Macroeconometrics
Problem Set 1

Nunzio Fallico 3072744
Chiara Locatelli 3070485
Matteo Ticli 3077833

23 October 2022
Università Bocconi

Exercise 1

(a)(b)

There are two methods for generating an $AR(1)$ process y_t with $\mathbb{E}(y_t) = 0$: one is using a *for loop* where each time we append to the y_t vector a new observation of the random variable and the other using the function *filter*.

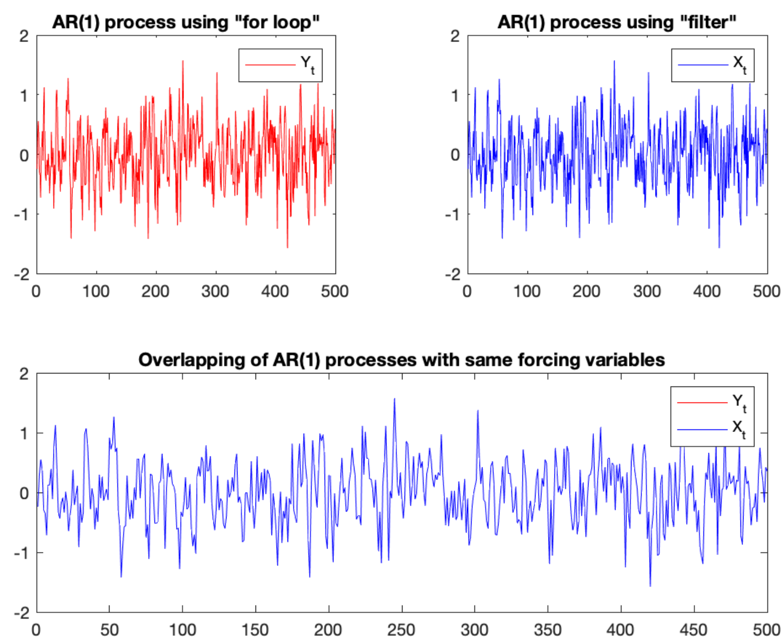

Figure 1: The *AR(1)* processes generated by *for loop* and *filter* deliver the same process

(c)

When the forcing variables are the same, then the output of the two approaches is the same. We can show this by simply plotting the two processes in the same graph (lower plot of Figure 1) and see that they indeed coincide.

Exercise 2

We want to generate the following *AR(1)* process:

$$y_t = \delta + \phi y_{t-1} + \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

We can rewrite it recursively as :

$$y_t = \phi^t y_0 + \delta \underbrace{(1 + \phi + \phi^2 + \phi^3 + ...)}_{\text{geometric series } = 1/(1-\phi)} + \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i} = \phi^t y_0 + \delta \frac{1}{1-\phi} + \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}$$

Since $\phi^t y_0 \to 0$ as $t \to \infty$ and $\mathbb{E}(\sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}) = 0$ by assumption, thus delivering:

$$\mathbb{E}(y_t) = \delta \frac{1}{1-\phi} \to \delta = [\mathbb{E}(y_t)]\,(1-\phi)$$

Given that $\mathbb{E}(y_t) = 3$ and $\phi = 0.4$, we can identify the drift $\delta = 1.8$.

We proceed by generating the y_t process as in Exercise 1, first by using a *for loop* and then the *filter* function. Using *filter* including in the code for the starting condition $y_1 = \phi y_0 + \delta + \varepsilon_1$ where $y_0 = 10$ delivers the same process as that of the *for loop*.

As we can see from the graph below, if the starting condition is far from the unconditional mean of the process, then this will take some time before stabilizing around it.

