Lecture 12: Time Series

Thomas Chadefaux

Contents

1 The Fundamental Difference Between Cross-Sectional and Time-Series Data

In cross-sections, we assume that we have some underlying population and we observe a sample from that population. The implicit assumption is that this sample is a random sample.

In general, we need $E[\varepsilon_i|x_i] = 0$ AND $E[\varepsilon_i|x_j] = 0$ for all i, j. In cross-sections usually we only make the assumption that $E[\varepsilon_i|x_i] = 0$, because random sampling takes care of $E[\varepsilon_i|x_j] = 0$.

In time series data, however, the assumption of random sampling from an underling 'population' does not make as much sense. Instead, we have a (single) process that we observe over time. As a result, we have to be more rigorous about our assumptions and make it explicit that we are assuming $E[\varepsilon_i|x_j] = 0$.

2 Strict Exogeneity

$E[\varepsilon_t|x_s] = 0 \; \forall t, s$. Why might this assumption fail? Two main reasons

- Suppose we are interested in the effect of an increase in GDP on the political orientation of voters:

  $$orientation_{it} = \beta_0 + \beta_1 GDP_{it} + \varepsilon_{it}$$

  It is likely that a change in GDP at time t will affect political orientation at time $t+1$ as well, so that $\varepsilon_{it} = f(GDP_{it-1})$. This violates the assumption that $E[\varepsilon_t|x_s] = 0$. Why exactly is violation of this assumption a problem? The intuition is that GDP_t is taking most of the credit for the variation in $orientation_t$, when really some of it is explained by GDP_{t-1}. Just as in the case of endogeneity, we have biased coefficients.

  Luckily, there is an easy fix for it: include a lagged IV:

  $$orientation_{it} = \beta_0 + \beta_1 GDP_{it} + \beta_1 GDP_{it-1} + \varepsilon_{it}$$

- A more severe problem: feed-forward effect. Suppose that GDP is affected by political orientation. Then $GDP_{t+1} = f(orientation_t) = g(\varepsilon_t)$. Intuitively this is a problem because GDP_t is taking all the credit, when really $orientation_{t-1}$ explains some of the variance in $orientation_t$ This is a more severe problem because it is problematic to include future values of the independent variable as regressors. Luckily in large sample, OLS remains unbiased, and we only need to *weak* exogeneity condition.

A particular problem in which this problem arises is the case of lagged dependent variables:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x + \varepsilon \tag{1}$$

Here, even if we are willing to assume that $E[\varepsilon_t|y_s] = 0$ for all $s \neq t$, it must be the case that $E[\varepsilon_t|y_t] \neq 0$. Why? Note that

$$\begin{aligned} Cov(\varepsilon_t, y_t) &= Cov(\varepsilon_t, \beta_0 + \beta_1 y_{t-1} + \beta_2 x) \\ &= Cov(\varepsilon_t, \varepsilon) \\ &= Var(\varepsilon_t) \\ &= \sigma^2 \neq 0 \end{aligned}$$

This means that b_{OLS} is biased. Luckily b_{OLS} is still consistent (i.e., it converges to β in large samples), provided we are willing to assume weak dependence. More on this later. [1]

[1] The conditions for asymptotic efficiency of OLS are somewhat looser than the Gauss-Markov conditions, because Gauss-Markov meant that OLS was unbiased even in small samples. If we only care about consistency, then the set of conditions is:

- Linearity
- Stationarity + weak dependence (more on this below)
- $E[\varepsilon_{it}, x_{it}] = 0$, aka weak exogeneity assumption. Note that this is less demanding than the strict exogeneity asssumptions.
- No perfect collinearity (same as for Gauss Markov)

3 Stationarity

We care about stationarity because it is a requirement for the consistency of b_{OLS}. What do we mean by stationarity? Three conditions:

