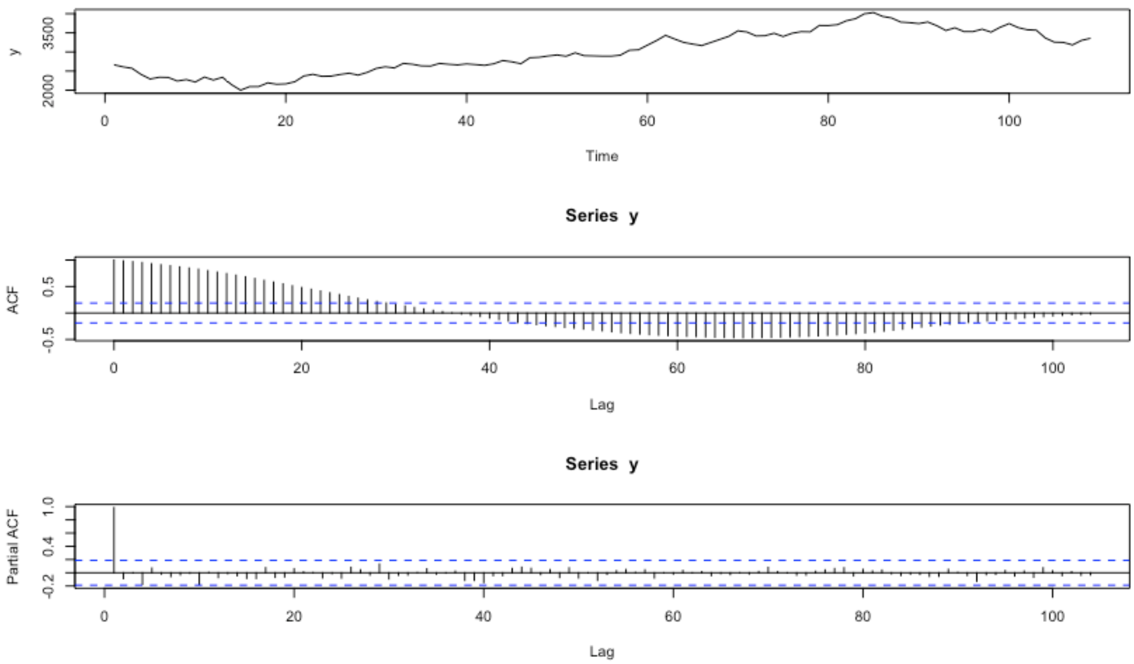
Forecasting Time Series

MBD 2019

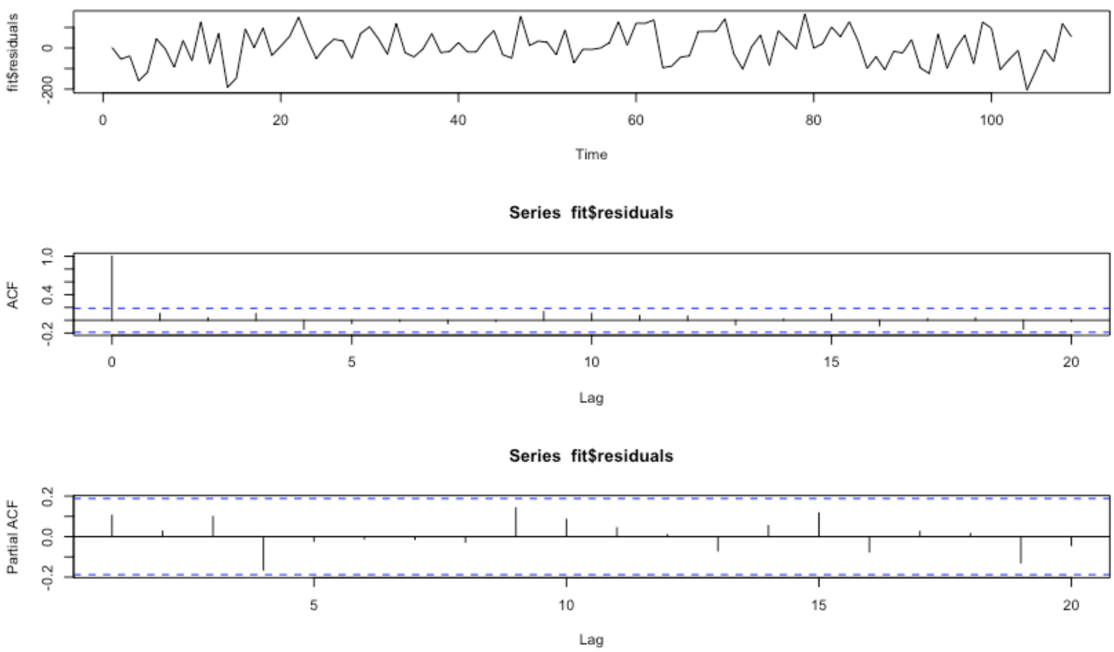
Table of contents

1. Find the best time series model for the variable "ibex"

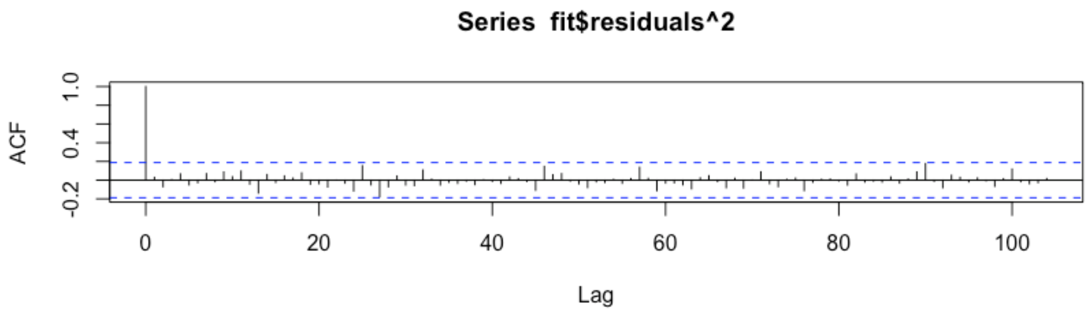
