Section 4: Time Series Models

Maria Carolina Caetano

February 9, 2006

Contents

1 General Stuff

1.1 Announcements

- Problem Set 1 is due today. Remember to leave it in my table before you leave the section. You can also put it in the mailbox of name GREENBAUM in 612 Evans until 4 p.m.
- Graded problem sets will be left in one of the doors of 608 Evans.

1.2 Section Preamble

We'll do exactly as we did in the last section: work our way through relaxing a certain assumption of the OLS model. This time we will try to relax the hypothesis of the data being i.i.d. What would happen if we didn't assume that the data we have is a sample drawn independently from the same distribution? Just as we did last time, we have to ask ourselves two questions:

1. Where was the iid assumption used?

Well, if you remember what we did last section, you will recall that we needed the iid assumption in order to apply the Law of Large Numbers, which helped prove consistency or convergence in probability in many cases, and also to apply the Central Limit Theorem, which gave us the asymptotic distribution of the OLS estimator. So, if we drop the iid assumption, we are going to have to go over all the results that we had before and see what we can do to still have consistency and an asymptotic distribution we can work with.

2. How can we remedy that? What can we do in order to still have a tractable model?

When we say iid it means INDEPENDENTLY and IDENTICALLY DISTRIBUTED. These are 2 assumptions. Exercise 13.5 in the problem set works with dropping the "identically distributed" part, while keeping the "independent" part. It still requires plenty of other conditions, but you should make sure that you have a good sense of what was going on in that exercise.

Independence is a very tricky assumption to drop, and it is exactly what we are doing today. We will still be assuming "identically distributed". Actually some kind if "identically distributed" that is a little different. We will require something called **Covariance Stationarity.** It is not as restrictive as "identically distributed", because we don't require that all the sample have the same distribution, only the same mean and variance. However, it is a stronger condition, because we impose a restriction in the relationship between the elements in the sample, through the covariance condition in the definition of stationarity. If we have that, we will still be able to get consistence and asymptotic normality. Then, if we impose even more structure in the relationship among the variables, for instance a Time Series kind of behavior, we can get clear asymptotic results for the OLS estimator, which happen to be very similar to the results we are already used to see.

2 Some math concepts and results

We define here only Covariance Stationarity, or Weak Stationarity. You can see the definition of Strong Stationarity in Prof. Powell's notes, since we won't need it for this section.

Definition 1 *A sequence of random variables $\{y_t\}$ is said to be **Weakly Stationary**, or **Covariance Stationary** if:*

$E(y_t) = \mu$

$Var(y_t) = \sigma^2 = \gamma_y(0)$
$Cov(y_t, y_s) = \gamma_y(t-s) = \gamma_y(|t-s|)$

In words, we want the mean and the variance to be constant (and in particular, they have to exist), and the covariance must depend only on the distance of the variables in time.

Definition 2 *(vectors) A sequence of random variables $\{y_t\}$ is said to be **Weakly Stationary**, or **Covariance Stationary** if:*
$E(y_t) = \mu$
$Var(y_t) = \Sigma = \Gamma_y(0)$
$Cov(y_t, y_s) = E[(y_t - \mu)(y_s - \mu)'] = \Gamma_y(t-s) = [\Gamma_y(s-t)]'$

Definition 3 *An stationary sequence of random variables $\{y_t\}$ is said to be **Weakly Ergodic** if*
$\frac{1}{T}\sum_{t=1}^{T} y_t \longrightarrow_p E(y_t) = \mu$

Theorem 4 *(Weak Ergodic Theorem) If $\{y_t\}$ is Weakly Stationary, with*
$E(y_t) = \mu$
$Cov(y_t, y_{t+s}) = Cov(y_t, y_{t-s}) = \gamma_y(s)$
and if:
$\sum_{s=-\infty}^{\infty} |\gamma_y(s)| < \infty$
then
$\overline{y}_T = \frac{1}{T}\sum_{t=1}^{T} y_t \longrightarrow_p \mu$*, which means that $\{y_t\}$ is Weakly Ergodic.*

The proof of this theorem is easy, but it may take a bit of your time if it's the first time you are dealing with time series. I suggest you go over this once in Prof. Powell's notes. It will be very useful to already have the intuition for this kind of manipulation if you plan to take Econ 241B.

