384.276
5	01APR94	FORECAST	5	1188.09	466.293	383.437
6	04APR94	FORECAST	6	1187.84	465.924	382.782
7	05APR94	FORECAST	7	1187.71	465.568	382.266
8	06APR94	FORECAST	8	1187.70	465.223	381.855
9	07APR94	FORECAST	9	1187.80	464.887	381.523
10	08APR94	FORECAST	10	1187.99	464.559	381.251

Fig. 1. Forecast Stepwise Autoregressive Model

Exponential Smoothing Model Model Prediction						
DJIA, AMEX, and Gpld						
Obs	day	_TYPE_	_LEAD_	dj_close	close	gold
1	28MAR94	FORECAST	1	1206.24	470.701	392.251
2	29MAR94	FORECAST	2	1209.31	470.841	393.396
3	30MAR94	FORECAST	3	1212.41	470.985	394.587
4	31MAR94	FORECAST	4	1215.53	471.133	395.824
5	01APR94	FORECAST	5	1218.67	471.286	397.107
6	04APR94	FORECAST	6	1221.83	471.443	398.435
7	05APR94	FORECAST	7	1225.02	471.603	399.810
8	06APR94	FORECAST	8	1228.23	471.769	401.231
9	07APR94	FORECAST	9	1231.46	471.938	402.697
10	08APR94	FORECAST	10	1234.72	472.111	404.210

Fig. 2. Exponential Smoothing Model Prediction

3 Build Models with PROC ARIMA

3.1 Build and Compare Time Series Analysis Models by PROC ARIMA

The ARIMA model has three parameters (p, d, q), where p refers to the order of the autoregressive part of the model, d refers to the number of sequence differences, and q refers to the number of average moving parts of the model. The AMEX index closing price series is shown in Fig. 3. After the first differencing, the sequence exhibits stability. After selecting an appropriate model, the sequence can be predicted [4, 5].

Through the analysis of the above results, the following conclusions are drawn:

(1) Parameter estimates, approximate standard errors, t-ratios, and lags for the specified model. The only parameters estimated are the mean (MU or constant) with a value of -0.15316, an approximate standard error of 0.44596 and a t-ratio of -0.34.
(2) The constant estimate represents the intercept parameter of the MA model adjusted for all AR parameters. If the AR parameter is not included in the model, the constant

A Stochastic Model of the First Difference of AMEX Index Closing Prices
process of ARIMA

Conditional Least Squares Estimation					
parameter	estimate	standard error	t value	approximate Pr>|t|	lag
MU	-0.15316	0.44506	-0.34	0.7347	0

constant estimation	-0.1532
variance estimation	3.7635
standard error estimate	1.93997
AIC	80.074
SBC	81.0185
number of residuals	19

Autocorrelation Checks for Residuals								
lag	Chi-square	degrees of freedom	Pr>Chi-square	Autocorrelation					
6	1.23	6	0.9754	0.072	-0.146	-0.097	0.005	0.079	-0.082
12	4.16	12	0.9804	0.077	-0.214	-0.083	-0.116	-0.051	0.035
18	13.06	18	0.788	-0.047	-0.129	-0.101	0.159	0.120	0.020

Fig. 3. A Stochastic Model of the First Difference of AMEX Index Closing Prices

Table 6. Model Comparison Results

Model Statistics	Stochastic Models with Trends (0,1,0)	random walk (1,1,0)	AR (1) (1,1,0)	MA (1) (0,1,1)	ARIMA (1,1,1)
Parameter Estimation	MU-0.15316 (-0.34)	N/A	MU-0.16306(-0.33)	MU-0.16186(-0.32)	MU-0.11633 (-0.24)
			AR1 0.07326(0.30)	MA1-0.09860(-0.41)	MA1-0.96017 (-5.46)
					AR1-0.81492 (-2.82)
variance	3.763501	3.588879	3.963963	3.95612	3.993331
AIC	80.07403	78.19862	81.97402	81.93639	82.9624
SBC	81.01847	78.19862	83.8629	83.82527	85,79572
Q					
lag6	123 (0.9754)	1.38	(0.9669)	117 (0.9479)	0.76 (0.9439)
Iagl2	4.16 (0.9804)	4.26	(0.9783)	411 (0.9665)	2.96 (0.9823)
Iagl8	13.06 (0.7880)	14.52	:(0.6945)	12.19 (0.7887)	10.51 (0.8386)

estimate and the mean parameter estimate are identical. If the model contains an AR(p) component, the constant in the output is estimated as.

(3) The goodness-of-fit statistics are variance, standard deviation, Akaike Information Criterion (AIC) and Schwartz-Bayesian Criterion (SBC). The better the estimated model fits, the smaller these statistics will be.

