Econometrics: Time Series

Diego López Tamayo [*] Based on MOOC by Erasmus University Rotterdam

Contents

"There are two things you are better off not watching in the making: sausages and econometric estimates." -Edward Leamer

[*] El Colegio de México, diego.lopez@colmex.mx

Time series

What is a time serie.

Look at Dates and Times in R Without Losing Your Sanity to understand how to use correctly date labes in R. Datasets to be used:

```
revenue <- read_csv(
  "https://raw.githubusercontent.com/diego-eco/diego-eco.github.io/master/downloads/dataset61.csv")
revenue$YEAR<- as.Date(paste0(revenue$YEAR, '-01-01'))

production <- read_csv(
  "https://raw.githubusercontent.com/diego-eco/diego-eco.github.io/master/downloads/dataset62.csv")
# We replace the weird "M" before months.
production <- rename(production, date=`YYYY-MM`)
production$date <- gsub("M","-",production$date)
production$date <- as.Date(as.yearmon(production$date))

dataset_training <- read_csv(
  "https://raw.githubusercontent.com/diego-eco/diego-eco.github.io/master/downloads/trainexer61.csv")
```

Time series data are a specific type of data that need a somewhat special treatment when using econometric methods. The specific aspect of time series variables is that they are sequentially observed. That is, one observation follows after another. The sequential nature of time series observations has important implications for modeling and especially for forecasting and this is different from the cross-sectional data that we have mostly looked at so far.

Think of the shoe size of your next-door neighbor. Now, it is quite unlikely that the very fact that someone lives next to you, implies that this person's shoe size has predictive value for yours. But with time series data, this is different.

Yesterday's sales level, likely has predictive value for today's sales level. Just like last month's inflation has for current inflation and your last year's disposable income for this year's.

A time series variable is observed at a **regular frequency.** This can be once per year, once per month, every day and sometimes, like in some areas of finance, even each millisecond. You can imagine that recent observations on a certain time series variable can have predictive value for future observations. If it is winter-like weather today, it will most likely be so tomorrow. When unemployment is high this month, it probably is still going to be high next month.

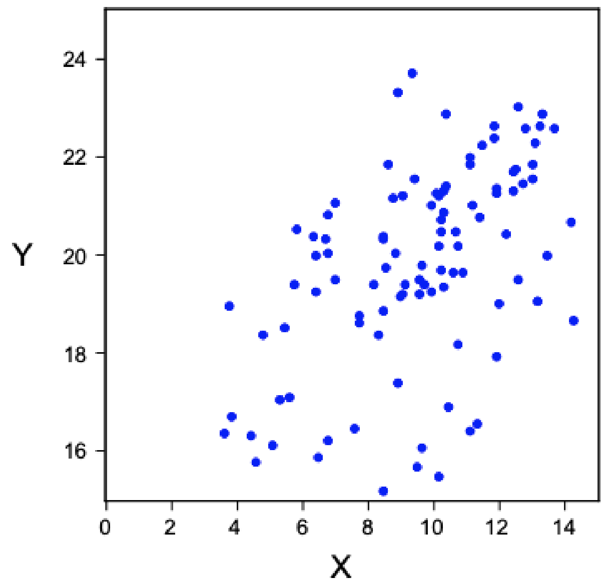
So in terms of regression models, you may want to include the past of a variable in order to predict its future. That is, to predict a new observation of y, you can use another variable X, but you can also think of using y one period lagged. $y(-1)$.

The inclusion of lagged values of the dependent variable in your regression model can also **prevent you from drawing spurious conclusions.** That is, you might think that another variable X helps to predict the variable of interest Y, while in reality, Y one period lagged predicts Y and X is irrelevant.

To illustrate this point consider two variables, X and Y, for which we know that the true data generating process is such that they depend with a factor 0.9 on their own previous value, whereas the variables X and Y are completely uncorrelated.

$$x_t = 1 + 0.9x_{t-1} + \epsilon_{x,t} \text{ and } y_t = 2 + 0.9y_{t-1} + \epsilon_{y,t}$$
$$\text{Two series completely uncorrelated } E(\epsilon_{y,t}, \epsilon_{x,s}) = 0 \forall t, s$$

A scatter of simulated Y and X variables with 100 observations may look like this.



Note that there seems to be some positive connection between the two, while we know that they are completely uncorrelated. You could be tempted to fit a simple regression model. Now suppose you would do so.

Dependent variable: Y (sample size $n = 100$)						
	Coef.	t-Stat.	p-value	Coef.	t-Stat.	p-value
Constant	15.99	23.45	0.000	2.91	2.87	0.005
X	0.40	5.78	$\boxed{0.000}$	0.07	1.53	$\boxed{0.129}$
Y(-1)	-	-	-	0.82	14.01	0.000
R-squared	0.254			0.753		

At the left-hand side of this table, you see that we estimate the slope parameter to be equal to 0.4 with a p-value of 0.000. So, this suggests that X has predictive value for Y. Now we know of course, this cannot be true given the way we created the data. The right-hand panel of the table shows what happens if we also include the Y variable one period lagged. The coefficient for this lagged variable is 0.82 and it is significant, whereas the coefficient of X is close to 0 and not statistically significant anymore.

You may now wonder whether we should have included not only X, but also X one period lagged. Consider the regression model where Y depends on Y one period lagged, X and also X one period lagged. Do X and its lag have any predictive power?

Dependent variable: Y (sample size $n = 100$)						
	Coef.	t-Stat.	p-value	Coef.	t-Stat.	p-value
Constant	2.88	2.83	0.006	2.69	2.66	0.009
Y(-1)	0.83	14.02	0.000	0.86	17.03	0.000
X	0.15	1.61	0.110	-	-	-
X(-1)	-0.09	-0.99	0.324	-	-	-
R-squared	0.756			0.747		

- Use F-test $F = \frac{(R_1^2 - R_0^2)/g}{(1-R_1^2)/(n-k)} \sim F_{(g,n-k)}$
- Number of restrictions: $g = 2$
- number of observations: $n = 100$
- number of parameters unrestricted model: $k = 4$
- values of R-squared: Unrestricted: $R_1^2 = 0.756$ and Restricted: $R_0^2 = 0.747$
- Substitute these values in formula for F-test: $F = 1.8 < 3.1$

- Joint effect of X and $X(-1)$ on Y is not significant

The larger model contains two extra variables, so the number of restrictions is two. We have 100 observations and the full model has 4 variables. The two R-squared values were reported in the table. Substituting these values in the familiar expression for the F-test gives a value of 1.8, which is smaller than the 5% critical value of 3.1. So even when we include X and one period lagged X, then these variables do not help to predict Y. Recall that the scatter of Y versus X was very suggestive, but proper analysis shows that pictures can sometimes fool us.

Example Airline revenue

Dataset: revenue

Simulated data set on yearly revenue passenger kilometers, 1975-2015 (estimation period 1976-2015, with pre-sample value for 1975).

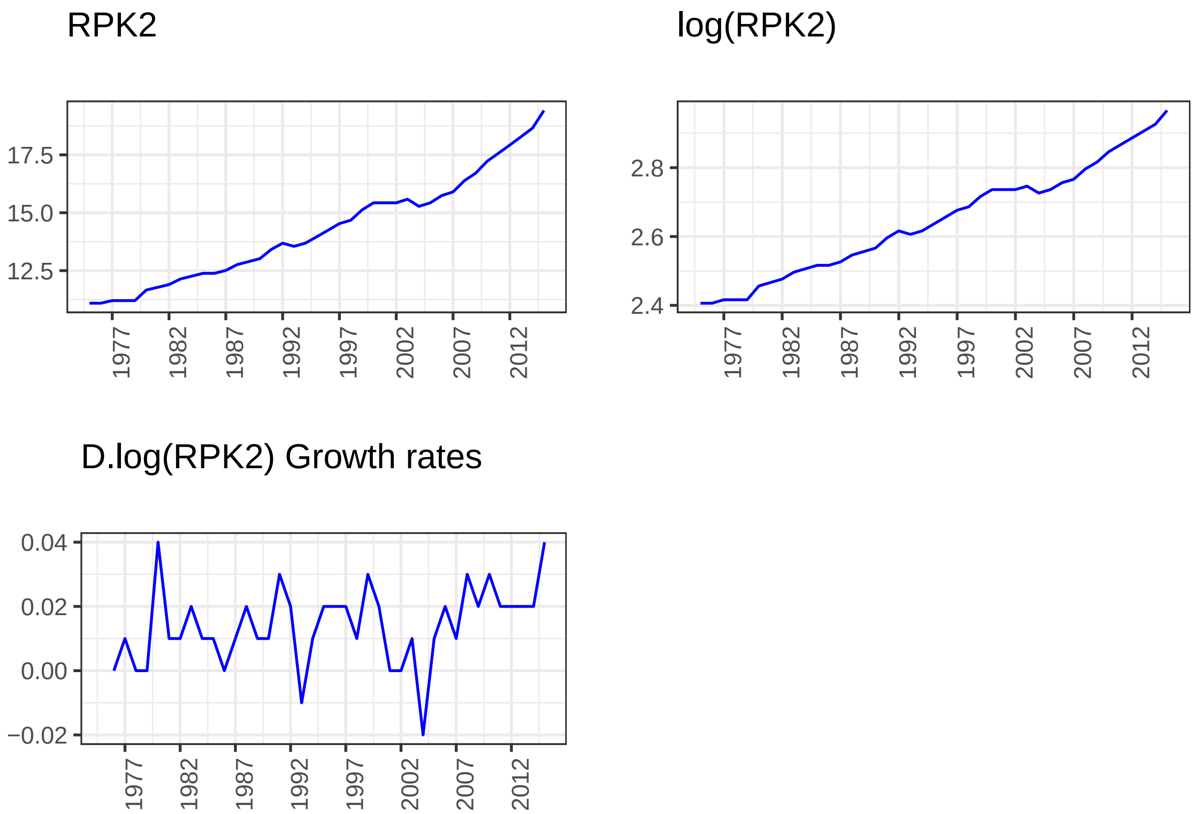
- RPK1: Revenue Passenger Kilometers of company 1 (1975-2015)
- RPK2: Revenue Passenger Kilometers of company 2 (1975-2015)
- X1: log(RPK1) (1975-2015)
- X2: log(RPK2) (1975-2015)
- DX1: first difference of X1, growth rate of RPK1 (1976-2015)
- DX2: first difference of X2, growth rate of RPK2 (1976-2015)
- Year: calendar year

Let us now look at how time series in economics and business can look like. Here is an example of passenger revenue data for an airline. The variable of interest is revenue passenger kilometers, which is the sum total over one year of the distance in kilometers traveled by each passenger on each flight of this airline company.

