Statistical Tools for Quantitative Risk Management

Assignment I: Time Series

Vincent Buekers
r0754046

G0Q24A
2019-2020

Contents

1 IBM Data

The first part of this report is concerned with analyzing a time series on the log-prices of the IBM stock, which has been recorded between 10/28/2002 and 3/18/2005. The time series consists of 602 measurements in total, measured with an approximate daily frequency. That is, some dates have not been recorded, yielding a rather irregular measurement frequency. However, these unrecorded days will not be considered missing data in the forthcoming analysis.

1.1 Exploratory Data Analysis

As a preliminary means of analyzing the data, a regular time-series plot is provided for both the time series in log-levels (Fig. 1, left) and in first log-differences (Fig. 1, right). The time series does not appear to be stationary at first glance. Moreover, there is an upward trend in the logged IBM stock over time. As such, the sample average (4.474) provides little insight into the series. would provide On the other hand, the series in differences does seem stationary. Whether or not the series is white noise should be more carefully assessed by means of a correlogram and and a Ljung-Box test.

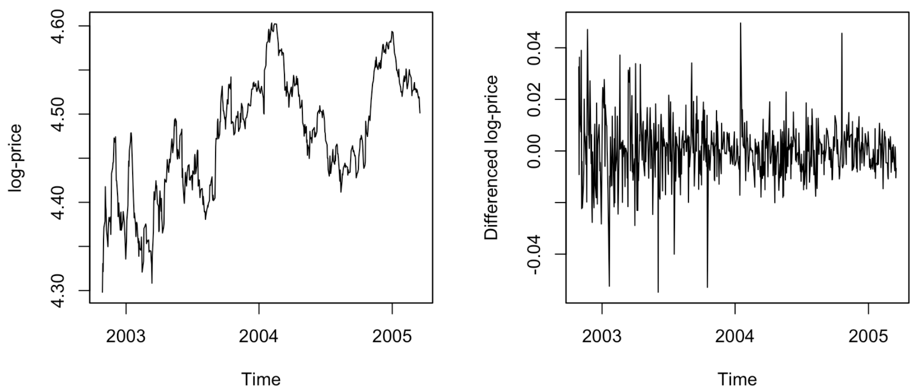

Figure 1: Time Series Plots, in levels (left) and in differences (right)

The results of the Ljung-Box test are $p < 2.2^{-16}$ and $p = 0.02461$ for the series in levels and first differences respectively. Hence, H_0 is rejected for both the series in levels and in differences at $\alpha = 0.05$. As such, neither of them can be considered white noise. Indee, the correlogram (Fig. 2, left) indicates that the series in levels has many significant autocorrelations which decay only very slowly. Conversely, the series in differences (right) only indicates 3 significant autocorrelations at lags 1, 5 and 13.

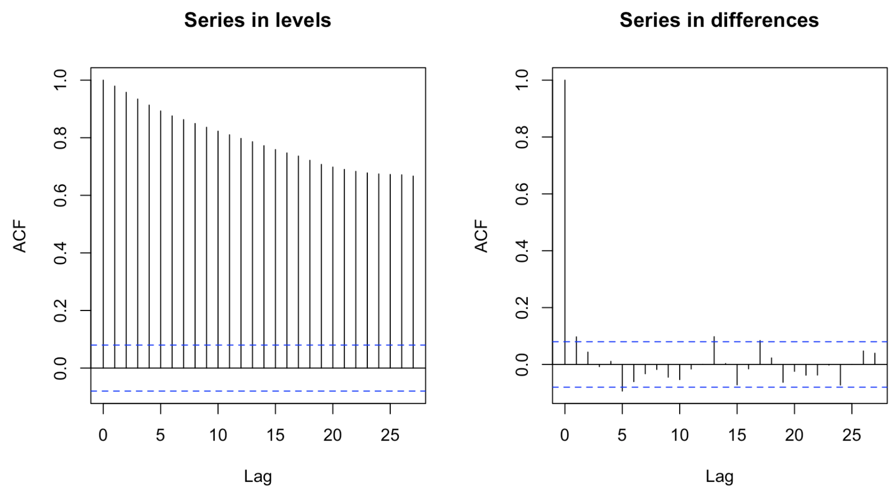

Figure 2: Correlogram, in levels (left) and in differences (right)

