EDHEC PhD Finance 2022 - Econometrics Homework

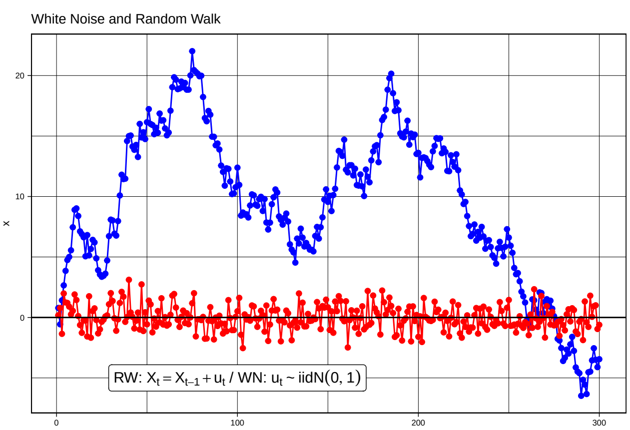
Giovanni Maffei, Timo Predoehl

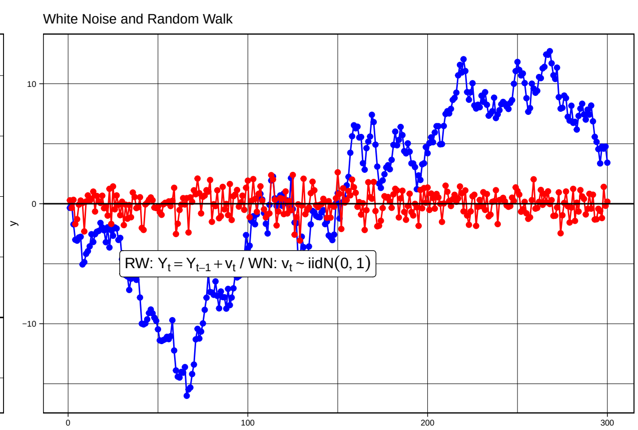
Question 1: Spurious Regressions (2 points)

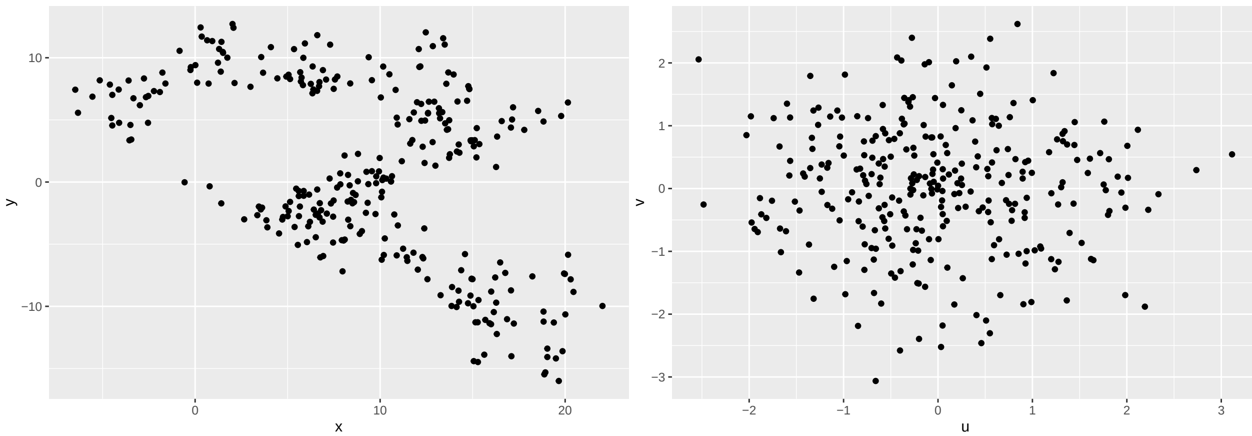
- **compulsory** (a) [1 Point]
  - i Replicate the analysis leading to Figure 14.1 in Davidson MacKinnon (2005, book) by running 100,000 MC, or simply 10,000 MC if your computer is slow (instead of 1 million), and using T = 6, 12, 60, 120, 240, 360, 480. (i) Compute also for each sample size T the distribution of the R^2s of the MC simulations with either 7 separate histograms, or one unique figure where you report on the y-axis the 5%, 10% 25%, 50%, 75%, 90% and 95% quantiles of the distributions of the simulated R^2, and on the x-axis you have T = 6, 12, 60, 120, 240, 360, 480.
  - ii Similarly (either with histograms, or with one plot of the quantiles) report the distributions of the estimates t-statistics for the test of the null H0 : $\beta_2 = 0$ and
  - iii their empirical rejection frequencies (that is the empirical size of the tests),which is exactly the figure 14.1 in Davidson MacKinnon (2005, book).
- (b) [1 Point] Based on the results obtained by answering to point (a) summarize the problems of spurious regressions in econometrics.