Theorem 5 *Given many regularity conditions (all satisfied by the processes we are going to be dealing with in this course),*
$\sqrt{T}\left(\overline{y}_T - \mu\right) \longrightarrow_d N\left(0, V_0\right)$
where $V_0 = \lim_{T\to\infty} Var\left(\sqrt{T}\left(\overline{y}_T - \mu\right)\right)$
$= \sum_{s=-\infty}^{\infty} \gamma_y(s)$

Definition 6 *A sequence of random variables* $\{\epsilon_t\}$ *is called* ***White Noise****, and denoted* $\epsilon_t \sim WN(\sigma^2)$ *if it is weakly stationary with:*

$E(\epsilon_t) = 0$

$Var(\epsilon_t) = \sigma^2$

$Cov(\epsilon_t, \epsilon_s) = 0, \;\; if \;\; t \neq s$

3 Time Series Models

Now we can go back to the OLS model. But now, instead of imposing that the data are iid, we are going to say that it obeys a specific Time Series Model. We will explore some basic models that you dealt with in class, and probably before that.

3.1 MA(1)

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

where $\epsilon_t \sim WN(\sigma^2)$ and $Cov(\epsilon_t, y_{t-s}) = 0$, if $s \geq 1$.

Now, this process is stationary, since:

$$\begin{aligned} E(y_t) &= \mu \\ Var(y_t) &= \sigma^2(1+\theta^2) \\ Cov(y_t, y_{t-s}) &= \gamma_y(s) = \sigma^2\theta, \text{ if } s = 1 \\ &= 0 \text{ if } s > 1 \end{aligned}$$

Since $\sum_{s=-\infty}^{\infty} |\gamma_y(s)| = |\sigma^2(1+\theta^2)| + 2\,|\sigma^2\theta| = \sigma^2(1+\theta^2) + \sigma^2\,|\theta| < \infty$, we know that:

- $\overline{y}_T = \frac{1}{T}\sum_{t=1}^{T} y_t \longrightarrow_p \mu$
- $\sqrt{T}\left(\overline{y}_T - \mu\right) \longrightarrow_d N\left(0, \sigma^2(1+\theta)^2\right)$

3.2 AR(1):

$$y_t = \alpha + \phi y_{t-1} + \varepsilon_t$$

where $\epsilon_t \sim WN(\sigma^2)$ and $Cov(\epsilon_t, y_{t-s}) = 0$, if $s \geq 1$.

If $|\phi| < 1$, this process is stationary, since:

$$\begin{aligned} E(y_t) &= \alpha + \phi E(y_{t-1}) \\ &\implies (1 - \phi L)E(y_t) = \alpha \\ &\implies E(y_t) = \frac{\alpha}{1 - \phi L} = \alpha \sum_{t=0}^{\infty} \phi^t L^t = \alpha \sum_{t=0}^{\infty} \phi^t = \frac{\alpha}{1 - \phi} \end{aligned}$$

$$\begin{aligned} Var(y_t) &= \phi^2 Var(y_{t-1}) + \sigma^2 \\ &\implies (1 - \phi^2 L)Var(y_t) = \sigma^2 \\ &\implies Var(y_t) = \frac{\sigma^2}{1 - \phi^2 L} = \sigma^2 \sum_{t=0}^{\infty} \left(\phi^2\right)^t L^t = \sigma^2 \sum_{t=0}^{\infty} \left(\phi^2\right)^t = \frac{\sigma^2}{1 - \phi^2} \end{aligned}$$

$$\begin{aligned} Cov(y_t, y_{t-1}) &= \phi Var(y_{t-1}) + Cov(\varepsilon_t, y_{t-1}) \\ &= \phi \frac{\sigma^2}{1 - \phi^2} \end{aligned}$$

so, by induction,

$$Cov(y_t, y_{t-s}) = \phi^s \frac{\sigma^2}{1 - \phi^2}$$

Since none of them depends on t, stationarity follows.