We have weekly seasonal data (s = 52). Let's look at ACF and PACF:

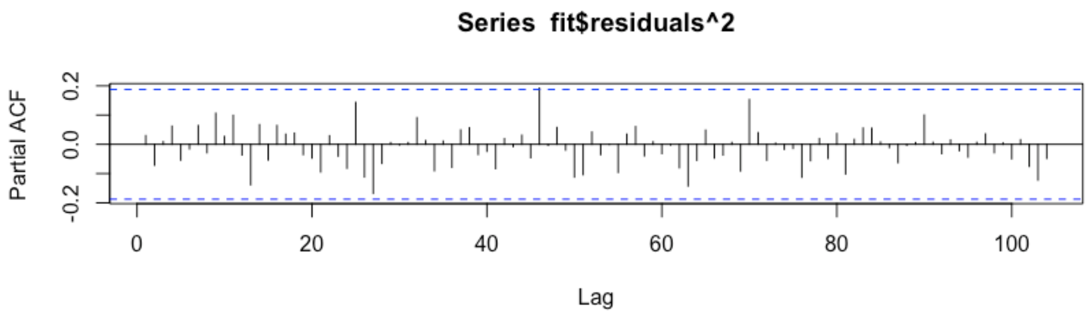


The data it is not stationary, we need 1 regular difference and no seasonal differences. Let's fit in arima(0,1,0) and look at the residuals:

