Econometrics

Spring semester 2024

Lecturer: Nicolás Blampied

Lecture 3:

Further Issues in Using OLS with Time Series Data

Contents

Lecture 3

Stationarity and nonstationary time series

When we talk about stationarity, we think of a stochastic process $\{x_t : t = 1, 2, \dots\}$ for which for every collection of time indices $1 \leq t_1 < t_2 < \dots < t_m$ the joint distribution of $(x_{t1}, x_{t2}, \dots, x_{tm})$ is the same as the joint distribution of $(x_{t1+h}, x_{t2+h}, \dots, x_{tm+h})$ for all integer $h \geq 1$.

- An implication of stationarity is that choosing m=1 and t_1=1, x_t has the same distribution as x_1 for all t = 2, 3, … In simple words, it means the sequence is identically distributed.
- This is not enough, stationarity requires that the joint distribution of (x_1, x_2) (the first two terms in the sequence) is the same as the joint distribution of any (x_t, x_{t+1}) for any $t \geq 1$.
- This places no restrictions on how x_t and x_{t+1} are related to one another; indeed, they may be highly correlated. Stationarity does require that the nature of any correlation between adjacent terms is the same across all time periods.

Stationarity and nonstationary time series

Why is stationarity so important?

- When we want to understand the relationship between variables, we need to have some kind of stability.
- Otherwise, if we allow this relationship to change over time arbitrarily, then it is not possible to understand much about it.
- **At a more tangible level, we could say that when a process is stationary, we know that the first and the second moments are stationary (mean and variance).**
- If a process is nonstationary, then we cannot expect to make any inference, we would have what we call “spurious regressions”.

Stationarity and nonstationary time series

Weakly dependent time series

- Apart from stationarity, we also need to assume some sort of weekly dependence. A stationarity time series process $\{x_t : t = 1,2, \ldots\}$ is said to be weakly dependent if x_t and x_{t+h} are almost independent as h increases.

- In simple words, we say that Corr(x_t, x_{t+h}) tends to 0 as h tends to infinite (asymptotically uncorrelated).

- This weak dependence, even if difficult to understand, replaces the assumption of random sampling -implying the law of large numbers LLN and the central limit theorem (CLT)-.

- Stationarity and weakly dependence are key to justify the use of OLS and be able to make inference.

Stationarity and nonstationary time series

Probably the most popular example of a weakly dependent process is the following:

$$y_t = \rho_1 y_{t-1} + e_t\ ,\ \ t = 1, 2, \dots \qquad (1)$$

The starting point in the sequence is y_0 (at t=0), and $\{e_t: t = 1, 2, \dots\}$ is an i.d.d sequence with zero mean and variance σ_e^2. Also, we assume that e_t are independent of y_0 and $E(y_0) = 0$.

This process is called an **autoregressive process of order one [AR(1)]**.

What do we need for this model to be weakly dependent?

We need the stability condition, this is $|\rho_1| < 1$. This implies that $\{y_t\}$ is a stable AR(1).

Stationarity and nonstationary time series

To fit an autoregressive model on R is quite simple… let just take a break and try it…

Remember our data for air passengers?

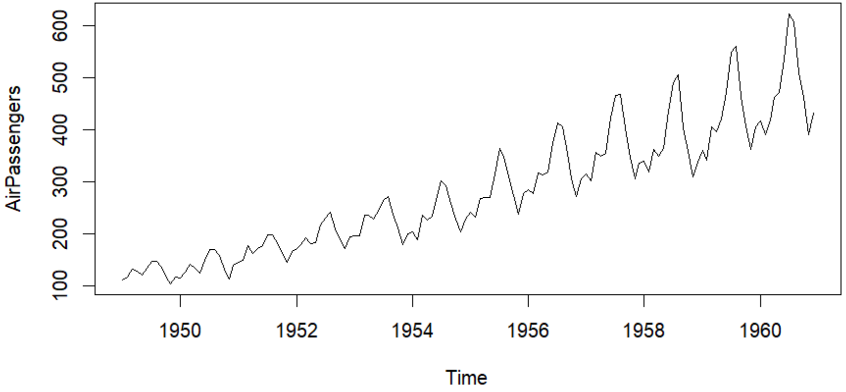

Figure 1: Air Passengers (retrieved from the R database)

So, what happen if we try to fit an AR(1) model into this data?

