EDHEC PhD Finance 2022 - Econometrics Homework

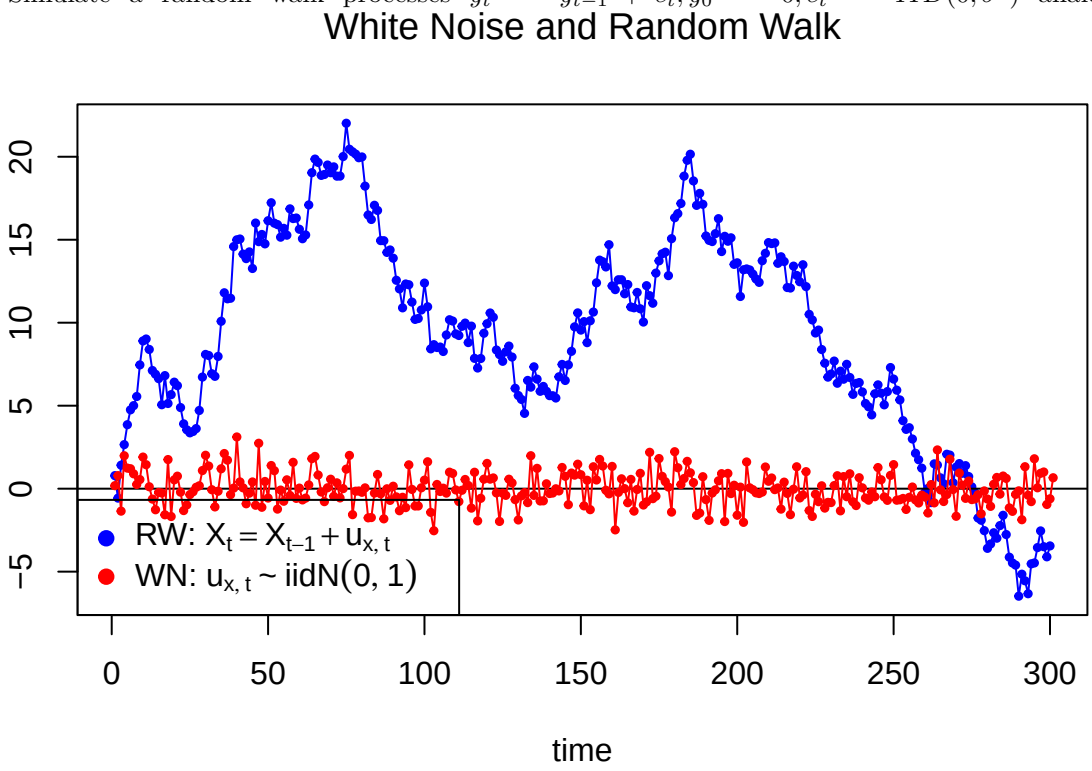
Giovanni Maffei, Timo Predoehl

Question 1: Spurious Regressions (2 points)

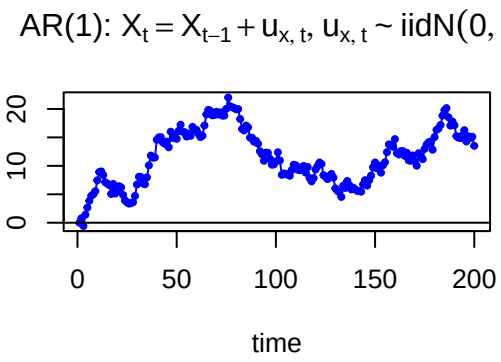
- **compulsory** (a) [1 Point]
  - i Replicate the analysis leading to Figure 14.1 in Davidson MacKinnon (2005, book) by running 100,000 MC, or simply 10,000 MC if your computer is slow (instead of 1 million), and using $T = 6, 12, 60, 120, 240, 360, 480$. (i) Compute also for each sample size T the distribution of the $R^2 s$ 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.