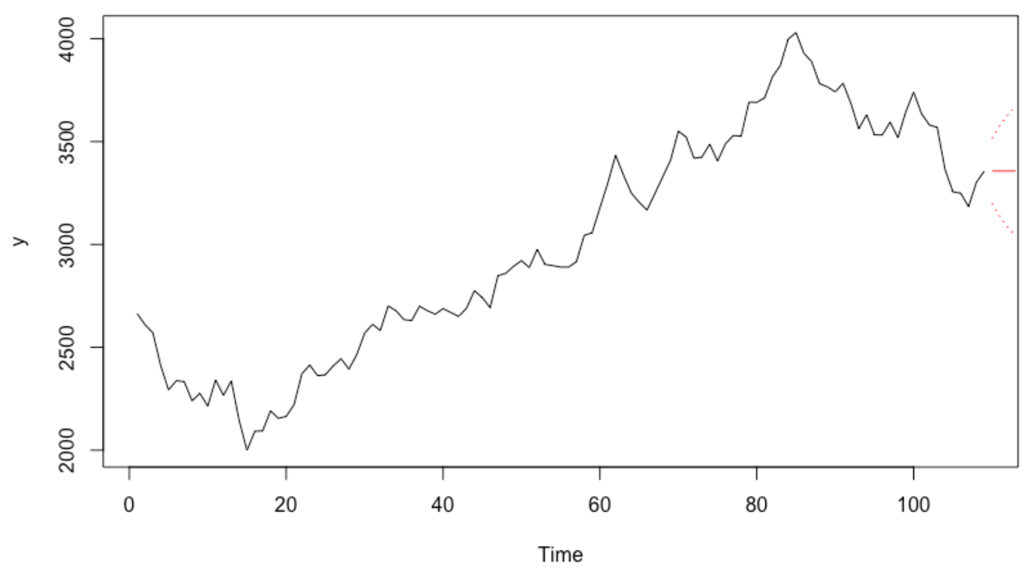


The residuals are White Noise (P-value = 1), they are normally distributed (Shapiro-Wilk normality test P-value = 0.4433) and are SWN (Box-Cox of residuals^2 P-value = 1), which can also be seen from the ACF and PACF of the squared residuals:





Thus, our best time model for this step will be arima(0,1,0), meaning that our best prediction is mean:



2. Find the best regression model for the dependent variable "ibex".

In order to proceed, let's take a look at correlation of the variables in our dataset:

```
                          Week       IBEX Exchange rate \x80/$ Short term rate Long term rate
Week                 1.0000000  0.9075622            0.8868654      -0.9293596     -0.7802327
IBEX                 0.9075622  1.0000000            0.8888655      -0.9318204     -0.9441591
Exchange rate \x80/$ 0.8868654  0.8888655            1.0000000      -0.8427291     -0.8675614
Short term rate     -0.9293596 -0.9318204           -0.8427291       1.0000000      0.8664739
Long term rate      -0.7802327 -0.9441591           -0.8675614       0.8664739      1.0000000
```

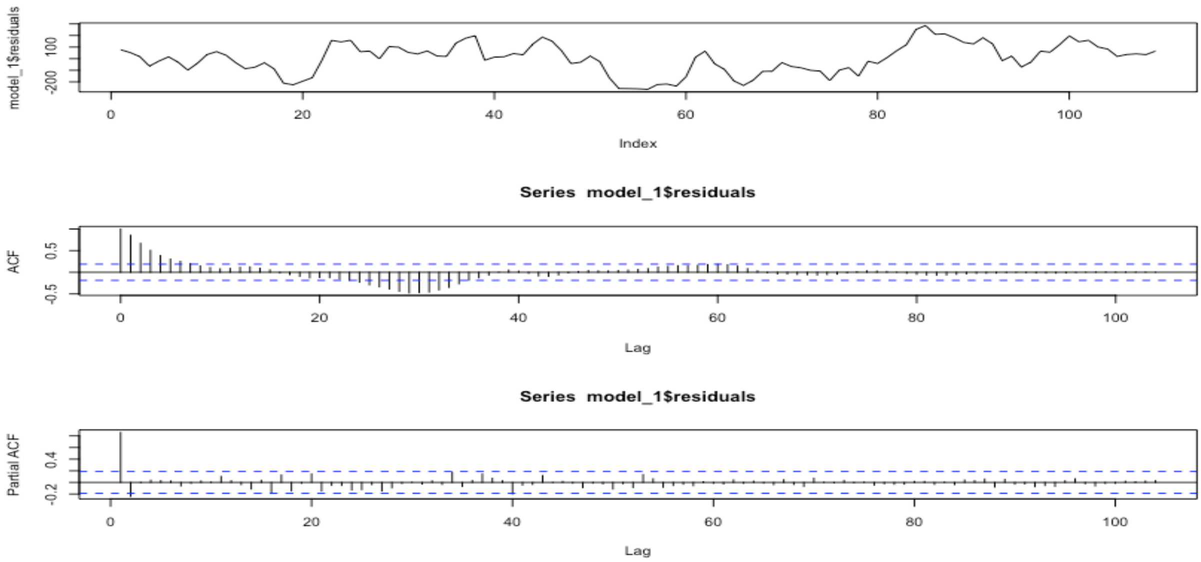
Clearly, everything is highly correlated, thus, we are expecting multicollinearity.
We're constructing a linear regression including Long Term Rate, Short Term Rate and Exchange rate as explanatory variables.

```
Coefficients:
                        Estimate Std. Error t value Pr(>|t|)    
(Intercept)              5231.68     376.91  13.881  < 2e-16 ***
`Exchange rate \\x80/$`   783.34     288.44   2.716  0.00773 ** 
`Short term rate`         -88.70      10.51  -8.444 1.84e-13 ***
`Long term rate`         -172.16      18.92  -9.098 6.45e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 129.3 on 105 degrees of freedom
Multiple R-squared:  0.9471,	Adjusted R-squared:  0.9455 
F-statistic: 626.1 on 3 and 105 DF,  p-value: < 2.2e-16
```

