

05 Regression with Time Series Variables

Andrius Buteikis, andrius.buteikis@mif.vu.lt
<http://web.vu.lt/mif/a.buteikis/>

Introduction

We have defined the autoregressive distributed lag $ADL(p, q)$ model as

$$\phi(L)Y_t = \alpha + \delta t + \beta(L)X_t + \epsilon_t$$

where

$$\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$$

$$\beta(L) = \beta_0 + \beta_1 L + \dots + \beta_q L^q$$

Estimation and interpretation of the $ADL(p, q)$ model depends on whether the series X and Y are stationary or not.

In this lecture we will assume that **both** X and Y have the same stationarity properties: either **both** are *stationary*, or **both** have a *unit root*.

If the properties of X differ from those of Y , it becomes difficult for X to explain Y . For instance, it is hard for a stationary series to explain the stochastic trend variation of a unit root series. In practice this means that, before running any time series regression, you should examine the univariate properties of the variables you plan to use.

In particular, you should carry out unit root tests along the lines described in the previous lectures for every variable in your analysis.

Time Series Regression when Y and X are Stationary

This section extends the topic presented in the previous lecture.

For the $AR(p)$ model, it proved convenient for both the OLS estimation and interpretation to rewrite the model with ΔY as the dependent variable. Similar considerations hold for the $ADL(p, q)$ model:

$$Y_t = \alpha + \delta t + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \beta_0 X_t + \dots + \beta_q X_{t-q} + \epsilon_t$$

which can be rewritten as:

$$\begin{aligned} \Delta Y_t &= \alpha + \delta t + \rho Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \dots + \gamma_{p-1} \Delta Y_{t-p+1} \\ &\quad + \theta X_t + \omega_1 \Delta X_t + \dots + \omega_q \Delta X_{t-q+1} + \epsilon_t \end{aligned}$$

where

$$\begin{aligned} \rho &= \phi_1 + \dots + \phi_p - 1 \\ \gamma_j &= -[\phi_{j+1} + \dots + \phi_p], \quad j = 1, \dots, p-1 \\ \theta &= \beta_0 + \beta_1 + \dots + \beta_q \\ \omega_i &= -[\beta_i + \dots + \beta_q], \quad i = 1, \dots, q \end{aligned}$$

It should be emphasized that this model is the same as that in the original form of the $ADL(p, q)$ - it has merely undergone a few algebraic manipulations. The second form is often more convenient than the first because it allows us to easier test the unit root hypothesis.

This model may look complicated, but it is still, nevertheless, just a regression model. That is, no new mathematical techniques are required for this model, which is, after all, still based on the simple equation of a straight line.

Example ADL(1, 1)

$$Y_t = \alpha + \delta t + \phi_1 Y_{t-1} \pm Y_{t-1} + \beta_0 X_t \pm \beta_1 X_t + \beta_1 X_{t-1} + \epsilon_t$$

yields:

$$\Delta Y_t = \alpha + \delta t + (\phi_1 - 1) Y_t + (\beta_0 + \beta_1) X_t - \beta_1 \Delta X_t + \epsilon_t$$

Example ADL(2, 2)

$$Y_t = \alpha + \delta t + \phi_1 Y_{t-1} \pm Y_{t-1} + \phi_2 Y_{t-2} \pm \phi_2 Y_{t-1} \\ + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} \pm \beta_2 X_{t-1} \pm (\beta_1 + \beta_2) X_t + \epsilon_t$$

yields:

$$\Delta Y_t = \alpha + \delta t + (\phi_1 + \phi_2 - 1) Y_{t-1} - \phi_2 \Delta Y_{t-1} \\ + (\beta_0 + \beta_1 + \beta_2) X_t - (\beta_1 + \beta_2) \Delta X_t - \beta_2 \Delta X_{t-1} + \epsilon_t$$

The coefficients of both models have their usual interpretation, however in economics the multiplier concept is more popular.

Coefficient and Short-run multiplier interpretation

We have already discussed how to interpret regression coefficients, placing special emphasis on *ceteris paribus* conditions. Recall that we made statements of the form:

The β_0 coefficient measures the influence of X_k on Y_k , *ceteris paribus*.

In the $ADL(p, q)$ model, such an interpretation can still be made, *but it is not that commonly done*.

In economics, a common way to interpret the coefficients is through the concept of a multiplier. Thus:

β_0 is called the **short-run** or **impact multiplier** it describes how a one unit size *one-moment* (i.e. **temporary**) increase of X_t affects Y_t .

However, it is common to focus on the **long-run**, or **total, multiplier**.

Long-run Multiplier Interpretation

Suppose that X and Y are in an **equilibrium** (steady state), i.e. they are not changing over time.

All of a sudden, X changes **permanently** to a new level, which is one unit higher than its previous value. This affects Y , which starts to change and **eventually** settling down *in the long-run to a new equilibrium value*.

The difference between the old and the new equilibrium values for Y can be interpreted as the **long-run** effect of X on Y and is the **long-run multiplier**.

This multiplier is often of great interest for policymakers who want to know the eventual effects of their policy changes in various areas. The long run multiplier measures the effect of this sort of change.

Let $\mathbb{E}(Y_t) = \mu_Y$ and $\mathbb{E}(X_t) = \mu_X$. By taking the expectations of both sides, it is possible to calculate the **long-run multiplier for the ADL(p, q) model**:

$$\begin{aligned}\mathbb{E}(Y_t) &= \alpha + \delta t + \phi_1 \mathbb{E}(Y_{t-1}) + \dots + \phi_p \mathbb{E}(Y_{t-p}) + \beta_0 \mathbb{E}(X_t) + \dots + \beta_q \mathbb{E}(X_{t-q}) \\ (1 - \phi_1 - \dots - \phi_p) \mu_Y &= \alpha + \delta \cdot t + (\beta_0 + \beta_1 + \dots + \beta_q) \mu_X\end{aligned}$$

The equilibrium model is then:

$$\mu_Y = \frac{\alpha}{1 - \phi_1 - \dots - \phi_p} + \frac{\delta}{1 - \phi_1 - \dots - \phi_p} \cdot t + \frac{\beta_0 + \beta_1 + \dots + \beta_q}{1 - \phi_1 - \dots - \phi_p} \cdot \mu_X$$

Now, if the process transfers to the new permanent level, say $X_{t+1} = X_{t+2} = \dots = X + 1$, this will change the equilibrium value of Y . The change of Y is called the **long-run** or total multiplier and it equals:

$$\frac{\beta_0 + \beta_1 + \dots + \beta_q}{1 - \phi_1 - \dots - \phi_p} = -\frac{\theta}{\rho}$$

In other words, only the coefficients on X_t and Y_{t-1} will matter for the long-run behavior in the rewritten ADL model for ΔY_t .

Marginal Effect Interpretation

In some cases, you might be interested in the effect of a temporary change in X (i.e. X starts at some original level, then increases by one unit for one period before going back to the original level the next). The **long-run multiplier does not measure the effect of this type of change**.

We can use the traditional 'marginal effect' interpretation of regression coefficients for such temporary changes (i.e. the short-run multiplier).

ADL(1,1) Example

To be more specific in distinguishing different multipliers we shall analyse the $ADL(1, 1)$ model in more detail (with assumption that $|\phi| < 1$). The model:

$$Y_t = \alpha + \phi Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t$$

can be rewritten in terms of X as follows:

$$(1 - \phi L) Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t$$

or:

$$\begin{aligned} Y_t &= (1 - \phi L)^{-1} (\alpha + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t) \\ &= (1 + \phi L + (\phi L)^2 + \dots) (\alpha + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t) \\ &= \frac{\alpha}{1 - \phi} + \beta_0 X_t + (\beta_0 \phi + \beta_1) X_{t-1} \\ &\quad + \phi (\beta_0 \phi + \beta_1) X_{t-1} + \phi^2 (\beta_0 \phi + \beta_1) X_{t-2} + \dots + \nu_t \\ &= \tilde{\delta} + \delta_0 X_t + \dots + \delta_s X_{t-s} + \dots + \nu_t \end{aligned}$$

$$Y_t = \tilde{\delta} + \delta_0 X_t + \dots + \delta_s X_{t-s} + \dots + \nu_t$$

- ▶ The coefficient $\delta_0 = \beta_0$ at X_t (or $\partial Y_t / \partial X_t$) is called the **short-run multiplier**.
- ▶ The coefficient $\delta_s = \phi^s(\beta_0\phi + \beta_1)$ at X_{t-s} (or $\partial Y_{t+s} / \partial X_t$) is the **s-period delay multiplier**.
- ▶ The value $\sum_{i=1}^s \delta_i$ is the **s-period interim multiplier**.
- ▶ $\sum_{i=0}^{\infty} \delta_i = \frac{\beta_0 + \beta_1}{1 - \phi}$ is the total or **long-run multiplier**.