The left-hand graph gives the actual total number of kilometers traveled. The middle graph is obtained when taking natural logs and the right-hand graph shows the yearly growth rates.

```
# We create the log and the growth rate of both series
revenue <- revenue %>% mutate(X1=log(RPK1),DX1=c(NA,diff(log(RPK1))),X2=log(RPK2),DX2=c(NA,diff(log(RPK

plot_a <- ggplot(data=revenue, aes(x=YEAR)) +
  geom_line(aes(y=RPK2),col="blue") +
  labs(x = "", y = "", title = "RPK2",
       subtitle = ("")) +
    scale_x_date(date_breaks = "5 year", date_labels = "%Y") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1),
        legend.position = c(.5, .20),
        legend.background = element_rect(fill = "transparent")) +
  scale_color_brewer(name= NULL, palette = "Dark2")

plot_b <- ggplot(data=revenue, aes(x=YEAR)) +
  geom_line(aes(y=X2),col="blue") +
  labs(x = "", y = "", title = "log(RPK2)",
       subtitle = ("")) +
    scale_x_date(date_breaks = "5 year", date_labels = "%Y") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1),
        legend.position = c(.5, .20),
        legend.background = element_rect(fill = "transparent")) +
  scale_color_brewer(name= NULL, palette = "Dark2")

plot_c <- ggplot(data=revenue, aes(x=YEAR)) +
```

```r
  geom_line(aes(y=DX2),col="blue") +
  labs(x = "", y = "", title = "D.log(RPK2) Growth rates",
       subtitle = ("")) +
    scale_x_date(date_breaks = "5 year", date_labels = "%Y") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1),
        legend.position = c(.5, .20),
        legend.background = element_rect(fill = "transparent")) +
  scale_color_brewer(name= NULL, palette = "Dark2")

grid.arrange(plot_a, plot_b, plot_c, nrow = 2)
```



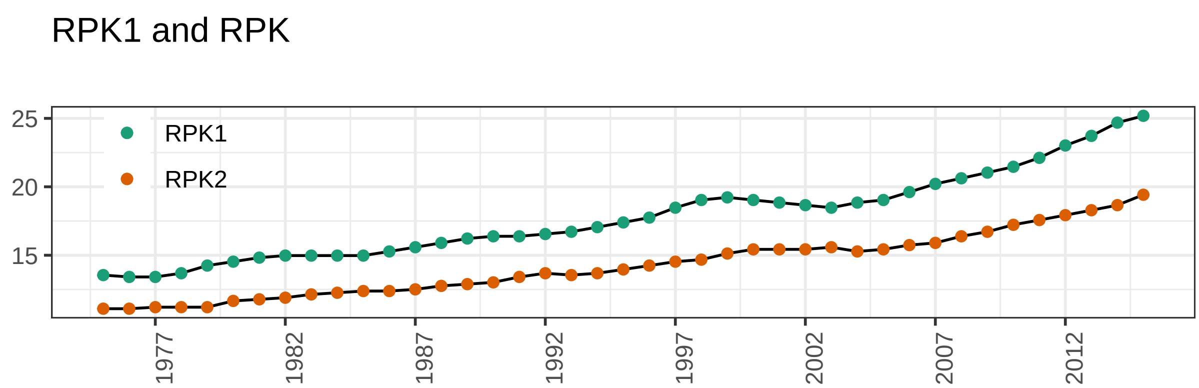
- RPK: Revenue Passenger Kilometers (in billions) yearly totals 1976-2015, trend somewhat exponential
- $log(RPK)$: more linear trend
- $D.log(RPK) = log(RPK_t) - log(RPK_{t-1}) \approx \frac{RPK_t - RPK_{t-1}}{RPK_{t-1}}$ yearly growth rate of RPK

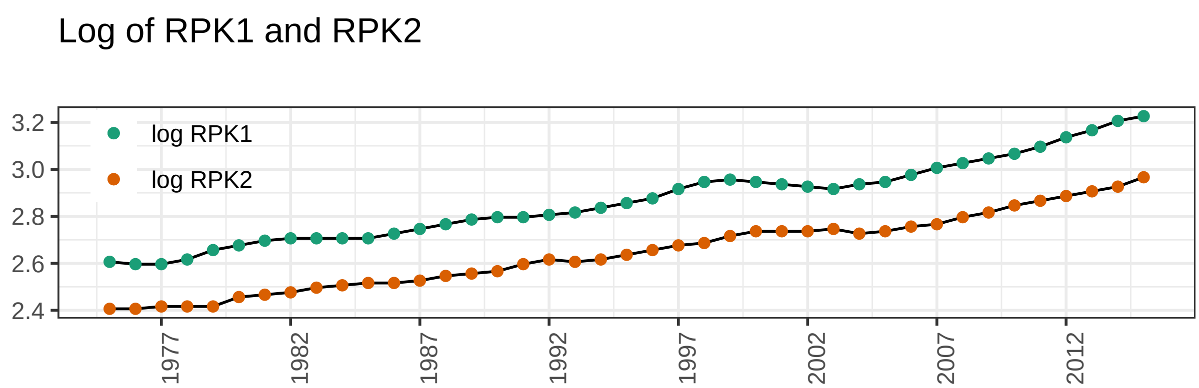
The raw data on the left seems somewhat exponentially increasing, whereas the trend for the log of the time series seems more linear. The yearly growth rates fluctuate between minus 2% and plus 4%. The two leftmost graphs show that the data have a pronounced upward trend. When this occurs, it is not reasonable to assume that the mean of the data is constant over time. In fact, the mean increases with each new observation.

In the next section, we will deal with this important issue in more detail, as for proper statistical analysis, we need data with constant mean. **A constant mean is one aspect of what we call stationarity.** For a stationary time series like in the $D.log(RPK1)$ graph here, we have a straightforward modeling strategy. But for non-stationary time series, we will first need to get rid of this non-stationarity.

This issue of trends is even more important when two time series show similar trending behavior. Look at this graph that depicts the revenue passenger kilometers of two airlines.

```r
plot_d <- ggplot(data=revenue, aes(x=YEAR)) +
  geom_line(aes(y=RPK2)) + geom_point(aes(y=RPK2,col="RPK2")) +
  geom_line(aes(y=RPK1)) + geom_point(aes(y=RPK1,col="RPK1")) +
  labs(x = "", y = "", title = "RPK1 and RPK",
       subtitle = ("")) +
    scale_x_date(date_breaks = "5 year", date_labels = "%Y") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1),
        legend.position = c(.1, .8),
        legend.background = element_rect(fill = "transparent")) +
  scale_color_brewer(name= NULL, palette = "Dark2")

plot_e <- ggplot(data=revenue, aes(x=YEAR)) +
  geom_line(aes(y=X2)) + geom_point(aes(y=X2,col="log RPK2")) +
  geom_line(aes(y=X1)) + geom_point(aes(y=X1,col="log RPK1")) +
  labs(x = "", y = "", title = "Log of RPK1 and RPK2",
       subtitle = ("")) +
    scale_x_date(date_breaks = "5 year", date_labels = "%Y") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1),
        legend.position = c(.1, .8),
        legend.background = element_rect(fill = "transparent")) +
  scale_color_brewer(name= NULL, palette = "Dark2")

grid.arrange(plot_d, plot_e, nrow = 2)
```





Clearly, they seem to have the same trend, especially when you take logs. This feature can be useful for forecasting in the following way. You may use both time series to estimate the common trend, then you can forecast the trend. And finally, derive the individual forecast for each of the airlines. In case of a single or univariate time series, you can use its own past to make forecasts. When you have several or multivariate time series like in this example, you can try to use the other series to improve your forecasts.

