Week 6: Time Series, Training Exercises

Coursera/Erasmus U., Econometric Methods and Applications

Anthony Nguyen

Training Exercise 6.1

Notes:

- This exercise uses the datafile `TrainExer61`and requires a computer.
- The dataset `TrainExer61` is available on the website.

Questions

The datafile `TrainExer61` contains values of four series of length 250. Two of these series are uncorrelated white noise series denoted by ε_{xt} and ε_{yt}, where both variables are NID(0,1), that is, normally and independently distributed standard normal random variables. The other two series are so-called random walks constructed from these two white noise series by $x_t = x_{t-1} + \varepsilon_{xt}$ and $y_t = y_{t-1} + \varepsilon_{yt}$, with starting values $x_1 = 0$ and $y_1 = 0$.

As ε_{xt} and ε_{yt} are independent for all values of t and s, the same holds true for all values of x_t and y_s. The purpose of this exercise is to experience that, nonetheless, the regression of y on x indicates a highly significant relation between y and xx if evaluated by standard regression tools. This kind of result is called 'spurious regression' and is caused by the trending nature of the variables x and y. The lesson we learn is that standard regression tools are not applicable if the variables contain trends similar to those of the random walks considered here.

(a) Use dataset `TrainExer61` to make the following graphs: the time series plot of x_t against time t, the time series of plot y_t against time t, and the scatter plot of y_t against x_t. What conclusion could you draw from these three graphs?

(b) To check that the series ε_{xt} and ε_{yt} are uncorrelated, regress ε_{yt} on a constant and ε_{xt}. Report the t-value and p-value of the slope coefficient.

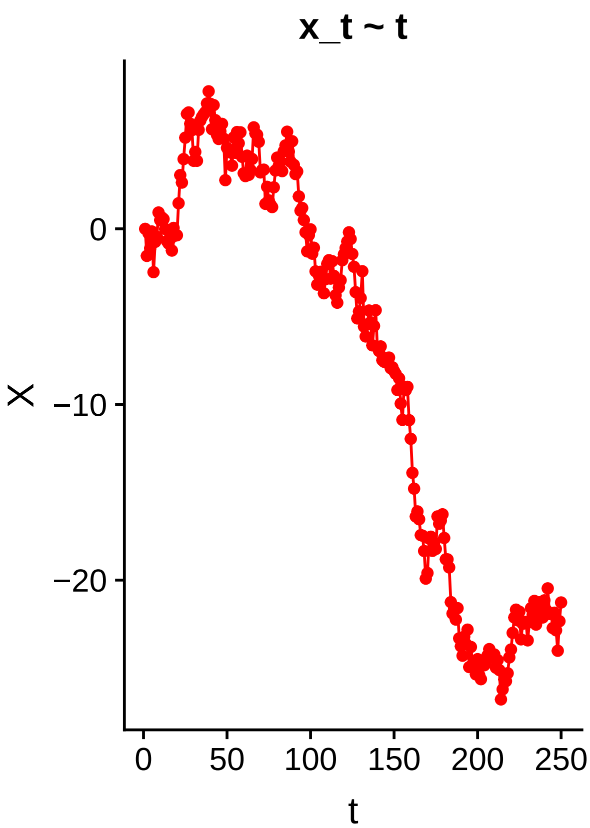
(c) Extend the analysis of part(b) by regressing ε_{yt} on a constant, ε_{xt} and three lagged values of ε_{yt} and of ε_{xt}. Perform the F-test for the joint insignificance of the seven parameters of ε_{xt} and the three lags of ε_{xt} and ε_{yt}. Report the degrees of freedom of the F-test and the numerical outcome of this test, and draw your conclusion. *Note:* The relevant 5% critical value is 2.0.

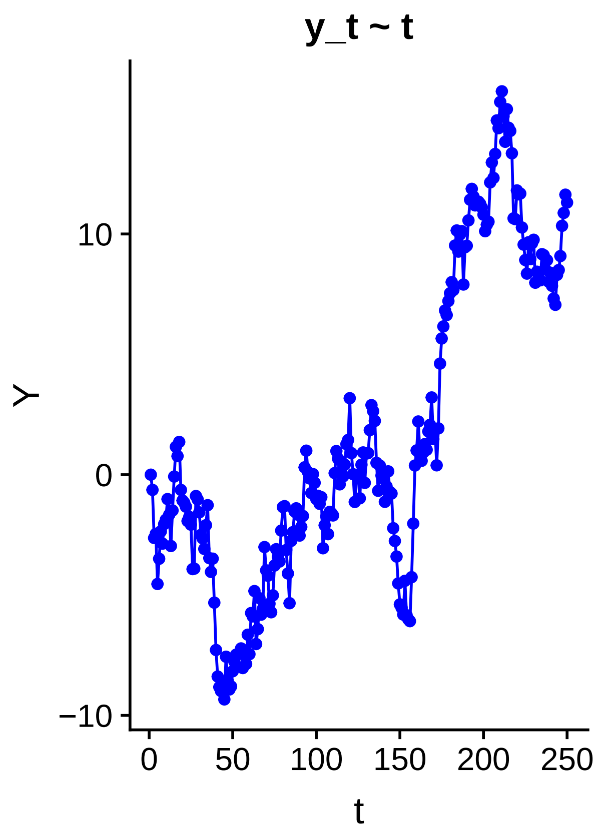
(d) Regress y on a constant and x. Report the t-value and p-value of the slope coefficient. What conclusion would you be tempted to draw if you did not know how the data were generated?