Stationarity and nonstationary time series

Well, this is what we get…

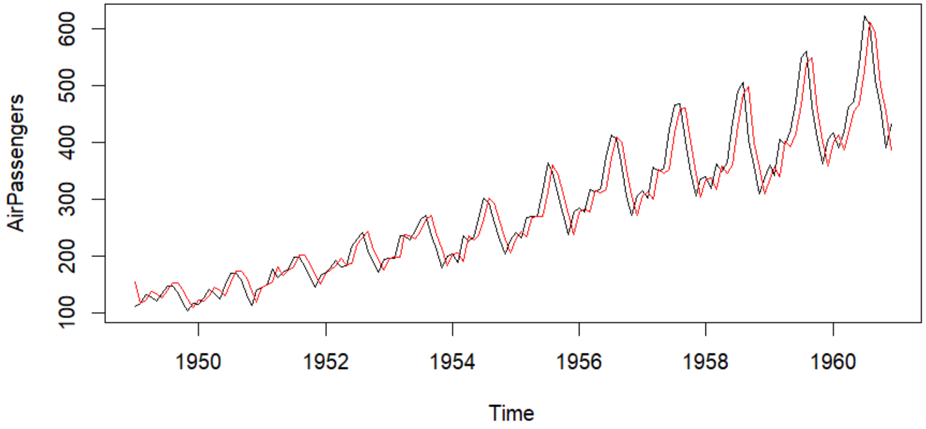

Figure 2: Air Passengers vs AR(1) fitted model

It looks like there is some room for improvement…

Do you think this would improve if we fit an AR(2)?

If we think of the value estimated for beta in the AR(1). Would this process be weakly dependent?

Asymptotic properties of OLS

- As we stated in Lecture 1, **sometimes not all the classical linear model assumptions hold**, in which case **we must appeal to large sample properties of OLS.**

- These assumptions, you will see, have a similar flavor to the ones we already know.

Assumption 1': Linearity and weak dependence

Here we assume the model is just as in Assumption 1, but we add that $\{(x_t, y_t): t = 1,2, \ldots\}$ is stationary and weakly dependent. In particular, the LLN and the CLT can be applied to sample averages.

- **The important extra assumption here is weakly dependence, which is required on both x_t and y_t and certainly puts restrictions on the joint distribution.**

Asymptotic properties of OLS

Assumption 2': No perfect collinearity

Do not worry, this is the same as in Lecture 1.

$$\mathrm{E}(u_t \,/\, X) = 0, \qquad t = 1, 2, \dots n.$$

Assumption 3': Zero conditional mean

The explanatory variables $x_t = (x_{t1}, x_{t2}, \dots x_{tk})$ *are contemporaneously exogenous. This means* $E[u_t/x_t] = 0$.

- Interestingly, this is a much weaker assumption than Assumption 3 from Lecture 1 because it puts no restrictions on the relationship between u_t and the explanatory variables on other periods.

- Since we assume stationarity, if exogeneity holds for one period, then it holds for all of them.

Asymptotic properties of OLS

Consistency of OLS

Remember that under Assumptions 1, 2 and 3 in Lecture 1 our estimator was unbiased? Well, here we talk about *consistency*.

Indeed, if Assumptions 1', 2' and 3' hold, we say that our OLS estimators are consistent. This means that plim $\widehat{\beta}_j = \beta_j$, $j = 0,1,...k$.

The simplest way to understand consistency is by thinking that our estimator tends to the true parameter as the sample increases and tends to infinite.

