EDHEC PhD Finance 2022 - Econometrics Homework

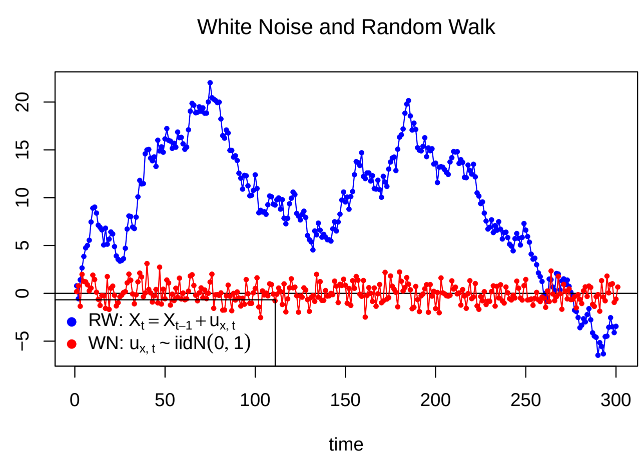
Giovanni Maffei, Timo Predoehl

Question 1: Spurious Regressions (2 points)

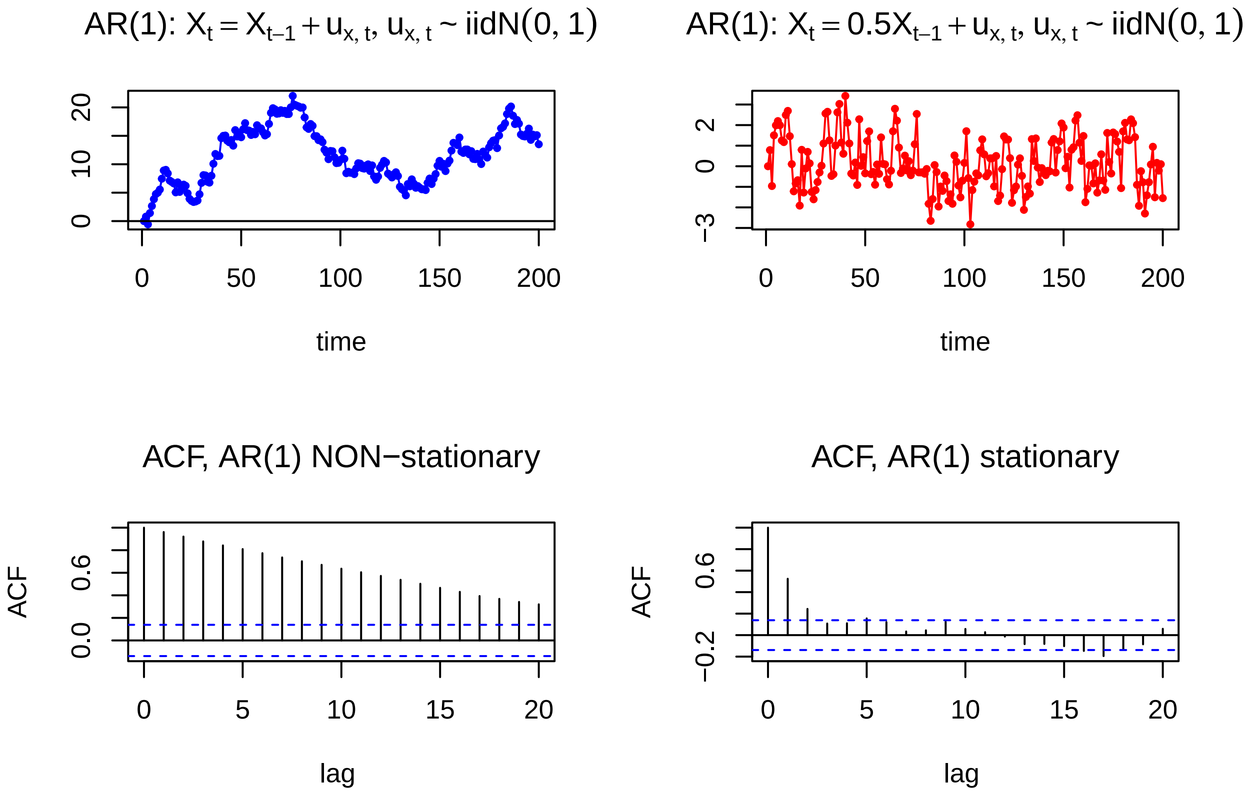
- **compulsory** (a) [1 Point]
  - i Replicate the analysis leading to Figure 14.1 in Davidson MacKinnon (2005, book) by running 100,000 MC, or simply 10,000 MC if your computer is slow (instead of 1 million), and using $T = 6, 12, 60, 120, 240, 360, 480$. (i) Compute also for each sample size T the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