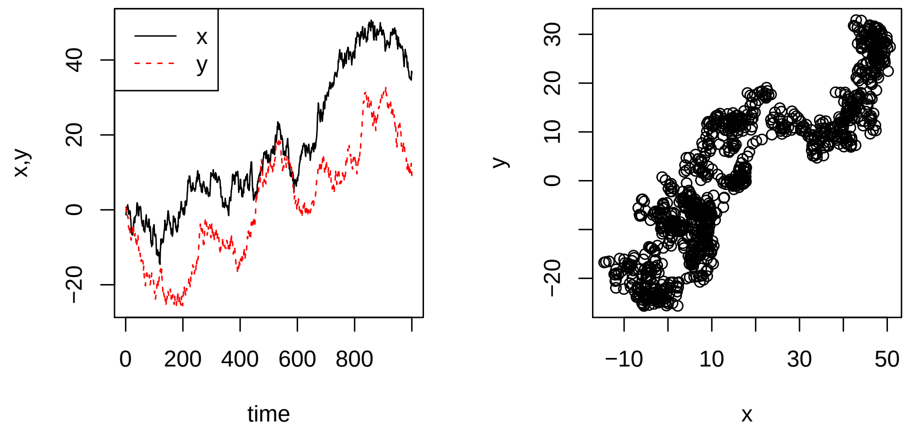
3.1 Mean Stationary

$E[x_t] = \mu$, i.e., it is a constant. It does not vary as a function of time. It varies around a stable value μ. Why do we care about this assumption? Suppose that x is trending up, whereas y remains constant. The problem then is that for low values of x, we might have $y = 2x$, but for high values of x we'll have $y = 0.5x$. I.e., there is a no way we can estimate a regression with constant coefficients. What if both are non-stationary? Again, the gap between the two variables might be changing over time.

Another important reason for wanting stationarity is the problem of *spurious regression* (Hendry 1980). Hendry tried to explain price levels over time, as a function of money supply. Finds an R^2 of 0.995. But then regresses price levels against a variable c_t, and finds an R^2 of 0.998, i.e., even better. BUT c_t is actually the cumulative rainfall in the UK, which obviously should have no effect on prices! But because both are increasing over time, it appears as if there was a relation between the two.[2]

[2] A rule of thumb for whether you have a problem of serial correlation: if $R^2 > DW$, where DW is the Durbin-Watson statistic, then you probably have a problem.

```r
x <- cumsum(rnorm(1000))
y <- cumsum(rnorm(1000))
```



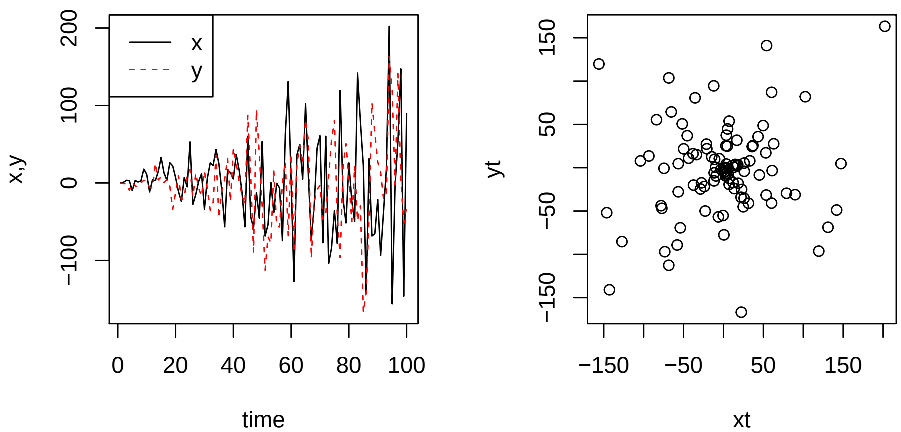
	Dependent variable:
	y
x	0.754*** (0.013)
Constant	−11.634*** (0.320)
R^2	0.781
Note:	*p<0.1; **p<0.05; ***p<0.01

3.2 Variance Stationary

$Var[x_t] = \sigma^2$, i.e., variance is also not a function of time. Why do we care? Suppose y_t is variance-stationary, but x is not. But if there really was a fixed coefficient relating x to y, then the variance of y should also be a function of x.