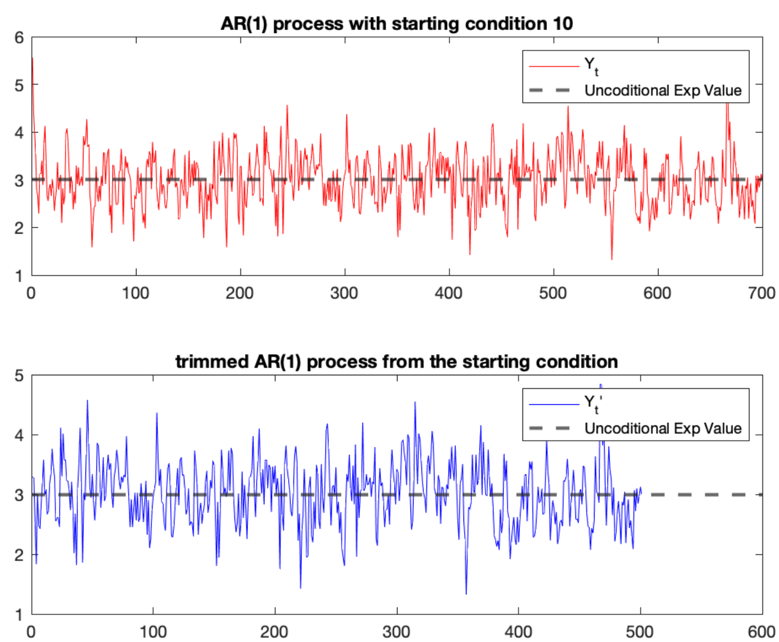

Figure 2: The sample path (lower plot) is a "proper" realization, high starting condition

In order to make sure that the sample path is a "proper" realization of the process, it is enough to increase the number of observations $N = 500$ by $T = 200$. The new time span is $T + N$, we simply disregard a considerable part at the beginning, say T, and plot the process from the $T^{th} + 1$ observation, as in the lower plot of Figure 2. The idea behind this is that because of a high starting value, during the first observations the process will be converging to the unconditional mean without assuming standard values of the underlining stationary process. However, at some point, it will stabilize around the unconditional mean, thus being proper for the process we want to simulate from.

Exercise 3

Similarly to Exercise 1, we can generate the process y_t, this time an $MA(1)$, by using either a *for loop* or the function *filter*. It is important that we use in both cases the same forcing variables, σ and θ, otherwise we would not get the same result. By plotting the processes generated with the two different methods on the same graph we see that they coincide, thus showing that when the forcing variables are the same, the output of the two approaches is the same also in the case of an $MA(1)$.

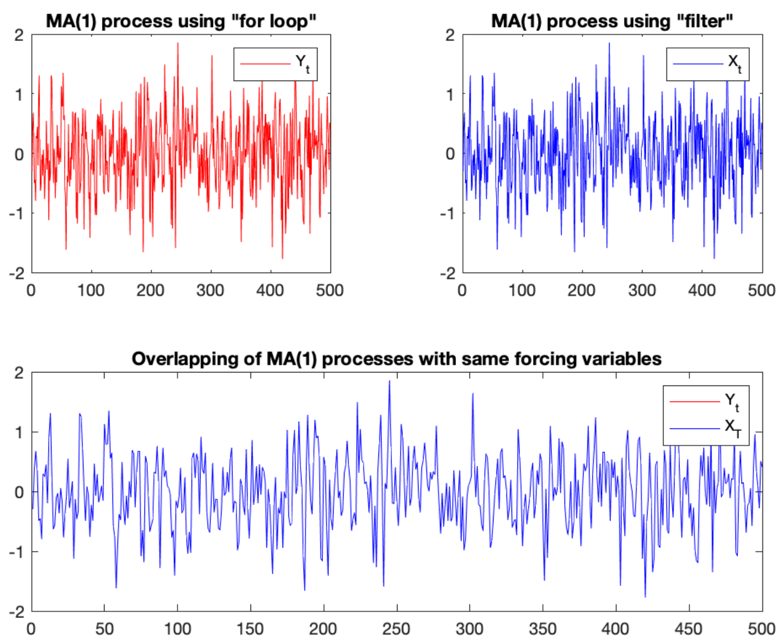

Figure 3: The *MA(1)* processes generated by *for loop* and *filter* deliver the same process

Exercise 4

The function takes as inputs T (number of observations) and the variance of the white noise forcing variables. Moreover, it takes as input either the coefficients or the roots of the AR and MA processes, depending on the value of the additional parameter *roots* that must be given as input too.
If *roots* $! = 1$, then the value of the inputs *ma* and *ar* must be the coefficients of the respective processes in the form [1, a, b, c], so that the unit coefficient must always be provided for both polynomial and a,b,c are the remaining coefficients of interest. In fact, if we would represent the ARMA process $y_t = 0.1y_{t-1} - 0.2y_{t-2} + 0.3\varepsilon_{t-1} + \varepsilon_t$, we must provide *ar*=[1 -0.1 0.2] and *ma*=[1 0.3].