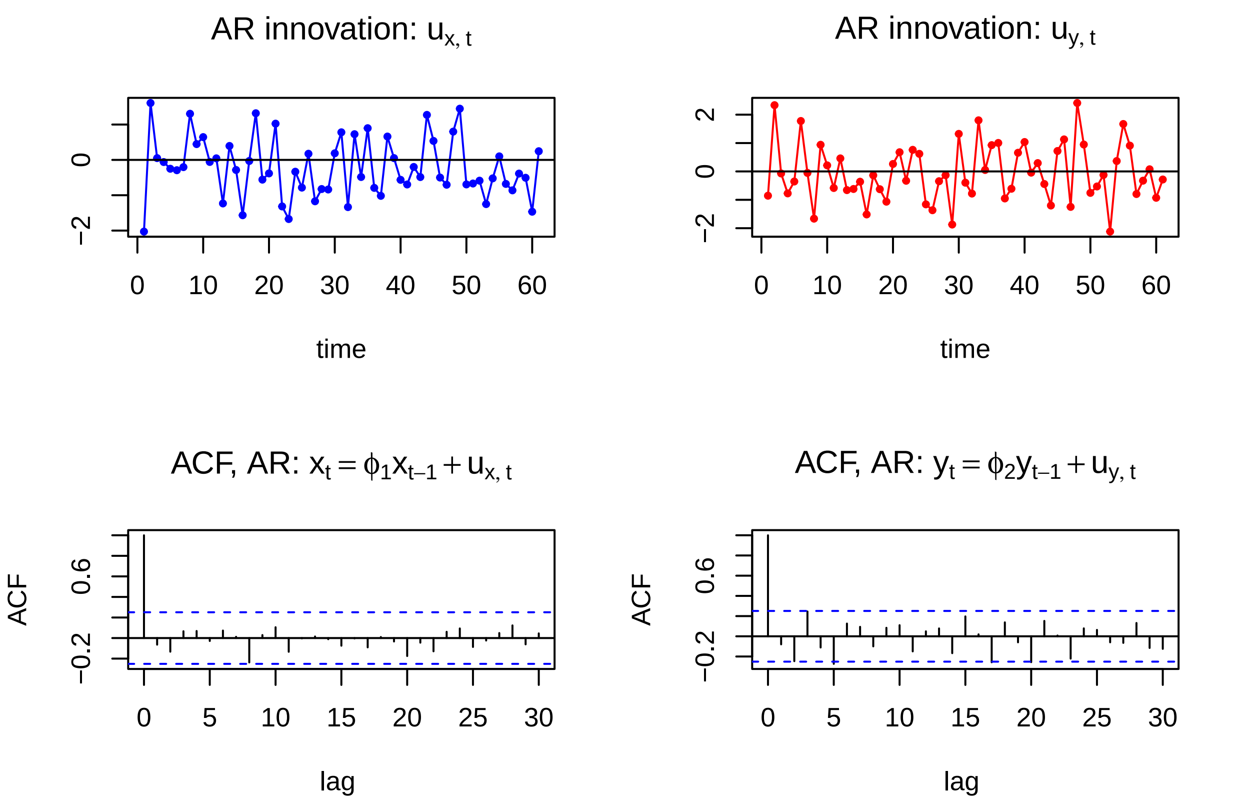


Simulate two AR(1) processes according to $x_t = \phi x_{t-1} + u_t$, with $\phi = 0.5$ and with $\phi = 1$. The ACFs of two processes indicate that in latter significant autocorrelation persists at least up to 20 lags, while the former autocorrelation becomes insignificant after 3 lags. Hence, in the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