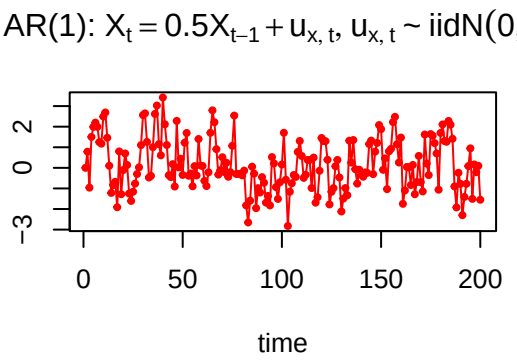
(a)(i) Replication Davidson McKinnon, 2005, fig 14.1

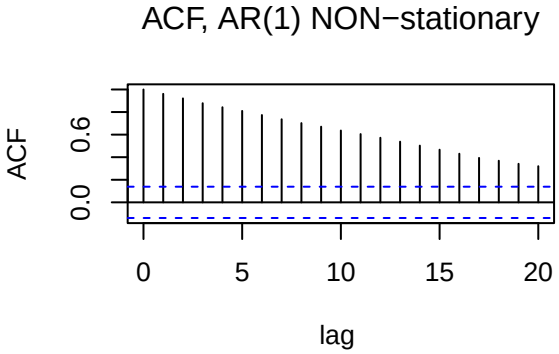
Simulate a random walk processes $y_t = y_{t-1} + e_t, y_0 = 0, e_t \sim IID(0, \sigma^2)$ analog to (14.03)

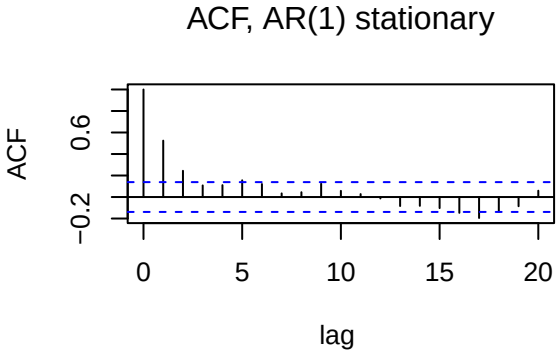


Simulate two AR(1) processes according to $x_t = \phi x_{t-1} + u_t$, with $\phi = 0.5$ and with $\phi = 1$. The ACFs of two processes indicate that in latter significant autocorrelation persists at least up to 20 lags, while the former autocorrelation becomes insignificant after 3 lags. Hence, in the case of AR(1), autocorrelation is a function of the parameter ϕ.



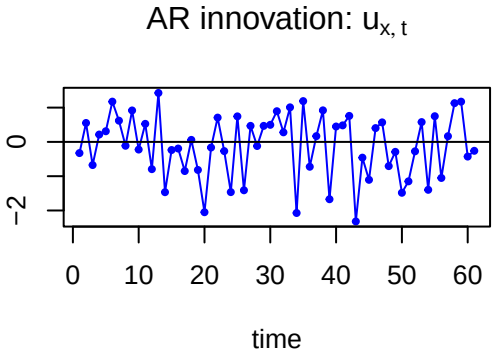


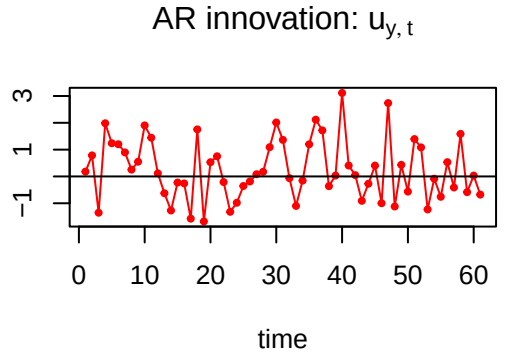


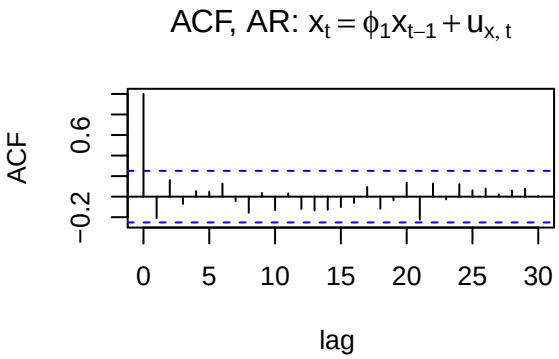


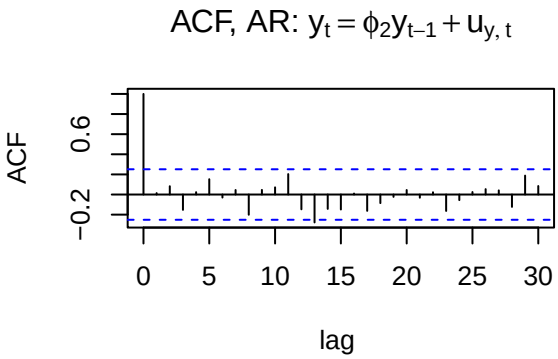
Define a function to produces N AR(1) simulations.

