HW1: Temporal Series

Marc Luque, NIU: 1567604
Sergi Cucala, NIU: 1570044
Carlo Sala, NIU: 1570775
Júlia Albero, NIU:1566550

Financial Engineering

Exercise 1

Given a time series, we say it is a stationary process if after some time their properties such as mean, variance and covariance have not changed. In other words if it is time-invariant, precisely if $\{Y_t\}_t$ is a time series sampled with cumulative distribution function $F(Y_{n_1},\ldots,Y_{n_k})$, it is strongly stationary if

$$F(Y_{n_1},\ldots,Y_{n_k}) = F(Y_{n_1+m},\ldots,Y_{n_k+m})\ \forall m \in \mathbb{N}$$

However, since a stationary process shows oscillations around the same mean-reversion level, for any real data one can guess just by the plot which ones of them are not stationary and giving a proper test to justify which ones are. In particular consider the time series of the data extracted from The 91-day treasury bill (r), The log of real GPD (y) and The inflation Rate (pi). Their respective plots are:

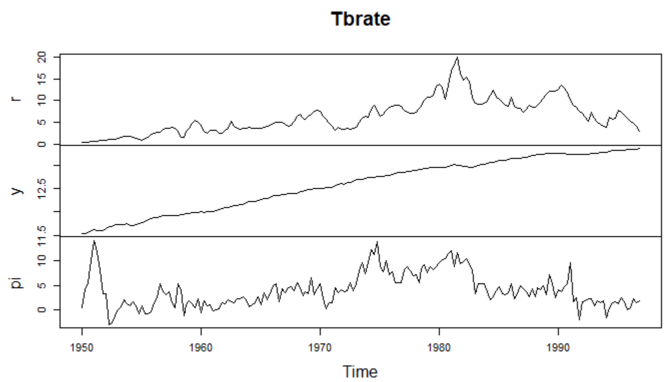


Figure 1: Time Series Plot

It is clear enough that, while the Inflation Rate shows some stationary properties, despite its huge amount of volatility and compared to the other time series, both the Treasury Bill and real log of GPD have a positive tendency, which means it is almost sure r and y are not stationary processes. In detail, the $y-$series increases monotonously over time which clearly exempts it from being stationary. Instead, the $r-$series drops drastically over the year 1981, after increasing over the first 30 years.
Furthermore, having a look at the ACF plots, one can notice that the processes may not be modeled as an $AR(p)$ or $MA(q)$ models. First of all, given that almost every graph pictures non-zero lags it is clear that they are not a *White Noise*, meaning they are not random. In addition, practically all of them show non geometric decays, which is a characteristic property every $AR(p)$ holds. Finally there is no lag q where the ACF drops drastically, which means it cannot be modeled as a $MA(q)$.

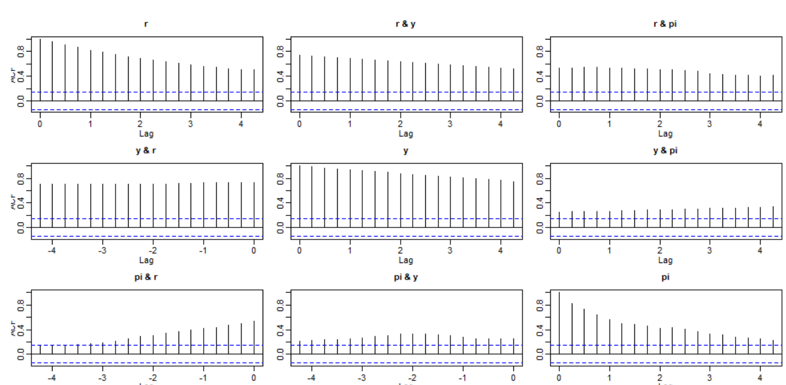


Figure 2: ACF plots

Despite all the signs of not being stationary could happen the temporal series need a more complex model to be patterned. The Dickey-Fuller test, a particular example of Unit Root Test, provide a clear answer. It is known that for $\{Y_t\}_t$, an $ARMA(p,q)$ process given by

