David R. Brillinger
Statistics 153

An Example of Forecasting an Economic Series

Introduction. **The Dow Jones Industrial Average (DJI) is an important indicator of the state of the New York Stock market. In this report we present the results of studying the ability to forecast it using the methods of Brockwell and Davis [1]. There follows a description of the DJI.**

Dow Jones Industrial Average.
The Dow, as it is popularly known, is probably the most widely watched indicator of American stock market movements. The Dow has many virtues: it is more than 100 years old, it is well known, and by including only 30 stocks, it is manageable. These stocks tend to be those of the largest, most established firms and represent a range of industries, too. Unfortunately, there are only 30 of them, and they are not always an ideal proxy for the thousands of stocks that make up the market as a whole. In recent years, broader indexes such as the Standard & Poor's 500 (for large companies), the Russell 2000 (for smaller companies) and the Wilshire 5000 (for an especially broad measure) have gained currency, in part due to the rising popularity of index investing. See [2].

The specific question is how well can we forecast the next 7 days of DJI values given the preceding 218 values. We will address this by developing forecasts and associated uncertainties.

The sections of the paper are Introduction, The Data, Preliminary Analysis, ARIMA Fitting, Assessment of the Fit, Forecasting the Series, Examining the Forecasts, Conclusions, References, Appendix.

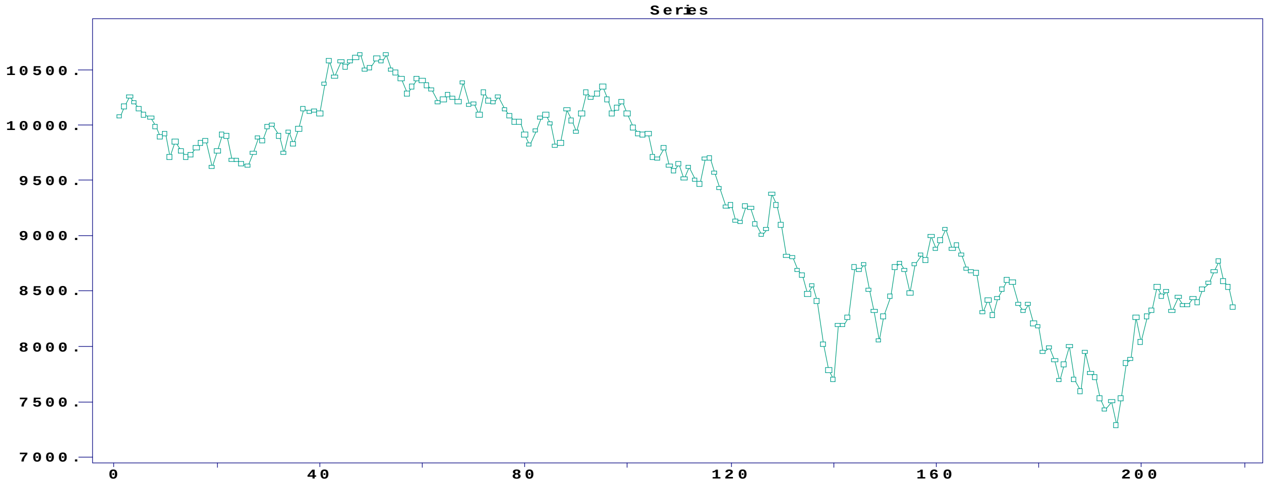
The Data. **The data used in the estimation and fitting part of our study are the daily closing values of the DJI from Jan. 28, 2002 up to Nov. 11, 2002. They were found at the website finance.yahoo.com by clicking on dow and then on historical prices.**

The data used in the assessment part of our work were for the next 7 days of values, i.e. Nov. 12-20. They were found on the same site.

ITSM>PROJECT>OPEN>UNIVARIATE
FILE>IMPORT FILE

A:dow.tsm

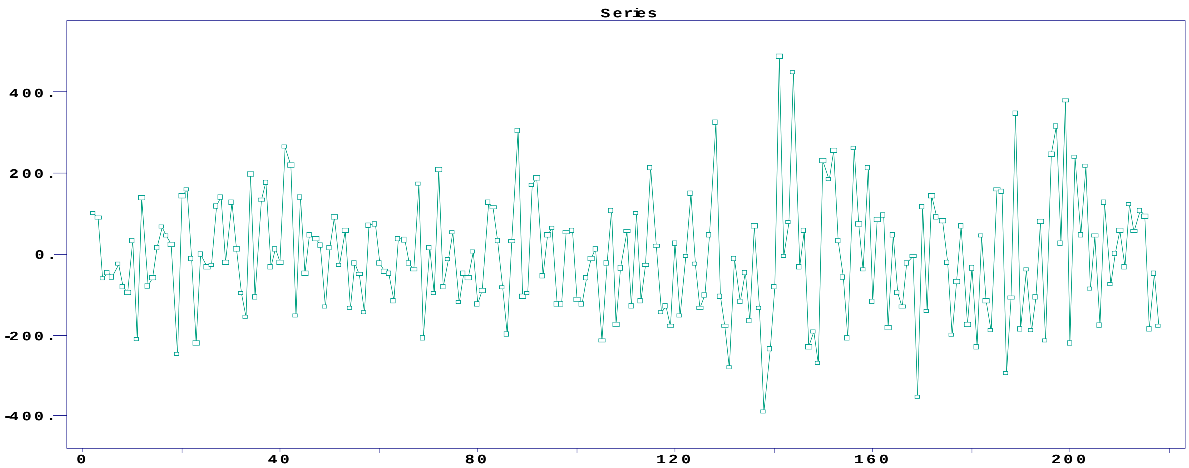
Preliminary Analysis. **We begin by graphing the data. In the figure below we see that the index fluctuated around the level of 10000 for approximately the first half of the study period. Then it fell to about 7800, recovered for a while, then sunk and recovered. It was falling again the last few days.**