(4) A list of test statistics (i.e. chi-square or Q-statistics) for the white noise hypothesis of the fitted model residuals. The null hypothesis is that the residuals are white noise. The p-value indicates that the null hypothesis cannot be rejected at the 0.05 significance level.

The above table only outputs random models, and Table 6 lists the results for these models for ease of comparison [6].

Overall, the random walk (0, 1, 0) model and the ARIMA (1, 1, 1) model are better than other models, so they are used as alternative models for the next comparison.

Obs	day	close	f_random	STD	l95_r	u95_r	RESIDUAL
15	18MAR94	472.96	470.63	1.89443	466.917	474.343	2.33
16	21MAR94	470.77	472.96	1.89443	469.247	476.673	-2.19
17	22MAR94	471.73	470.77	1.89443	467.057	474.483	0.96
18	23MAR94	473.38	471.73	1.89443	468.017	475.443	1.65
19	24MAR94	469.66	473.38	1.89443	469.667	477.093	-3.72
20	25MAR94	468.43	469.66	1.89443	465.947	473.373	-1.23
21	28MAR94	.	468.43	1.89443	464.717	472.143	.
22	29MAR94	.	468.43	2.67913	463.179	473.681	.
23	30MAR94	.	468.43	3.28126	461.999	474.861	.
24	31MAR94	.	468.43	3.78887	461.004	475.856	.
25	01APR94	.	468.43	4.23608	460.127	476.733	.
26	04APR94	.	468.43	4.64040	459.335	477.525	.

Obs	day	close	f_arima	STD	l95_a	u95_a	RESIDUAL
15	18MAR94	472.96	470.924	1.99833	467.008	474.841	2.03553
16	21MAR94	470.77	472.805	1.99833	468.888	476.721	-2.03456
17	22MAR94	471.73	470.390	1.99833	466.473	474.307	1.33997
18	23MAR94	473.38	472.023	1.99833	468.106	475.940	1.35685
19	24MAR94	469.66	473.127	1.99833	469.210	477.044	-3.46706
20	25MAR94	468.43	469.151	1.99833	465.235	473.068	-0.72141
21	28MAR94	.	468.529	1.99833	464.612	472.445	.
22	29MAR94	.	468.237	3.03825	462.282	474.192	.
23	30MAR94	.	468.263	3.66632	461.078	475.449	.
24	31MAR94	.	468.031	4.29896	459.605	476.457	.
25	01APR94	.	468.009	4.77909	458.642	477.376	.
26	04APR94	.	467.816	5.26775	457.491	478.140	.

Fig. 4. Different models of AMEX index closing prices correspond to forecast values

3.2 Comparing the Random Walk (0, 1, 0) Model and the ARIMA (1, 1, 1) Model

3.2.1 Forecasting Using PROC ARIMA

You can use PROC ARIMA to predict the future value of the time series, use the ARIMA (0,1,0) model and ARIMA(1,1,1) model with a certain trend to predict the future value of the AMEX index closing price, and output the predicted value (as Fig. 4) [7, 8].

The table above outputs the actual and predicted values, standard errors, 95% upper and lower confidence limits, and residuals for the two models, whose point estimates differ only slightly. The predicted value of the random walk model on March 28, 1994 is 468.43, while the predicted value of the ARIMA (1, 1, 1) model is slightly lower than the former. The prediction confidence interval (upper bound minus lower bound) of the random walk model (0, 1, 0) is smaller than that of the ARIMA (1, 1, 1) model. This is because ARIMA (0, 1, 0) has fewer parameters to estimate and the standard error STD is small, so if the predictions are very similar, the ARIMA (0, 1, 0) random walk model should be chosen.

3.2.2 AMEX Prediction with PROC ARIMA

The examples in this section examine the closing prices of the AMEX index in early 1994. A time series model fits the series. Through the figure below, it is believed that the sequence may show a certain trend during the period from February 28 to March 25.

As can be seen from the chart below, the sequence reached 473.38 points on March 23rd, then fell to 469.66 points on the 24th, and extended to 468.43 points on the 25th. The series of predicted values for the random walk model in the table below will remain at 468.43 in the short term. A very important question is whether the series reached a turning point on March 23rd that created a new trend, or whether the apparent dip from March 23rd to 25th was just random fluctuations.

If the previous trend was still valid, the series would be expected to follow the predicted values in the graph below, but still between the upper and lower confidence limits. The rule of thumb for technical analysis is that a sequence is valid until there is sufficient evidence that a new trend will be established [9, 10].