Asymptotic properties of OLS

Assumption 4': Homoskedasticity

The errors are contemporaneously homoscedastic, which means that $Var(u_t|x_t) = \sigma^2$.

Assumption 5': No serial correlation

$$\mathbf{Var}(u_t/X) = \mathbf{Var}(u_t/X) = \sigma^2, \mathbf{t = 1, 2, \ldots, n}$$

For all $t \neq s,\ E(u_t u_s|x_t x_s) = 0$.

$$\mathbf{Corr}(u_t, u_s\ /X) = 0$$

- It looks like these conditions are less restrictive than in Lecture 1.
- For instance, in Assumption 4' the restriction holds only for time t.
- The same happens with Assumption 5', since we condition only on the variables coinciding with u_t and u_s.

Asymptotic properties of OLS

Asymptotic normality of OLS

Under assumptions 1’-5’, the OLS estimators are asymptotically normally distributed. Further, the usual OLS standard errors, t statistics and F statistics are asymptotically valid.

Highly persistent time series

- **What happen if we violate some of the assumptions we saw in Lecture 1 and the data is not weakly dependent?**

- Basically, we are asking ourselves what should we do if our data is highly persistent or strongly dependent.

- **Remember that we just saw that the stability condition for an AR(1) process requires that $|\rho_1| < 1$?**

- **What would happen if $|\rho_1| = 1$?**

Highly persistent time series

- **A quite popular persistent time series is the so-called Random Walk (RW):**

$$y_t = y_{t-1} + e_t \,, \; t = 1, 2, \ldots \qquad (1)$$

- This process is called RW because y_t depends on a zero mean random variable independent from y_{t-1}.
- This makes the value at each t completely random.
- This is why the process is often referred as a drunkard's walk.
- So, how does this process look like?

Highly persistent time series

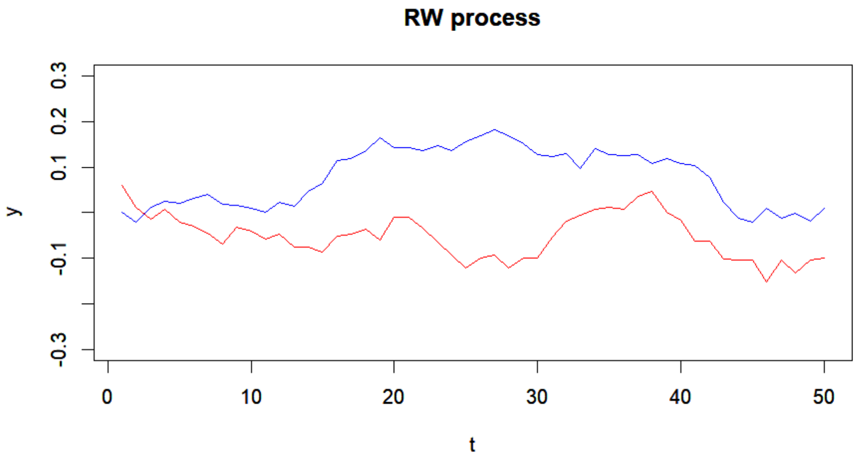

Figure 3: Two different realizations of a random walk

Let us move to R for a second and see whether we get the same simulations…

Highly persistent time series

- The problem with random walk process is that they are highly persistent.

- In consequence, it becomes impossible to make inference, since the best guess for tomorrow's value of the variable (or in 30 years time), will always be today.

- The problem is that, since $|\rho_1| = 1$, then the importance of today's values for the variable remain determinant while with $|\rho_1| < 1$, the weight of today values decrease over time.

- When we talk about RW processes, we are in fact talking about a specific case of what we called **unit root processes!**

- **So, what do we have to do when we work with highly persistent data?**

Transformation on highly persistent time series

First differences

- If a non-stationary process is integrated of order 1 -I(1)-, that means that the first difference of that process will be stationary.