All the variables are significant, let's check VIF to see if we have multicollinearity (VIF above 5 are considered to be high):

```
`Exchange rate \\x80/$`       `Short term rate`        `Long term rate` 
               4.670654                4.635314                5.431246 
```

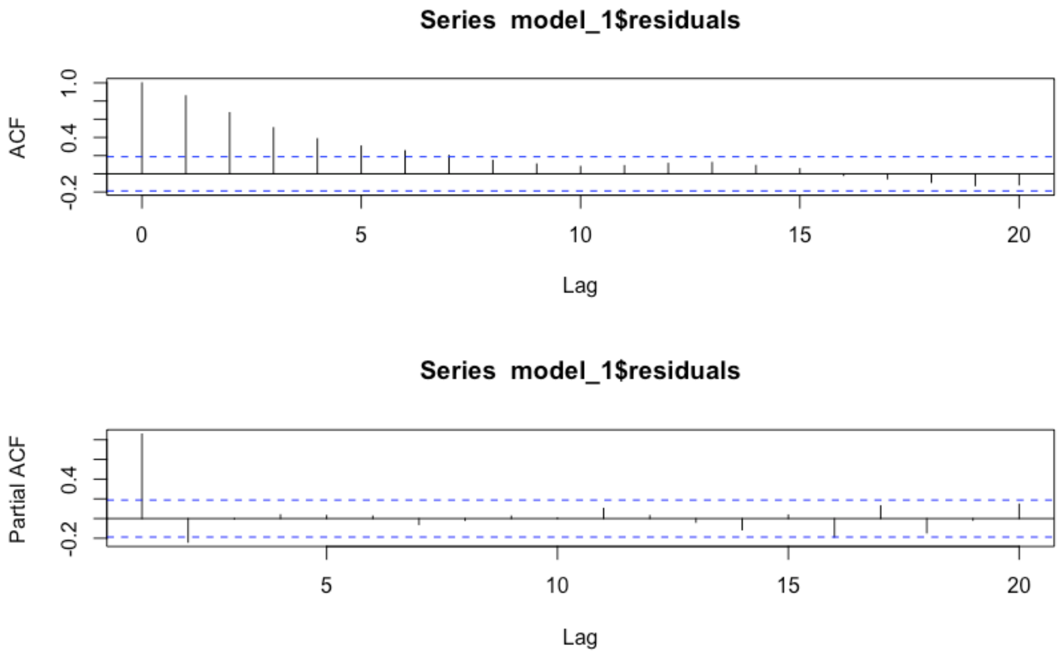
So, we do have multicollinearity. Checking whether the residuals are white noise:



Box-Cox test P-value is < 2.2e-16, thus, our residuals are not white noise, but we cannot deal with that in this step. Other than that, the data is stationary, we don't need regular or seasonal differences.