In other words:

- ▶ The **short-run** multiplier is the *immediate* Y_t 's response to a momentary change in X : $X_1 = \dots = X_{t-1} = X$, $X_t = X + 1$ (i.e. X increases by one in period t);
- ▶ The **long-run** multiplier is the *eventual* Y_t 's, as $t \rightarrow \infty$, response to a permanent change in X : $X_1 = \dots = X_{t-1} = X$,
 $X_t = X_{t+1} = \dots = X + 1$

Example: The Effect of Financial Liberalization on Economic Growth

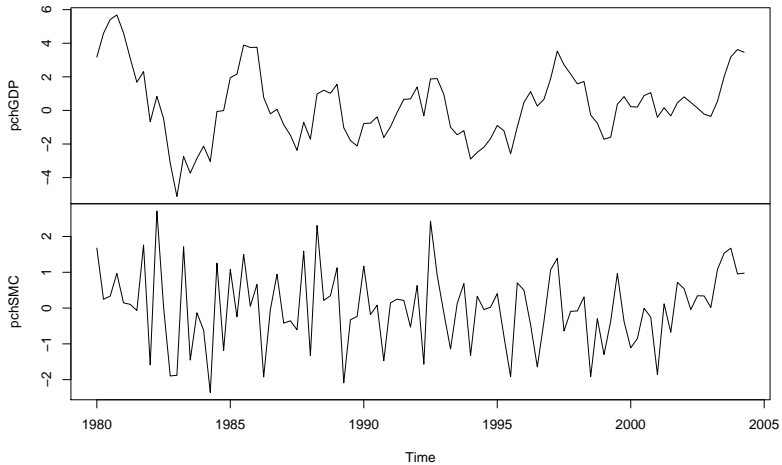
Researchers in the field of international finance and development are interested in whether financial factors can play an important role in encouraging growth in a developing country.

The purpose of this example is to investigate this issue empirically using time series data from a single country. Dataset contains data from Country A for 98 quarters on GDP growth and a variable reflecting financial liberalization: the expansion of the stock market. In particular, the dependent and explanatory variables are:

- ▶ $Y = \text{pchGDP}$ - the percentage change in GDP;
- ▶ $X = \text{pchSMC}$ - the percentage change in total stock market capitalization.

["http://uosis.mif.vu.lt/~rlapinskas/\(data%20R&GRETl/liberal.xls"](http://uosis.mif.vu.lt/~rlapinskas/(data%20R&GRETl/liberal.xls))

liberal



```
colMeans(liberal)
```

```
##      pchGDP      pchSMC  
## 0.29677662 0.01256935
```

The mean of these two variables is 0.30% and 0.01% per quarter, indicating that stock markets in Country A have not expanded by much **on average**.

Note, however, that this average hides wide variation. In some quarters market capitalization increased considerably, while in other quarters it decreased.

Assuming that both variables are stationary, we can estimate an $ADL(2, 2)$ model using OLS. Remember that, if the variables in a model are stationary, then the standard regression quantities (e.g. OLS estimates, p-values, confidence intervals) can be calculated in an ordinary way.

For now, we note that the ADL(2,2) model has insignificant variables:

$$\Delta pchGDP_t = \alpha + \delta t + \rho \cdot pchGDP_{t-1} + \gamma_1 \Delta pchGDP_{t-1} \\ + \theta pchSMC_t + \omega_1 \Delta pchSMC_t + \omega_2 \Delta pchSMC_{t-1} + \epsilon_t$$

##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	-5.8072	5.9112	-0.9824	0.3286
## time(pchGDP)	0.0029	0.0030	0.9836	0.3280
## L(pchGDP, 1)	-0.1196	0.0126	-9.4641	0.0000
## L(d(pchGDP), 1)	0.7942	0.0310	25.6277	0.0000
## pchSMC	0.1254	0.0481	2.6054	0.0108
## L(d(pchSMC), 0:1)0	0.8375	0.0438	19.1106	0.0000
## L(d(pchSMC), 0:1)1	0.0022	0.0218	0.1028	0.9183

Generally, using the formula for the **long-run multiplier** of an ADL(p,q) model, we can see that its OLS estimate is

$$-\theta/\rho = -(0.1254/(-0.1196)) = 1.0485$$

There are different ways of expressing this information verbally (remember that the dependent and explanatory variables are percentage changes).

Interpretation Examples

1. On average, market capitalization in Country A has been increasing by 0.01% per quarter and GDP by 0.30% per quarter. If Country A's total stock market capitalization increases by 1.01% in each month (i.e., increase by one unit from 0.01 each month to 1.01 each month), then in the long run its GDP should start increasing by 1.3485% per quarter (i.e., the initial 0.30 plus the long run multiplier of 1.0485).
2. The long run multiplier effect of financial liberalization on GDP growth is 1.0485%.
3. If X permanently increases by 1%, the equilibrium value of Y will increase by 1.0485%.

The statistical information, though, indicates that this might not be a good model, since some of the explanatory variables are not significant (e.g. the p-values for the coefficients on ΔX_{t-1} and the time trend both imply insignificance at the 5% level).

This raises the issue of lag length selection in the $ADL(p, q)$ model.

Note on $ADL(p, q)$ Lag Selection

We will note that the strategy for selecting q in the regression model with lagged explanatory variables and the strategy for selecting p in the $AR(p)$ model can be combined.

There is no general convention about whether you should first select p , then q , then decide whether the deterministic trend should be included, or make another ordering (e.g., select q , then p , then trend or select q , then trend, then p , etc.).

As long as you are careful, you will not be led too far wrong in selecting a good model.

The final ADL model looks like:

##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	0.0071	0.0204	0.3496	0.7275
## L(pchGDP, 1)	-0.1201	0.0114	-10.5213	0.0000
## L(d(pchGDP), 1)	0.7997	0.0297	26.9430	0.0000
## pchSMC	0.1244	0.0416	2.9905	0.0036
## d(pchSMC)	0.8393	0.0365	23.0218	0.0000

which can be written as:

$$\Delta pchGDP_t = 0.0071 + (-0.1201) \cdot pchGDP_{t-1} + 0.7997 \cdot \Delta pchGDP_{t-1} \\ + 0.1244 \cdot pchSMC_t + 0.8393 \cdot \Delta pchSMC_t + \epsilon_t$$

the **long-run** multiplier $-\theta/\rho = -0.1244/(-0.1201) = 1.035803$ and the equilibrium model is:

$$pchGDP = 0.0071 + 1.035803 \cdot pchSMC$$

thus, if *pchSMC* increased by 1 to a new permanent level, *pchGDP* will eventually increase by 1.035803.

Time Series Regression when Y and X are Trend Stationary: Spurious Regression

Consider the regression:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \epsilon_t$$

and assume that some (or all) of the variables have a linear trend, e.g.:

$$X_{1t} = \alpha_{10} + \alpha_{11}t + \omega_{1t}$$

In this case the seemingly (or, spuriously) significant coefficients of the (spurious) regression can be obtained not because the response 'truly' depends on the predictive variables, but because of the trends 'hidden' in these variables - **both** Y and X increase because of a trend, but Y increases not because X increases.

To clear the 'true' dependence of Y on X , we can act two-fold:

1. Augment the initial equation with a linear, or polynomial, trend:

$$Y_t = \gamma_0 + \gamma_1 X_{1t} + \gamma_2 X_{2t} + \delta t + \epsilon_t$$

or

2. Rewrite in initial equation in the form of deviations from the trend:

$$\tilde{Y}_t = \beta_0 + \gamma_1 \tilde{X}_{1t} + \gamma_2 \tilde{X}_{2t} + \epsilon_t$$

where, for example:

$$\tilde{X}_{1t} = X_{1t} - (\hat{\alpha}_{10} + \hat{\alpha}_{11}t)$$

Note that the γ coefficients, which describe the 'true' dependence between X and Y , are the same in both methods.

Example #1:

We will generate two series:

$$Y_t = \beta_0 + \beta_1 t + \omega_{1t}, \quad \omega_{1t} \sim WN$$

$$X_t = \alpha_0 + \alpha_1 t + \omega_{2t}, \quad \omega_{2t} \sim WN$$

And we will create two regression models:

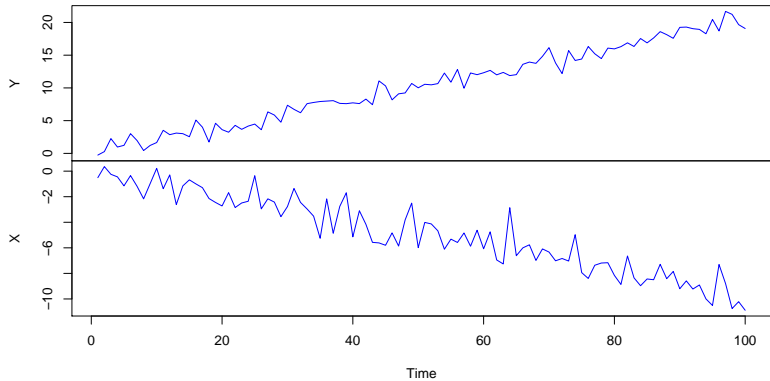
$$Y_t = \gamma_{10} + \gamma_{11} X_t + \epsilon_{1t}$$

$$Y_t = \gamma_{20} + \gamma_{21} X_t + \delta t + \epsilon_{2t}$$

```
set.seed(123)
n = 100
Y <- 0.1 + 0.2 * 1:n + rnorm(n)
X <- 0.3 - 0.1 * 1:n + rnorm(n)
Y <- ts(Y)
X <- ts(X)
```

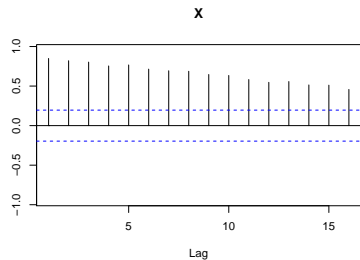
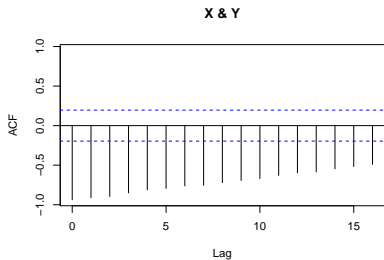
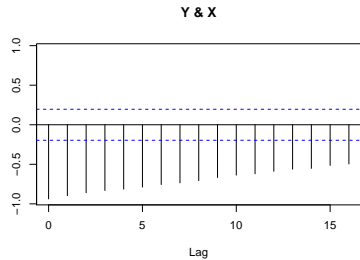
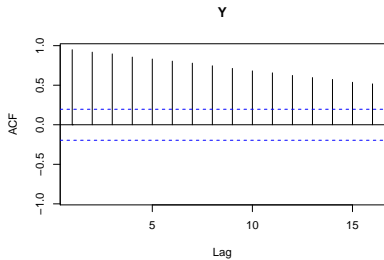
```
plot.ts(data.frame(Y, X), col = 4)
```

data.frame(Y, X)