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

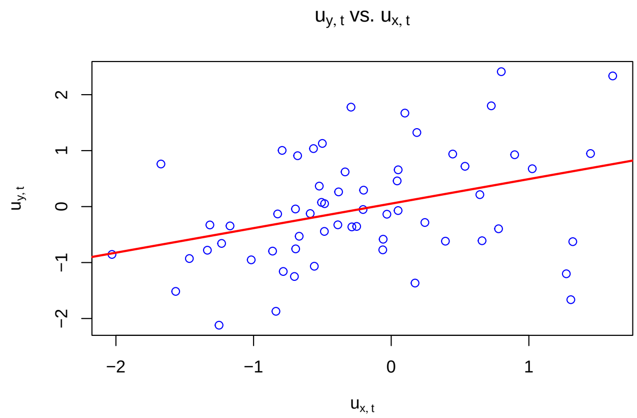
Plot the simulated error terms of the two AR(1)s, that is $u_{x,t}$ and $u_{y,t}$. The respective ACFs indicate that the error terms are not autocorrelated.



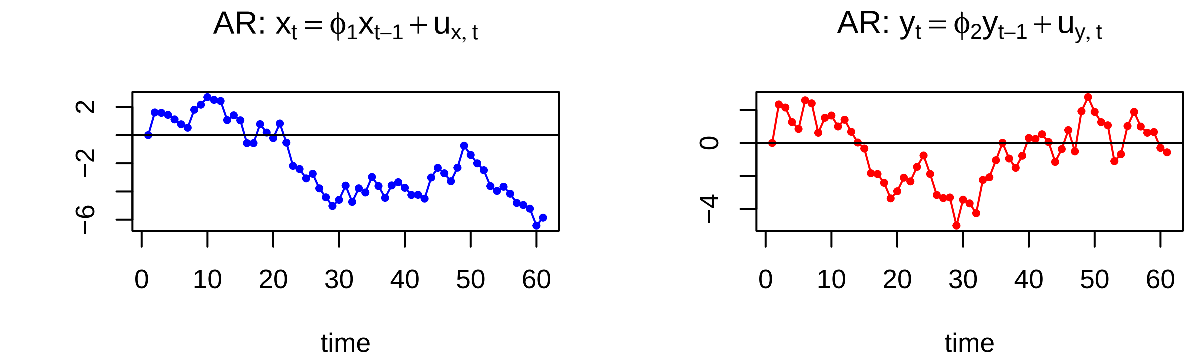
Scatterplot of $u_{x,t}$ and $u_{y,t}$. The error terms, while individually and randomly generated, appear to to correlated, which is confirmed by the regression of $u_{y,t}$ on $u_{x,t}$. The resulting t_β is significant.

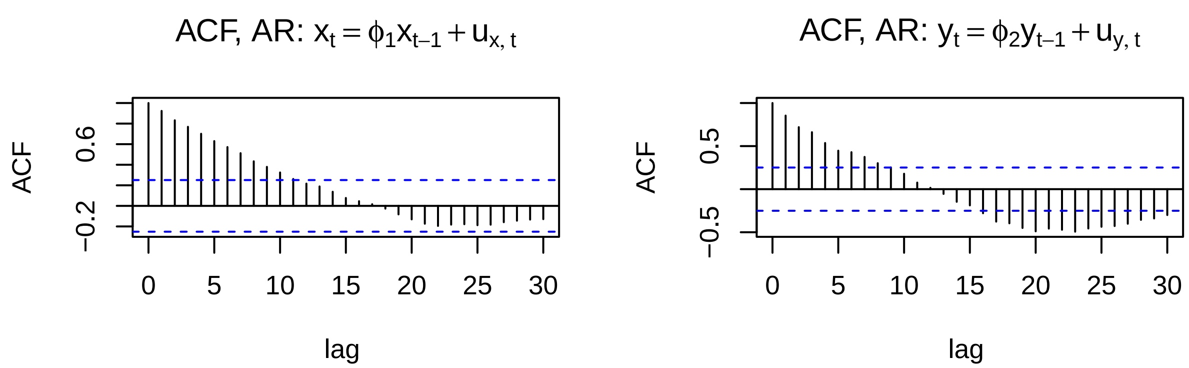
```
##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                             u_sim[, 2]
## -----------------------------------------------
## u_sim[, 1]                   0.438***
##                               (0.147)
##
## Constant                       0.053
##                               (0.125)
##
## -----------------------------------------------
## Observations                    61
## R2                             0.131
## Adjusted R2                    0.116
## Residual Std. Error       0.943 (df = 59)
## F Statistic            8.895*** (df = 1; 59)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