If we instead set *roots* $= 1$, then we must provide the roots representing the polynomial of interest. For example, if we have an AR polynomial in the form $1 - 0.5L - 0.4L^2$, then the corresponding roots that must be passed to the functions are $ar =$ [-2.3252, 1.0752], that will give back the original coefficients of the AR polynomial ([1, -0.5, -0.4]) once that the inverse of ar is given as an object of the function *poly* in Matlab.

Exercise 5

(a)

We perform an OLS regression including for the constant on an $AR(1)$, which hereafter we will generate using the *filter* function. Saving the OLS estimators and plotting them on a graph, we get their empirical distributions.

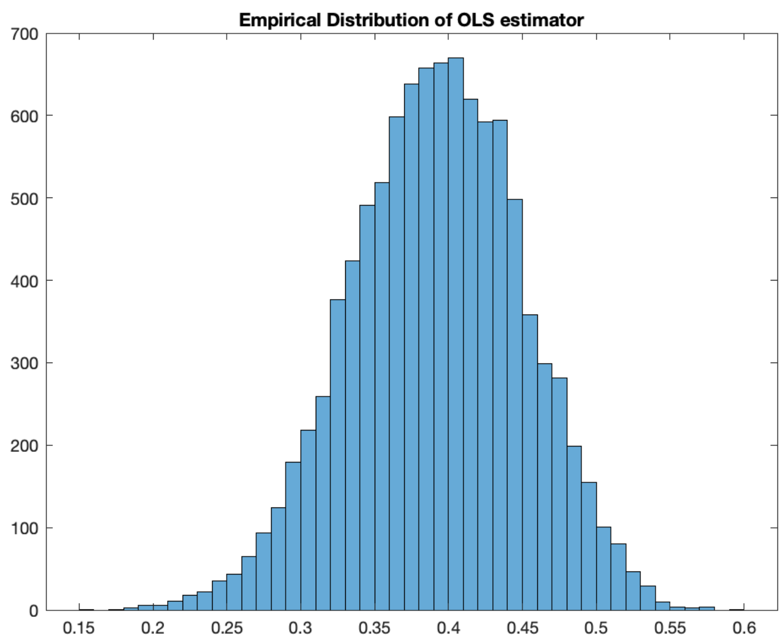

Figure 4: Empirical distribution of OLS estimators

(b)

We now test the null hypothesis $H_0: \ \phi = 0$, against the two-sided alternative $H_1: \ \phi \neq 0$. However, recall that when we generated our process y_t we used $\phi = 0.4$. We thus expect the null hypothesis to be rejected most of the times, and indeed this is the case as it is 99.98%.