$$Y_t - \mu = \sum_{i=1}^{p} \phi_i (Y_{t-i} - \mu) + \varepsilon_t + \sum_{j=1}^{q} \delta_j \varepsilon_{t-q}$$

is stationary if and only if all roots of the polinomial $1-\sum_{i=1}^{p} \phi_i x^i$ have absolute value greater than one. Particularly, the Dickey-Fuller tests has null hypothesis that there is at least one unit root, hence the process is not stationary. Studying the results one can finally conclude that none of the processes are not stationary due their respective $p-$values (greater than 0.05).

```
adf.test(Tbrate[,1])
=> p-value = 0.6075
adf.test(Tbrate[,2])
=> p-value = 0.9873
adf.test(Tbrate[,3])
=> p-value = 0.1788
```

Despite the results, let's consider studying their differentiate temporal series and see if they can now be modeled as an $ARMA(p,q)$ process. This time the Dickey-Fuller test provides the following results

```
adf.test(diff_rate[,1])
=> p-value = 0.01
adf.test(diff_rate[,2])
=> p-value = 0.01
adf.test(diff_rate[,3])
=> p-value = 0.01
```

Which means that none of their respective roots are in the unit disk, hence the differentiated processes appear stationary. Viewing the data and ACF plots, we can assume that clearly now they have more stationary properties, with a general loss of volatility.

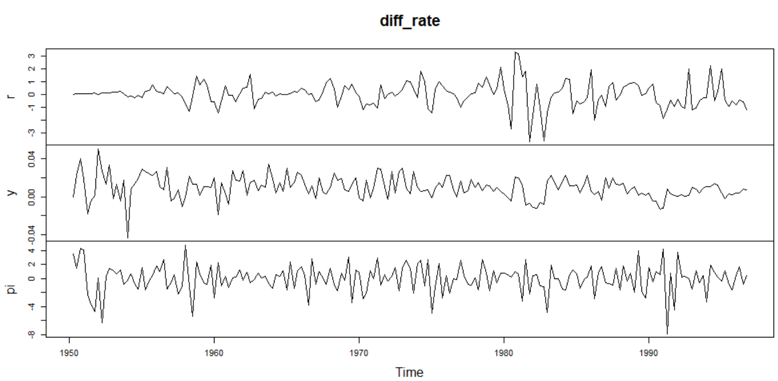


Figure 3: Differentiated temporal series

In addition the correlation between lags display in general autocorrelations of values between 0 and 0.05, which means they are not random but an $ARMA(p,q)$ model is most suited to them. In particular, the ACF plots that present the series alone (not a mix of them) share that after one lag the acf is significantly 0 and their tails seem to be narrow.

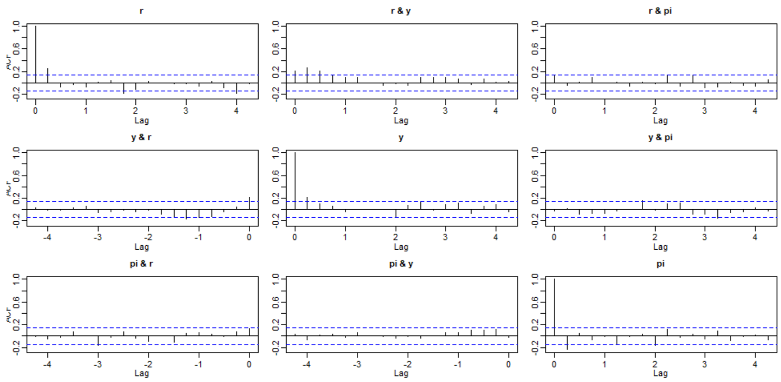


Figure 4: ACF of differentiated temporal series

In conclusion, they seem to behave like a $ARIMA(0,1,1)$ according to ACF plots.