Both of the series were generated independently, however because of the deterministic trend, it might appear that they are correlated:

```
suppressPackageStartupMessages({library("forecast")})  
Acf(data.frame(Y, X))
```



```
suppressPackageStartupMessages({library("dynlm")})
model1 <- dynlm(Y ~ X)
model2 <- dynlm(Y ~ X + time(X))
round(summary(model1)$coefficients, 4)
```

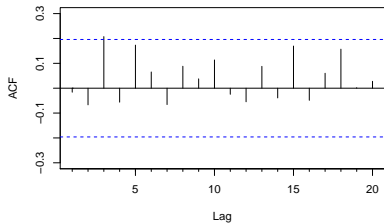
```
##           Estimate Std. Error  t value Pr(>|t|)
## (Intercept)   1.1684    0.3996   2.9237  0.0043
## X            -1.8779    0.0703 -26.7157 0.0000
```

```
round(summary(model2)$coefficients, 4)
```

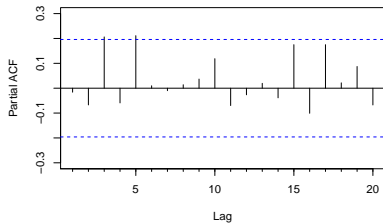
```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.0653    0.1850  0.3529  0.7249
## X            -0.0545    0.0958 -0.5687  0.5708
## time(X)       0.1972    0.0098 20.1151 0.0000
```

- ▶ We can see that if we do not include the trend coefficient, we have a spurious regression as it may seem that Y depends on X .
- ▶ If, however, we do include a trend component, we actually see that the coefficient of X is not significant.

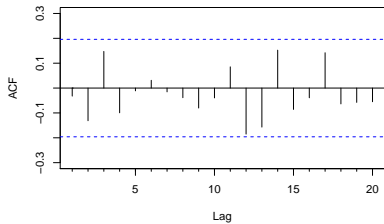
Series model1\$residuals



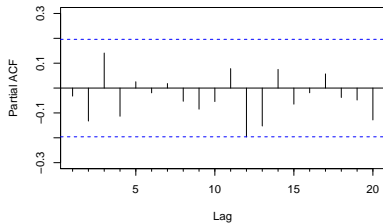
Series model1\$residuals



Series model2\$residuals



Series model2\$residuals



Example #2:

We will generate two series:

$$Y_t = \alpha_0 + \delta_1 t + \beta_0 X_t + \omega_{1t}, \quad \omega_{1t} \sim WN$$

$$X_t = \alpha_1 + \delta_2 t + \omega_{2t}, \quad \omega_{2t} \sim WN$$

And we will create two regression models:

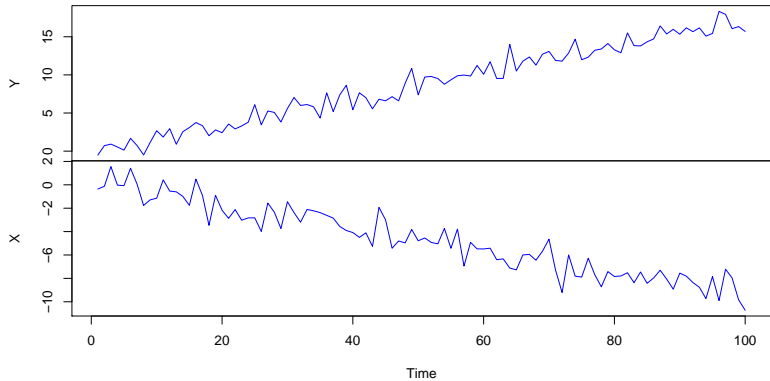
$$Y_t = \gamma_{10} + \gamma_{11} X_t + \epsilon_{1t}$$

$$Y_t = \gamma_{20} + \gamma_{21} X_t + \delta t + \epsilon_{2t}$$

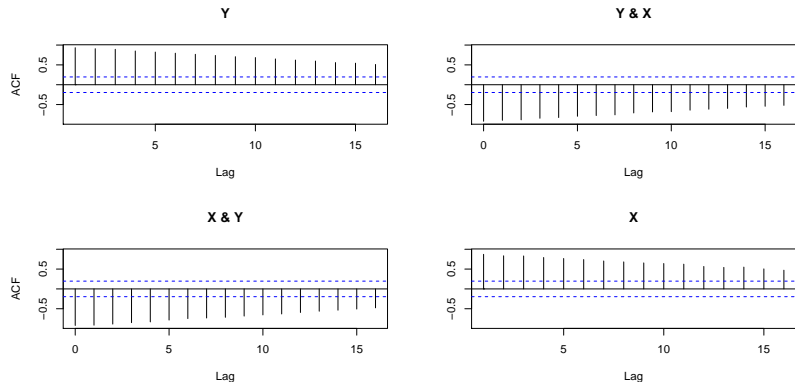
```
set.seed(123)
n = 100
X <- 0.3 - 0.1 * 1:n + rnorm(n)
Y <- 0.1 + 0.2 * 1:n + 0.3 * X + rnorm(n)
Y <- ts(Y)
X <- ts(X)
```

```
plot.ts(data.frame(Y, X), col = 4)
```

data.frame(Y, X)



```
suppressPackageStartupMessages({library("forecast")})
Acf(data.frame(Y, X))
```



Similarly to the previous example, the series appear correlated. The difference is that in this case they are generated as dependent variables. **But Y should positively depend on X .** From the cross-correlation plots of $Y&X$ (or $X&Y$) we see that the correlation between these variables is **negative** - opposite of what we would expect. This is because of the negative trend effect on X .

```
suppressPackageStartupMessages({library("dynlm")})
model1 <- dynlm(Y ~ X)
model2 <- dynlm(Y ~ X + time(X))
round(summary(model1)$coefficients, 4)
```

```
##           Estimate Std. Error  t value Pr(>|t|)
## (Intercept)   1.2644     0.375   3.3718  0.0011
## X             -1.5946     0.068 -23.4646  0.0000
```

```
round(summary(model2)$coefficients, 4)
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.1532     0.1977 -0.7751  0.4401
## X             0.2390     0.1072  2.2291  0.0281
## time(X)      0.1973     0.0110 17.9635  0.0000
```

We can see that if we do not include the trend coefficient, then the **signs** of the *immediate effect*, $\hat{\beta}_0$ is incorrect. If we do include the trend component, then the signs are correctly estimated.

So, even if there is no spurious regression, if both X and Y have trends - excluding the trend component results in an **omitted variable bias**.

Time Series Regression when Y and X Have Unit Roots: Spurious Regression

For the remainder of this lecture, we will assume that Y and X have unit roots.

In practice, of course, you would have to test whether this was the case using the previously introduced sequential procedure and/or Dickey-Fuller (or any other unit root) test.

We begin by focusing on the case of regression models without lags, then proceed to similar models to the $ADL(p, q)$ model.

Suppose we are interested in estimating the following regression:

$$Y_t = \alpha + \delta t + \beta X_t + \epsilon_t$$

If Y and X contain unit roots, then OLS estimation of this regression can yield results which are **completely wrong**.

For instance, even if the true value of β is 0, OLS can yield an estimate $\hat{\beta}$ which is very different from zero. Statistical tests (using the t-ratio or p-value) may indicate that β is not zero.

Furthermore, if $\beta = 0$, then R^2 should be zero. In fact, the R^2 will often be quite large. To put it another way: **if Y and X have unit roots, then all the usual OLS regression results might be misleading and incorrect.**

This is the so-called **spurious regression problem**. We do not have the statistical tools to prove that this problem occurs, but it is important to stress the practical implications:

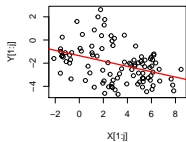
With the one exception of cointegration (which will be presented later on), **you should never run a regression of Y on X if the variables have unit roots.**

Example of a spurious regression

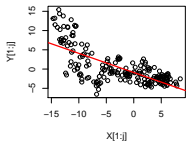
We will illustrate the above statements: let X and Y be two independent random walks with $\delta = 0$ (as it follows from independence - true $\beta = 0$). However, the OLS estimate $\hat{\beta}$ differs from zero and, as its p-value is almost always less than 0.05, we would mistakenly “conclude” that β is almost always ‘significant’. As such, **do not trust OLS in such a case!**

```
N <- 5000 # The length of a random walk
set.seed(231)
Y <- ts(cumsum(rnorm(N)))
X <- ts(cumsum(rnorm(N)))
v1 <- c(seq(from = 100, to = 1100, by = 100), 5000)
par(mfrow = c(3, 4))
for(j in v1){
  plot(X[1:j], Y[1:j], main = paste0("N = ", j))
  lm.fit <- lm(Y[1:j] ~ X[1:j])
  abline(lm.fit, col = "red", lwd = 1.5)
}
```