Exercise 6

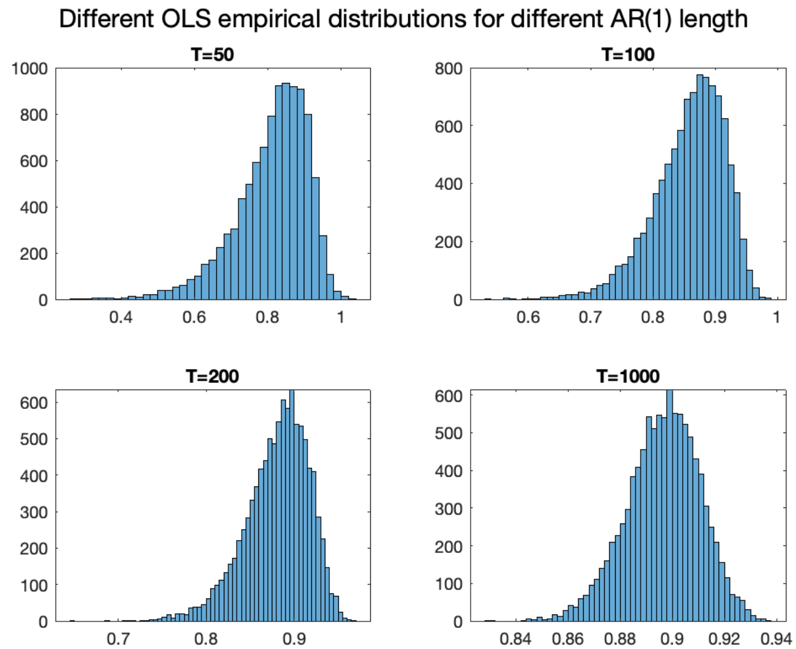

Figure 5: Empirical distributions of OLS estimators of an *AR(1)* for different sample sizes

In the graph above are plotted the empirical distributions of the OLS estimators in the case of an AR(1) with $\phi = 0.9$ for different sample sizes, $T = \{50; 100; 200; 1000\}$. It can be noticed that in all the graphs, the distribution is always centered at around 0.9 (just 1 parameter). Moreover, we see that for smaller sample sizes the empirical distribution of the OLS estimators is right-skewed, while as $T \to \infty$ it becomes less and less skewed (and the variance of the OLS estimator decreases), tending to the distribution of a normal with mean different from zero (more precisely, even more centered at 0.9). The result is in line with what we expect following Central Limit Theorem reasoning in a certain way.

Exercise 7

Recall that when we perform an OLS regression on

$$x_t = ax_{t-1} + \nu_t$$

the convergence value of the OLS estimator, given in our case the data generating process for x_t is $MA(1)$ with $\theta = 0.6$, is

$$a_{\text{OLS}} = \frac{Cov(x_t, x_{-1})}{Var(x_t)} = \frac{\theta}{1+\theta^2} = \frac{0.6}{1+0.36} = 0.4412$$

If we plot the empirical distribution of the OLS regressors we can see that it tends to that of a normal with mean different from zero. More specifically, the mean of the estimators for $T = 250$ is 0.4396, while that for $T = 10000$ is 0.4412. Hence, we conclude that as the sample size increases, $T \to \infty$, the mean of the distribution converges to the true value of the OLS regressor $a_{\text{OLS}} = 0.4412$.

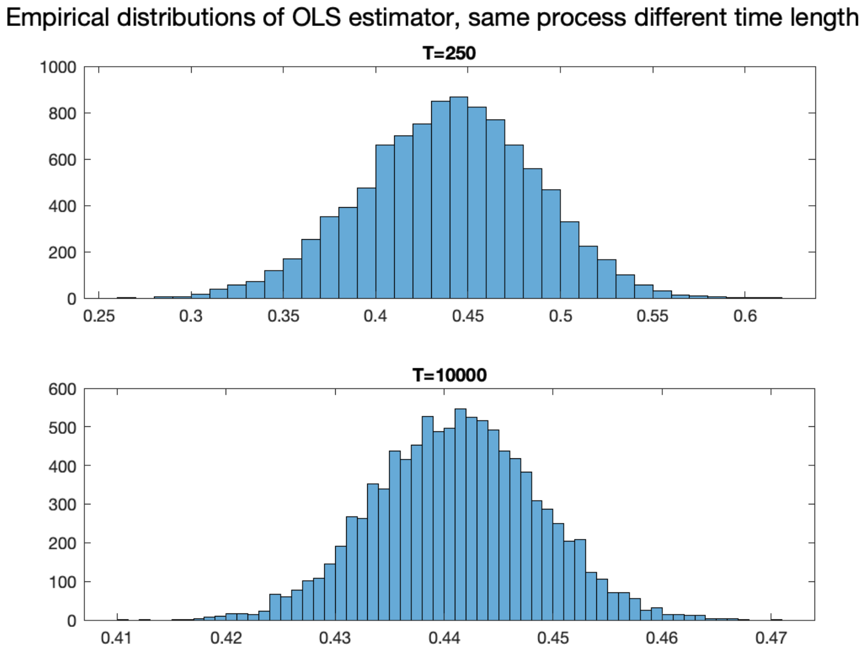

Figure 6: The mean of the distribution converges to the true value of the OLS estimator

Exercise 8

(a)

We generate 10000 observations of an $AR(1)$ process with $\phi = 1$ and $T = 250$ (important value for the later comparison of the critical levels of the empirical distribution found and the one tabulated by Dickey Fuller). We perform an initial OLS regression on $AR(1)$ and we store the values of the regression coefficients that can be seen in the upper left plot of Figure 7.