Let's see the best-fitting model obtained by *auto.arima()*, the built-in fucntion of R that provides this answer, with AIC (Akaike Information Criterion) for the $r-$data:

```
Series: Tbrate[,1]
ARIMA(0,1,1)
...
AIC=495.3    AICc=495.37    BIC=501.76
```

In particular, the order of differentiating chosen is 1 and that is the same order we got from the previous analysis for that data. We also observe that this test minimized AIC as its goodness-of-fit criterion. Lastly, note that changing the criterion to BIC (Bayesian Information Criterion) does not change the best-fitting model.
ACF plot of residuals shows no residual autocorrelation:

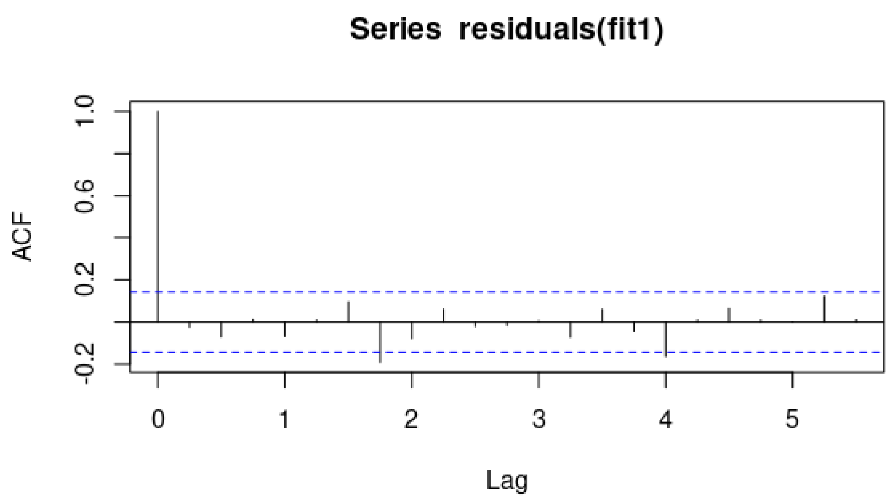


Figure 5: ACF of residuals

Moreover, Box text also tells us to accept the null hypothesis that residuals are independent from each other:

```
Box-Ljung test
data:  residuals(fit1)
X-squared = 13.017, df = 10, p-value = 0.2227
```

Which means there is no information left in residuals of the original temporal series, hence the model proposed is quite good to forecast futures values and give predictions.

Exercise 2

Considering $\{\varepsilon_t\}_t$ a white-noise $WN(0, \sigma_\varepsilon^2)$ and the constants μ an $\{\phi_i\}_{i=0}^p$, the mean and memory respectively. We say $\{Y_t\}_t$ is and $AR(1)$ model if

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + \varepsilon t$$

Rewriting the equation, we obtain the equivalent $Y_t = (1-\phi)\mu + \phi Y_{t-1} + \varepsilon t$, whom we can easily extract the value of ϕ. Furthermore, we say an $AR(1)$ process is stationary if and only if $|\phi| < 1$. In our case the AR model is

$$Y_t = 5 - 0.55 Y_{t-1} + \varepsilon_t$$

hence $|\phi| = |-0.55| < 1$, meaning our model is stationary. The sample mean, variance and covariance are easily derived from the model which statistics are

- $\mathbb{E}(Y_i) = \mu = \dfrac{5}{1-\phi} = 3.225806$
- $\gamma(0) = Var(Y_i) = \dfrac{\sigma_\varepsilon^2}{1-\phi^2} = \dfrac{(1.2)^2}{1-\phi^2} = 1.72043$
- $\gamma(i-j) = Cov(Y_i, Y_j) = \dfrac{\sigma_\varepsilon^2}{1-\phi^2}\phi^{|i-j|} = 1.72043 \cdot (-0.55)^{|i-j|}$

Where $\sigma_\varepsilon^2 = 1.2$.