3. Find the best regression model with time series errors for the dependent variable "ibex"

Let's try to deal with our residuals with the help of time series model.



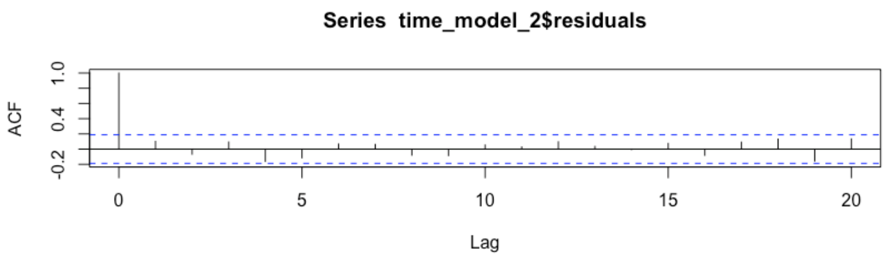
Looks like our fit may be AR(2), let's try it:

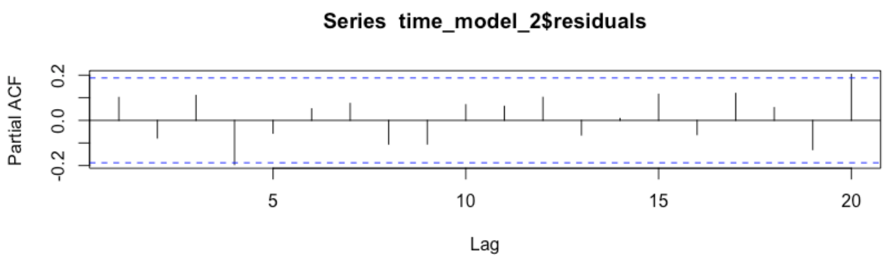
```
Coefficients:
         ar1      ar2  intercept  Exchange rate \x80/$  Short term rate  Long term rate
      1.0795  -0.1239  4493.4243              958.2677         -10.7181       -192.6842
s.e.  0.0989   0.1022   394.7952              355.0949          15.4530         23.7137

sigma^2 estimated as 3191:  log likelihood = -595.67,  aic = 1205.33
```

Our AR(2) parameter is not significant, as well as Short term rate. Let's account for that:

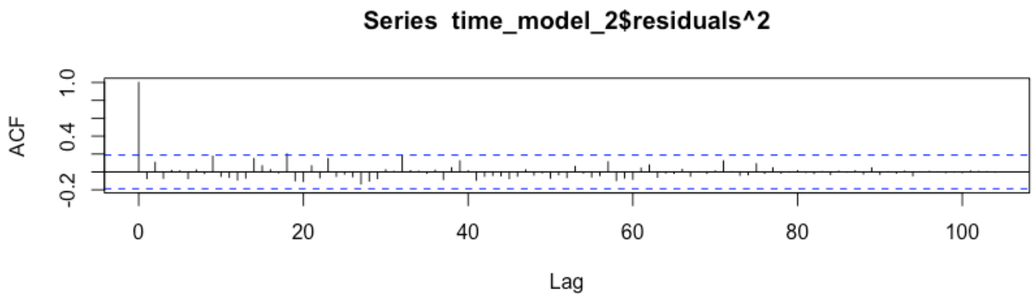
```
Coefficients:
         ar1  intercept  Exchange rate \x80/$  Long term rate
      0.9665   4572.186              881.0670       -204.1894
s.e.  0.0233    393.453              351.1813         21.3818

sigma^2 estimated as 3357:  log likelihood=-596.47
AIC=1202.93   AICc=1203.51   BIC=1216.39
```