(b)

Subsequently, we perform the following regression:

$$\Delta y_t = \alpha + \rho y_{t-1} + \varepsilon_{t-1} \tag{1}$$

and we construct the following one sided test of hypothesis:

$$H_0 : \rho = \phi - 1 = 0 \ \text{ vs } \ H_1 : \rho < 0$$

Using the Standard Normal distribution we constructed the *t-statistics* of a t-test, for each iteration of the simulation, using the canonical form:

$$t_{\hat{\rho}} = \frac{\hat{\rho} - \rho_0}{\sqrt{Var(\hat{\rho})}}$$

Using the critical value of the Standard Normal distribution we reject the null hypothesis (H_0), 45.05% of the times which indeed is very high compared to the result we would expect using a confidence interval of 95% that entails a rejection of only 5% of the cases. This peculiar aspect is related to the

fact that we should use the critical values tabulated by Dickey Fuller to compare a test statistic of a regression of the form expressed in equation (1) and not the ones of a Standard Normal distribution. Not using the Dickey Fuller distribution, would lead to a rejection of the *t-statistics* much more often, thus rejecting the presence of a unit root more frequently than what we should actually do.
As it can be seen in the lower plot of Figure 7 the distribution of the t-test appears to be almost symmetric nonetheless it is not centered at 0 (mean value of the Standard Normal), convergence value of the *t-statistics* of a canonical t-test that leverages a Standard Normal Distribution.

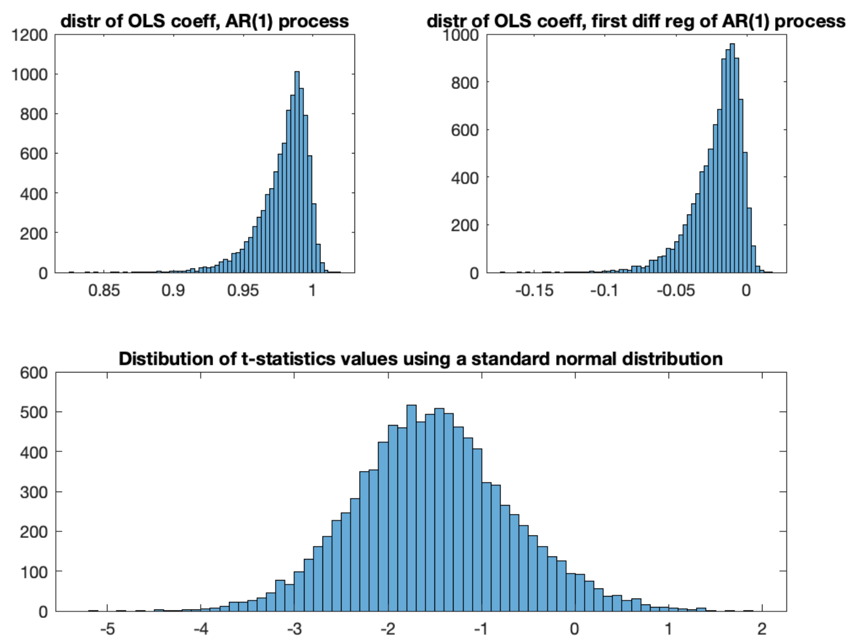

Figure 7: Distributions of OLS coefficient and distribution of t-statistics

(c)

We now compare the values of the percentiles of the empirical distribution found for the t-test and the one tabulated by Dickey Fuller.

Distributions Percentiles				
	1%	2.5%	5%	10%
Emp. values	-3.4427	-3.1282	-2.8481	-2.5626
Dickey Fuller	-3.45	-3.14	-2.88	-2.58

Table 1: Comparison of the empirical percentiles with the percentiles tabulated Dickey Fuller

As we can see from the table, the percentiles values of the empirical distribution are very close to the ones presented by Dickey Fuller. Assuming that we perform a simulation with more iterations, it would be likely to obtain values much more closer to the ones presented in the DF distribution and reported in Table 2.