Exercise 3

Let us now consider the general $AR(p)$ model, given $\{\varepsilon_n\}_n$ a $WN(0, \sigma_\varepsilon^2)$, and the constants μ and $\{\phi_i^p\}_{i=0}$, then $\{Y_t\}_t$ is an $AR(p)$ model if

$$Y_t - \mu = \sum_{i=1}^{p} \phi_i (Y_{t-i} - \mu) + \varepsilon_t$$

In particular, for $p = 3$ we have $Y_{t+1} - \mu = \phi_1(Y_t - \mu) + \phi_2(Y_{t-1} - \mu) + \phi_3(Y_{t-2} - \mu) + \varepsilon_{t+1}$, we would like to forecast the values of Y_{t+1} and Y_{t+2}. Assuming known $Y_1, \ldots, Y_3$ the parameters ϕ_i, and using the estimators for mean $\hat{\mu}$. Despite the unknown value of ε_{t+1}, it is independent of the past and present Y_t, hence we can use its mean 0 as the best prediction of itself. Finally our forecasting model is

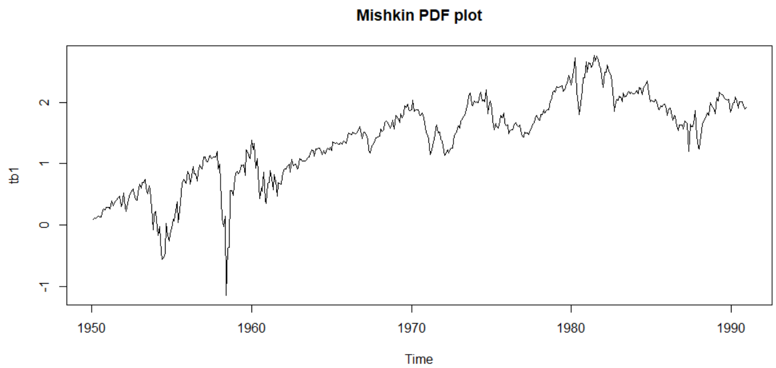
$$\widehat{Y_{t+1}} = \hat{\mu} + \phi_1(Y_t - \hat{\mu}) + \phi_2(Y_{t-1} - \hat{\mu}) + \phi_3(Y_{t-3} - \hat{\mu})$$

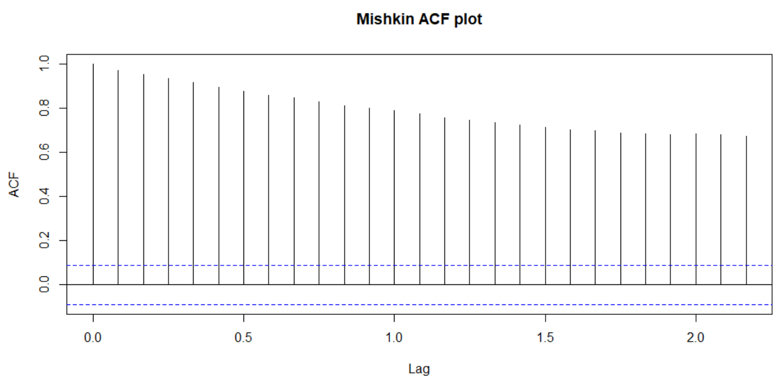
Given $Y_t = 99,\ Y_{t-1} = 103,\ Y_{t-2} = 102,\ \mu = 104,\ \phi_1 = 0.4\ \phi_2 = 0.25,\ \phi_3 = 0.1$, we obtain

$$Y_{t+1} = 101.55,\ Y_{t+2} = 101.67$$

Exercise 4

Given the monthly observations over the years $1950 - 1990$ of the american Price Index, let's proceed as done previously to search the best fit model for the time series. Firstly let's consider the time series and ACF plot to determine the amount of differencing needed to obtain a stationary process.