Example Industrial Production

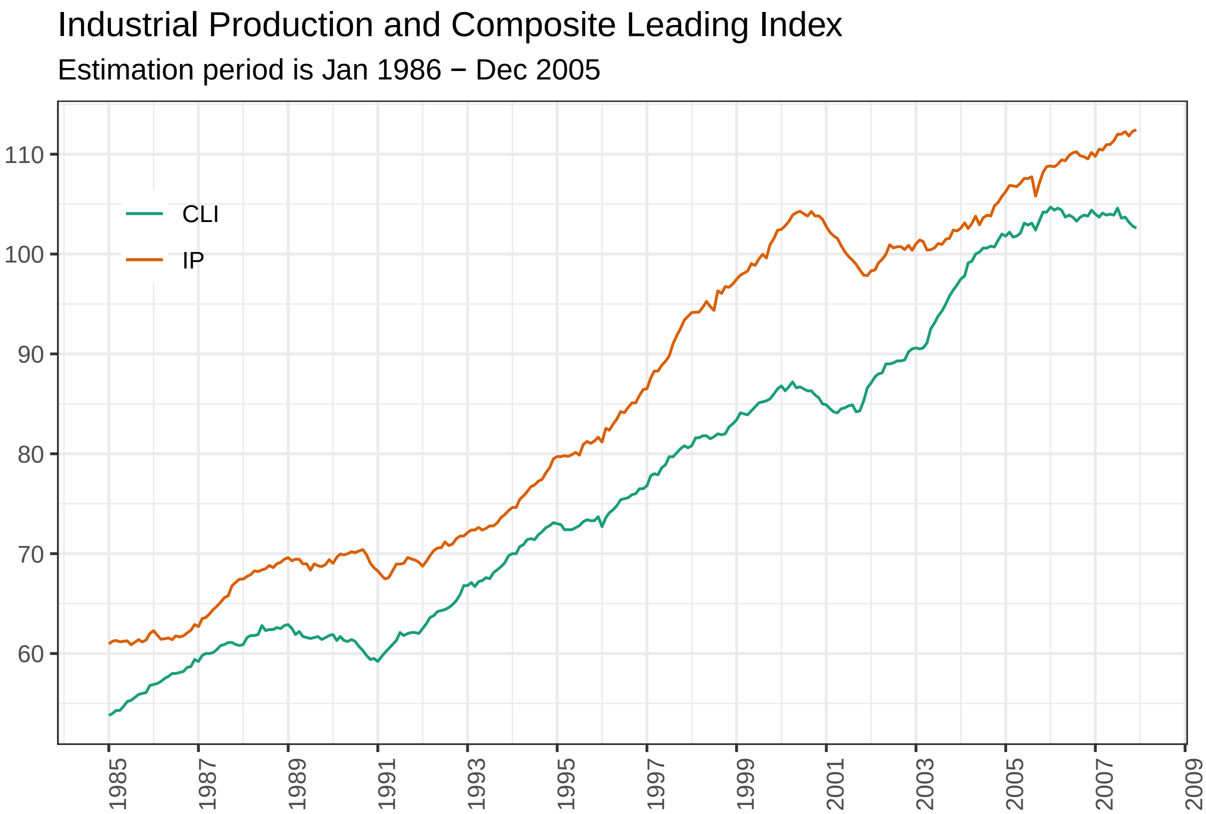
Dataset: production

Data set on Industrial Production and the Composite Leading Index for the USA, monthly data Jan 1985 - Dec 2007 (Source: Conference Board, USA). Estimation period is Jan 1986 - Dec 2005 (pre-sample values in 1985). Forecast evaluation period is Jan 2006 - Dec 2007.

- CLI: Composite Leading Index (based on 10 leading indicators)
- IP: Industrial Production (index, seasonally adjusted)
- LOGCLI: logarithm of CLI
- LOGIP: logarithm of IP
- GRCLI: monthly growthrate of CLI, first difference of LOGCLI
- GRIP: monthly growthrate of IP, first difference of LOGIP

Here is another pair of time series that are clearly related over time. These are the monthly industrial production index for the United States of America and the so-called **composite leading indicator or CLI.**

```
plot_f <- ggplot(data=production, aes(x=date)) +
  geom_line(aes(y=IP,col="IP")) +
  geom_line(aes(y=CLI,col="CLI")) +
  labs(x = "", y = "", title = "Industrial Production and Composite Leading Index ",
       subtitle = ("Estimation period is Jan 1986 - Dec 2005")) +
    scale_x_date(date_breaks = "2 year", date_labels = "%Y") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1),
        legend.position = c(.1, .8),
        legend.background = element_rect(fill = "transparent")) +
  scale_color_brewer(name= NULL, palette = "Dark2")
plot_f
```



The CLI is constructed by The Conference Board based on a set of ten variables like manufacturer's new orders, stock prices and consumer expectations. All these variables are forward looking. And therefore, they are believed to have predictive value for future macroeconomic developments. And for that reason, it may be useful to consider the CLI in case you want to forecast a variable like industrial production.

As with the airlines, the trends in industrial production and the Composite Leading Index seem to follow a similar pattern, which here associates with the business cycle. In our last section on time series, you will see if industrial production can indeed be predicted by means of this index.

Example spurious regression

Dataset: dataset_training

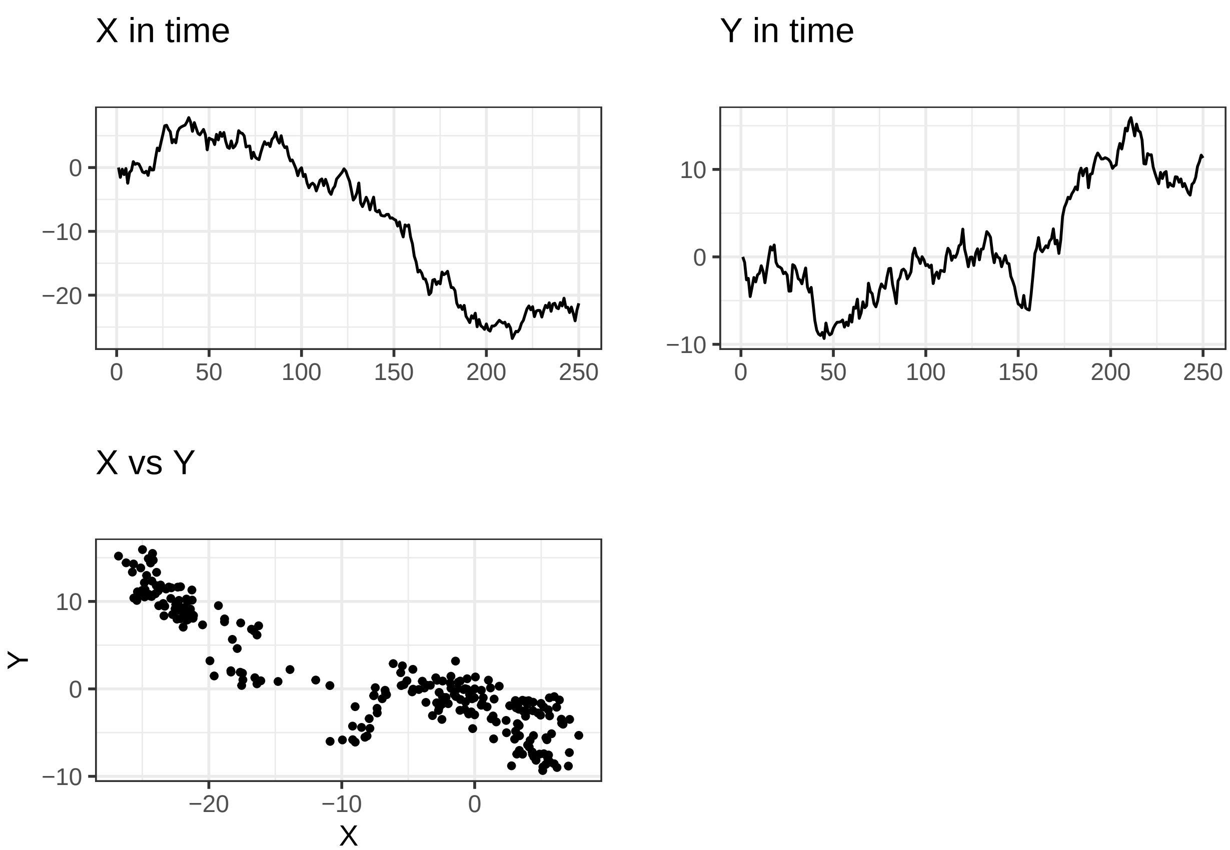
- epsx: sample of 250 values from normally and independently white noise with mean 0 and variance 1 (independent of ϵ_{yt})
- epsy: sample of 250 values from normally and independently distributed white noise with mean 0 and variance 1 (independent of ϵ_{xt})
- x: random walk generated from epsx: $x_1 = 0$, and $x_t = x_{t-1} + \epsilon_{xt}$
- y: random walk generated from epsy: $y_1 = 0$, and $y_t = y_{t-1} + \epsilon_{yt}$

The datafile contains values of four series of length 250. Two of these series are uncorrelated white noise series denoted by $\epsilon_{x,t}$ and $\epsilon_{y,t}$ where both variables are $NID(0,1)$ and $E(\epsilon_{y,t}, \epsilon_{x,s}) = 0 \forall t, s$. The other two series are so-called random walks constructed from these two white noise series by $x_t = x_{t-1} + \epsilon_{xt}$ and $y_t = y_{t-1} + \epsilon_{yt}$.

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_t The purpose of this exercise is to experience that, nonetheless, the regression of y on x indicates a highly significant relation between y and x 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.