Exercise 9

(a)

We construct a Monte Carlo experiment with a random walk with drift. We recall that the Dickey Fuller test assumes the following characterization:

$$x_t = \alpha + \beta t + \phi x_{t-1} + \varepsilon_t$$

Under the joint null hypothesis of a unit root we have $H_0 : \phi = 1, \quad \beta = 0$. However, we can rewrite it for clearness and we perform the following regression:

$$\Delta x_t = \alpha + \beta t + \rho x_{t-1} + \varepsilon_t$$

whose null hypothesis of existence of unit root thus becomes $H_0 : \rho = \phi - 1 = 0, \quad \beta = 0$. We proceed with the experiment by simulating 10000 times the random walk plus drift stochastic process of length $T = 250$. Moreover, we evaluate the performance of the DF test using the *F-statistics* that allows to test for joint restrictions. In order to do so, we compare the percentiles of the empirical distribution of the *F-statistics* and the ones tabulated by Dickey Fuller.

Distributions Percentiles				
	90%	95%	97.5%	99%
Emp. values	5.2765	6.2820	7.1970	8.3724
Dickey Fuller	5.39	6.3400	7.2500	8.4300

Table 2: Comparison of the empirical percentiles with the percentiles tabulated Dickey Fuller

As it can be seen, the percentiles of the Dickey Fuller test are very similar to the ones of the produced empirical distribution which indeed highlights the correspondence between our procedure and the one performed by Dickey and Fuller. Moreover, we have assessed the probability of rejecting the null hypothesis H_0 using the 95% significance level of the DF distribution in order to assess the performance of the test and to check whether or not our expectations are met. We obtain that 4.75% of the times we reject the null hypothesis that there is the presence of a unit root. The results obtained are in line with our expectations, as we should reject in about 5% of the cases.

(b)

We use now the 95% critical level of an $F(2, 246)$ distribution since we use the ratio of two F distributions, to calculate the $F - statistics$ using the canonical form:

$$F = \frac{\frac{RSS_1 - RSS_2}{(p_2 - p_1)}}{\frac{RSS_2}{(T-1-p_2)}}$$

The degrees of freedom at the numerator are 2 because the parameters of the unrestricted model are $p_1 = 1$ (α) while the parameter of the restricted model are $p_2 = 3$ (α, t, x_{t-1}) thus simply $p_2 - p_1 = 2$. The degrees of freedom of the denominator are 246 since the length of Δx_t is 249 while the parameters of the unrestricted model are p_2 thus $249 - 3 = 246$. RSS_1 and RSS_2 are respectively the residual sum of squares of the restricted and unrestricted regressions performed.
If we reject the null hypothesis using the p-value of the standard F distribution, we simply reject the *F-statistics*, whose distribution can be checked in Figure 8, that have a p-value less than 0.05. In conclusion, if we use the wrong distribution to check whether or not we shall reject the *F-statistics* we incur in over rejecting the instances, in fact we reject the null H_0 44.16% of times.

Moreover, we know that the chi-squared distribution is the limiting distribution[1] of the F distribution as the denominator degrees of freedom is reasonably larger, after a normalization[2]. More precisely, being x the degrees of freedom of the numerator and y the ones of the denominator, we can approximate F(x,y) when y is large enough (when y tends to infinity) with a chi-squared distribution χ^2(x). So in our case, we can say that our distribution can be asymptotically approximated as a $\chi^2(2)$. Consequently, if we use the p-value generated from the χ^2 distribution, we simply get a rejection of the statistic in 45.09 % of the cases, in line with the proportion of the sample rejected using an F-distribution and respective p-value less than 0.05. Once more, we emphasize the fact that we should deploy the test of hypothesis using the correct distribution tabulated by Dickey and Fuller.