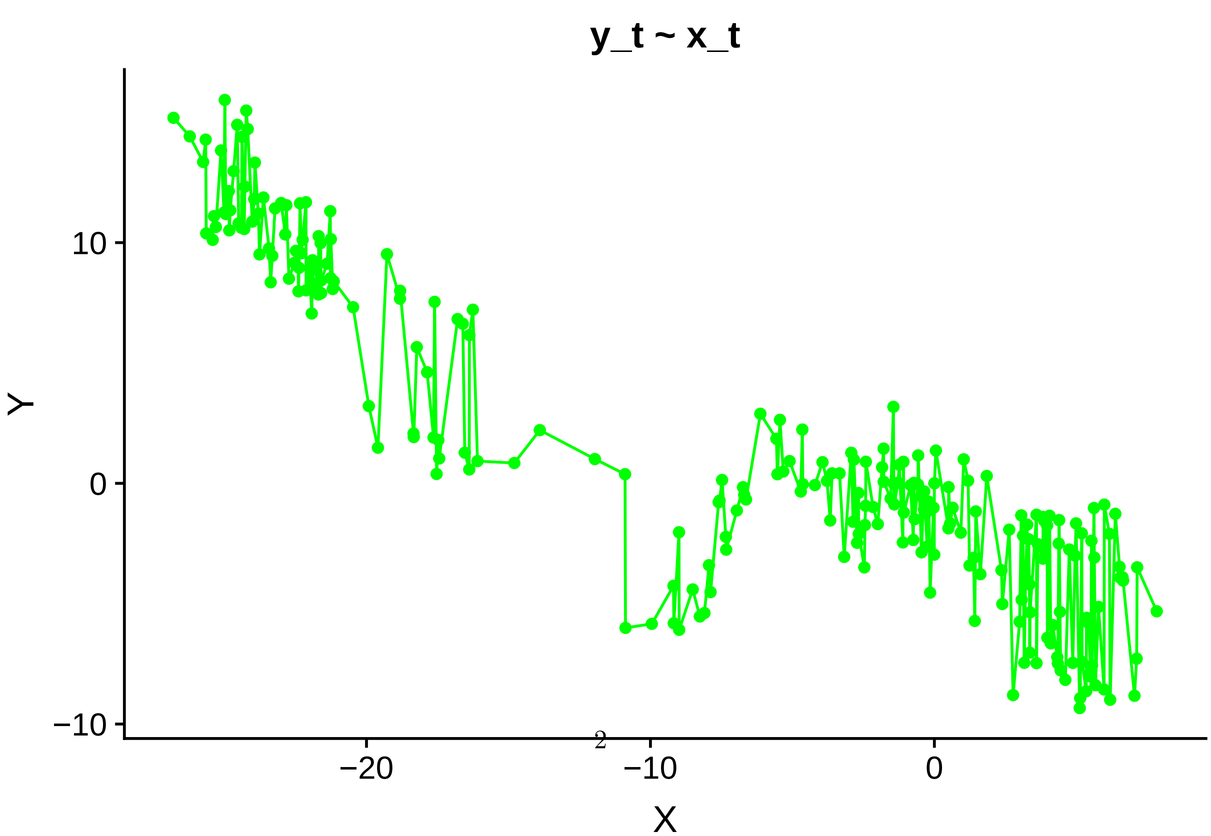
(e) Let e_t be the residuals of the regression of part (d). Regress e_t on a constant and the one-period lagged residual e_{t-1}. What standard assumption of regression is clearly violated for the regression in part (d)?

Answers

(a) Use dataset `TrainExer61` to make the following graphs: the time series plot of x_t against time t, the time series of plot y_t against time t, and the scatter plot of y_t against x_t. What conclusion could you draw from these three graphs?







Looking at the three graphs, we can see that the first two graphs move randomly up and down, are examples of "random walk".

In the third graph, when we look at the graph of y_t against x_t, it looks like there is a negative correlation between the two variables, even though we know this is **not** the case.