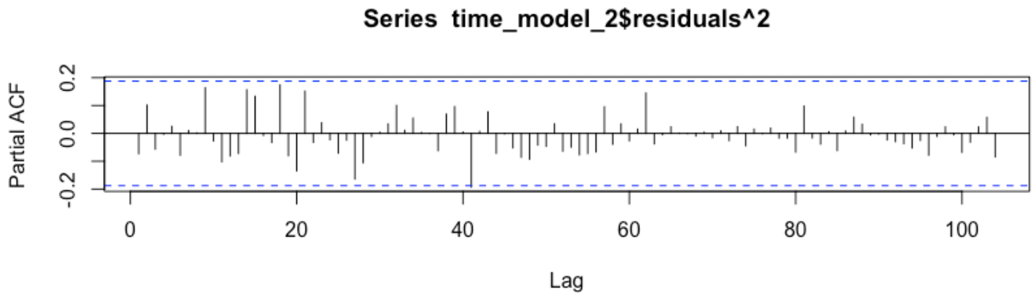
Ensuring that residuals are white noise graphically and with a formal test (P-value = 0.2876):





Everything is within limits, residuals are WN. P-value for Shapiro test is 0.4977, the residuals are normally distributed, they are also SWN (p-value for residuals^2 is 1):





Our final model has zero lags, 2 explanatory variables, on the contrary with the regression model without time series errors which had 4 variables and zero lags.
The final equation is:

$$\widehat{IBEX}_t = \underset{(393,45)}{4572,18} + \underset{(351,18)}{958,27} \times Exchange\_rate_t - \underset{(21,38)}{204,19} \times Long\_term\_rate_t + e_t$$

Where:

$$e_t = a_t + 0.97 \times a_{t-1}$$

with:

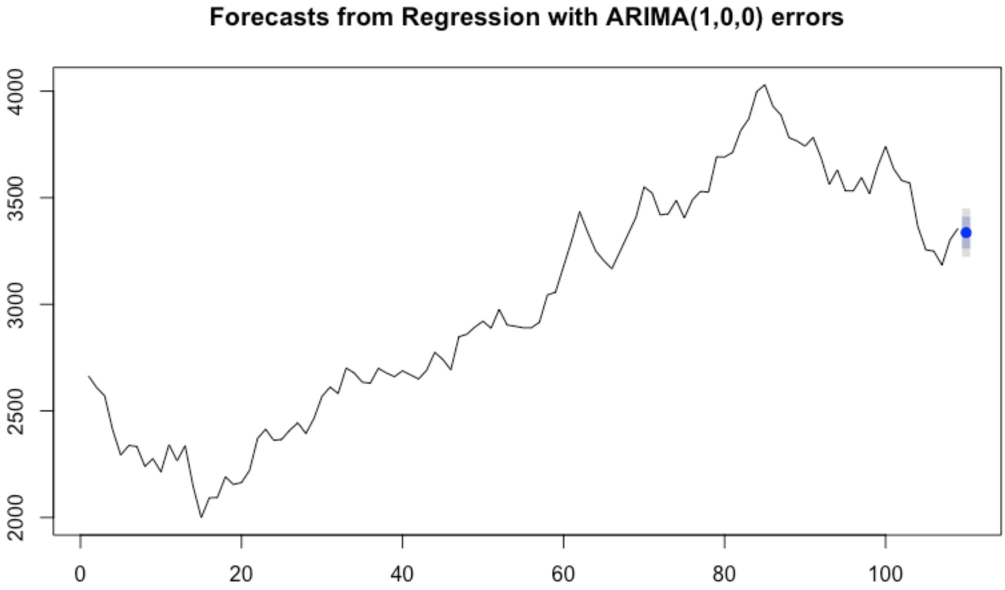
$$\delta_a = 0.0233$$

4. The best model and point prediction

Let's look at the sigma to figure out which of the 3 models is better:

Model 1: ARIMA(0,1,0) – sigma = 80.41
Model 2: linear regression – sigma = 129.32
Model 3: ARIMA(1,0,0) with time series errors - sigma = 56.87

So, the best model is the model 3.
Point prediction for the Short Term Rate = 10.76 and Exchange Rate = 10.76 is (3336.309 +/- 1.96 * 57.94198) , which is (3222.74; 3449.88)



Forecasts from Regression with ARIMA(1,0,0) errors
2000
2500
3000
3500
4000
0
20
40
60
80
100