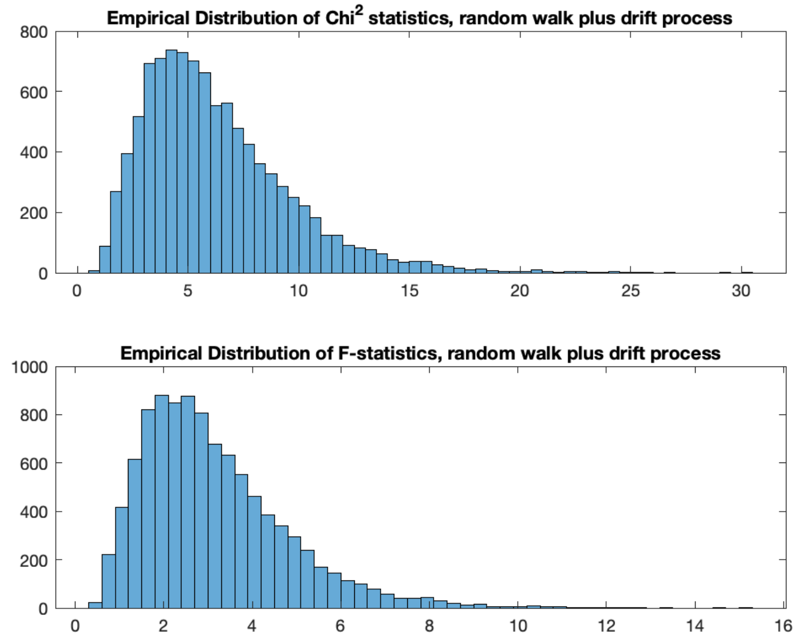

Figure 8: Empirical distribution of χ^2-statistics and F-statistics in a random walk plus drift process

(c)

Generating data from a deterministic time trend that assumes the following form:

$$x_t = \alpha + \beta t + \varepsilon_t$$

which can be seen to be non stationary because of the time trend component (in fact, this process doesn't comply with the definition of stationarity).
Considering as H_0 the existence of a unit root as requested, we perform a simulation on 10000 observations and we see that, using the correct value of 95% significance level tabulated by Dickey Fuller (6.34 in this case), we reject the null hypothesis with probability = 1. The aforementioned result is justified by the fact that the *F-statistics* are extremely high in all cases (check Figure 9, in the majority of cases *F-statistics* > 100), distant from the rejection level of 6.34, tabulated by Dickey Fuller, in this peculiar case.

[1] For further references visit this page.
[2] chi-squared = (numerator degrees of freedom) * F.

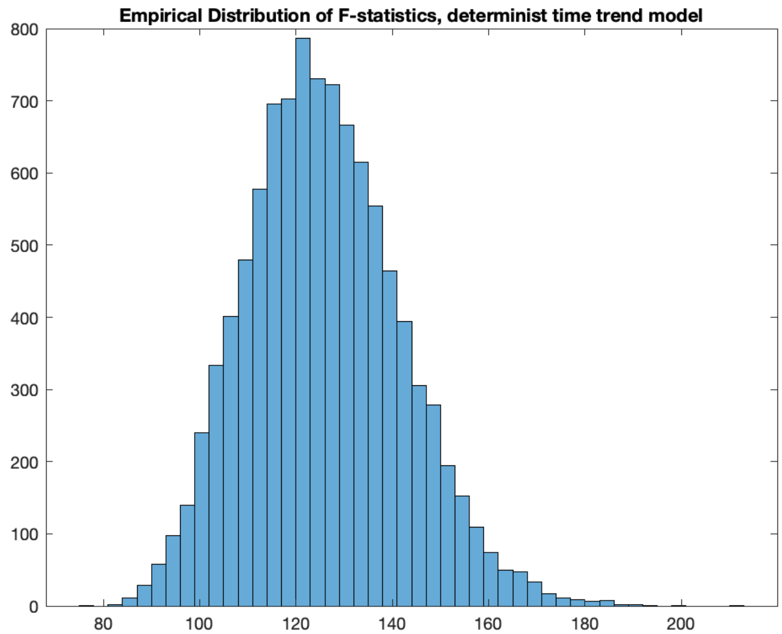

Figure 9: Empirical distribution of F-statistics in a deterministic time trend process