- a) Graph the time series plot of x_t against time t, the time series plot of y_t against time t, and the scatter plot of y_t against x_t. What conclusion could you draw from these three graphs?

```r
# We add an index column to the dataset for the time t
dataset_training <- dataset_training %>% mutate(time = row_number())
plot_1 <- ggplot(data=dataset_training, aes(x=time)) +
  geom_line(aes(y=X)) +
  labs(x = "", y = "", title = "X in time",
       subtitle = ("")) +
  theme_bw() +
  theme(legend.background = element_rect(fill = "transparent")) +
  scale_color_brewer(name= NULL, palette = "Dark2")
plot_2 <- ggplot(data=dataset_training, aes(x=time)) +
  geom_line(aes(y=Y)) +
  labs(x = "", y = "", title = "Y in time",
       subtitle = ("")) +
  theme_bw() +
  theme(legend.background = element_rect(fill = "transparent")) +
  scale_color_brewer(name= NULL, palette = "Dark2")
plot_3 <- ggplot(data=dataset_training, aes(x=X,y=Y)) +
  geom_point(shape=20) +
  labs(x = "X", y = "Y", title = "X vs Y",
       subtitle = ("")) +
  theme_bw() +
  theme(legend.background = element_rect(fill = "transparent")) +
  scale_color_brewer(name= NULL, palette = "Dark2")
grid.arrange(plot_1, plot_2, plot_3, nrow = 2)
```



The two variables X,Y have completly random movements up and down. And the scatter plot seems to have a negative relation, so we could use X to forecast Y, but we know that this is not the case, the scatterplot is **misleading** in this sense.

- 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.

We use the summ() function to output our regression.

```
lm1 <- lm(EPSY ~ EPSX, data=dataset_training)
summ(lm1, digits = 3)
```

Observations	250
Dependent variable	EPSY
Type	OLS linear regression

F(1,248)	1.736
R^2	0.007
Adj. R^2	0.003

	Est.	S.E.	t val.	p
(Intercept)	0.031	0.064	0.484	0.629
EPSX	-0.088	0.067	-1.318	0.189

Standard errors: OLS

The t-value of the coefficient is around -1.32 and the p-value around 0.19, this shows that ϵ_{xt} and ϵ_{yt} have no significant relation.

- 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.

```
lm2 <- lm(EPSY ~ lag(EPSY,1) + lag(EPSY,2) + lag(EPSY,3) + EPSX + lag(EPSX,1) + lag(EPSX,2) + lag(EPSX,
summ(lm2, digits = 3)
```

Observations	247 (3 missing obs. deleted)
Dependent variable	EPSY
Type	OLS linear regression

F(7,239)	0.546
R^2	0.016
Adj. R^2	-0.013

The $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = \gamma_6 = \gamma_7 = 0$ and the degrees of freedom of F test $df = (g, n-k)$ where g is the number of parameter restrictions on the null, n is the number of observations and k is the number of variables in the unrestricted model. In this case we have:

- $g = 7$ All 7 restrictions equal to 0.
- $n = 247$ Because 3 observarions are lost because the 3 lag values, so the first available observation is in $t = 4$.
- $k = 8$ Due to the 7 coefficient plus the constant term.

	Est.	S.E.	t val.	p
(Intercept)	0.046	0.066	0.698	0.486
lag(EPSY, 1)	0.025	0.064	0.387	0.699
lag(EPSY, 2)	-0.016	0.065	-0.244	0.807
lag(EPSY, 3)	-0.047	0.064	-0.734	0.464
EPSX	-0.097	0.069	-1.405	0.161
lag(EPSX, 1)	0.020	0.070	0.284	0.777
lag(EPSX, 2)	-0.060	0.070	-0.857	0.392
lag(EPSX, 3)	0.009	0.068	0.138	0.890

Standard errors: OLS

We can see the F-statistic at the top of the output or calculate it by hand $F = \frac{(R_1^2 - R_0^2)/g}{(1-R_1^2)/(n-k)} \sim F_{(g,n-k)}$ where $R_0^2 = 0$ because is a model with only a constant term. $F = 0.55$ and as is smaller of the critical value of 2. We do NOT reject the H_0. This is correct as the value of ϵ_{yt} is independent of all other observations.

- 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?

```
lm3 <- lm(Y ~ X, data=dataset_training)
summ(lm3, digits = 3)
```

Observations	250
Dependent variable	Y
Type	OLS linear regression

F(1,248)	1090.611
R²	0.815
Adj. R²	0.814

	Est.	S.E.	t val.	p
(Intercept)	-2.487	0.214	-11.606	0.000
X	-0.515	0.016	-33.024	0.000

Standard errors: OLS

It seems by looking at the large t-value of X that X has a relevant explanatory power over Y. We know that this is not the case, so the regression is misleading, due to the trending nature of both variables. Look again at the scatterplot of a), it happens that X moves downward for long periods as Y moves upwards for long periods. This is why it seems to be a negative relation.

- 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)?

```
# We add the residuals of lm3 into the dataset
dataset_training <- dataset_training %>% mutate(lm3.res = resid(lm3))
lm4 <- lm(lm3.res ~ lag(lm3.res,1), data=dataset_training)
summ(lm4, digits = 3)
```

This coefficient is significant at 99%, this shows that the residuals are very strongly correlated. Therefore violates the standar regression assumption A7 that the error terms should be uncorrelated.

Observations	249 (1 missing obs. deleted)
Dependent variable	lm3.res
Type	OLS linear regression

F(1,247)	1457.056
R^2	0.855
Adj. R^2	0.854

	Est.	S.E.	t val.	p
(Intercept)	0.001	0.067	0.008	0.993
lag(lm3.res, 1)	0.925	0.024	38.171	0.000

Standard errors: OLS

Representing time series

Time series models typically are constructed with two main objectives. First, we want to describe the key properties of the time series data. In particular, the nature of the trend and the correlations with past values. And second, we may want to exploit these features to make forecasts of future observations.

The time series of interest is denoted as y, with the subscript t indicating the observation in period t. n refers to the number of available observations, or the length of the time series.

$$\text{Time serie :}$$
$$y_t \text{ where } t = 1, .., n \text{ is the time index.}$$

Stationarity

Let us begin with a very important issue, that of stationarity. A time series y is called **stationary, if its mean, variance, and covariances with past observations are constant over time.** Stationarity is an important condition that needs to be satisfied before we can even start thinking about designing a meaningful model for a given time series. Intuitively, if for each new observation of the time series properties like the mean and variance change, then we cannot reliably model such data, let alone provide reliable forecasts.

$$y_t \text{ is stationary if :}$$
$$mean = E(y_t) = \mu \text{ is fixed and same for all t}$$
$$autocovariance = E[(y_t - \mu)(y_{t-k} - \mu)] = \gamma_k \text{ is same for all t}$$

The autocovariances gamma k measure how strongly related observations at different points in time are. In terms of forecasting, they indicate whether past observations can be useful to make predictions of future observations.

Note that when all these autocovariances are zero, then the past carries no predictive value for the future. We call such a time series white noise.

$$\text{Special case : } \gamma = k = 0 \forall k$$
$$y_t \Rightarrow \text{ The time serie is White Noise}$$

Note that this relates to assumption A5 for regression models. A white noise time series cannot be predicted from its own past, and the only useful prediction is the mean of the variable itself.

A5. Uncorrelated error terms: $E(\epsilon_i \epsilon_j) = 0 \forall i \neq j$

Now this brings us to the **essence of time series modeling**. Our aim is to design a model that distills information from the past for forecasting. The model is deemed successful or adequate, if after all this distillation, there is nothing left that is informative for prediction. That is, the residuals are white noise.

Purpose of modeling time series: Create a time series model such that residuals are white noise.

In all what follows, we follow the usual convention to write a white noise variable as epsilon.

White noise

Uncorrelated series with mean zero : ϵ_t

Autoregressive model

Sometimes we call this ϵ_t the error but we also use the word **shock** to indicate that epsilon is something new to the variable y. A simple and popular time series model is the **autoregressive model**. An autoregression of order 1, or briefly **AR(1)**, is a model where the current observation of y in period t is explained by the previous observation of y in period t minus 1. This simple model provides a nice way to illustrate the relevance of stationarity.

$$\textbf{AR(1)}\ y_t = \alpha + \beta y_{t-1} + \epsilon_t$$

If the slope parameter β lies between -1 and +1, the effects of past shocks ϵ_t die out. So, the more distant in the past, the less impact those shocks have on current values of the variable y_t. This is a typical property of a stationary time series.

$$\textbf{Stationarity if}\ -1 < \beta < 1$$

$$y_t = \alpha + \beta y_{t-1} + \epsilon_t = \alpha + \beta(\alpha + \beta y_{t-2} + \epsilon_{t-1}) + \epsilon_t$$

$$y_t = \alpha(1+\beta) + \epsilon_t + \beta\epsilon_{t-1} + \beta^2 y_{t-2} = \alpha(1+\beta) + \epsilon_t + \beta\epsilon_{t-1} + \beta^2(\alpha + \beta y_{t-3} + \epsilon_{t-2}) = ...$$

$$y_t = \alpha \sum_{j=0}^{t-2} \beta^j + \sum_{j=0}^{t-2} \beta^j \epsilon_{t-j} + \beta^{t-1} y_1$$

$$\text{For } t \to \infty \text{ we get } \beta^{t-1} y_1 \to 0 \text{ and } y_t = \frac{\alpha}{(1-\beta)\sum_{j=0}^{\infty} \beta^j \epsilon_{t-j}}$$

Later, we will see that stationarity is lost if beta is equal to 1. The first order autoregression assumes that current y can be predicted by 1 period lagged y, but of course it might also be that 1 period lagged y and also 2 period lagged y are useful for predicting the current observation.

$$\textbf{AR(2)}\ y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \epsilon_t$$

In fact, the number of lags can run up to p, giving rise to the so-called AR(p) model.

$$\textbf{AR(p)}\ y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + +\beta_p y_{t-p} + \epsilon_t$$

Consider again the autoregression of order 1, where the current value of the white noise series epsilon is uncorrelated with the past of y.

$$\textbf{AR(1)}\ y_t = \alpha + \beta y_{t-1} + \epsilon_t$$
$$\epsilon_t \text{ uncorrelated with } y_{t-k} \forall k$$

If $\beta = 1$ then y_t can not be stationary because:

- If $\alpha \neq 0$ and $\beta = 1$, then y_t cannot have a fixed mean.