Plot the simulated error terms of the two AR(1)s, that is $u_{x,t}$ and $u_{y,t}$. The respective ACFs indicate that the error terms are not autocorrelated.









Scatterplot of $u_{x,t}$ and $u_{y,t}$. The error terms, while individually and randomly generated, appear to to correlated, which is confirmed by the regression of $u_{y,t}$ on $u_{x,t}$. The resulting t_β is significant.

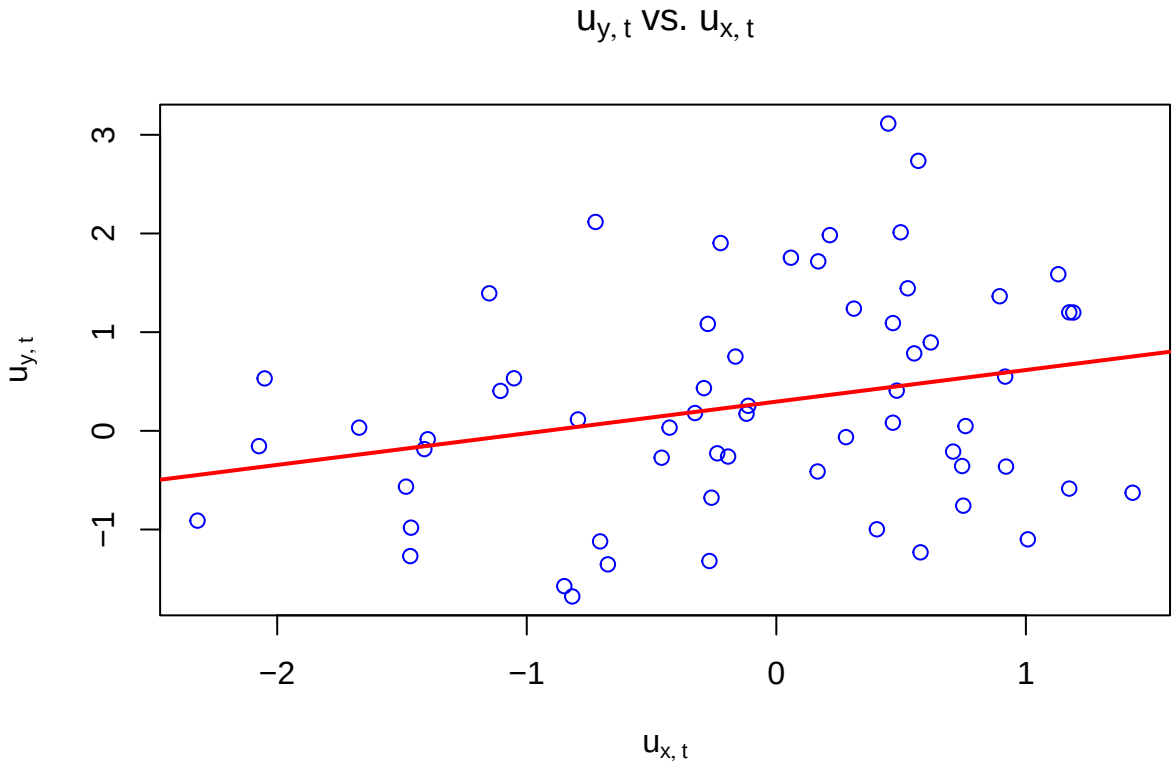
```
par(mfrow=c(1,1))
plot(x = u_sim[,1], y = u_sim[,2],
     col = "blue", lwd =1,
     xlab=TeX("$u_{x,t}$"), ylab=TeX("$u_{y,t}$"),
     main=TeX('$u_{y,t}$ vs. $u_{x,t}$'))
reg1  <- lm(u_sim[,2] ~ u_sim[,1])
stargazer(list(reg1),type="text")
```

```
##
## ================================================
##                         Dependent variable:
##                     ---------------------------
##                              u_sim[, 2]
## ------------------------------------------------
## u_sim[, 1]                     0.321**
##                                (0.151)
##
## Constant                       0.296**
##                                (0.139)
##
## ------------------------------------------------
## Observations                     61
## R2                              0.071
## Adjusted R2                     0.055
## Residual Std. Error        1.080 (df = 59)
## F Statistic             4.481** (df = 1; 59)
## ================================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

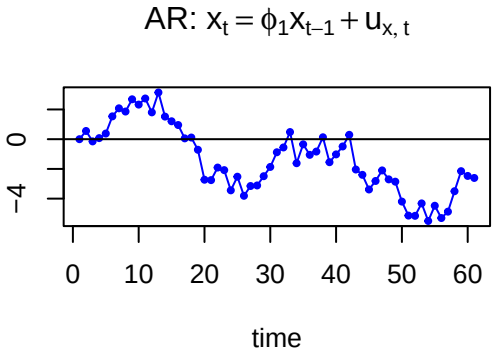