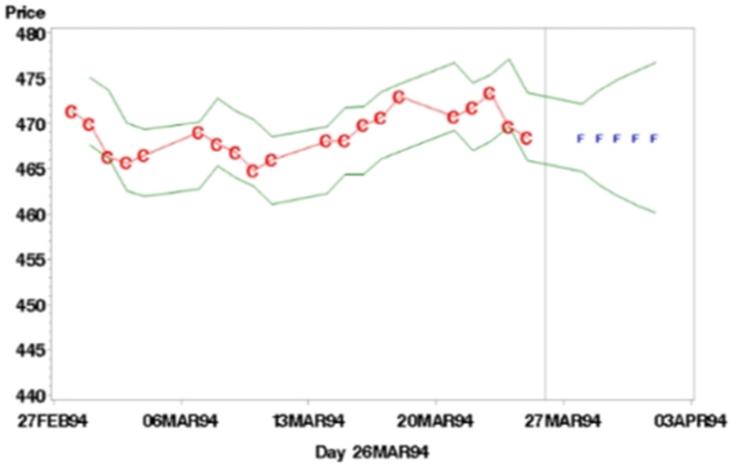

Fig. 5. AMEX index closing price forecast (1)

3.3 95% Confidence Interval Plot for Random (0, 1, 0) Model

The following three graphs generated by PROC GPLOT show the predicted xvalue of the series and the upper and lower confidence limits of the predicted value day by day (as Fig. 5, Fig. 6 and Fig. 7). The first graph is from March 26, and the second graph is added to the AMEX closing price on March 30. Technical analysis charts usually have multiple interpretations. Before using PROC GPLOT, merge the data sets through the DATA step and delete unnecessary values. PROC SORT is used to ensure that these observations are plotted in the proper order.

3.4 95% Confidence Interval Plot for a Stochastic (0, 1, 0) Model with Added Information

Add to the actual closing data on March 30, 1994.

3.5 95% Confidence Interval Plotting for the Random (0, 1, 0) Model for More Information

Adding to the actual closing data for April 11, 1994.

In a time series analysis model, if the trend is still valid, the series will be carried forward with predicted values and confidence limits. The above chart shows that the actual closing price on the 28th was lower than the predicted value and lower than the lower confidence limit. This observation is a strong indication that the sequence has changed. The chart above shows that the actual closing prices on the 29th, 30th and 31st continued to fall and were well below the lower confidence limit.

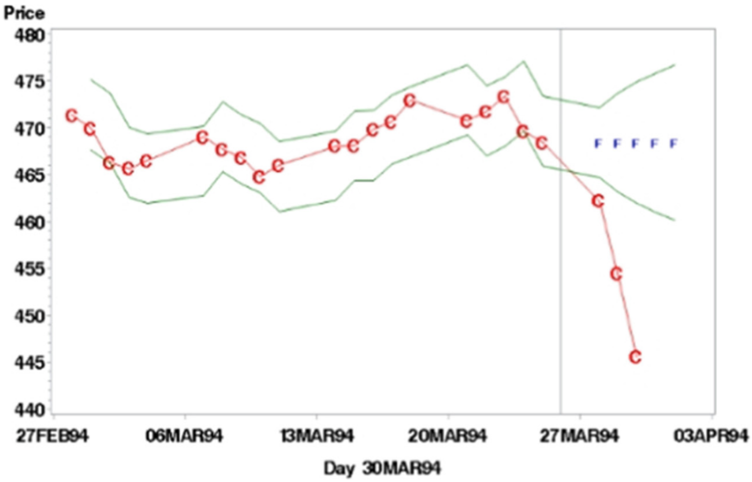

Fig. 6. AMEX index closing price forecast (2)

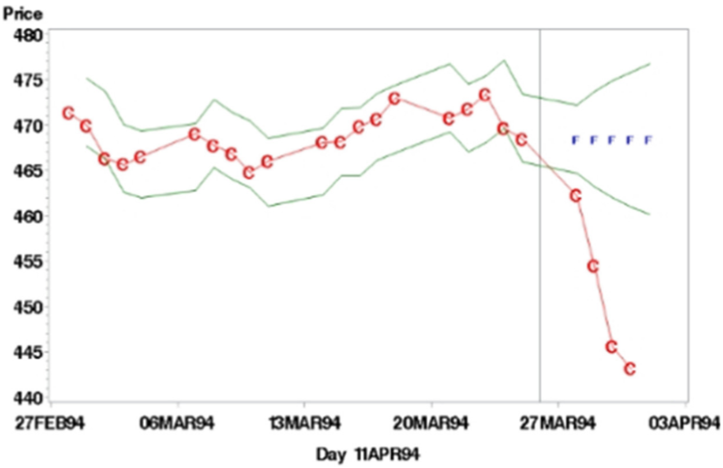

Fig. 7. AMEX index closing price forecast (3)

4 Tests for AMEX Predicted Values

The above model can be used to predict the AMEX index closing price (CLOSE). This example uses the second intervention model to predict the value of the CLOSE variable on March 29 and March 30.