Another reason to care is that we can also face the issue of spurious regression here. Suppose this time that both y and x are nonstationary in variance. Then it might appear that there is a strong relationship between the two, when really there is not:

```
xt <- yt <- NULL
for(t in 1:100){xt <- c(xt, rnorm(1, sd = t)) }
```



	Dependent variable:
	yt
xt	0.124 (0.087)
Constant	−2.677 (5.230)
R^2	0.020
Note:	*p<0.1; **p<0.05; ***p<0.01

3.3 Covariance Stationary

$Cov[x_t, x_{t+h}] = f(h)$, but NOT a function of time. Why do we care? Suppose that y_t is covariance stationary, but x_t is not. If x_t really changes in its structure over time, but y_t does not, then how can y_t really be related to it?

In summary, we like stationarity for two reasons. The first, which we talked about here, is that there can be no linear relationship between y and x if stationarity is violated. I.e., our assumption of constant coefficients cannot hold without it. The second reason is more theoretical and I will not cover it here: we need stationarity for the law of large number and the central limit theorem to hold, which we need for inference

4 Weak Dependence

We care about weak dependence because it is another requirement for the consistency of b_{OLS}. Weak dependence means that:

$$Corr(x_t, x_{t+h}) \longrightarrow 0$$

as n goes to ∞. Two examples:

- Suppose that we have an the following process i.e.:

  $$y_t = \varepsilon_t + \theta\varepsilon_{t-1}$$

  Then

  $$y_{t+1} = \varepsilon_{t+1} + \theta\varepsilon_t$$

  But note that both y_t and y_{t-1} are a function of ε_t, and so $Corr(y_t, y_{t+1}) \neq 0$. But $Corr(y_t, y_{t+2}) = 0$. I.e., weak dependence is satisfied.[3]

- $$y_t = \rho y_{t-1} + \varepsilon_t.$$

  Then as long as $|\rho < 1|$, the process is weakly dependent. BUT if $|\rho < 1|$, then the condition is not satisfied (in fact, we are then dealing with a *random walk*).

[3] Note that this is an example of an MA(1) process—more on this below.

Why do we care about weak dependence?

5 MA(1), AR(1) and I(1) processes

5.1 MA(1) processes

$$y_{t+1} = \varepsilon_{t+1} + \theta\varepsilon_t$$

We call this process a Moving Average of order 1 (order 1 because I only have one lagged error term). E.g., votes in midterm elections, where ε_t might be the change in votes in the past election. Or change in oil price.

Note that the effect of the error propagates over time.

Note also that MA(1) processes are stationary and satisfy the weak dependence condition. That means that we can run OLS on them.

5.2 AR(1) processes

$$y_t = \rho y_{t-1} + \varepsilon_t.$$

This is an "autoregressive process" of order 1 (hence AR(1)). Note that in the MA(1) process, the shock only lasts 2 periods. Here the shock has an impact for a much longer time.

Note that assuming that $E[y_0] = 0$ and $|\rho| < 1$, then AR(1) is stationary in mean and variance and covariance.

5.3 AR(1) vs MA(1)

How can you determine whether you have an AR(1) process, or an MA(1) process, or maybe something else? First, you should plot the series. First, for both, you need to have a constant mean and variance. To differentiate between the 2, look at the covariance structure. In an MA(1) process,

$$corr(x_t, x_{t+h}) = \begin{cases} \frac{\theta}{1+\theta^2} & \text{if } h = 1 \\ 0 & \text{if } h > 1 \end{cases} \tag{2}$$