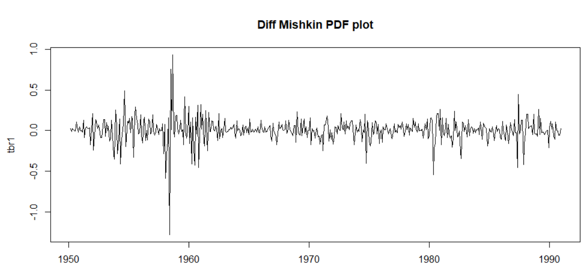


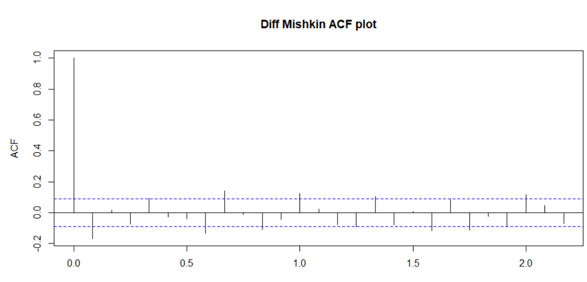


The data plot displays extreme volatility and increasing behaviours. In addition, the lags of the autocovariance function surpass the 0.05 level, which are clear signs of being a non stationary process. Precisely, the Dickey-Fuller test provides the following $p-$value

```
adf.test(tb1)
=>p-value = 0.08319
```

which implies there is at least one unit root, hence the process is not stationary as expected. Let's proceed and study a first derivation of the series.

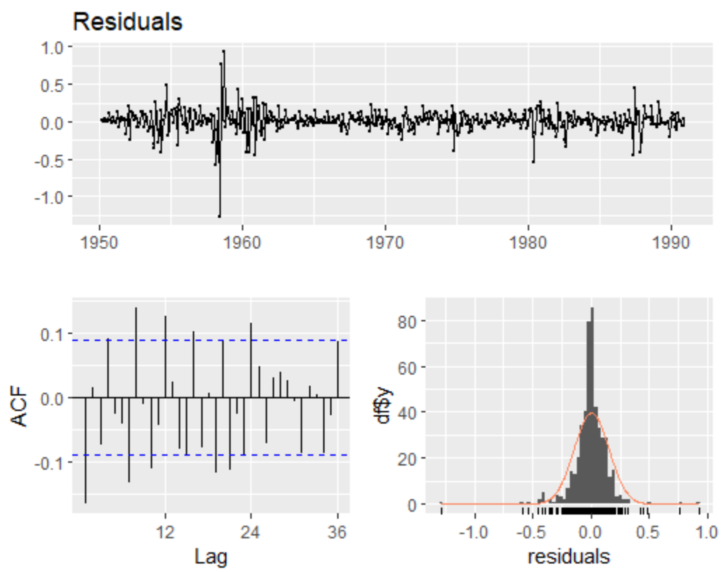




On one hand, from the temporal series plot, the series seems to smooth its volatility after one differentiation, despite some punctual peaks it may appear right before the 60 decade. On the other hand, even if the tails do not decrease, the ACF plot displays minor levels of correlation between one lag from the previous, that remain mostly on the $(-0.05, 0.05)$ interval. Therefore, we must say the temporal series can be modeled after one differentiation as a known process. To be precise, using the R built-in function *auto.arima*() to determine the model that fits best to our data, with both the BIC and AIC criterion respectively, we get that the temporal series best fitted with

```
Series: tb1
ARIMA(0,1,1)
Series: tb1
ARIMA(3,1,5)
```

Which it fulfils the expected level of differentiating. The last step to know if the model best explains the series behaviour, is to study the residuals one. Given $\{Y_t\}_t$ the time series and $\{\widehat{Y}_t\}_t$ their best fit model, then the residuals $e_t = Y_t - \widehat{Y}_t$ are useful in checking whether a model has adequately captured the information in the data. A good forecasting method will yield residuals behaving like a $WN(0, \sigma_\varepsilon^2)$, but mainly are uncorrelated, have zero mean. Our residuals information relies on the following graphics.



Where we observe that the mean of the residuals is mostly zero, despite showing some peaks right before the 1960. However their correlation sings given by the ACF plot, displays some non zero values, which could mean the lags are not uncorrelated. Additionally, they might not behave normally distributed, due the amount of outliers it represents. Therefore the residuals are not completely a white noise, meaning they might not work as a predictive intervals tool, but to forecast some future Y_{t+k} values it might work.