```
summary(reg1)
```

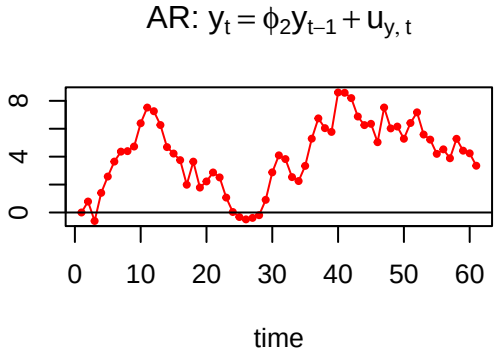
```
##
## Call:
## lm(formula = u_sim[, 2] ~ u_sim[, 1])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.71839 -0.80795 -0.02899  0.64651  2.67481
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.2958     0.1393   2.123   0.0379 *
## u_sim[, 1]    0.3205     0.1514   2.117   0.0385 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.08 on 59 degrees of freedom
## Multiple R-squared:  0.07058,    Adjusted R-squared:  0.05483
## F-statistic: 4.481 on 1 and 59 DF,  p-value: 0.03851
```

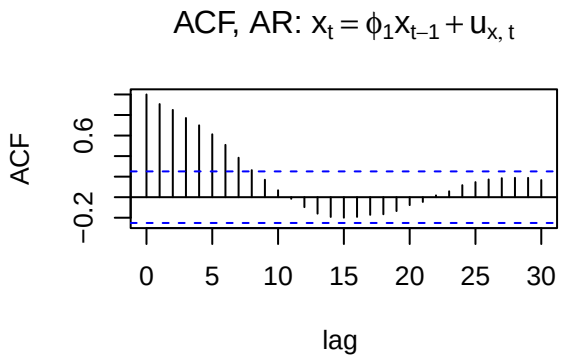
```r
abline(reg1, col="red", lwd=2)
```

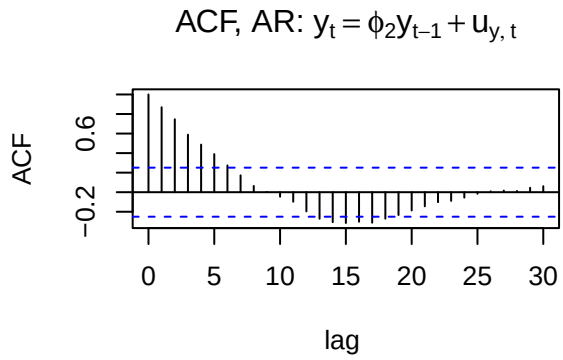


Compare ACF of the two AR processes. Autocorrelation is present and takes a fading wave pattern with significance at the initial 5-10 lags and then again around 15 lags.









Run a regression of two AR(1) processes.

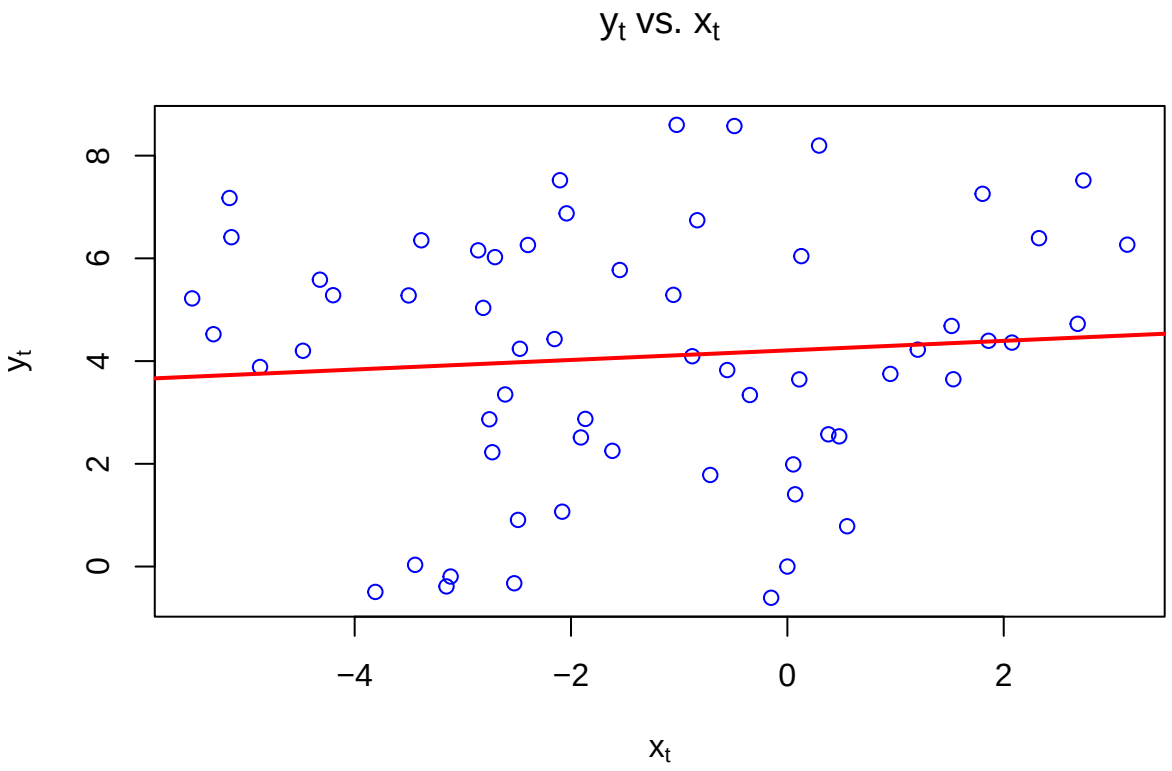
```
################################################
# scatterplot of y_t vs. x_t
################################################
par(mfrow=c(1,1))
plot(x = AR_sim[,1], y = AR_sim[,2],
     col = "blue", lwd =1,
     xlab=TeX("$x_{t}$"), ylab=TeX("$y_{t}$"),
     main=TeX('$y_{t}$ vs. $x_{t}$'),   )