1.2 Modelling

In this part of the analysis, the appropriate model should be identified. Constructing a parsimonious model based on the series in levels is quite unlikely. Also, the ADF-test resulted in $p = 0.1285$, indicating a unit root exists. It indeed turned out that a simple AR(1) model had a root very close to 1, namely 1.011. This concludes the analysis for the series in levels. The series in first log-differences should allow for some viable options to model the IBM stock. The ADF-test resulted in $p < 2.2^{-16}$, indicating the series in first differences is stationary.

Based on Fig. 3, it would seem an ARIMA(1,1,0) is a plausible structure since there is a significant partial autocorrelation at lag 1 (i.e. AR(1)) Although there is a significant autocorrelation at lag 1, it is possibly just due to chance. Indeed, the first order Moving average term was not significantly different from 0 upon evaluation of an ARIMA(1,1,1). However, The first order autoregressive term from the ARIMA(1,1,0) model is significantly different from 0. The coefficient is estimated as $\hat{\phi}_1 = 0.0991$ with a corresponding standard error of 0.0408. This result is in agreement with the ADF-test since $|\hat{\phi}_1| < 1$. Also, the root of the characteristic equation is 10.092.

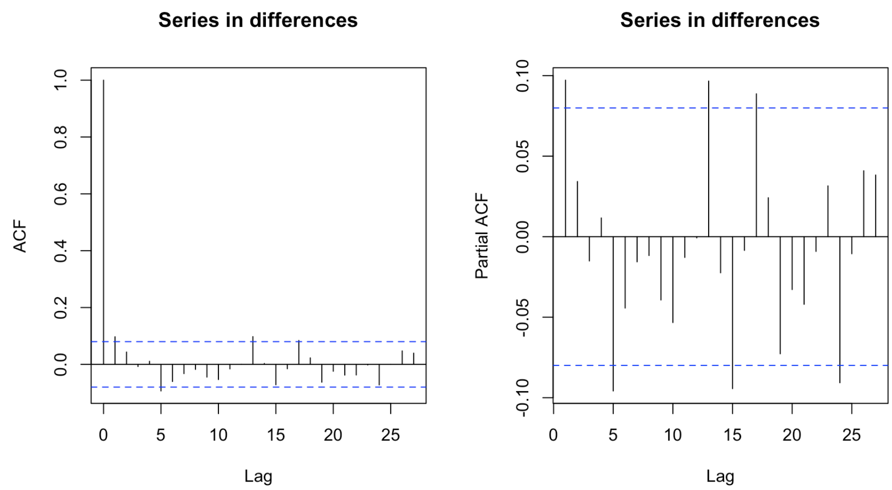

Figure 3: ACF (left) and PACF (right) of series in first differences

1.3 Residuals

Next, the model adequacy can be evaluated by inspecting the residuals of the ARIMA(1,1,0) model. Fortunately, these do seem to satisfy the characteristics of a white noise process. Hence, there is evidence for the validity of the model since there should no longer be an informative structure left within these residuals. The Ljung-Box test also gives a p-value of 0.2867, indicating the residuals are indeed white noise and thus reinforcing the former statements.