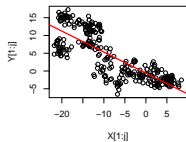
N = 100



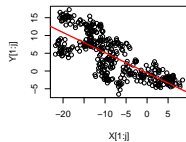
N = 200



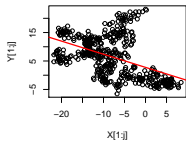
N = 300



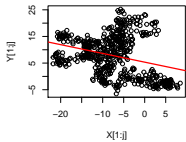
N = 400



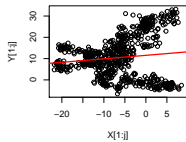
N = 500



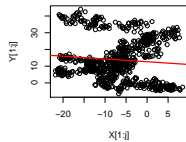
N = 600



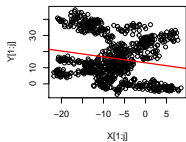
N = 700



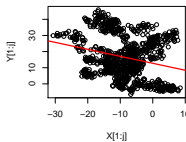
N = 800



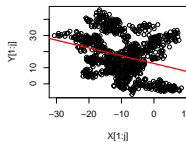
N = 900



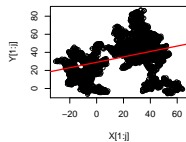
N = 1000



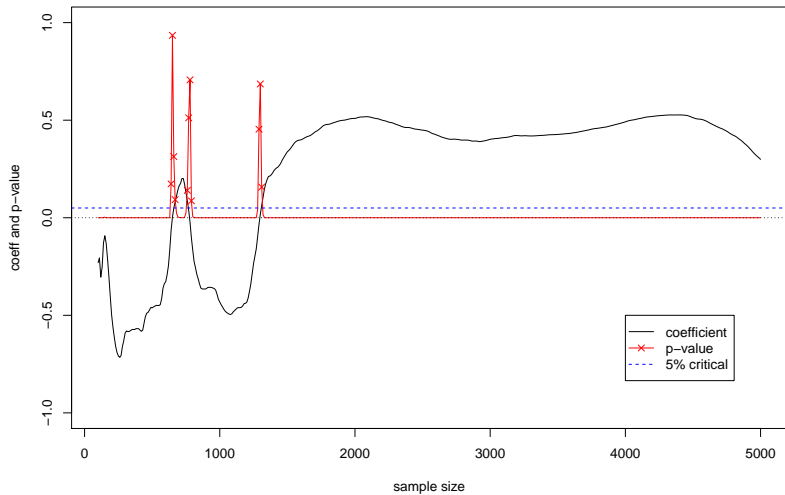
N = 1100



N = 5000



coefficient and p-value estimated via OLS



We will soon see that the reason of these wrong conclusions is the fact that $\epsilon_t \sim I(1)$. Here are some properties of popular statistics in that 'bad' case where Y_t , X_t and ϵ_t are $I(1)$:

- ▶ DW statistics (tests the presence of autocorrelation in the residuals) tend to 0, when $T \rightarrow \infty$;
- ▶ R^2 tends to a random limit;
- ▶ $\hat{\beta}^{OLS}$ is not consistent;
- ▶ t - statistics on $\hat{\beta}^{OLS}$ diverge - need to be divided by \sqrt{T} to obtain a random variable with well defined distribution.

Because both Y_t and X_t contain a stochastic trend, the OLS estimator may indicate a significant correlation between the two series, even if they are completely unrelated. **The problem is that ϵ_t is nonstationary.**

In short - **do not** regress Y on X in such a case.

However, the question remains - what to do if X_t and Y_t have unit roots and are not cointegrated?

The general advice is to rethink your model. For example, instead of working with

$$Y_t = \alpha + \beta X_t + \epsilon_t$$

analyze the model

$$Y_t = \alpha + \phi Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t$$

In this case the OLS estimators are consistent - no spurious regression problem arises because the parameters can be selected as $\phi = 1$ (in the ADL model case ϕ is used to determine whether the model is stable) and $\beta_0 = \beta_1 = 0$ such that the error term is $I(0)$, even if $Y_t \sim I(1)$ and $X_t \sim I(1)$.

Another possibility is to create an ADL model for *stationary* differences ΔY_t and ΔX_t , for example:

$$\Delta Y_t = \alpha + \phi \Delta Y_{t-1} + \beta_0 \Delta X_t + \beta_1 \Delta X_{t-1} + \epsilon_t$$

If, for example, Y_t is the log of GDP and X_t is the log of the price level, then ΔY_t and ΔX_t are the growth rate and inflation, respectively, and the above equation has a nice interpretation.

If we include the lagged coefficients and **estimate via OLS**:

```
round(coef(summary(dynlm(Y ~ L(Y, 1) + L(X, 0:1)))), 5)
```

##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	0.04771	0.03041	1.56885	0.11675
## L(Y, 1)	0.99898	0.00075	1339.42121	0.00000
## L(X, 0:1)0	0.01108	0.01398	0.79204	0.42837
## L(X, 0:1)1	-0.01153	0.01399	-0.82436	0.40977

Or if we estimate a model on variable differences:

```
round(coef(summary(dynlm(d(Y) ~ d(X)))), 5)
```

##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	-0.00074	0.01403	-0.05294	0.95778
## d(X)	0.01136	0.01398	0.81261	0.41648

The coefficient significance is inline with what we would expect (the estimate of ϕ is also close to 1).

Time Series Regression when Y and X Have Unit Roots: Cointegration

The one time where you do not have to worry about the spurious regression problem occurs when X and Y are cointegrated.

This case not only overcomes the spurious regression problem, but it also provides some nice financial intuition.

Cointegration has received a great deal of attention recently in the financial literature and many theoretical finance models imply cointegration should occur, so it is worthwhile to discuss the topic in detail.

Some intuition on cointegration

Consider the errors of a regression model:

$$\epsilon_t = Y_t - \alpha - \beta X_t$$

It is clear that the errors are just a *linear* combination of X and Y . However, if both X and Y exhibit nonstationary unit root behavior, we would expect that the error would also exhibit nonstationary behavior.

The error does indeed usually have a unit root. Statistically it is this unit root in the error term that causes the spurious regression problem.

However, it is possible that the unit roots in Y and X 'cancel each other out' and the resulting error is stationary, i.e. both X and Y trend together. In this special case (called cointegration) the spurious regression problem vanishes and it is valid to run a regression of Y on X .

Definition

If Y and X have unit roots but some linear combination of them, $\gamma_1 Y_t + \gamma_2 X_t$, is (trend) stationary, then we say that Y and X are cointegrated.

In other words: $Y_t \sim I(1)$ and $X \sim I(1)$ are cointegrated if they share a **common trend** such that $\gamma_1 Y_t + \gamma_2 X_t \sim I(0)$.

Note that the coefficients γ_1 and γ_2 are not uniquely defined if $\gamma_1 Y_t + \gamma_2 X_t$ is stationary, then, for example $Y_t + (\gamma_2/\gamma_1)X_t$ is also stationary.

How to make sure that such constants γ_1 and γ_2 exist? Stationarity means low variability around the mean, therefore, to best approximate Y_t in terms of X_t , we can use OLS.

In other words¹: If $Y_t \sim I(1)$ and $X \sim I(1)$ and there exists such $\tilde{\gamma}$ such that $Z_t = Y_t - \tilde{\gamma}X_t \sim I(0)$, then X_t and Y_t are cointegrated with:

- ▶ $\tilde{\gamma}$ - called the **cointegrating parameter**,
- ▶ $(1, -\beta)^\top$ - called the **cointegrating vector**.

¹See: Verbeek, Marno - A guide to modern econometrics 5e, (2017)

Example

Let:

$$\begin{cases} Y_t = 1.3X_t + u_t \\ u_t = 0.6u_{t-1} + \epsilon_{1t} \\ X_t = X_{t-1} + \epsilon_{2t}, \quad \epsilon_{it} \sim i.i.d.N(0, 1) \end{cases}$$

Here both X_t and Y_t are quite irregular random walks, but the linear combination: $1 \cdot Y_t - 1.3 \cdot X_t = u_t = 0.6u_{t-1} + \epsilon_{1t}$ is a stationary AR(1) process.

Thus, Y_t and X_t cointegrate (they have a common stochastic trend).

In practice, we have to test the residuals of the regression $Y_t = \alpha + \beta X_t + u_t$ for stationarity.

In the real world, it is unlikely that a financial system will ever be in precise equilibrium since shocks and unexpected changes to it will always occur. However, departures u_t from equilibrium should not be too large and there should always be a tendency to return to equilibrium after a shock occurs. Hence, if a financial model which implies an equilibrium relationship exists, then we should observe Y and X as being cointegrated.

Common sense tells you that, if two assets are close substitutes for one another, then their prices should not drift too far apart. After all, if one asset becomes much more expensive than a similar asset, then investors will sell the first asset in order to buy the cheaper alternative. But if many investors are selling the expensive asset, then its price should drop. And if many investors are buying the cheap asset its price would rise. Thus, the prices of the expensive and cheap assets would move closer to one another. Many financial theories formalize this intuition to imply different cointegrating relationships.

In short, financial theory suggests that cointegrating relationships between many different financial time series should exist. Hence, it is important to test whether cointegration is present (i.e., to see whether financial theory holds in practice) and, if it is present, to estimate models involving cointegrated variables.

Estimation and Testing with Cointegrated Variables

As mentioned above, if Y and X are cointegrated, then the spurious regression problem does not apply; consequently, we can run an OLS regression of Y on X and obtain valid results.

Furthermore, the coefficient from this regression is the long run multiplier. Thus, **to the extent that interest centers on the long run multiplier**, then estimation with cointegrated variables is very easy.

The test for cointegration described here is referred to as the **Engle - Granger test**, after the two econometricians who developed it.