reg2  <- lm(AR_sim[,2] ~ AR_sim[,1])
stargazer(list(reg2),type="text")
```

```
##
## ================================================
##                          Dependent variable:
##                      ---------------------------
##                              AR_sim[, 2]
## ------------------------------------------------
## AR_sim[, 1]                    0.093
##                               (0.143)
##
## Constant                      4.208***
##                               (0.372)
##
## ------------------------------------------------
## Observations                    61
## R2                             0.007
## Adjusted R2                    -0.010
## Residual Std. Error       2.479 (df = 59)
## F Statistic             0.424 (df = 1; 59)
## ================================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

```
summary(reg2)
```

```
##
## Call:
## lm(formula = AR_sim[, 2] ~ AR_sim[, 1])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.8017 -1.7177  0.1307  1.8230  4.4862
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.20802    0.37243  11.299   <2e-16 ***
## AR_sim[, 1]  0.09304    0.14292   0.651    0.518
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.479 on 59 degrees of freedom
## Multiple R-squared:  0.007132,   Adjusted R-squared:  -0.009697
## F-statistic: 0.4238 on 1 and 59 DF,  p-value: 0.5176
```

```
abline(reg2, col="red", lwd=2) # display regression line on scatterplot (y~x)
```



Compare the regression results

```
stargazer(list(reg1,reg2), type = "text")
```

```
##
## ===============================================================
##                                   Dependent variable:
##                               ----------------------------
##                                  u_sim[, 2]    AR_sim[, 2]
##                                     (1)            (2)
## ---------------------------------------------------------------
## u_sim[, 1]                        0.321**
##                                   (0.151)
##
## AR_sim[, 1]                                       0.093
##                                                  (0.143)
##
## Constant                          0.296**       4.208***
##                                   (0.139)        (0.372)
##
## ---------------------------------------------------------------
## Observations                        61             61
## R2                                0.071          0.007
## Adjusted R2                       0.055         -0.010
## Residual Std. Error (df = 59)     1.080          2.479
## F Statistic (df = 1; 59)         4.481**         0.424
## ===============================================================
## Note:                          *p<0.1; **p<0.05; ***p<0.01
```

```
# Magics: you can save tabel directly in latex !!!!!!!!!!!!!!!!!!
# stargazer(list(reg1,reg2),
#           align=TRUE, type = "text", no.space = TRUE,
#           title = "Table X", out = "texoutput/fit.tex")
```