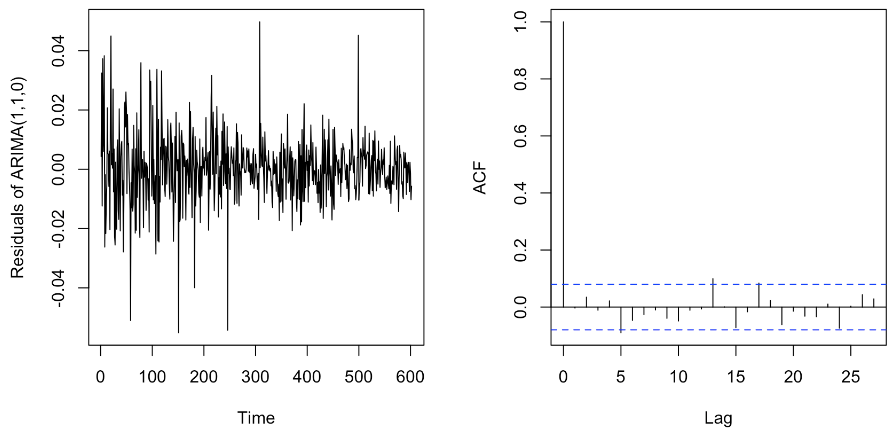

Figure 4: Residuals of ARIMA(1,1,0)

1.4 Forecast

The model can now be used to predict the last 5 observed values of our training data. To that end, the model has been retrained without considering these same observations. The difference with the former estimate is negligible as it is very small. The predictions for these particular observations and their corresponding true values are given below in Table 1. Moreover, Fig. 5 should provide a visual depiction of the predictions (blue), with corresponding control limits (red) and observed values (black). It can be seen that the predictions become more inaccurate as the forecast horizon increases, as well as diverging forecast limits.

LCL	Expected	Observed	UCL
4.503227	4.525685	4.518207	4.548143
4.492310	4.525667	4.519794	4.559024
4.484051	4.525665	4.518421	4.567279
4.477169	4.525665	4.508047	4.574161
4.471148	4.525665	4.501299	4.580181

Table 1: Forecasts with corresponding lower contol limit, upper control limit and observed values

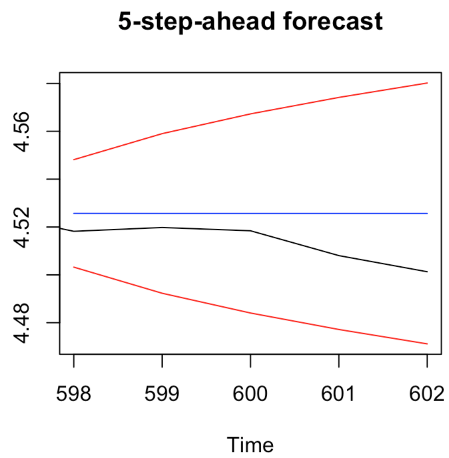

Figure 5: 5-step ahead forecast

2 Simulated Data

Unlike the IBM stock, the data are now purely simulated. Nevertheless, the same means of analysis are applicable to the 200 simulated points in this data set. Evidently, there is no information on the measurement frequency of this time series.

2.1 Exploratory Data Analysis

The sample average of the data is calculated to be about -0.155. Visually, the series indeed seems to be approximately centered around zero on average (Fig. 6, left). It is difficult to assess stationarity from this figure, yet the series in first differences does appear stationary at first glance (Fig. 6, right). The Ljung-Box test respectively results in $p < 2.2^{-16}$ and $p = 2.377^{-16}$. Consequently, neither one of the series can be considered a white noise process. Furthermore, the correlogram it is very clear this is not a white noise process. The Ljung-Box test also rejects H_0 ($p < 2.2^{-16}$).