```
##
## Call:
## lm(formula = u_sim[, 2] ~ u_sim[, 1])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.2890 -0.5590 -0.0190  0.4804  2.0071
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.05336    0.12516   0.426  0.67143
## u_sim[, 1]   0.43838    0.14699   2.982  0.00415 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9435 on 59 degrees of freedom
## Multiple R-squared:  0.131,  Adjusted R-squared:  0.1163
## F-statistic: 8.895 on 1 and 59 DF,  p-value: 0.004152
```



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.



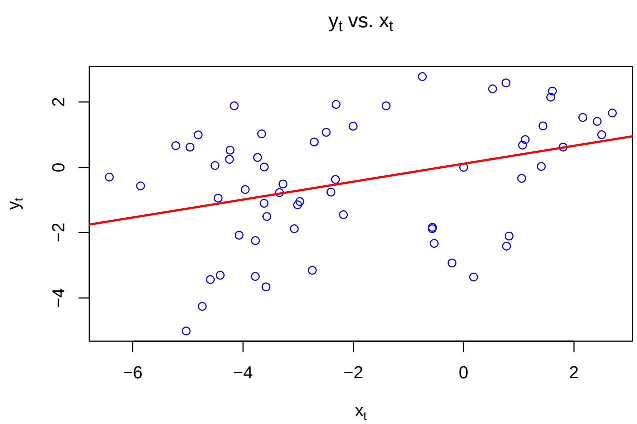


Run a regression of two AR(1) processes.

```
##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                             AR_sim[, 2]
## -----------------------------------------------
## AR_sim[, 1]                  0.274***
##                               (0.091)
##
## Constant                       0.110
##                               (0.290)
```

```
##
## -----------------------------------------------
## Observations                      61
## R2                               0.132
## Adjusted R2                      0.118
## Residual Std. Error        1.777 (df = 59)
## F Statistic             8.997*** (df = 1; 59)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

```
##
## Call:
## lm(formula = AR_sim[, 2] ~ AR_sim[, 1])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7395 -1.1474  0.2007  1.4091  2.9137
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.11036    0.29011    0.38  0.70500
## AR_sim[, 1]  0.27438    0.09148    3.00  0.00396 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.777 on 59 degrees of freedom
## Multiple R-squared:  0.1323, Adjusted R-squared:  0.1176
## F-statistic: 8.997 on 1 and 59 DF,  p-value: 0.003956
```



Compare the regression results using stargazer:

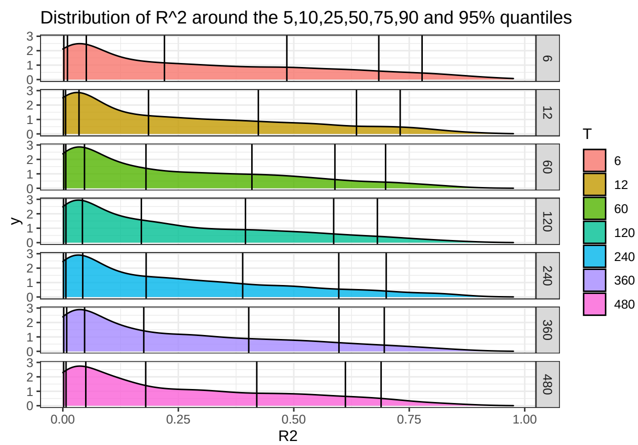
```
##
## ==========================================================
##                               Dependent variable:
##                               ----------------------------
##                                 u_sim[, 2]    AR_sim[, 2]
##                                    (1)            (2)
## ----------------------------------------------------------
## u_sim[, 1]                       0.438***
##                                  (0.147)
##
## AR_sim[, 1]                                    0.274***
```

```
##                                                  (0.091)
##
## Constant                          0.053          0.110
##                                   (0.125)        (0.290)
##
## -----------------------------------------------------------
## Observations                        61             61
## R2                                0.131          0.132
## Adjusted R2                       0.116          0.118
## Residual Std. Error (df = 59)     0.943          1.777
## F Statistic (df = 1; 59)         8.895***       8.997***
## ===========================================================
## Note:                          *p<0.1; **p<0.05; ***p<0.01
```

Now lets do the 1,000,000 simulations of 2 AR(1)s, regress them on each other and store the regression results in a matrix for further analysis.

```
## 23.417 sec elapsed
```

Plot histogram of results



The rate of rejection of $H_0 : \beta = 0$ increases with the sample size T.