(b) To check that the series ε_{xt} and ε_{yt} are uncorrelated, regress ε_{yt} on a constant and ε_{xt}. Report the t-value and p-value of the slope coefficient.

```
##                Estimate Std. Error    t value  Pr(>|t|)
## (Intercept)  0.03119474 0.06445519  0.4839756 0.6288303
## EPSX        -0.08791399 0.06671517 -1.3177510 0.1888030
```

After performing the regression, we can see that the regression ε_{xt} has the following values:
t-value = -1.318
p-value = 0.189

(c) Extend the analysis of part(b) by regressing ε_{yt} on a constant, ε_{xt} and three lagged values of ε_{yt} and of ε_{xt}. Perform the F-test for the joint insignificance of the seven parameters of ε_{xt} and the three lags of ε_{xt} and ε_{yt}. Report the degrees of freedom of the F-test and the numerical outcome of this test, and draw your conclusion. *Note:* The relevant 5% critical value is 2.0.

Here, the question is asking us to perform the following regression:

$$\varepsilon_{yt} = \text{constant} + \gamma_1\varepsilon_{xt} + \gamma_2\varepsilon_{xt-1} + \gamma_3\varepsilon_{xt-2} + \gamma_4\varepsilon_{xt-3} + \gamma_5\varepsilon_{yt-1} + \gamma_6\varepsilon_{yt-2} + \gamma_7\varepsilon_{yt-3}$$

```
##
## Call:
## lm(formula = EPSY ~ EPSX + lag(EPSX, 1) + lag(EPSX, 2) + lag(EPSX,
##     3) + lag(EPSY, 1) + lag(EPSY, 2) + lag(EPSY, 3), data = TrainExer61)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6912 -0.6563 -0.0319  0.6930  2.9320
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.046154   0.066095   0.698    0.486
## EPSX         -0.097274   0.069233  -1.405    0.161
## lag(EPSX, 1)  0.019901   0.070094   0.284    0.777
## lag(EPSX, 2) -0.059641   0.069598  -0.857    0.392
## lag(EPSX, 3)  0.009456   0.068390   0.138    0.890
## lag(EPSY, 1)  0.024786   0.064125   0.387    0.699
## lag(EPSY, 2) -0.015816   0.064698  -0.244    0.807
## lag(EPSY, 3) -0.047213   0.064340  -0.734    0.464
##
## Residual standard error: 1.015 on 239 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.01573,    Adjusted R-squared:  -0.0131
## F-statistic: 0.5457 on 7 and 239 DF,  p-value: 0.7992
```

The F-test for the join insignificance of the 7 parameters would test to see if:

$$H_0 = \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = \gamma_6 = \gamma_7 = 0$$

The test statistic is given by: $F(g, n-k)$
Where $g =$ the number of parameter restrictions in H_0, n is the number of observations, and k is the number of variables in the unrestricted model.

```
H_0 <- c("EPSX=0", "lag(EPSX, 1)=0", "lag(EPSX, 2)=0", "lag(EPSX, 3)=0", "lag(EPSY, 1)=0", "lag(EPSY, 2
linearHypothesis(mod2, H_0)
```

```
## Linear hypothesis test
##
## Hypothesis:
## EPSX = 0
## lag(EPSX,0
## lag(EPSX, 2) = 0
## lag(EPSX, 3) = 0
## lag(EPSY,0
## lag(EPSY, 2) = 0
## lag(EPSY, 3) = 0
##
## Model 1: restricted model
## Model 2: EPSY ~ EPSX + lag(EPSX, 1) + lag(EPSX, 2) + lag(EPSX, 3) + lag(EPSY,
##     1) + lag(EPSY, 2) + lag(EPSY, 3)
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    246 250.07
## 2    239 246.14  7    3.9344 0.5457 0.7992
```

Here we can see that our F-test statistic has degrees of freedom equal to: F(7, 239), and a value of 0.5457, with a p-value of 0.799, which means that we cannot reject H_0, the null hypothesis.

(d) Regress y on a constant and x. Report the t-value and p-value of the slope coefficient. What conclusion would you be tempted to draw if you did not know how the data were generated?

```
##               Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) -2.4866496 0.21425034 -11.60628 3.669633e-25
## X           -0.5149409 0.01559274 -33.02440 9.020776e-93
```

Looking at the t-value (-33.024) and the p-value (0.000) for the X coefficient, one would conclude that there is a negative correlation between Y and X, and that this relationship is highly significant. This is an example of a 'spurious regression', as we demonstrated in part(a), where we showed that this relationship is misleading.

(e) Let e_t be the residuals of the regression of part (d). Regress e_t on a constant and the one-period lagged residual e_{t-1}. What standard assumption of regression is clearly violated for the regression in part (d)?

```
##                     Estimate Std. Error      t value      Pr(>|t|)
## (Intercept)    0.0005562279 0.06738631  0.008254316  9.934207e-01
## lag(e_mod3, 1) 0.9251405179 0.02423649 38.171397911 1.408281e-105
```

Looking at the t-value and p-value from this regression, we can see that the residuals are very strongly correlated, which violates the standard regression assumption that the error terms should be uncorrelated.