When the intercept alpha is not 0, then the mean will change by alpha for every new observation.

$$E(\epsilon_t) = 0 \text{ so } \mu = E(y_t) = \alpha + E(y_{t-1}) + 0 = \alpha + \mu \neq \mu$$

- If $\alpha = 0$ and $\beta = 1$, then y_t cannot have a fixed variance.

And when the intercept alpha is zero then the variance of the observations increases over time.

$$y_t = y_{t-1} + \epsilon_t \text{ so } (y_t - \mu) = (y_{t-1} - \mu) + \epsilon_t \text{ uncorrelated.}$$

$$E[(y_t - \mu)^2] = E[(y_{t-1} - \mu)^2] + E[\epsilon_t^2] > E[(y_{t-1} - \mu)^2]$$

Moving average model

Another useful time series model includes past shocks as explanatory variable. When you look at the epsilons as forecast errors, you can learn from these errors by taking them into account when making new forecasts. The so called **first order moving average model, or MA(1)**, includes epsilon 1 period lagged.

$$\textbf{MA(1)}\ y_t = \alpha + \epsilon_t + \gamma \epsilon_{t-1}$$

As ϵ_t is uncorrelated with it's own past and future, this model implies that y_t is correlated with y_{t-1} period lagged but not with more distant lags. y_{t-k} for $k = 2, 3...$

We can generalize this model to a moving average model of order q, which includes q lag forecast errors.

$$\textbf{MA(q)}\ y_t = \alpha + \epsilon_t + \gamma_1 \epsilon_{t-1} + ... + +\gamma_q \epsilon_{t-q}$$

ARMA model

It is also possible to combine the two models, which gives rise to an ARMA(1,1), if p and q are both equal to 1, or an ARMA(p,q), if these orders take different values.

$$\textbf{ARMA(p,q)}\ y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + \beta_p y_{t-p} + \epsilon_t + \gamma_1 \epsilon_{t-1} + ... + +\gamma_q \epsilon_{t-q}$$

Two autoregressive equations.

Moving average (MA) terms may arise when two autoregressive processes are related. Let $\epsilon_{x,t}$ and $\epsilon_{y,t}$ be be two mutually independent white noise processes, and let $y_t = \gamma x_t + \epsilon_{y,t}$ and $x_t = \delta x_{t-1} + \epsilon_{x,t}$. We can derive the orders p and q for the ARMA model for y_t (that does not include x_t).

- Consider that $y_t - \delta y_{t-1} = \gamma(x_t - \delta x_{t-1}) + \epsilon_{y,t} - \delta \epsilon_{y,t-1}$
- So we can express
  $$y_t = \delta y_{t-1} + \gamma \epsilon_{x,t} + \epsilon_{y,t} - \delta \epsilon_{y,t-1}$$
  Notice that this is an AR(1) order proces p=1, and error $w_t = \gamma \epsilon_{x,t} + \epsilon_{y,t} - \delta \epsilon_{y,t-1}$ is a MA(1) because $E(w_t w_{t-1}) = -\delta Var(\epsilon_{y,t-1})$ and $E(w_t w_{t-2}) = E(w_t w_{t-3}) = E(w_t w_{t-4}) = ... = 0$

We see that now y depends on y 1 period lagged, on the shock to x, the shock to y and, this is crucial, also the one period lagged shock to y. So the autoregressive order is one, whereas the moving average order is also one. Hence, **correlation, across joint autoregressive time series can lead to individual time series models of the ARMA type.**

(Partial) Autocorrelation Function

The time series models that we have discussed so far, autoregression and moving average and their combination, imply specific correlation properties of the time series. This relation can be reversed, that is, when you see certain properties of the data in the real world you can decide which model to use. And the **autocorrelation** is a very useful tool for this purpose.

k-th order sample autocorrelation coefficient :

$$ACF_k = cor(y_t, y_{t-k}) = \frac{\sum_{t=k+1}^{n}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=k+1}^{n}(y_t - \bar{y})^2}$$

For example, when the data are generated by a moving average model of order q, then the sample autocorrelations after lag q, will all be close to zero. For example, when the data are generated by a moving average model of order q, then the sample autocorrelations after lag q, will all be close to zero.

- If y_t is MA(q), then $ACF_k \approx 0$ for all $k > q$.

Next to autocorrelations, we also have the concept of **partial autocorrelations.** These account for the fact that the observations of y at time t y_t, and at time t minus two y_{t-2}, may seem to be correlated due to the fact that they both are related to the observation of y at t minus one y_{t-1}.

The sample partial autocorrelations follow from regressions of y on its own past values. If in this regression, the k-th lag coefficient is insignificant for values larger than p, then this suggests to use an AR model of this order p.

- k-th order sample partial autocorrelation coefficient: $PACF_k$ k is the OLS coefficient b_k in regression model:

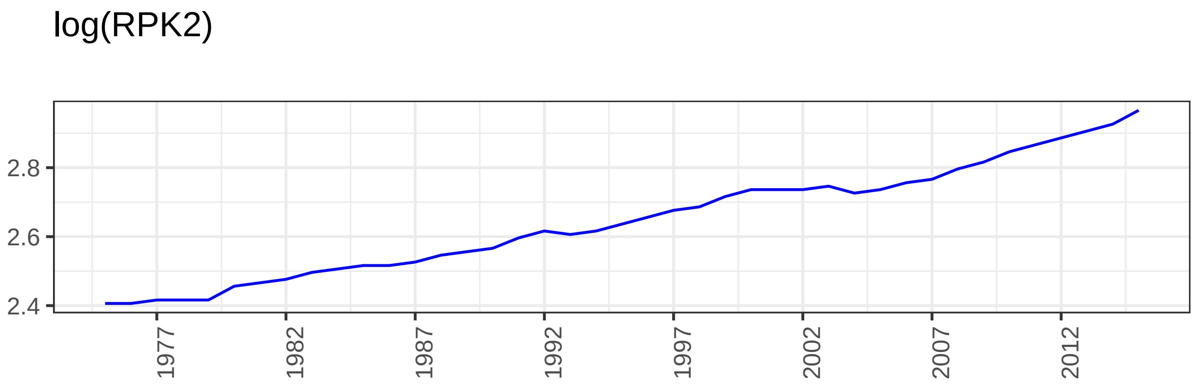
$$y_t = \alpha + \beta_1 y_{t-1} + ... + \beta_{k-1} y_{k-1} + \beta_k y_{t-k} + \epsilon_k$$

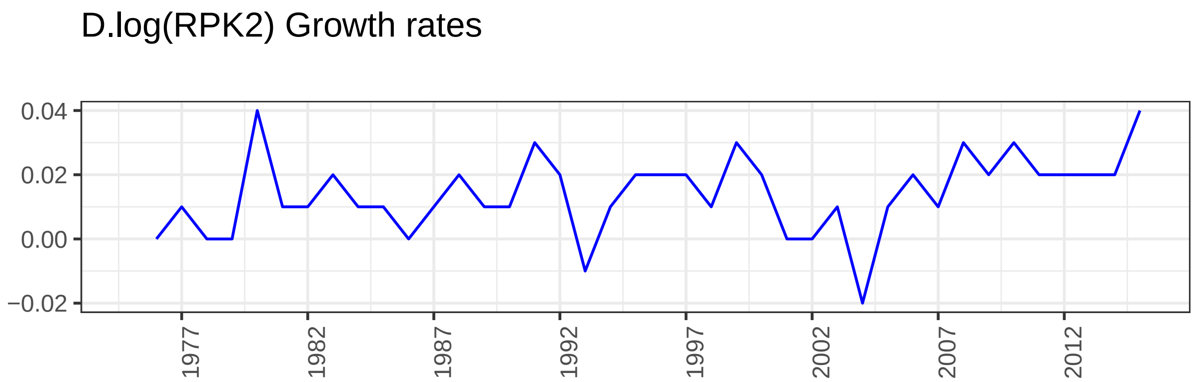
- If y_t is AR(p), then $PACF_k \approx 0$ for all $k > p$. The 95% confidence bounds around autocorrelations and partial autocorrelations are marked by plus and minus two divided by the square root of the sample size n. - 5% critical value: not significant if $-\frac{2}{\sqrt{n}} < (P)ACF < \frac{2}{\sqrt{n}}$.