It is based on the regression of Y on X . Remember that, **if cointegration occurs, then the errors from this regression will be stationary**. Conversely, **if cointegration does not occur, then the errors will have a unit root**.

Given the close relationship between the errors and the residuals, it is reasonable to examine the properties of the residuals in order to investigate the presence of cointegration.

Engle - Granger test for cointegration

In the previous lectures we discussed testing for a unit root in a time series variable. Here, we test for a unit root in the **residuals** using the same techniques. In particular, the test for cointegration involves the following steps:

0. Carry out a Dickey-Fuller test on the null hypothesis that Y and X each has a unit root. If both time series are $I(0)$ - standard regression analysis will be valid. If they are both integrated to the same order (usually $I(1)$) - proceed to the next step
1. Run a regression of Y on X and save the residuals;
2. Carry out a unit root test on the residuals (**without including a constant and a deterministic trend**) ;
3. If the unit root hypothesis is **rejected**, then conclude that Y and X are **cointegrated**. However, if the unit root hypothesis is **accepted**, then conclude that cointegration **does not occur**.

Thus, if Y_t and X_t are cointegrated in $Y_t = \alpha + \beta X_t + \epsilon_t$, the error term is $I(0)$. If not, ϵ_t will be $I(1)$. Hence, one can test for the presence of a cointegration relationship by testing for a unit root in the OLS residuals e_t .

It seems that this can be done by using the Dickey-Fuller tests of the previous chapter. For example, one can run the regression:

$$\Delta e_t = \rho e_{t-1} + \sum_{j=1}^p \beta_j \Delta e_{t-j} + u_t, \quad u_t \sim WN$$

on OLS residuals rather than an observed time series.

Because the OLS estimators 'choose' the residuals in the cointegrating regression to have as small variance as possible, even if the variables are not cointegrated, the OLS estimator will make the residuals *look* as stationary as possible (because these are the assumptions under which OLS estimation is carried out).

Thus, using standard DF or ADF tests, we may reject the null hypothesis of nonstationarity too often. As a result, the appropriate critical values are more negative than those for the standard Dickey-Fuller tests.

If e_t is not appropriately described by a first order autoregressive process, one should add lagged values of Δe_t to its equation leading to the augmented Dickey-Fuller (ADF) tests, with the same asymptotic critical values. This test can be extended to test for cointegration between three or more variables. If more than a single X_t variable is included in the [cointegrating regression](#)

$$Y_t = \alpha + \delta t + \beta X_t + \epsilon_t,$$

the critical values shift further to the left.

Number of X's in the rhs of eq.	No deterministic terms in eq.	Only a constant in eq.	Constant and a trend in eq.
1	-2.76	-3.37	-3.80
2	-3.27	-3.77	-4.16
3	-3.74	-4.11	-4.49
4	-4.12	-4.45	-4.74
5	-4.40	-4.71	-5.03

Thus if the ADF test statistic $\hat{\rho}/se(\hat{\rho})$ is closer to zero than respective critical value, we do not reject the unit root hypothesis.

Note also that in the Dickey-Fuller test, we test the hypothesis that $\rho = 0$. (i.e. unit root).

In the cointegration test, we use the Dickey-Fuller methodology but cointegration is found if we **reject** the unit root hypothesis for the *residuals*.

In other words, the null hypothesis in the Engle-Granger test is **no cointegration** and we conclude *cointegration is present only* if we **reject** this hypothesis.

Furthermore, note the following:

- ▶ If all variables are stationary at level, i.e. if $X_t, Y_t \sim I(0)$ - there is no **long run relationship (i.e. co-movement)**, but a short run relationship may exist - there is no need for cointegration estimation.
- ▶ While the **long-run relationship** may not exist - this does not mean that the same applies to the **s-period interim multiplier**, which describes the effect (which may be quickly declining to zero) of a (usually permanent) change in X on Y after s periods.

Time Series Regression when Y and X are Cointegrated: The Error Correction Model (ECM)

You may encounter instances where unit root tests indicate that your **time series have unit roots**, but the Engle-Granger test indicates that the series **are not cointegrated** (they have no long-run relationship).

That is, the series may not be trending together and may not have an equilibrium relationship. In these cases, you should not run a regression of Y on X due to the spurious regression problem.

The presence of such characteristics suggests that you should rethink your basic model and include other explanatory variables.

Instead of working with Y and X themselves you could, for example, take their differences (since they have unit roots - their differences should be stationary).

In other words, you may wish to estimate the ADL model for stationary differences:

$$\Delta Y_t = \phi + \gamma_1 \Delta Y_{t-1} + \dots + \gamma_{p'} \Delta Y_{t-p'} + \omega_0 \Delta X_t + \dots + \omega_{q'} \Delta X_{t-q'} + \epsilon_t$$

Although such models can be estimated, it is important to understand that they describe only short-run interactions between Y_t and X_t (there is no equilibrium, therefore there is no sense to speak about long run multiplier).

Furthermore, if the variables are **cointegrated**, these models are not satisfactory, because they ignore the long-run relationship between the variables Y_t and X_t .

A principal feature of cointegrated variables is that their time paths are influenced by the extent of any deviation from long-run equilibrium. After all, if the system is to return to the long-run equilibrium, the movements of at least some of the variables must respond to the magnitude of the disequilibrium.

The ECM

The dynamic model implied by this discussion is one of **error correction**. In an error-correction model, the short-term dynamics of the variables in the system are influenced by the deviation from equilibrium . Furthermore, there exists the Granger Representation Theorem that says that if Y and X are cointegrated, then the relationship between them can always be expressed as an ECM.

A common expression for ECM in two variables (Y and X) is a single equation form:

$$\Delta Y_t = \psi + \lambda u_{t-1} + \omega_0 \Delta X_t + \epsilon_t$$

where u_{t-1} is the error obtained from cointegrating regression of Y on X , i.e. $u_{t-1} = Y_{t-1} - \alpha - \beta X_{t-1}$. Note that if we knew u_{t-1} , then the ECM is just a regression model.

For the derivation of the ECM, see the slides on Regressions with time lags.

We can also generalize the above expression to:

$$\Delta Y_t = \psi + \delta t + \lambda u_{t-1} + \tilde{\gamma}_1 \Delta Y_{t-1} + \dots + \tilde{\gamma}_p \Delta Y_{t-p} + \tilde{\omega}_0 \Delta X_t + \dots + \tilde{\omega}_q \Delta X_{t-1} + \epsilon_t$$

Thus, if Y_t and X_t both have unit roots - regress in differences, but if they are cointegrated - add an error correction term.

Some sophisticated statistical techniques have been developed to estimate the ECM, but the simplest thing to do is merely to replace the unknown errors by the residuals from the regression of Y on X , i.e. replace u_{t-1} by \hat{u}_{t-1} .

That is a simple technique based on two OLS regressions (also known as the Engle and Granger 2-step approach) proceeds as follows:

1. Run a regression of Y on X and save the residuals²:
$$u_t = Y_t - \hat{\alpha}_0 - \hat{\alpha}_1 X_t.$$
2. Run a regression of ΔY_t on all regressors from the above equation, including the residuals from the previous step, which are lagged by one period.

²OLS will not only be valid, but in fact super consistent as the estimate converges to the true parameter much faster than with conventional asymptotics.

ECM Derivation Outline

Assume that the long-run relationship is defined as: $Y = \tilde{\alpha} + \tilde{\beta}X$. Taking the expectation from the ADL(p, q) (with **stationarity**: $X_t, Y_t \sim I(0)$) yields:

$$\mu_Y = \alpha + \sum_{i=1}^p \phi_i \mu_Y + \sum_{j=0}^q \beta_j \mu_X \iff \mu_Y = \frac{\alpha}{1 - \sum_{i=1}^p \phi_i} + \frac{\sum_{j=0}^q \beta_j}{1 - \sum_{i=1}^p \phi_i} \mu_X$$

with

$$\tilde{\alpha} = \frac{\alpha}{1 - \sum_{i=1}^p \phi_i}, \quad \tilde{\beta} = \frac{\sum_{j=0}^q \beta_j}{1 - \sum_{i=1}^p \phi_i}$$

Then, the general dynamic relationship can be rewritten as:

$$\begin{aligned} Y_t &= \alpha + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{j=0}^q \beta_j X_{t-j} + \epsilon_t \\ &= \tilde{\alpha} \left[1 - \sum_{i=1}^p \phi_i \right] + \left[1 - \frac{\alpha}{\tilde{\alpha}} - \sum_{i=2}^p \phi_i \right] Y_{t-1} + \sum_{i=2}^p \phi_i Y_{t-i} \\ &\quad + \beta_0 X_t + \left[\left(1 - \sum_{i=1}^p \phi_i \right) \tilde{\beta} - \beta_0 - \sum_{j=2}^q \beta_j \right] X_{t-1} + \sum_{j=2}^q \beta_j X_{t-j} + \epsilon_t \end{aligned}$$

(alternatively, we can derive this by subtracting Y_{t-1} from both sides and do some rewriting, but the above derivation has an economic interpretation)