- Forget about the technicalities, think of our random walk:

$$y_t = y_{t-1} + e_t \,, \ \mathrm{t} = 1, 2, \ldots \qquad (2)$$

- It is quite straightforward to see that the first difference of this process is stationary. Look:

$$y_t - y_{t-1} = \Delta y_t = e_t \,, \ \mathrm{t} = 1, 2, \ldots \qquad (3)$$

Transformation on highly persistent time series

- When we suspect that our series are non-stationary, it is a good idea to differentiate it, so to get a stationary process.

- If we work with logs, for example, we get:

$$\Delta \log(y_t) = \log(y_t) - \log(y_{t-1}) \qquad (4)$$

$$\Delta \log(y_t) = (y_t - y_{t-1})/y_{t-1} \qquad (5)$$

- This means that, quite simply, we can directly work with the rates of variation of a variable.

Transformation on highly persistent time series

- Another good thing of working with differences, is that it removes any trend in the data.

- Remember when we talked about detrending? We said we could include a trend in the regression.

- Well, working with first differences is another good way for detrend the data.

- Why don’t we see an example?

Transformation on highly persistent time series

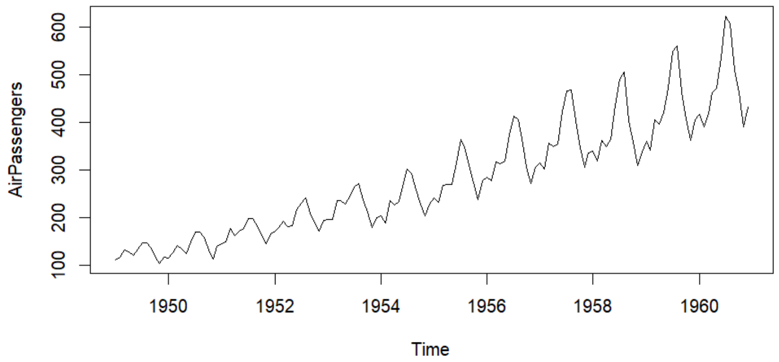

Figure 4: Air Passengers (retrieved from the R database)

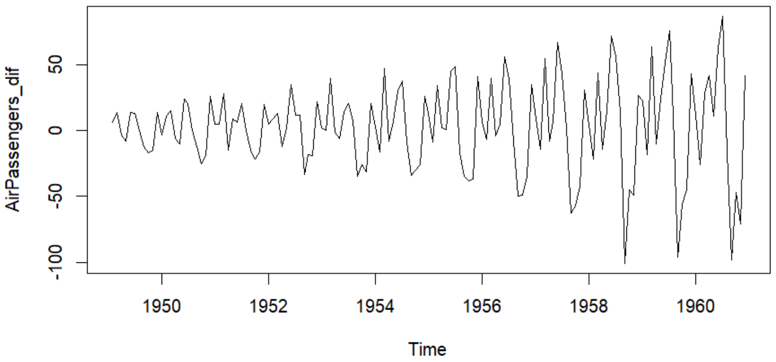

Figure 5: Air Passengers (first differences)

Transformation on highly persistent time series

- Normally, when working with time series, you may find the data in levels and in differences.
- Check, for instance, https://fred.stlouisfed.org/series/GDP.
- In these cases, you may decide whether to work with differences or not, depending on the stationarity of the series.
- Is there any formal procedure to check the stationarity of the series?
- Let us check a popular unit root test…

Unit root tests

- First, we need to understand that there are many unit root tests we could perform…
- Believe me that a good unit root test is the Augmented Dickey-Fuller test (ADF).
- The Dickey-Fuller test the null hypothesis that a unit root process is present in our data.
- Basically, if we run the test and we reject the null hypothesis, then we can be fairly sure that the series is stationary, and we can run our models.
- Let us study the ADF formally…

Unit root tests

- Think of the following AR(1) process:

$$y_t = \alpha + \rho y_{t-1} + \varepsilon_t \qquad (6)$$

- The ADF test consist of running the following equation:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \ldots + \delta_{p-1} \Delta y_{t-1+1} + \varepsilon_t \qquad (7)$$

Where α is a constant, β is a coefficient on a time trend, p is the lag order of the autoregressive process and $\delta_1 = \rho$-1. Note that when α=0 and β=0, then we are in the presence of a typical RW process.