Back to airlines example

Let us return now to the airline revenue passenger kilometers data. The data show a trend. And the growth rates do not.

```
grid.arrange(plot_b, plot_c, nrow = 2)
```

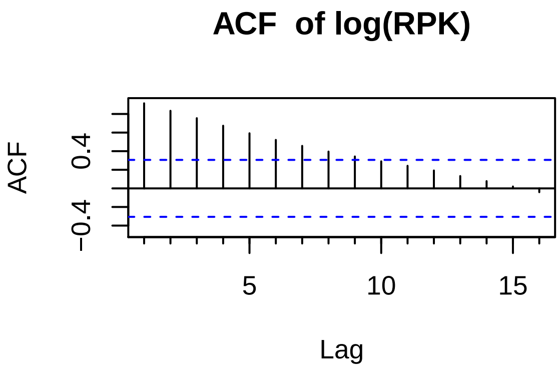


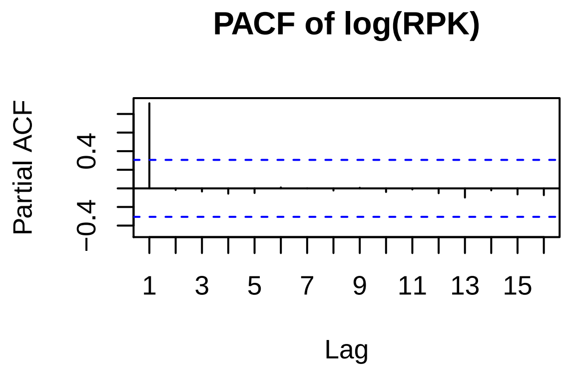


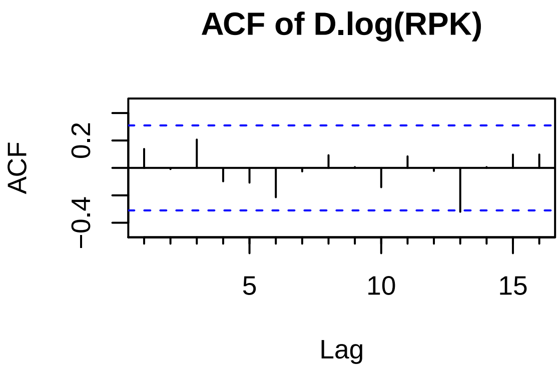
- log(RPK) is not stationary.
- first difference of log(RPK) (yearly growth rate) is stationary.

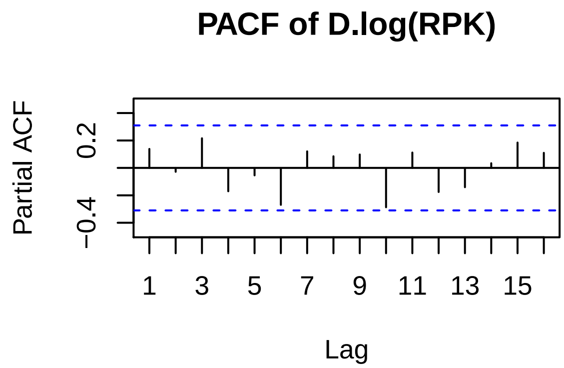
First we create the time series object in R And we use use the library forecast to plot the Acf and Pacf

```
# We create the time series object
ts1 <- ts(revenue[,-1], frequency = 1,start = 1975)
par(mfrow=c(2,2))
Acf(ts1[,4],main = "ACF  of log(RPK)")
Pacf(ts1[,4],main = "PACF of log(RPK)")
Acf(ts1[,6],main = "ACF of D.log(RPK)")
Pacf(ts1[,6],main = "PACF of D.log(RPK)")
```









The autocorrelations of the log series show a very slowly decaying pattern. And the partial autocorrelations are only large at lag 1. The values for the growth rates are not significant, as the number of observations is 39 and 2 divided by the square root of 39 is about 0.3.

Trends

Next, let us pay attention to the important issue of trends. Several trend models are available.

First, you have what is called a random walk. This is an autoregressive model, but with a slope parameter equal to 1.

$$y_t = y_{t-1} + \epsilon_t \text{ random walk, stochastic trend, no clear direction}$$

When the intercept alpha is unequal to 0, then we get trending data that looks like the airline's data, or the industrial production index considered before.

$$y_t = \alpha + y_{t-1} + \epsilon_t(\alpha \neq 0): \text{ stochastic trend}$$

When the model also contains a deterministic trend term, beta times t, then you get an explosive trend pattern.

$$y_t = \alpha + \beta t + y_{t-1} + \epsilon_t(\beta \neq 0): \text{ stochastic (explosive) trend}$$

If no lagged value of y is included, this gives a fully deterministic trend model.

$$y_t = \alpha + \beta t + \epsilon_t(\beta \neq 0): \text{ deterministic trend}$$

And if the lagged y term has a parameter smaller than 1, then this still results in a deterministic trend, but without random walk aspects.

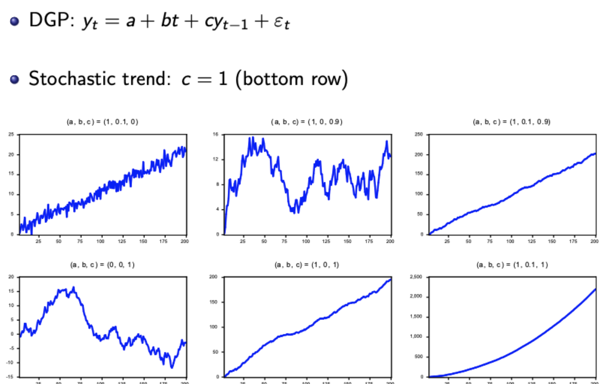
$$y_t = \alpha + \beta t + \gamma y_{t-1} + \epsilon_t (\beta \neq 0, \mid \gamma \mid < 1): \text{ deterministic trend}$$

The main notion here, is that a **stochastic trend, that is where the autoregressive parameter is equal to 1**, can only be silenced by transforming the data, by taking their first difference. This is denoted by the symbol Delta.

$$y_t = \alpha + y_{t-1} + \epsilon_t \text{ then } \Delta y_t = y_t - y_{t-1} = \alpha + \epsilon_t$$

Example deterministic and stochastic trend

To give you some visual impression of how the parameters in a time series model determine how the data will look like, consider the following graphs of artificially generated data.



Clearly the parameters matter a lot in how the data look like. This notion will be exploited in the actual analysis of real data. You look at the data and you see certain properties and from these properties, you can get ideas on the models that might be useful to fit and forecast this time series.

Cointegration

If two time series share the same stochastic trend, we say that they are cointegrated. In the next section, we will consider this in more detail.

- Sometimes: x_t and y_t each have stochastic trend, but $y_t - cx_t$ is stationary for some value of c.
- Cointegration (common stochastic trend)

Suppose that $z_t = z_{t-1} + \epsilon_{z,t}$ is unobserved, whereas $x_t = \alpha_1 + \gamma_1 z_t + \epsilon_{x,t}$ and $y_t = \alpha_2 + \gamma_2 z_t + \epsilon_{y,t}$ are observed, where $\epsilon_{z,t}, \epsilon_{x,t}, \epsilon_{y,t}$ are white noise processes. Show that x_t and y_t are cointegrated, and find the value of c for which $y_t - cx_t$ is stationary.

$$\gamma_1 y_t - \gamma_2 x_t = (\gamma_1 \alpha_2 - \gamma_2 \alpha_1) + (\gamma_1 \epsilon_{y,t} - \gamma_2 \epsilon_{x,t})$$

Where $\epsilon_t = \gamma_1 \epsilon_{y,t} - \gamma_2 \epsilon_{x,t}$ is white noise, therefore stationary.

$\gamma_1 y_t - \gamma_2 x_t = \gamma_1 (y_t - \frac{\gamma_2}{\gamma_1} x_t)$ So $c = \frac{\gamma_2}{\gamma_1}$.

We use the fact that y and x share the same stochastic trend z. A specific linear combination of y and x does not include that trend anymore.

Example of autocorrelation

If y_t is a stationary process with mean μ, then the k-th order autocovariance is defined as $\gamma_k = E[(y_t - \mu)(y_{t-k} - \mu)]$. In particular, the variance is $\gamma_0 = E(y_t - \mu)^2$. The k-th order autocorrelation is defined as $\rho_k = \frac{Cov(y_t, y_{t-k})}{Var(y_t)}$.

The AR(1) model is:

$$y_t = \alpha + \beta y_{t-1} + \epsilon_t$$

We know $E(\epsilon_t) = 0$ and ϵ_t is uncorrelated with all values of y_s for all $s < t$.

- a) Show that the mean of the AR(1) model is equal to $\mu = \frac{\alpha}{1-\beta}$

The expected value of y_t is equal to $E(y_t) = E(\alpha + \beta y_{t-1} + \epsilon_t) = \alpha + \beta E(y_{t-1}) + E(\epsilon_t) = \mu = \alpha + \beta\mu + 0$
And that means that $\beta \neq 1$

$$\mu = \frac{\alpha}{1-\beta}$$

- b) Define $z_t = y_t - \mu$. Show that $z_t = \beta z_{t-1} + \epsilon_t$ and that $Var(z_t) = \frac{\sigma^2}{(1-\beta^2)}$.