With some rewriting we get the following:

$$\begin{aligned}
 Y_t - Y_{t-1} &= \tilde{\alpha} \left[1 - \sum_{i=1}^p \phi_i \right] - \left[1 - \sum_{i=1}^p \phi_i \right] Y_{t-1} - \sum_{i=2}^p \phi_i Y_{t-1} \\
 &\pm \sum_{k=3}^p \left(Y_{t-k+1} \sum_{j=k}^p \phi_j \right) + \sum_{i=2}^p \phi_i Y_{t-i} \\
 &+ \left[1 - \sum_{i=1}^p \phi_i \right] \tilde{\beta} X_{t-1} + \beta_0 X_t - \beta_0 X_{t-1} - \sum_{j=2}^q \beta_j X_{t-1} \\
 &\pm \sum_{m=3}^q \left(X_{t-m+1} \sum_{j=m}^q \beta_j \right) + \sum_{j=2}^q \beta_j X_{t-j} + \epsilon_t \\
 &= \sum_{j=1}^{p-1} -[\phi_{j+1} + \dots + \phi_p] \Delta Y_{t-j} + \beta_0 \Delta X_t + \sum_{i=1}^{q-1} -[\beta_i + \dots + \beta_q] \Delta X_{t-i} \\
 &- \left[1 - \sum_{i=1}^p \phi_i \right] \left(Y_{t-1} - \tilde{\alpha} - \tilde{\beta} X_{t-1} \right) + \epsilon_t
 \end{aligned}$$

which we can finally write as:

$$\Delta Y_t = \gamma_1 \Delta Y_{t-1} + \dots + \gamma_p \Delta Y_{t-p+1} + \beta_0 \Delta X_t + \omega_2 \Delta X_{t-1} + \dots + \omega_q \Delta X_{t-q+1} \\ - \left[1 - \sum_{i=1}^p \phi_i \right] \left(Y_{t-1} - \tilde{\alpha} - \tilde{\beta} X_{t-1} \right) + \epsilon_t$$

The model can be estimated in a number of ways (as long as the variables are either cointegrated or do not have a unit root):

- ▶ by multiplying out the terms in parenthesis;
- ▶ by estimating the nonlinear equation via non-linear least squares (`nls(...)` function in R);
- ▶ by substituting the long-run difference with the residuals from its **cointegrating regression**: $\hat{u}_{t-1} = Y_{t-1} - \hat{\tilde{\alpha}} - \hat{\tilde{\beta}} X_{t-1}$. The cointegrating vector can be estimated (super)consistently from the cointegrating regression.

Relationship between Y_t , ΔY_t and the ECM

- ▶ **The ADL(p, q) model for the levels** (i.e. X_t and Y_t):

$$Y_t = \alpha + \delta t + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \beta_0 X_t + \dots + \beta_q X_{t-q} + \epsilon_t$$

- ▶ **The ADL(p, q) model for the levels rewritten for the differences** (i.e. ΔX_t and ΔY_t):

$$\begin{aligned} \Delta Y_t = & \alpha + \delta t + \rho Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \dots + \gamma_{p-1} \Delta Y_{t-p+1} \\ & + \theta X_t + \omega_1 \Delta X_t + \dots + \omega_q \Delta X_{t-q+1} + \epsilon_t \end{aligned}$$

- ▶ **The Error-Correction form of the ADL(p, q) model:**

$$\begin{aligned} \Delta Y_t = & \alpha + \delta t + \lambda \hat{u}_{t-1} + \tilde{\gamma}_1 \Delta Y_{t-1} + \dots + \tilde{\gamma}_{p-1} \Delta Y_{t-p+1} \\ & + \tilde{\omega}_1 \Delta X_t + \dots + \tilde{\omega}_q \Delta X_{t-q+1} + \epsilon_t, \quad \text{where } \hat{u}_{t-1} = Y_{t-1} - \hat{\alpha} - \hat{\beta} X_{t-1} \end{aligned}$$

- ▶ If $X_t, Y_t \sim I(1)$ and $u_t \sim I(1)$, then there is no long-run relationship and we set $\lambda = 0$;
- ▶ If $X_t, Y_t \sim I(0)$, then the process can always be written in an error-correction form. But the concept of cointegration can be applied to $I(1)$ time series only. **The presence of a cointegrating vector can be interpreted as the presence of a long-run equilibrium relationship.**
- ▶ If $\alpha \neq 0$ (short-run) in addition to $\hat{\alpha} \neq 0$ (long-run), then the long-run equilibrium corresponds to a steady state growth with $\Delta Y_t = \Delta X_{t-1} \neq 0$.

The Long-run and Short-run Multipliers

Now we shall get back to the material which was earlier exposed. First, assume that **both** Y_t and X_t are **stationary** .

1. Let us consider two processes Y_t and X_t , bound together by the **static** equation:

$$Y_t = \alpha + \beta X_t + \epsilon_t$$

If the values of Y_t and X_t do not change: $\dots = X_{T-1} = X_T = X$, $\dots = Y_{T-1} = Y_T = Y = \alpha + \beta X$, and $\epsilon_t = 0$, we say that X_t and Y_t are in **equilibrium**.

Now assume that in time $T + 1$ the variable X_t changes its value to $X_{T+1} = X_{T+2} = \dots = X + 1$ - the (X_t, Y_t) system will move to a new equilibrium but, generally speaking, it could take some time. A (usually not very realistic) **static system is remarkable in the sense that it reaches the new equilibrium in no time** :

$$Y_{T+1} = \alpha + \beta X_{T+1} = \alpha + \beta \cdot (X + 1) \Rightarrow Y_{T+1} - Y = \beta$$

$$Y_{T+h} = \alpha + \beta X_{T+h} = \alpha + \beta \cdot (X + 1) \Rightarrow Y_{T+h} - Y = \beta, \quad h \geq 2$$

Here, β is both the short-run (when $h = 1$) and the long run (when $h \rightarrow \infty$) multiplier.

2. Transition to the new equilibrium is slower in **dynamic** case. Assume that DGP is described by a DL(1) process:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t$$

Now:

$$Y_{T+1} = \alpha + \beta_0 X_{T+1} + \beta_1 X_T = \alpha + \beta_0 \cdot (X + 1) + \beta_1 X$$

$$\Rightarrow Y_{T+1} - Y = \beta_0$$

$$Y_{T+h} = \alpha + \beta_0 X_{T+h} + \beta_1 X_{t+h-1} = \alpha + (\beta_0 + \beta_1) \cdot (X + 1)$$

$$\Rightarrow Y_{T+h} - Y = \beta_0 + \beta_1, \quad h \geq 2$$

i.e., the short-run (or impact) multiplier is β_0 and the long-run multiplier is $\beta_0 + \beta_1$. Note that $\beta_0 + \beta_1$ is also the slope of the equilibrium equation which can be obtained from the original dynamic equation as $t \rightarrow \infty$, $X_{t-1} \rightarrow X$ and $X_t \rightarrow X$: $\lim_{t \rightarrow \infty} Y_t = Y = \alpha + (\beta_0 + \beta_1)X$.

3. Assume that the DGP is described as ADL(1, 1) process:

$$Y_t = \alpha + \phi Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t$$

In equilibrium, at moment T we have: $Y = \alpha + \phi Y + \beta_0 X + \beta_1 X$ or

$$Y_T = \frac{\alpha}{1 - \phi} + \frac{\beta_0 + \beta_1}{1 - \phi} \cdot X$$

Thus:

$$\begin{aligned} Y_{T+1} &= \alpha + \phi Y_T + \beta_0(X + 1) + \beta_1 X \\ &= \alpha + \phi \cdot \frac{\alpha}{1 - \phi} + \phi \cdot \frac{\beta_0 + \beta_1}{1 - \phi} \cdot X + (\beta_0 + \beta_1)X + \beta_0 \\ &= \frac{\alpha}{1 - \phi} + \frac{\beta_0 + \beta_1}{1 - \phi} \cdot X + \beta_0 = Y_T + \beta_0 \end{aligned}$$

$$\implies Y_{T+1} - Y_T = \beta_0$$

i.e., the short-run multiplier equals β_0 .

To find the limit $Y_{T+h} - Y_T$ as $h \rightarrow \infty$, repeat the reasoning as before on the ADL model by adding up the partial derivatives of Y_t, Y_{t+1}, \dots by X_t which implies that the long-run multiplier equals $(\beta_0 + \beta_1)/(1 - \phi)$ for the ADL(1,1) (see the previous lecture on ADL models).

4. The $ADL(1, 1)$ process can be rewritten as the error correction model (ECM):

$$Y_t = \alpha + \phi Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t \pm Y_{t-1} \pm \beta_0 X_{t-1}$$

...

$$\begin{aligned}\Delta Y_t &= \alpha + \beta_0 \Delta X_t + (\phi - 1) \left(Y_{t-1} - \frac{\beta_0 + \beta_1}{1 - \phi} X_{t-1} \right) + \epsilon_t \\ &= \beta_0 \Delta X_t + (\phi - 1) \left(Y_{t-1} - \frac{\alpha}{1 - \phi} - \frac{\beta_0 + \beta_1}{1 - \phi} X_{t-1} \right) + \epsilon_t\end{aligned}$$

where the short-run and long-run features of the dynamic relationship are modeled separately. Here:

- ▶ β_0 describes the short-run relationship;
 - ▶ $(\beta_0 + \beta_1)/(1 - \phi)$ - the long-run multiplier;
 - ▶ $\phi - 1$ - speed of adjustment to equilibrium described by the error correction term.
5. Let us assume that the above equations are not in levels but in logarithms. Then respective multipliers are called short-run and long-run **elasticity**.

Now, assume that **both** Y_t and X_t are $I(1)$ processes, i.e. they have a unit root.