Note that the hypothesis we test is whether γ=0 (stationarity), against the alternative of $\gamma < 0$ (non-stationarity).

Unit root tests

Note that the implementation of the ADF allows for three options:

- Include a constant or not
- Include a trend
- Include lags

A visual inspection of the data is key in order to choose the right selection and understand with what kind of data we are dealing.

Let us move to R for a bit and prepare for our seminar…

ARIMA models

- ARIMA models are quite popular univariate timer series models.
- Based on the Box-Jenkins theoretical framework.
- In order to understand ARIMA models, we need to think about three parameters (p, d and q):

ARIMA(p, d, q)

Autoregressive part

Integration (differencing)

Moving average

ARIMA models

- Each parameter tells us the order of the process.
- Remember that working with time series requires the data to be stationary.
- In consequence, if the data to is not stationary, when working with typical functions R will perform this operation for you.
- Only then the data is usable you will be able to identify the other two parameters.
- Do not worry, we will go over a typical ARIMA model on R together! However, let us see some examples:

ARIMA models

Some examples:

- **ARIMA(1,0,0)** $\longrightarrow$ $y_t = \beta_0 + \beta_1 y_{t-1} + u_t$
- **ARIMA(1,0,1)** $\longrightarrow$ $y_t = \alpha_0 + \beta_1 y_{t-1} + \delta_1 u_{t-1} + u_t$
- **ARIMA(0,0,1)** $\longrightarrow$ $y_t = \alpha_0 + \delta_1 u_{t-1} + u_t$
- **ARIMA(2,0,1)** $\longrightarrow$ $y_t = \alpha_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \delta_1 u_{t-1} + u_t$
- **ARIMA(1,0,2)** $\longrightarrow$ $y_t = \alpha_0 + \beta_1 y_{t-1} + \delta_1 u_{t-1} + \delta_2 u_{t-2} + u_t$

ARIMA models

ARIMA models are quite flexible and allow for different combinations

We need to inspect the series to have an idea of the correct specification

Some functions help us choose the right structure

Through a function that employs the Kalman-Filter, we can have good forecasts!

ARIMA models

Univariate models

Multivariate models

ETS

ARIMA models

Vector autoregressive models (VAR)

Structural VAR

Main takeaways

- What is stationarity? What is weekly dependence? Why are they important?
- What does it mean asymptotic consistency?
- What is a R-W process? Is it stationary? What is the problem about this process?
- What test can we use to test stationarity? What is the equation for this test?
- How can we deal with non-stationary data?
- When do we use ARIMA models? What do their parameters represent?

Seminar 3

Seminar 3

- Using data on CPI (you can retrieve it from our course materials), proceed to:

1. Load the data and plot the time series. Describe it briefly in terms of the existence of trends, constants and preliminary ideas of stationarity (1 point).

2. Perform an ADF test and report the results. Perform the test both with and without a trend, and using 0, 1 and 12 lags. Interpret them briefly and justify which test applies in your opinion (1 point).

3. Differentiate the data and compare it to the original series. Plot both of them and describe them (1 point).

4. Perform an ADF on the new differentiated series. Explain whether now the series is stationary. (1 point).

5. Remember the AirPassenger dataset from last week? Well, estimate an ARIMA model using the auto.arima() function. Use both the quick and the more complex option (using stepwise, trace and approximation options for your convenience). What model does the ARIMA function suggest? Do the results change when using these latter options? Produce a forecast for the next year with both models identified. (1 point)

6. Deliver the codes and a brief report before March 15. (1 point)