(a)(i) Replication Davidson McKinnon, 2005, fig 14.1

Define 2 random walk processes x_t and y_t analog to (14.03):





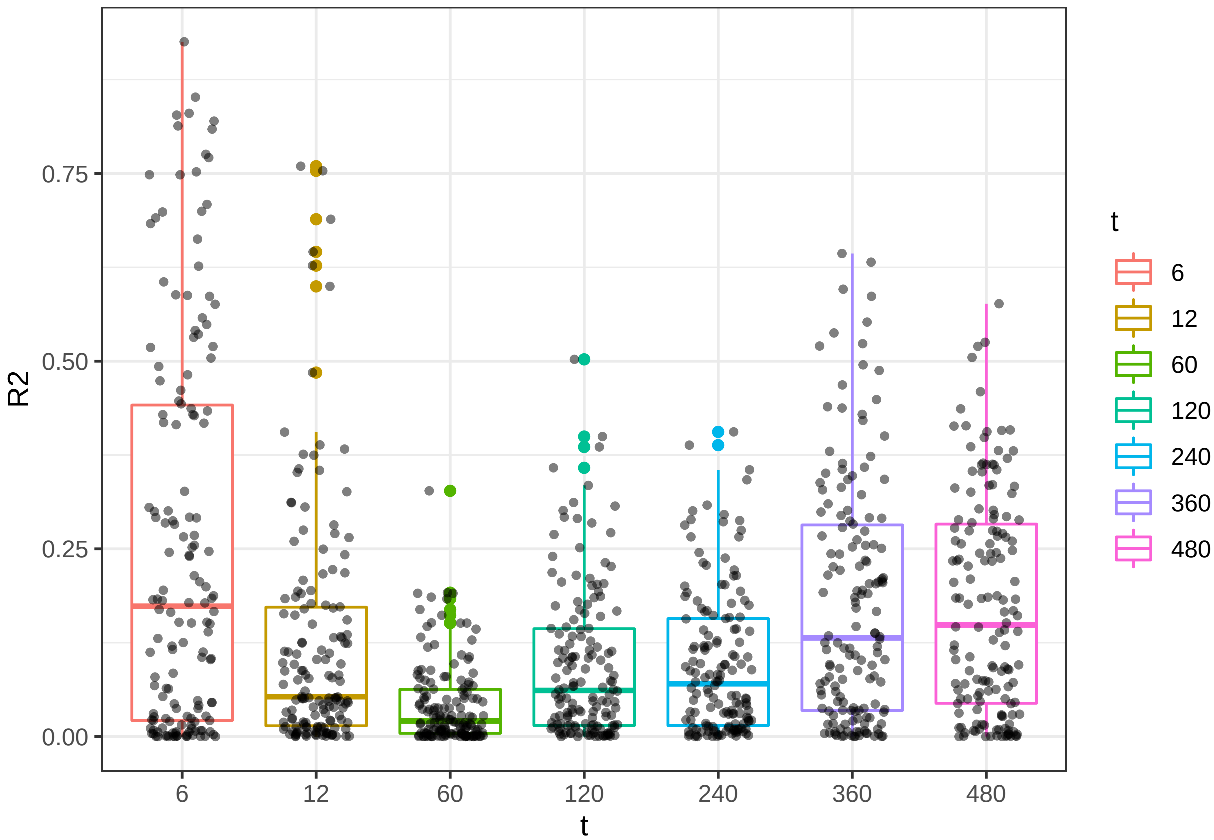


We run a simple linear regression analog to (14.12) in the form of $y_t = \beta_1 + \beta_2 x_t + v_t$ and expect it to yield both R^2 as well as β_2 insignificantly different from 0. The hypotheses $H_0^1 : \beta_2 = 0$ and $H_0^2 : R^2 = 0$ will be tested by running 50 simple linear regressions on each sample size n ranging from 6 to 480, yielding a total of 100,000 regression outcomes. The 50 runs per sample will be based on individually seeded random walks using the command 'set.seed(1000+n)'. For each regression, the t-test $t = \frac{\beta_2 - 0}{\sigma_{\beta_2}} > 1.96$ indicates a rejection of the hypothesis. The rejection rate (no. of t-test > 1.96 / 50) will then be plotted against n in replication of figure 14.1.

```
T <- c(6, 12, 60, 120, 240, 360, 480)
result <- c(t=NULL, R2=NULL, t_val=NULL, t_sig=NULL)
for (t in T){
  rejected_r2 <- NULL
  rejected_t <- NULL
  for (s in 1:150) {
    # n = 20; x0 = 0; mu = 0; sd = 1
    Seed <- t+s+100
    xt <- RW(Seed = Seed, N = t, x0 = 0, mu = 0, sd = 1)
    yt <- RW(Seed = 1/Seed, N = t, x0 = 0, mu = 0, sd = 1)
    su <- summary(lm(yt$x ~ xt$x))
    result <- rbind(result,c(t, su$r.squared, su$coefficients[2,3], ifelse(su$coefficients[2,3]>1.96,1,
  }
}
colnames(result) <- c("t", "R2", "t_value", "t_sig")

result <- as_tibble(result) %>%
  mutate(t = as.factor(t))

graph <- ggplot(result, aes(x = t, y = R2, colour = t)) +
  geom_boxplot(notch = FALSE) +
  geom_jitter(size = 1, alpha = 0.5, width = 0.25, colour = 'black') +
  theme_bw()
graph
```



0.75
0.50
0.25
0.00
R2
6
12
60
120
240
360
480
t
t
6
12
60
120
240
360
480