- Regression analysis applied to non-stationary time series can produce misleading results (spurious regression), therefore (if the series are **not cointegrated**, i.e., they are not trending together) we should **not** look for a long-run multiplier in equilibrium equation:

$$Y_t = \alpha + \beta X_t + \epsilon_t \text{ -- NOT VALID IF NO COINTEGRATION}$$

Instead, to find the short-run multiplier, assume that process for stationary differences can be expressed as and $ADL(1, 1)$:

$$\Delta Y_t = \phi \Delta Y_{t-1} + \beta \Delta X_t + \beta_1 \Delta X_{t-1} + \epsilon_t$$

For example, say $y_t = \log(GDP_t)$ and $x_t = \log(P_t)$ (stock exchange index) are both integrated but **not cointegrated**. In this case, regression:

$$y_t = \alpha + \beta x_t + \epsilon_t$$

is **senseless** but $ADL(p, q)$ equation for stationary GDP growth rate Δy_t and stock returns Δx_t can be reasonable (the coefficient β_0 is the short-run multiplier).

Unfortunately, the model on the differences then removes any information about the long-run from the model - are we above or below the equilibrium ?

7. If Y_t and X_t are **cointegrated**, i.e., the equilibrium relationship **exists**, the ECM equation:

$$\Delta Y_t = \phi + \lambda(Y_{t-1} - \alpha - \beta X_{t-1}) + \omega_0 \Delta X_t + \epsilon_t$$

contains both:

- ▶ The short-run multiplier ω_0 ;
- ▶ The long-run multiplier β .

Summary

The static (or long-run) model: $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t}$ depicts an equilibrium relationship such that for given values of the right-hand-side variables and their long-run impact on Y (i.e. β_i), there is no reason for the response to be at any other value than Y_t .

However, it is also of interest to consider the short-run evolution of the variables under consideration, especially since equilibrium (i.e., the **steady-state**) may rarely be observed. The major reason why relationships are not always in equilibrium centers on the inability of economic agents to adjust to new information instantaneously.

- ▶ There are often substantial costs of adjustment (both pecuniary and non-pecuniary) which result in the current value of the dependent variable Y being determined not only by the current value of some explanatory variables X_t but also by their past values X_{t-1}, \dots, X_{t-q} .
- ▶ In addition, as Y evolves through time in reaction to current and previous values of X , past (i.e. lagged) values of itself, Y_{t-1}, \dots, Y_{t-p} will also enter the short-run (dynamic) model.

This inclusion of lagged values of the dependent variable as regressors is a means of simplifying the form of the dynamic model, which would otherwise tend to have a large number of highly correlated lagged values of X , and lead to the problem of multicollinearity (when two or more of the predictors in a regression model are highly correlated) with high R^2 but imprecise parameter estimates and low t – values, even though the model may be correctly specified. A very simple dynamic model of **short-run adjustments** is:

$$Y_t = \alpha_0 + \gamma_0 X_t + \gamma_1 X_{t-1} + \alpha_1 Y_{t-1} + \epsilon_t, \quad \epsilon_t \sim WN$$

(here variables in lower case are in logarithms).

- ▶ The parameter γ_0 denotes the short-run reaction of Y_t to a change in X_t and not the long-run effect that would occur if the model were in equilibrium (the latter is defined by $Y_t = \beta_0 + \beta_0 X_t$).
- ▶ So in the long-run, the elasticity between Y and X is $\beta_1 = (\gamma_0 + \gamma_1)/(1 - \alpha_1)$, assuming that $|\alpha_1| < 1$ (necessary if the short-run model is to converge to a long-run solution).

The dynamic model represented in the previous slide is easily generalized to allow for more complicated, and often more realistic adjustment processes by increasing the lag-lengths. However, there are several potential problems with this form of the dynamic model.

1. Multicollinearity.
2. some (if not all) of the variables in a dynamic model are likely to be non-stationary, since they enter in levels (this leads to the potential problem of spurious regression).

A solution might be to re-specify the dynamic model in (first) differences. However, this then removes any information about the long-run from the model (are we above or below the equilibrium?) and consequently is unlikely to be useful for forecasting purposes. A more suitable approach is to adopt the **error-correction (ECM)** formulation of the dynamic model:

$$\Delta Y_t = \gamma_0 \Delta X_t + (1 - \alpha)(Y_{t-1} \tilde{\beta}_0 - \tilde{\beta}_1 X_{t-1}) + \epsilon_t$$

An advantage of the ECM is that all the terms in the model are stationary so standard regression techniques are valid, assuming cointegration and that we have estimates of β_0 and β_1 .

The simple depicted ECM can be generalized in several directions. First, increasing the lag-length p and/or q results in additional lagged first differences:

$$A(L)\Delta Y_t = B(L)\Delta X_t - (1 - \pi) [Y_{t-1} - \tilde{\beta}_0 - \tilde{\beta}_1 X_{t-1}] + \epsilon_t$$

where $\pi = \alpha_1 + \dots + \alpha_p$.

If more than one X enters the model, we have to consider the possibility for more than one cointegrating relationship to exist. This leads to the ECM in multivariate form, VECM, which will be explored in the next topic.

Remark (1): Model Equivalence

Consider a **static** model with autoregressive errors:

$$\begin{cases} Y_t &= \alpha + \beta_0 X_t + u_t \\ u_t &= \phi u_{t-1} + \epsilon_t, \quad |\phi| < 1, \epsilon_t \sim WN \end{cases}$$

The process can be re-written as a **dynamic** model:

$$\begin{cases} u_t &= Y_t - \alpha - \beta_0 X_t \\ u_t &= \phi u_{t-1} + \epsilon_t \end{cases}$$

$$Y_t = \alpha(1 - \phi) + \phi Y_{t-1} + \beta_0 X_t - \beta_0 \phi X_{t-1} + \epsilon_t$$

which we can write as an ADL(1, 1):

$$Y_t = \tilde{\alpha} + \phi Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t$$

After estimating the model via OLS we would need to test the nonlinear restriction:

$$H_0 : \beta_1 = -\beta_0 \phi$$

This can be achieved by creating two models - one without parameter restrictions, estimated via OLS and another - with parameter restrictions. Both can be estimated via `nls()` and compared with the `anova()` function.

Remark (2): Cross-Correlation Functions

While examining the relationship between two time series variables Y_t and X_t , the series Y_t may depend on the past values of X . Similarly to the ACF for single time series, the **sample cross correlation function (CCF)** may be employed to identify, which lags of X_t may be useful in predicting Y_t by calculating $\text{Corr}(Y_t, X_{t+h})$, where $h = 0, \pm 1, \pm 2, \dots$

- ▶ When X_{t+h} , $h < 0$ (past values) significantly correlate with Y_t , we may say that X **leads** Y ;
- ▶ When X_{t+h} , $h > 0$ (future values) significantly correlate with Y_t , we may say that X **lags** Y ;

In some problems, the goal may be to identify which variable is leading and which is lagging.

```
X <- axtsa::rec  
Y <- axtsa::soi
```

If we want to examine which X_{t-k} to include in our model for Y_t - we need to examine cross-correlations where X is **leading** Y .

```
# custom cross-correlation
cr <- function(x, y, h = 1){
  mux <- mean(x)
  muy <- mean(y)
  dx <- mean((x - mux)^2)
  dy <- mean((y - muy)^2)
  n <- length(x)
  #
  if(h < 0){
    h <- -h
    x <- x[-c(length(x) - (h-1):0)]
    y <- y[-c(1:h)]
  }else{
    x <- x[-c(1:h)]
    y <- y[-c(length(y) - (h-1):0)]
  }
  cxy <- sum((x - mux)*(y - muy)) / n
  cxy / sqrt(dx * dy)
}
```

Compare the custom, with the built-in results:

```
# X leads Y
round(cr(X, Y, h = -6), 3)
```

```
## [1] -0.232
```

```
# X lags Y
round(cr(X, Y, h = 6), 3)
```

```
## [1] -0.599
```

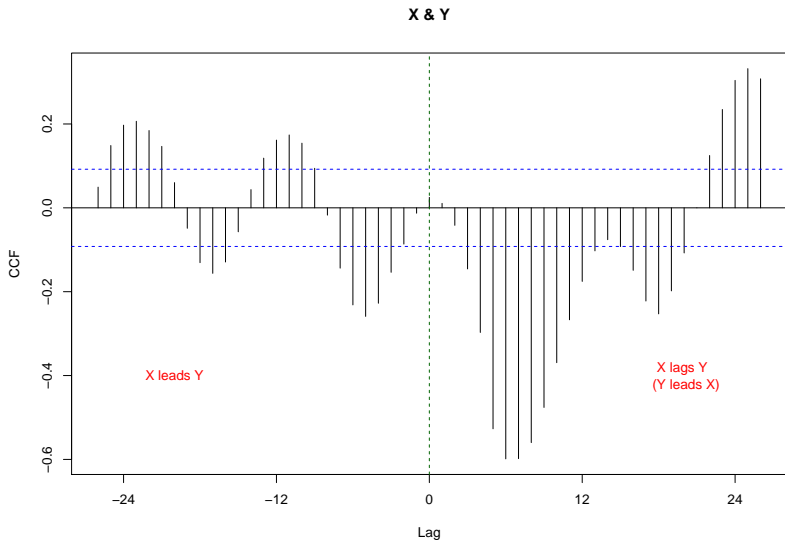
```
forecast::Ccf(X, Y, lag.max = 6, plot = FALSE)
```

```
##
## Autocorrelations of series 'X', by lag
##
##      -6      -5      -4      -3      -2      -1      0      1      2      3      4
## -0.232 -0.259 -0.228 -0.154 -0.086 -0.013  0.025  0.011 -0.042 -0.146 -0.297
##      5      6
## -0.527 -0.599
```

Specifying $Ccf(X, Y)$ means that X will be **leading** Y for $h < 0$.