In part a) we see that $\alpha = \mu(1-\beta) = \mu - \beta\mu$. If we substitute this in the AR(1) equation we get:

$$y_t = \mu - \beta\mu + \beta y_{t-1} + \epsilon_t$$
$$y_t - \mu = \beta(y_{t-1} - \mu) + \epsilon_t$$

And as we can define $z_t = y_t - \mu$

$$z_t = \beta z_{t-1} + \epsilon_t$$

The expected value of $E(z_t) = E(y_t - \mu) = 0$ and the variance of $Var(z_t) = E(z_t - E(z_t))^2 = E(z_t^2)$ So if we use the definition of z_t we obtain:

$$Var(z_t) = E(z_t^2) = E((\beta z_{t-1} + \epsilon_t)^2) = \beta^2 E(z_{t-1}^2) + E(\epsilon_t^2) + 2\beta E(z_{t-1}\epsilon_t)$$

We use the fact that $E(z_{t-1}\epsilon_t) = 0$

$$Var(z_t) = \beta^2 Var(z_t^2) + \sigma^2 + 0$$

$$Var(z_t) = \frac{\sigma^2}{(1-\beta^2)}$$

That means that that $-1 < \beta < 1$

- c) Use the idea of part (b) to show that the autocorrelations of y_t are equal to $\rho_k = \beta^k$

Because y_t and z_t have the same autocovariances, we compute those of z_t, it is simplier because the mean of z_t is zero. The first order autocovariance γ_1

$$\gamma_1 = E(z_t z_{t-1}) = E(\beta z_{t-1} + \epsilon_t)z_{t-1} = \beta E(z_{t-1}^2) + E(\epsilon_t z_{t-1}) = \beta\gamma_0 + 0 = \beta\gamma_0$$

Hence the first order autocorrelation is:

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \beta$$

And the second order autovoariance γ_2 is equal:

$$\gamma_2 = E(z_t z_{t-2}) = E(\beta z_{t-1} + \epsilon_t)z_{t-2} = \beta E(z_{t-1}z_{t-2}) + E(\epsilon_t z_{t-2}) = \beta\gamma_1 + 0 = \beta(\beta\gamma_0) = \beta^2\gamma_0$$

Hence the second order autocorrelation is:

$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \beta^2$$

And similarly, the k-th order autocovariance is equal to:

$$\gamma_k = E(z_t z_{t-k}) = E(\beta z_{t-1} + \epsilon_t) z_{t-k} = \beta E(z_{t-1} z_{t-k}) + E(\epsilon_t z_{t-k}) = \beta\gamma_{k-1} + 0 = \beta\gamma_{k-1}$$

Which is equal to:

$$\gamma_k = \beta^k \gamma_0$$

And the k-th order autocorrelation:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \beta^k$$

- d) Argue that stationarity requires that $-1 < \beta < 1$

The correlations are always $-1 < \beta < 1$ so $\mid \beta \mid \leq 1$, furthermore $\beta = 1$ is exclueded in a) and $\beta = -1$ excluded in part b).

Specification and estimation

In this section, you will learn which steps to take to specify time series models, and to estimate parameters in such models. **Stationarity is crucial here**. And therefore, you should take care of any non-stationarity right at the start. Once a stationary series is obtained after proper transformation, you can use the autocorrelation (AC) and partial autocorrelation (PAC) functions to specify a first-guess model.

We start with the major motivation for using time series, that is, forecasting. Forecasts are based on a model that properly summarizes past information.

$$\text{Past values of time series model} \rightarrow \text{ Model } \rightarrow \text{ Forecast future values}$$

This past information can concern the own past of the dependent variable y, or possibly also the past of an explanatory factor, x. For notational convenience, we use PY and PX to denote this past information.

- In a univariate model, the forecast is based only on the past of the dependent variable y itself.
- And if we use also an explanatory factor x, then the forecast is a function of the past of both x and y.

$$\begin{aligned} &\text{Notation :} \\ &\qquad y_t : \text{ time series of interest } (t = 1, ..., n) \\ &\qquad x_t : \text{ time series possible explanatory factor (restrict to one)} \\ &\qquad Py_{t-1} : [y_{t-1}, y_{t-2}, ..., y_1] \text{ past information on y at time t} \\ &\qquad Px_{t-1} : [x_{t-1}, x_{t-2}, ..., x_1] \text{ past information on x at time t} \end{aligned}$$

$$\begin{aligned} \text{Univariate time series forecast model } &: \hat{y}_t = F(Py_{t-1}) \\ \text{Forecast model with explanatory factor } &: \hat{y}_t = F(Py_{t-1}, Px_{t-1}) \end{aligned}$$

Of course, we want to use the past information in an optimal way, such that there is no predictive value anymore in the errors that we make. Indeed, we wish to arrive at a forecast error that is uncorrelated with the information in PY and PX.

- Aim: Optimal use of past information to get best forecasts.
- Wish: Forecast error $\epsilon_t = y_t - \hat{y}_t$ uncorrelated with past information.

Note that if the forecast error would be predictable, then some relevant information is still missing that could be used to improve the forecast.

Univariate time series model

Let us start with a univariate time series model. Here the forecast for y is a function of past observations of y only.

$$\text{Univariate time series forecast model: } : \hat{y}_t = F(Py_{t-1})$$

Now we first have to decide on the function F. And although many functions F can be chosen, a popular choice is the **linear function**. When the lagged information in PY is limited at lag p, the well-known autoregressive model of order p emerges. AR(p).

The true value of y is the forecast plus an error term that is equal to the forecast error. And together, this gives rise to the AR(p) model.

$$\hat{y}_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + +\beta_p y_{t-p}$$

$$y_t = \hat{y}_t + \epsilon_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + \beta_p y_{t-p} + \epsilon_t$$

Because ϵ_t is white noise that is, the future values of the epsilons cannot be predicted in a linear way. This can be seen in the following way:

Consider forecasts from an autoregression of order p, which says that there is useful information in the past until, and including, p observations ago $\hat{y}_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + +\beta_p y_{t-p}$. And consider the forecast error epsilon, which should be uncorrelated with the past of y. $\epsilon_t = y_t - \hat{y}_t$ uncorrelated with uncorrelated with y_s for all s < t.

You can show that, in this situation, the epsilon process is white noise, that is, the future values of the epsilons cannot be predicted in a linear way.

- Without loss of generality, consider case s < t.
- $\epsilon_s = y_s - \alpha - \sum_{j=1}^{p} \beta_j y_{s-j}$ is a linear function of y_r, $r \leq s < t$
- e_t is uncorrelated with y_r for all $r < t$, so also uncorrealted with ϵ_s.

The crucial step here is that **epsilon is a linear function of current and past values of the dependent variable.**

Estimation of AR and ARMA

To estimate the parameters in an **AR model of order p**, we use the same ideas as in linear regression, where an optimal strategy is to **minimize the sum of squared errors**, which, of course, now are the forecast errors. So, you can use ordinary least squares.

$$\text{Forecast error: } : \epsilon_t = y_t - \hat{y}_t = y_t - \alpha - \sum_{j=1}^{p} \beta_j y_{t-j}$$

$$\textbf{To minimize via OLS: } : \sum_{t=p+1}^{n} \epsilon_t^2$$

As a moving average model **MA(q) includes also lagged forecast errors that are still unknown before estimation**, we have to resort to the method of **maximum likelihood in case of ARMA models**.

ADL(p,r) model

The usefulness of least squares extends to the case where the time series model also includes the past of an explanatory factor x. And again, here the popular choice is to use a linear forecast function.

$$\text{Forecast: } : \hat{y}_t = F(Py_{t-1}, Px_{t-1})$$

$$\text{find F such that } \epsilon_t = y_t - \hat{y}_t \text{ uncorrelated with } Py_{t-1}, Px_{t-1}$$

And this model that includes the lags of y and also the lags of x, when there are p lags of y, and r lags of x, we usually call this the autoregressive distributed lag model of order p and r, or shortly **ADL(p,r)**.

$$\text{ADL(p,r) } : \hat{y}_t = \alpha + \beta_1 y_{t-1} + ... + \beta_p y_{t-p} + \gamma_1 x_{t-1} + ... + \gamma_r x_{t-r}$$

Also for this ADL model, we can use the least squares method to estimate the parameters.

$$\textbf{To minimize via OLS: } : \sum_{t=m+1}^{n} \epsilon_t^2 \text{ where m=max(p,r)}$$

Granger causality

The ADL model is particularly useful to examine what is called Granger causality, named after the Nobel Laureate Sir Clive Granger. This idea of causality builds on the idea of forecastability. That is, when the past of one variable is helpful to predict the future of another, you might consider that as some form of causality.

What you do is to construct two ADL models. One for the variable y, and another for the variable x. Note that both models include the past of the dependent variable itself plus the past of the other variable.

$$y_t = \alpha + \sum_{j=1}^{p} \beta_j y_{t-j} + \sum_{j=1}^{r} \gamma_j x_{t-j} + \epsilon_t$$

$$x_t = \alpha^* + \sum_{j=1}^{p^*} \beta_j^* x_{t-j} + \sum_{j=1}^{r^*} \gamma_j^* y_{t-j} + \epsilon_t^*$$

-x_t helps to predict y_t if $\gamma_j \neq 0$ for some j -y_t helps to predict x_t if $\gamma_j^* \neq 0$ for some j

If some of the gamma parameters in the ADL model for y are different from 0, then the past of x helps to predict the future of y. And vice versa, in case the gamma star parameters are non-zero in the ADL model for x.