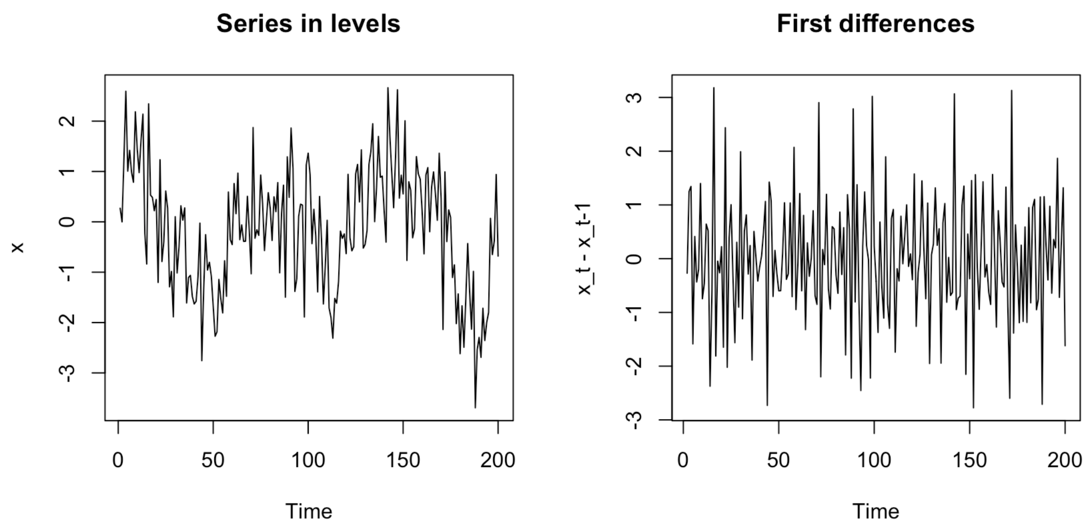

Figure 6: Time series plot (left) and ACF (right)

2.2 Modelling

Since the series is not a white noise process, there ought to be some dynamic present that can be translated into a model. While it is not immediately how to decide on the MA for the series in levels (Fig. 7, top left), the PACF (Fig. 7, bottom left) is more indicative for the AR term. More specifically, there seem to be 2 considerable partial autocorrelations. For comparative purposes with regards to the forecasts, the model for the series in differences will be the same as that of the series in levels, even though the (partial) correlogram would suggest otherwise.

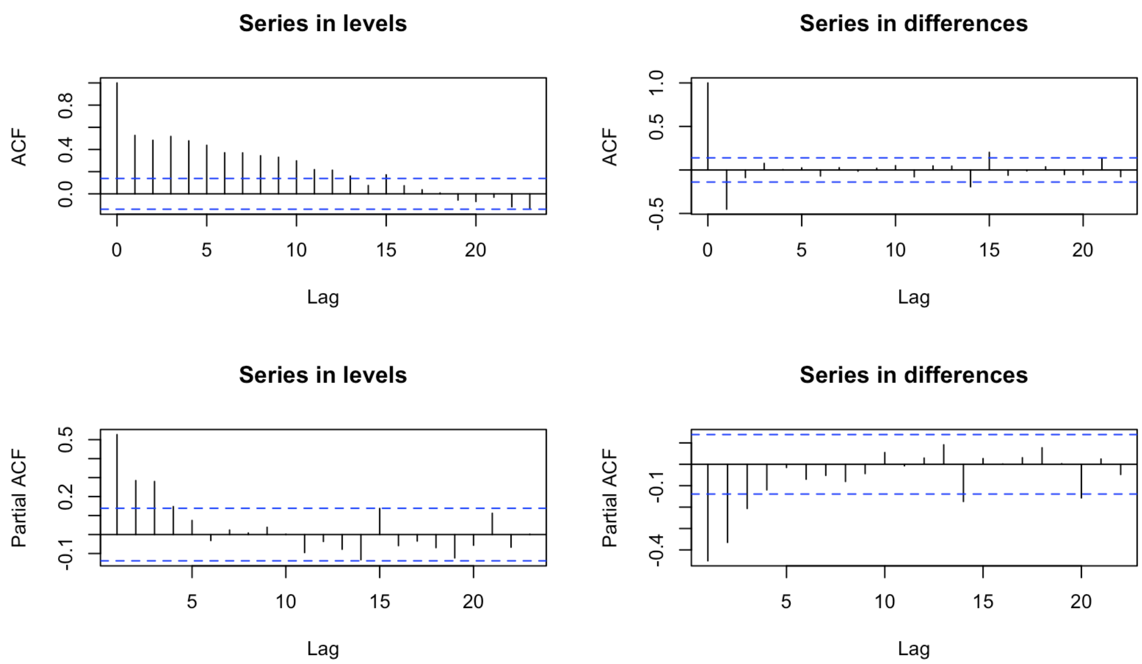

Figure 7: (Partial) Correlograms

To make a less heuristic decision, several models will be evaluated in terms of their AIC and BIC values. The order to be considered for both the moving average and the autoregressive polynomial is 2. This gives way to $3^2 = 9$ amount of combinations, listed below in Table 2. Both the AIC and BIC are minimal for the ARMA(1,1) model.