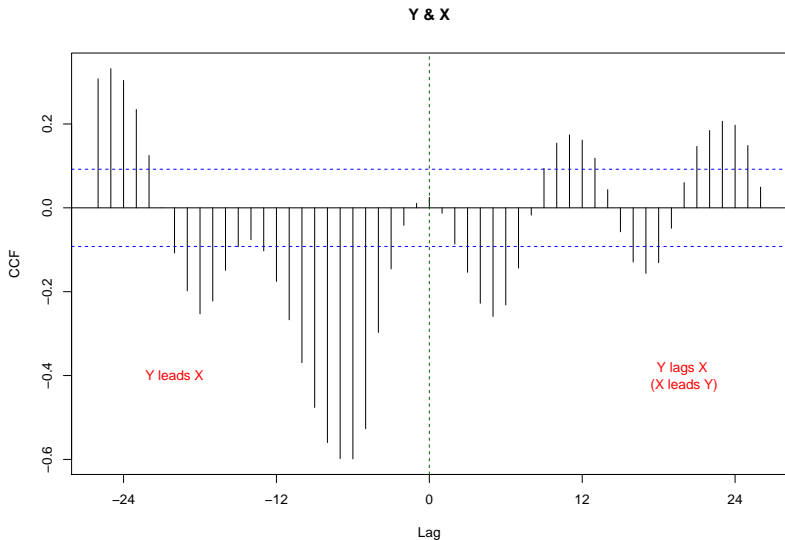
To calculate the $\widehat{\text{Corr}}(X_{t+h}, Y_t)$ use the forecast `::Ccf()` function in R:

```
forecast::Ccf(X, Y)
abline(v = 0, lty = 2, col = "darkgreen")
text(x = -20, y = -0.4, labels = "X leads Y", col = "red")
text(x = 20, y = -0.4, labels = "X lags Y \n (Y leads X)", col = "red")
```



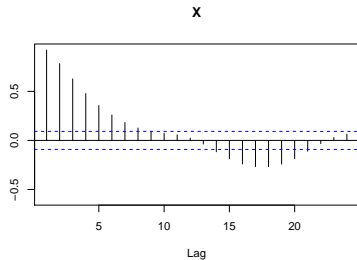
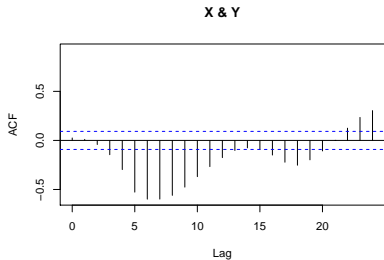
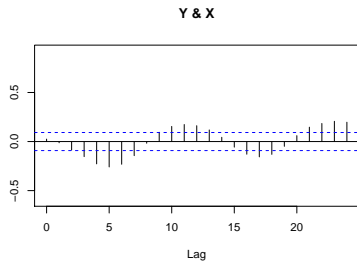
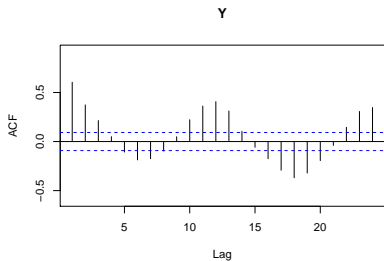
To calculate the $\widehat{\text{Corr}}(Y_{t+h}, X_t)$ use the forecast `::Ccf()` function in R:

```
forecast::Ccf(Y, X)
abline(v = 0, lty = 2, col = "darkgreen")
text(x = -20, y = -0.4, labels = "Y leads X", col = "red")
text(x = 20, y = -0.4, labels = "Y lags X \n(X leads Y)", col = "red")
```



We can get the ACF and CCF using:

```
forecast::Acf(cbind(Y, X))
```

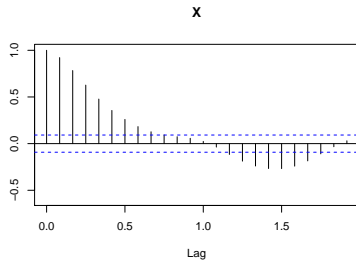
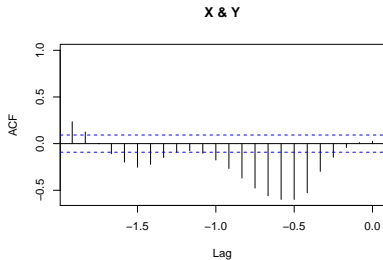
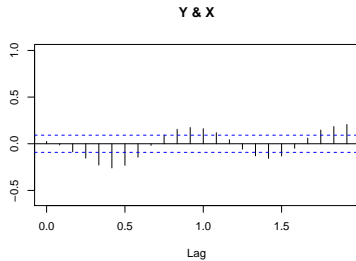
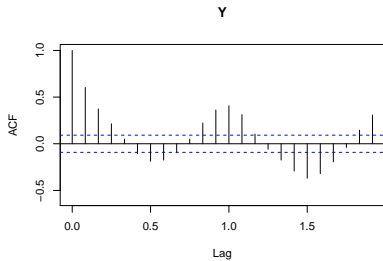


▶ Y & X - X leads Y - $\widehat{\text{Corr}}(X_t, Y_{t+h}) = \widehat{\text{Corr}}(X_{t-h}, Y_t), h > 0;$

▶ X & Y - Y leads X - $\widehat{\text{Corr}}(Y_t, X_{t+h}) = \widehat{\text{Corr}}(Y_{t-h}, X_t), h > 0;$

Note the difference in CCF plots with stats::acf()!

```
acf(cbind(Y, X))
```

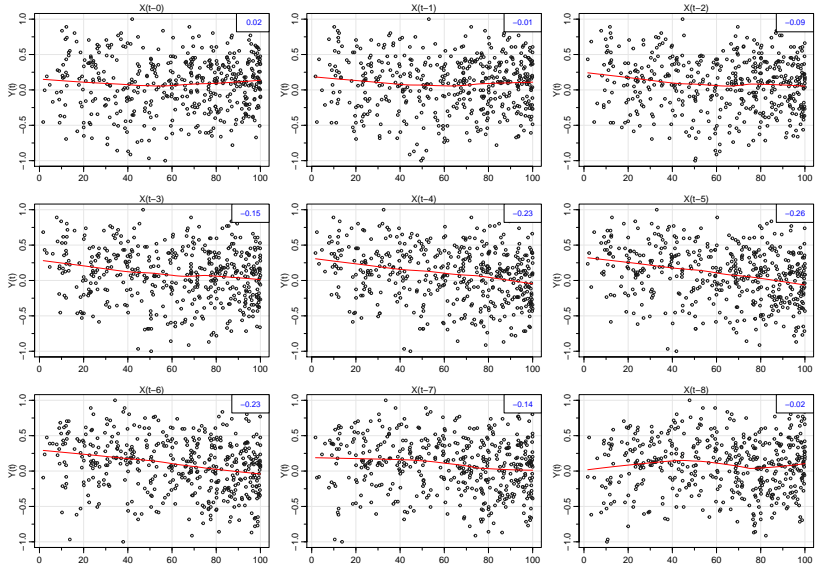


▶ Y & X - X leads Y - $\widehat{\text{Corr}}(X_t, Y_{t+h}) = \widehat{\text{Corr}}(X_{t-h}, Y_t), h > 0$;

▶ X & Y - Y leads X - $\widehat{\text{Corr}}(X_{t+h}, Y_t), h < 0$;

Scatterplots can also be employed:

```
astsa::lag2.plot(X, Y, max.lag = 8)
```



Significant lags could be selected by including the most dominant (i.e. lags with the most negative and/or the most positive correlation). Examining the coefficient significance may help further determine, whether some lags are insignificant.

Another option is **pre-whitening** the X_t variable, which is achieved by fitting an appropriate ARIMA model (usually with $q = 0$) on X_t . Then the series Y is filtered using the coefficients from the model on X_t as a `stats::filter()`. Finally, a **CCF** is plotted on the pre-whitened X_t model residuals and the filtered Y_t series.

The CCF pattern is affected by the underlying time series structures of the two variables and the trend each series has. It often is helpful to either de-trend (if it is a stationary process) and/or take into account the univariate ARIMA structure of X_t and Y_t before graphing the CCF.

Remark (3): Seasonality In Time Series Regression

We can look back at the Seasonal ARMA model to get an idea on which coefficients could be included in the model to account for seasonal correlation.

- ▶ If we look at SARIMA(0,0,0) \times (1,0,0)_s

$$(1 - \Phi_1 L^s) Y_t = \epsilon_t \iff Y_t = \Phi_s Y_{t-s} + \epsilon_t$$

A similar model can also be applied to a time series regression. For example, if we assume that the seasonal correlation comes both from Y and X . The model would look like:

$$Y_t = \phi_s Y_{t-s} + \beta_0 X_t + \beta_s X_{t-s} + \epsilon_t$$

- ▶ If we look at a SARIMA(1,0,0) \times (1,0,0)_s

$$(1 - \Phi_1 L^s)(1 - \phi L) Y_t = \epsilon_t$$

$$\iff Y_t = \phi_1 Y_{t-1} + \Phi_s Y_{t-s} - \phi_1 \Phi_s Y_{t-s-1} + \epsilon_t$$

$$\iff Y_t = \phi_1 Y_{t-1} + \Phi_s Y_{t-s} + \tilde{\delta} Y_{t-s-1} + \epsilon_t, \quad \tilde{\delta} = -\phi_1 \Phi_s$$

we could estimate a similar model for the time-series regression:

$$Y_t = \phi_1 Y_{t-1} + \phi_s Y_{t-s} + \phi_{s+1} Y_{t-s-1} + \beta_0 X_t + \beta_s X_{t-s} + \beta_{s+1} X_{t-s-1} + \epsilon_t$$