The series seems to be wandering about rather like a random walk. This suggests working with the series of differences, i.e. the daily changes. This is common practice when dealing with stock prices, indeed the newspaper seem to always list the changes as well as the prices.

The next figure graphs the daily changes, that is the first difference of the original data. One notices that this series is much more noise like. The mean level appears approximately constant. There are a couple of periods of extra-variability in the last third of the plot. Overall it appears reasonable to proceed as if the series were stationary.

Transform>Difference
   Enter lag 1



Some basic statistics of the data set are:

# of Data Points =   217

Sample Mean = -7.9007
Sample Variance = .207931E+05
Std.Error(Sample Mean) = 8.821715

ARIMA Fitting. The class of ARIMA processes, see [1], proves particularly appropriate when forecasting series that may be made approximately stationary by differencing. As suggested above we fit a model assuming the first differences are stationary, i.e. an ARMA(p,1,q).

The “best” model is obtained by examining models with p and q <= 5 and looking for the minimum of the AIC criterion, see [1]. The commands and results follow.

Model>Estimation>Autofit
   Subtract mean

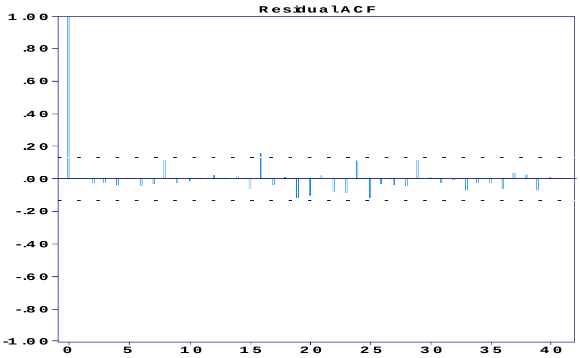
```
========================================
ITSM::(Maximum likelihood estimates)
========================================

Method: Maximum Likelihood


ARMA Model:
X(t) = Z(t) - .04158 Z(t-1)

WN Variance = .207590E+05

MA Coefficients
      -.041581

Standard Error of MA Coefficients
       .069858

(Residual SS)/N = .207590E+05

AICC = .277702E+04
BIC  = .277292E+04

-2Log(Likelihood) = .277296E+04

Accuracy parameter = .100000E-08

Number of iterations = 6

Number of function evaluations = 15701

Optimization stopped with gradient near zero.
```

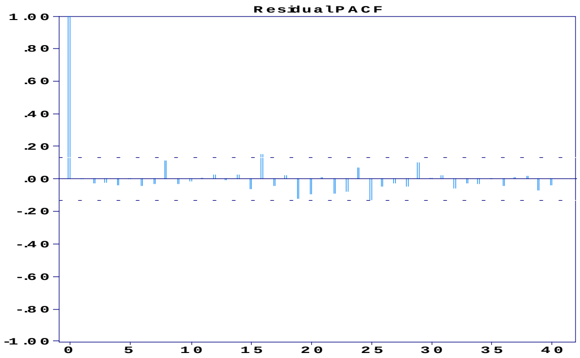
The model obtained is an ARIMA(0,1,1). We are not concerned with the uncertainties of the coefficients because we will be using the model for forecasting, not interpretation.

Assessment of the Fit. Before proceeding to make forecasts it is essential to assess the fit of the estimated model. We do this by looking at the estimated acf and pacf and some particular goodness of fit statistics.

All the values of the estimated acf and pacf are within the +- 2 standard error limits so we are not led to reject the hypothesis the the noise series of the ARIMA model is not white.

Statistics>Residual Analysis>ACF/PACF





Statistics>Residual Analysis>Tests of Randomness

The results of some specific goodnesss-of-fit tests suggsted in [1] follow. Using the 5% level of significance the only significant result concerns the normality of the error series.

Normality is assumed in developing the maximum likelihood estimates of the parameters of the ARIMA model, but it has been found that the estimates are robust to modest non-normality [1].