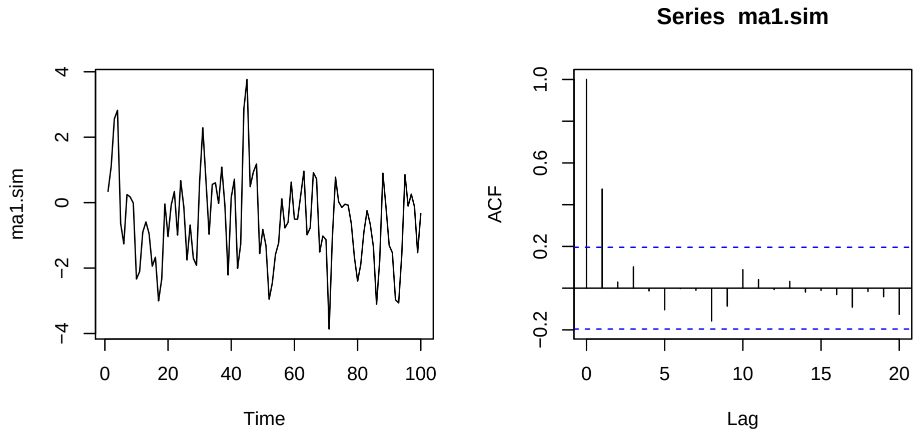
In an AR(1) process,

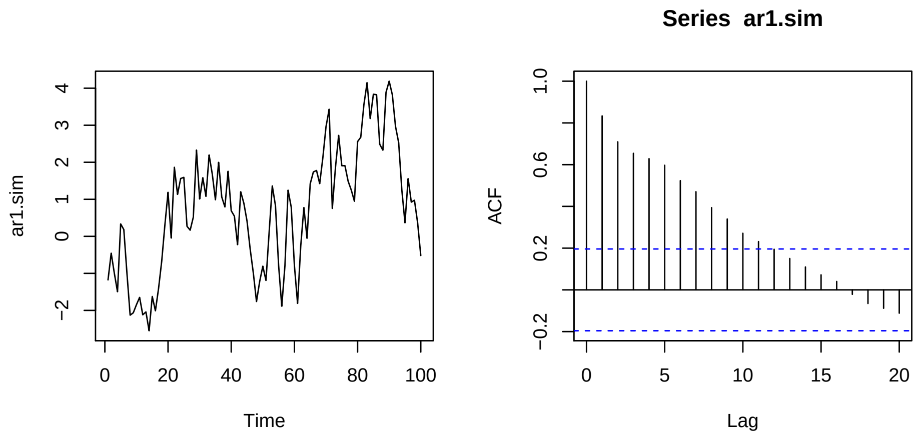
$$corr(x_t, x_{t+h}) = \rho^h \tag{3}$$

So to decide between the two processes, look at their correlogram. The MA(1) process should have 0 correlation beyond 1, whereas the AR(1 will still have correlation).

```
#Simulate an MA(1) process
ma1.sim <- arima.sim(model=list(ma=c(.9)),n=100)
plot(ma1.sim)
acf(ma1.sim)

#Simulate an AR(1) process
ar1.sim <- arima.sim(model=list(ar=c(.9)),n=100)
plot(ar1.sim)
acf(ar1.sim)
```





5.4 How to test for stationarity?

Dickey-fuller test and augmented Dickey-fuller test (not covered here)

5.5 I(1) processes: Highly persistent time series

One example is a random walk. Practical examples: interest rate, GDP, etc. These are variables that are NOT weakly dependent and NOT stationary. What to do? Take the difference. If I then get a stationary Time series, then I am dealing with a I(1) process (integrated of order 1).

6 Cointegration

We discussed the fact that regressing a non-stationary TS on another is usually a bad idea. But what if we can find a parameter γ such that once x is multiplied by γ, then the difference between the two processes is stationary. I.e.,

$$y - \gamma X_t = I(0)$$

6.1 Testing for cointegration

Run the regression, get estimates. If x and y are cointegrated, then the residuals should be I(0). So apply a Dickey-fuller test on residuals.

6.2 Why not use difference regression?

i.e., why not just take the first difference of both variable, to make them I(0)? A stable relationship in levels implies a stable relationship in differences, but the reverse is not necessarily true. Also, sometimes we are interested in levels.