Econ 722: Time Series Assignment 4

Name(s):Benjamin O. Harrison

1. In this exercise, we will conduct a Monte Carlo experiment that studies the phenomenon of spurious regression. The steps required are:

   (a) Run the algorithm (i)-(iii). Use the t-statistics from (iii) to test the null hypothesis that $\beta = 0$. What is the R2 for this regression?

   Answer
   From the algorithm, the DGP for Y_t and X_t is given as :

   $$\boxed{\begin{aligned} Y_t &= Y_{t-1} + e_t \\ X_t &= X_{t-1} + a_t \end{aligned}}$$

   where $e_t \sim \mathcal{N}(0,1)$ and $a_t \sim \mathcal{N}(0,1)$ for T=100. The regression$Y_t = \beta X_t + u_t$ provides the following results.
   The Adjusted R^2 of this regression is given as $R^2 = 0.0246$. The hypothesis test for

Variable	**Coefficient**	**Std. Error**	**t**	**Pr(>|t|)**	**Lower 95%**	**Upper 95%**
Intercept	3.24301	0.602144	5.39	<1e-06	2.04808	4.43795
x	-0.121019	0.0646305	-1.87	0.0641	-0.249276	0.00723867

Table 1: Regression results for $y \sim 1 + x$

   the significant of the coefficient shows that the coefficient is significant since the p-value is very small. This means that the corresponding variable X has a statistically significant effect on the dependent variable Y even tho Y and X are random walk processes.

   (b) Repeat a) 1000 times, saving each value of R2 and the t-statistics. Construct a histogram of the t-statistics and the R2. In which fraction of your 1000 simulated data sets the t-statistics exceed 1.96 in absolute value? Is this what you were expecting? Comment.

   Answer
   By repeating or constructing the OLS regression 1000 times, we extract the R^2 and the $T-statistics$ for each simulation and obtain the histogram below in Figure (1)

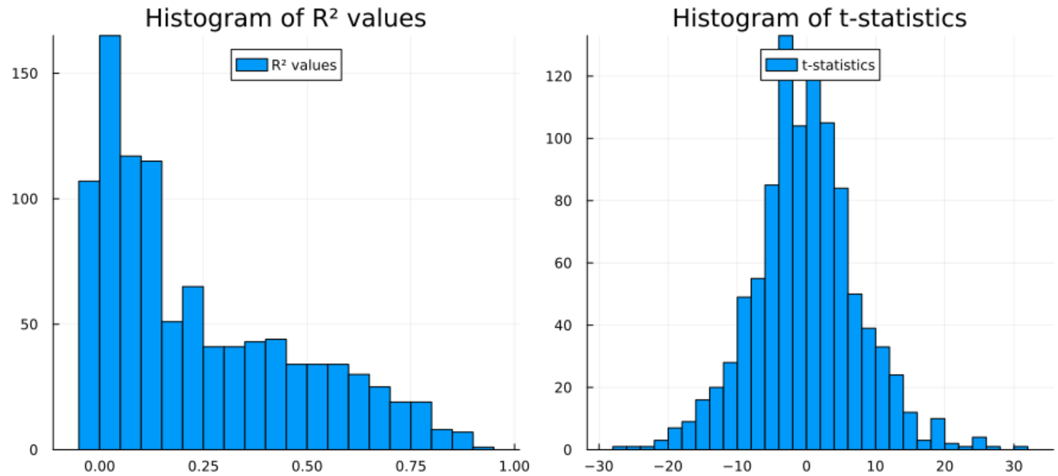

Figure 1: The distribution of $T - stat$ and R^2 estimates using 1000 MC simulations.

We then compute the fraction of the distribution for which the T-statistics exceeds 1.96. This fraction is given as $p = 0.751$. This is quite unexpected since there is no relationship between the underlying process because they both follow independent stochastic processes i.e. random walks. However, the fraction p = 0.75 tells us that we obtain a significant relationship between these two processes 75% of the time. This illustrates the concept of spurious regression where two independent random processes tend to have significant relationships or trends but are completely independent.

(c) Repeat b) for different numbers of observations, T = 50 and T = 200. As the sample size increase is that fraction changing? Which number does it seem to approach? Which number should it approach? Discuss.

Answer

By increasing the sample sizes we obtain that the fraction of the distribution for the t-stat exceeds 1.96 increases and seems to converge to 1 as in Table (2). This shows that as the sample size increases the higher the probability of finding a significant relationship between Y and X.

T	**Fraction Exceeding 1.96**
20	0.51
50	0.672
100	0.757
200	0.847
500	0.861
1000	0.915

Table 2: Fraction of t-statistics exceeding 1.96 for different values of T

2. Do the same simulations you have in the previous problem but this time play around with the DGP and try different values for 2and 2. First, try both equal to 0.99 , then 0.95 then 0.9 and then 0.8 . Comments on your results.

   Answer:
   Consider different combination of $\rho_1 = [0.99, 0.95, 0.9, 0.8]$ and $\rho_2 = [0.99, 0.95, 0.9, 0.8]$, such that the DGP is now

   $$\begin{aligned} Y_t &= \rho_1 Y_{t-1} + e_t \\ X_t &= \rho_2 X_{t-1} + a_t \end{aligned}$$

   We obtain the results in table (3) which shows the OLS coefficient, the R-squared, and the T-statistic for different combinations of ρ_1 and ρ_2.

ρ_1	ρ_2	β	R^2	**t_stat**
0.99	0.99	-0.253163	0.205691	-5.16107
0.99	0.95	-0.0671947	-0.00411377	-0.770977
0.99	0.90	0.390716	0.0287065	1.9814
0.99	0.80	0.155889	0.000921049	1.04464
0.95	0.99	-0.513305	0.211232	-5.2452
0.95	0.95	0.570944	0.255309	5.9111
0.95	0.90	0.207693	0.0283006	1.97062
0.95	0.80	-0.532849	0.0902175	-3.28895
0.90	0.99	-0.121814	0.0229923	-1.82478
0.90	0.95	0.547003	0.332112	7.0872
0.90	0.90	-0.340545	0.0490416	-2.47093
0.90	0.80	-0.46789	0.211607	-5.2509
0.80	0.99	-0.266865	0.20017	-5.07704
0.80	0.95	0.21162	0.0391799	2.24432
0.80	0.90	0.129121	0.00435239	1.19698
0.80	0.80	-0.021754	-0.00949668	-0.262056

Table 3: Regression results for different ρ_1 and ρ_2 values

However, we run the 1000 simulations and computed the fraction of the simulations that exceeds 1.96 but maintained T = 100, we obtained that for $\rho_1 = \rho_2 = [0.99, 0.95, 0.9, 0.8]$. This is interesting as even tho the coefficient increases and is close to 1 does not change the fraction too much. That is the fraction is about 0.77.

ρ_1	**Fraction Exceeding 1.96**
0.99	0.773
0.95	0.763
0.90	0.749
0.80	0.761

Table 4: Fraction of t-statistics exceeding 1.96 for different ρ_1 values

3. Do the same simulations you have in the previous problem but this time play around with the DGP and now generate the two variables as having a deterministic trend. What happens now?