==
ITSM::(Tests of randomness on residuals)
==

Ljung - Box statistic = 17.791 Chi-Square (20), p-value = .60120

McLeod - Li statistic = 72.155 Chi-Square (21), p-value = .00000

 # Turning points = .14600E+03~AN(.14333E+03,sd = 6.1851), p-value = .66636

 # Diff sign points = .10700E+03~AN(.10800E+03,sd = 4.2622), p-value = .81450

Rank test statistic = .11346E+05~AN(.11718E+05,sd = .53459E+03), p-value = .48652

Jarque-Bera test statistic (for normality) = 9.2430 Chi-Square (2), p-value = .00984

Order of Min AICC YW Model for Residuals = 0

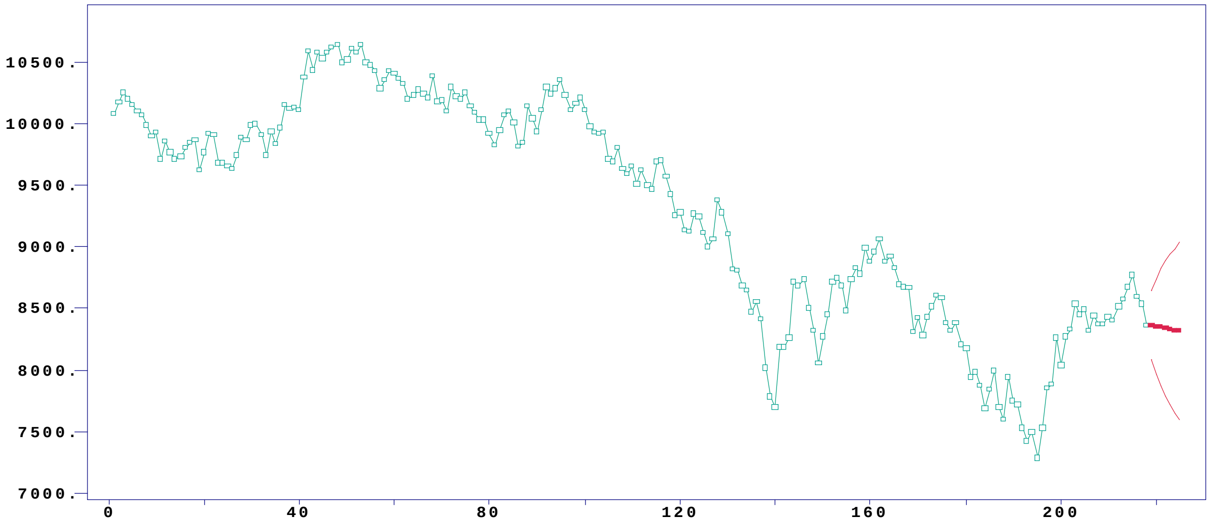
It is interesting that two of the tests lead to rejection of the white noise hypothesis. The McLeod-Li test is sensitive to non-normality of the residuals, as is the Jarque-Bera. (The histogram of the rescaled

residuals is provided in the Appendix. The histogram looks a bit skewed.)

Forecasting the Series. The future values of an ARMA may be forecast using the program ITSM [1]. The results follow.

It is noticeable that the uncertainty limits are quite broad. Given the variability of the series seen in the first Figure this is not surprising.

Forecasting>ARMA
Enter number 7
Plot 95 percent bounds



INFO
====================
ITSM::(ARMA Forecast)
====================

Step	Prediction	sqrt(MSE)	Approximate 95 Percent Prediction Bounds Lower	Upper
1	.83582E+04	.14408E+03	.80758E+04	.86406E+04
2	.83503E+04	.19957E+03	.79592E+04	.87415E+04
3	.83424E+04	.24269E+03	.78668E+04	.88181E+04
4	.83345E+04	.27922E+03	.77872E+04	.88818E+04
5	.83266E+04	.31150E+03	.77161E+04	.89371E+04
6	.83187E+04	.34074E+03	.76509E+04	.89865E+04
7	.83108E+04	.36765E+03	.75902E+04	.90314E+04

Examining the Forecasts. The actual values of the DJI for the following 7 days are:

8386.00, 8398.49, 8542.13, 8579.09, 8486.57, 8474.78, 8623.01

Each lies within the approximate 95% limits, so the forecasting may be viewed as successful. However the uncertainty limits of the forecasts are so broad that these results may not prove too useful in practice.

Conclusions. **We have fit an ARIMA(2,1,1) to the series and the fit seemed reasonable. The predictions made using the model had high uncertainty and were not contradicted by the actual values of the DJI.**

References.

[1] Brockwell, P. J. and Davis, R. A. (2002). *Introduction to Time Series and Forecasting, Second Edition*. **Springer, New York.**

[2] //moneycentral.msn.com/investor/glossary/glossary.asp?TermID=130

Appendix.