Figure 8 is the predicted value of the random walk model (0, 1, 0) without using the intervention model before.

The actual values, predicted values, 95% confidence intervals, and known residuals are listed above for March 29. In the above table, using the information on March 25, the stochastic model was fitted to obtain the closing price of the AMEX index for the week

Obs	day	close	f_random	STD	l95_r	u95_r	RESIDUAL
15	18MAR94	472.96	470.63	1.89443	466.917	474.343	2.33
16	21MAR94	470.77	472.96	1.89443	469.247	476.673	-2.19
17	22MAR94	471.73	470.77	1.89443	467.057	474.483	0.96
18	23MAR94	473.38	471.73	1.89443	468.017	475.443	1.65
19	24MAR94	469.66	473.38	1.89443	469.667	477.093	-3.72
20	25MAR94	468.43	469.66	1.89443	465.947	473.373	-1.23
21	28MAR94	.	468.43	1.89443	464.717	472.143	.
22	29MAR94	.	468.43	2.67913	463.179	473.681	.
23	30MAR94	.	468.43	3.28126	461.999	474.861	.
24	31MAR94	.	468.43	3.78887	461.004	475.856	.
25	01APR94	.	468.43	4.23608	460.127	476.733	.
26	04APR94	.	468.43	4.64040	459.335	477.525	.

Fig. 8. AMEX index closing price forecast based on stochastic (0, 1, 0) model without intercept term

of March 28 to be 468.43; the actual values of March 28 and 29 were 462.21 and 454.43, Then the predicted value of the random walk model for March 30 and 31 (468.43) is questionable. Because the intervention model forecast includes two other actual series values and takes advantage of the downtrend from March 23rd, it should be more precise [6, 11].

5 Conclusion

To sum up, the intervention model is relatively accurate relative to the random walk model (0, 1, 0) and the ARIMA model, but there is also a certain prediction bias. When you are satisfied with the analysis and forecast of the series, the next practical operation is to sell stocks that continue to decline and buy and hold stocks that are going to rise. When buying and selling stocks, you can control the situation through the choice of buying stocks, buy stocks that are forecast to rise at any time, and sell stocks that are forecast to fall, so as to reap benefits in the ups and downs.

References

1. Burges C. A Tutorial on Support Vector Machines for Pattern Recognition[J]. Data Mining and Knowledge Discovery, 1998, 2.
2. Wang Xiaopeng, Cao Guangchao, Ding Shengxi.Analysis, modeling and prediction of precipitation time series on the Qingnan Plateau based on Box-Jenkins method[J]. Mathematical Statistics and Management,2008(04):565-570.DOI: https://doi.org/10.13860/j.cnki.sltj.2008.04.001.
3. Chi Qishui. Analysis on the growth trend of China's oil consumption——Prediction and analysis based on ARIMA model[J].Resource Science,2007(05):69–73.
4. Bai Yingshan. Prediction and Analysis of CSI 300 Index Based on ARIMA Model[D].South China University of Technology,2010.
5. Ding Wenjuan. Comparison of ARIMA Model and LSTM Model Based on Stock Prediction[J].Industrial Control Computer,2021,34(07):109–112+116.

6. Hou Lu. Short-term analysis and forecast of oil price based on ARIMA model[D]. Jinan University,2009.
7. Gong Guoyong. Application of ARIMA Model in Shenzhen GDP Forecast[J]. Practice and Understanding of Mathematics,2008(04):53-57.
8. Hua Peng, Zhao Xuemin. Application of ARIMA Model in GDP Forecasting of Guangdong Province[J]. Statistics and Decision-Making,2010(12):166-167.DOI: https://doi.org/10.13546/j.cnki.tjyjc.2010.12.016.
9. Li Shengbiao. Analysis and prediction of commodity housing prices in Lanzhou City based on ARIMA model [J]. Gansu Science and Technology,2014,30(21):93–94+66.
10. Xu Liping, Luo Mingzhi. Short-term analysis and forecast of gold price based on ARIMA model[J]. Finance and Economics,2011(01):26-34.
11. Wu Yuxia, Wen Xin. Short-term stock price forecast based on ARIMA model[J]. Statistics and Decision-Making, 2016(23): 83-86. DOI: https://doi.org/10.13546/j.cnki.tjyjc.2016.23.051.