p	q	AIC	BIC
0	0	647.0848	652.3730
0	1	608.0511	618.6276
0	2	598.1463	614.0111
1	0	584.2564	610.6978
1	1	**550.5125**	**582.2421**
1	2	552.2342	589.2520
2	0	569.5693	617.1637
2	1	552.3220	605.2047
2	2	553.4086	611.5795

Table 2: AIC and BIC for ARMA(p,q)

The resulting coefficient estimates for the non-integrated model are $\hat{\phi}_1 = 0.9424$ and $\hat{\theta}_1 = -0.6757$ with corresponding standard errors of 0.0294 and 0.0606 respectively. Consequently, the model terms can be considered significantly different from 0. The ARMA(1,1) model turns out to have a root close to 0 (1.061). Hence, this model should very closely resemble a random walk process. Indeed, the ADF-test confirms this finding ($p = 0.2378$). Also, the AR(1) coefficient estimate is indeed close to 1 in absolute value. In an analysis involving actual data it would hence be best to model the series in differences instead. Upon evaluation of an ARIMA(1,1,1) model, the unit root of the AR term is indeed well above unity. The ADF-test also rejects non-stationarity ($p < 0.01$).

2.3 Forecast

Lastly, a 20-step ahead forecast needs to made for the simulated data. This time, the model is not retrained as it is desired to predict the future rather than already observed data points. As it turns out, whether or not the series is taken in differences has little effect on the forecast for the 20-step ahead forecast. This concludes the analysis.

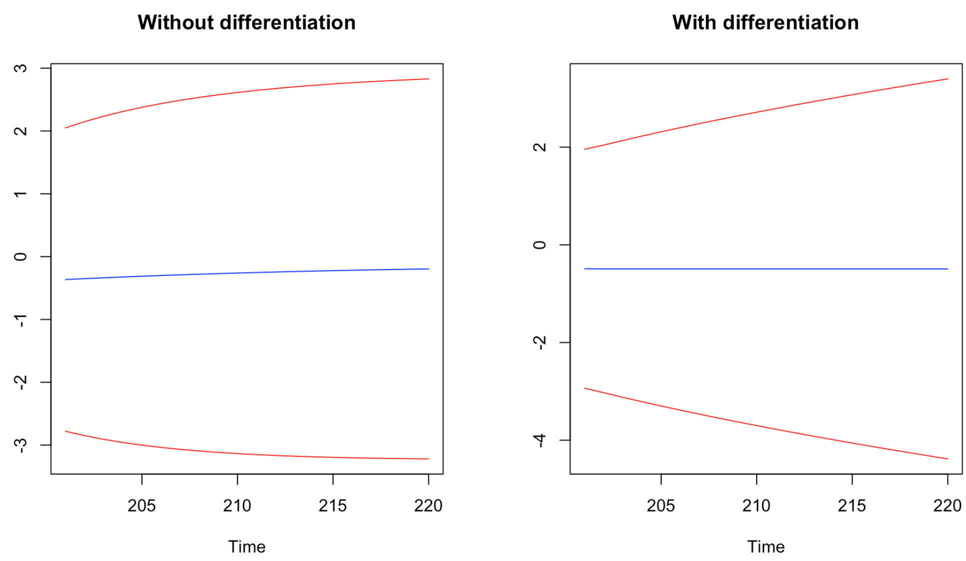

Figure 8: 20-step ahead forecast with and without differentiation