In case of such of non-zero parameters, you can say that **one variable is Granger causal to the other**. For example, we may find that x is Granger causal for y, but not the other way around. This indicates that the past of x can be helpful for predicting y. But for forecasting x, only its own past is relevant.

- x_t is Granger causal for y_t if it helps to predict y_t, whereas y_t does not help to predict x_t.

You can check the significance of, for example, the γ_j^* coefficients in the ADL model for x by means of the familiar F-test. And, conveniently, as these two models only include lag variables, you can still estimate the parameters using least squares for each of the two equations separately.

$$\text{Test } H_0 : \gamma_j^* = 0 \forall j = 1, ..., r^* \sim F - test$$

Consequences of non-stationarity

The first thing that needs to be done in modeling is to make sure that the time series you wish to analyze is stationary. The reason that we need stationarity is that this is required for proper statistical analysis. So, one should first somehow test if the variable of interest is stationary.

- Regression assumption A2 not satisfied: regressors y_{t-j} are random.
- Standard OLS t- and F-tests hold true in large enough samples provided all variables in equation are stationary.

So, one should first somehow test if the variable of interest is stationary. To do this, we can again make use of a time series model. For example, in the case of an autoregression of order 1, we know that when the parameter is equal to 1, then the time series is not stationary. And this suggests that we can test for stationarity by testing the value of this parameter.

$$\begin{aligned}&\text{Test for stationarity}:\\ &\textbf{AR(1)}\ y_t = \alpha + \beta y_{t-1} + \epsilon_t \text{ test } H_0 : \beta = 1 \text{ against} H_1 : -1 < \beta < 1\end{aligned}$$

As we are familiar with statistical tests for parameters to be equal to 0, we usually rewrite the AR(1) model by subtracting the one period lagged y_{t-1} from both sides of the equation.

$$\begin{aligned}&\text{Rewrite AR(1)}:\\ &\Delta y_t = y_t - y_{t-1} = \alpha + (\beta - 1)y_{t-1} + \epsilon_t = \alpha + \rho y_{t-1} + \epsilon_t\\ &\text{Where } \rho = (\beta - 1)\end{aligned}$$

Unit root test

Now, the ρ parameter can be tested to be equal to 0 using a t-test. As we are interested in the parameter rho being 0, or smaller than 0, we reject non-stationarity when the t- ratio is more negative than minus 2.9. Note, that this is not the usual value of $t_{0.95} = -1.65$. And this is due to the fact that **under the null hypothesis, y is non-stationary, which gives rise to a different statistical theory.**

$$\begin{aligned}&\text{Test for stationarity}:\\ &\Delta y_t = \alpha + \rho y_{t-1} + \epsilon_t \text{ test } H_0 : \rho = 0 \text{ against } H_1 : \rho < 0\\ &\text{Reject } H_0 : \text{non-stationarity if } t_{\hat{\rho}} < -2.9\end{aligned}$$

For an AR(2) model we follow the same process, it is convenient to write the AR(2) model as a mixture of variables in first differences and in levels.

$$\begin{aligned}&\text{Rewrite AR(2)}: \ y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \epsilon_t\\ &\Delta y_t = y_t - y_{t-1} = \alpha + (\beta_1 - 1)y_{t-1} + \beta_2 y_{t-2} + \epsilon_t\\ &\text{Sum and substract } \beta_2 y_{t-1}\\ &\Delta y_t = \alpha + (\beta_1 + \beta_2 - 1)y_{t-1} - \beta_2 y_{t-1} + \beta_2 y_{t-2} + \epsilon_t\\ &\Delta y_t = \alpha + (\beta_1 + \beta_2 - 1)y_{t-1} - \beta_2(y_{t-1} - y_{t-2}) + \epsilon_t\\ &\Delta y_t = \alpha + (\beta_1 + \beta_2 - 1)y_{t-1} - \beta_2 \Delta y_{t-1} + \epsilon_t\\ &\text{we get}:\\ &\Delta y_t = \delta + \rho y_{t-1} + \gamma \Delta y_{t-1} + \epsilon_t\\ &\text{Where } \delta = \alpha, \rho = (\beta_1 + \beta_2 - 1), \gamma = -\beta_2\end{aligned}$$

Augmented Dickey Fuller

This rewriting of the AR(p) models provides the basis of the so-called Dickey-Fuller test. The test equation can either include a deterministic trend, or not, and the choice is usually based on the visual impression of the data.

The inclusion or exclusion of the deterministic trend term, βt, matters for the relevant 5% critical value. When the trend term is not included, the critical value for the t-test on rho is $t_{\hat{\rho}} < -2.9$, as we saw before. But when the trend is included, it becomes $t_{\hat{\rho}} < -3.5$.

$$\text{If data NO clear trend direction :}$$
$$\Delta y_t = \alpha + \rho y_{t-1} + \gamma_1 \Delta y_{t-1} + ... + \gamma_L \Delta y_{t-L} + \epsilon_t$$

$$\text{Test without deterministic trend}$$
$$\text{Reject } H_0 : \text{non-stationarity if } t_{\hat{\rho}} < -2.9$$

$$\text{If data has clear trend direction :}$$
$$\Delta y_t = \alpha + \beta t + \rho y_{t-1} + \gamma_1 \Delta y_{t-1} + ... + \gamma_L \Delta y_{t-L} + \epsilon_t$$

$$\text{Test without deterministic trend}$$
$$\text{Reject } H_0 : \text{non-stationarity if } t_{\hat{\rho}} < -3.5$$

In practice, we decide on the number of lags in the autoregression by testing for correlation in the residuals, or by using a model selection criteria.

- Choice lag L: serial correlation check, or AIC/BIC. See (Partial) Autocorrelation Function.

When the autoregression has more than one lag, the Dickey-Fuller test is usually called the Augmented Dickey-Fuller test, abbreviated as ADF.

Summary

So, now, how should you proceed to specify a time series model?

1. First you perform an ADF test. And when you can reject a unit root, or non-stationarity, this means that y_t is stationary, and you can model the series y_t without further transformation. But when it is not rejected, you should take the first difference, and continue with Δy_t.
2. Next, you can use OLS to estimate the parameters in an autoregression AR(p), or when you wish to consider an autoregressive distributed lag model ADL(p,r), you perform unit root tests for both series y and x, and proceed with levels or with first differences.

There is, however, one exceptional, and practically relevant case, namely, when y and x are not stationary, **but a linear combination of the two variables is.** When y_t and x_t are cointegrated.

Cointegration

Two variables are called cointegrated when a linear combination is stationary. This can only occur when y_t and x_t have the same stochastic trend.

$$y_t, x_t \text{ are are cointegrated if both series are non-stationary, but a linear combination is stationary : } y_t - cx_t$$

You might, then, interpret this linear combination $y_t = cx_t$ as the **long-run equilibrium**, which is an attractive concept in economics.

Test for cointegration

A simple test for cointegration is the Engle-Granger test. This test amounts to regressing y_t on an intercept and x_t to estimate the long-run equilibrium relationship between these variables.

$$\begin{aligned} &\text{Engle-Granger test :} \\ &\qquad \text{Step 1 :} \quad \text{OLS in } y_t = \alpha + \beta x_t + \epsilon_t \to b \text{ and residuals } e_t \\ &\qquad \text{Step 2 :} \quad \text{Cointegrated if ADF test on } e_t \text{ rejects non-stationarity} \\ &\qquad\qquad \Delta e_t = \alpha + \rho e_{t-1} + \gamma_1 \Delta e_{t-1} + ... + \gamma_L \Delta e_{t-L} + w_t \end{aligned}$$

The residuals from this regression can then be interpreted as deviations from the equilibrium, and if the equilibrium relation actually exists, that is if y_t and x_t actually are cointegrated, the residuals should be stationary. And this can be examined, again, using the ADF test as before.

Because the test is now applied to residuals instead of an actually observed time series, the distribution of the ADF test is different. The 5% critical value for the relevant t-test is now minus 3.4, or even minus 3.8 if the trend term is included.

$$\begin{aligned} &\text{Engle-Granger test :} \\ \Delta e_t = \alpha + (\beta t) + \rho e_{t-1} &+ \gamma_1 \Delta e_{t-1} + ... + \gamma_L \Delta e_{t-L} + w_t \\ &\text{Test (with) or without deterministic trend} \\ \text{Reject } H_0 : \text{non-stationarity} &= \text{cointegrated if } t_{\hat{\rho}} < -3.4 \text{ or } (-3.8) \end{aligned}$$

In case of cointegration, the ADL model can be written in the so-called error correction format, which includes lagged difference variables and the stationary linear combination between y_t and x_t. The term error correction follows from the notion that deviations from the long-run equilibrium get corrected by the β_1 parameter in forecasting the changes of y_t.

$$\text{ECM : if } x_t, y_t \text{ cointegrated, estimate :}$$

$$\Delta y_t = \alpha + \beta_1 (y_{t-1} - b x_{t-1}) + \beta_2 \Delta y_{t-1} + \beta_3 \Delta x_{t-1} + \epsilon_t$$

$$\text{Or more lags for } \Delta y_t, \Delta x_t$$