Also, if $|\phi| < 1$, we know that $\sum\limits_{s=-\infty}^{\infty} |\gamma_y(s)| = \sum\limits_{s=-\infty}^{\infty} \phi^s \frac{\sigma^2}{1-\phi} = \frac{\sigma^2}{1-\phi} \sum\limits_{s=-\infty}^{\infty} \phi^s = \frac{\sigma^2}{(1-\phi)^2} < \infty$, which implies that the process is ergodic, so:

- $\overline{y}_T = \frac{1}{T} \sum\limits_{t=1}^{T} y_t \longrightarrow_p \frac{\alpha}{1-\phi}$

- $\sqrt{T}\left(\overline{y}_T - \frac{\alpha}{1-\phi}\right) \longrightarrow_d N\left(0, \frac{\sigma^2}{(1-\phi)^2}\right)$
  and the variance was calculated doing:

$$\lim_{T\to\infty} \quad Var\left(\sqrt{T}\left(\overline{y}_T - \frac{\alpha}{1-\phi}\right)\right) = \sum_{s=-\infty}^{\infty}\gamma_y(s) =$$
$$= \frac{\sigma^2}{1-\phi^2} + 2\frac{\sigma^2}{1-\phi^2}\frac{\phi}{(1-\phi)} = \frac{\sigma^2}{(1-\phi)^2}$$

3.3 ARMA(1,2) example:

Suppose the model is of the kind:

$$y_t = 1 + \frac{1}{2}y_{t-1} + \varepsilon_t + 2\varepsilon_{t-1}$$

where $\epsilon_t \sim WN(\frac{3}{4})$ and $Cov(\epsilon_t, y_{t-s}) = 0$, if $s \geq 1$.
Then, if the process is stationary:

$$\begin{aligned} E(y_t) &= 2 \\ Var(y_t) &= 5 \\ Cov(y_t, y_{t-s}) &= \gamma_y(1) = \left(\frac{1}{2}\right)^{s-1}\frac{25}{4} \end{aligned}$$

So, since $|\phi| = 1/2 < 1$, we know that $\sum_{s=-\infty}^{\infty}|\gamma_y(s)| = 5+2\sum_{s=1}^{\infty}\left(\frac{1}{2}\right)^{s-1}\frac{25}{4} = 5+\frac{25}{2}\sum_{s=0}^{\infty}\left(\frac{1}{2}\right)^{s} = 30 < \infty$, which implies that the process is ergodic, so:

- $\overline{y}_T = \frac{1}{T}\sum_{t=1}^{T} y_t \longrightarrow_p 2$
- $\sqrt{T}\left(\overline{y}_T - 2\right) \longrightarrow_d N\left(0, 30\right)$

4 Estimation:

Here we arrive at the core of this section, which is the answer to question 2 in the beginning of the notes. How can we still do hypothesis testing when we estimate the parameters using OLS? Well, if we had a model written like this:

$$y_t = \phi y_{t-1} + \gamma x_t + \varepsilon_t + \theta\epsilon_{t-1}$$

where $\epsilon_t \sim WN(\sigma^2)$ and $Cov(\epsilon_t, y_{t-s}) = 0$, if $s \geq 1$, and the x_t are iid just like we had before. Then the traditional OLS case is just a particular case with $\phi = \theta = 0$. This means that we relaxed the iid hypothesis and situated OLS inside a more flexible category. These models are called ARMAX.

Very well, if we estimate all the coefficients by OLS, how could we test hypothesis? We need an asymptotic distribution! This is a harder thing to get than in the old OLS model, but the results exist and in general they will look very close to the results we already know. We will go over an example of an AR(1) process, so you can figure out how things work in time series, but the AR(p) cases are very similar, and the result will be the same, as you can see in Prof. Powell notes.

4.1 AR(1)

Estimating ϕ

Suppose now that we are in the AR(1) model, but we will suppose $\alpha = 0$, to make matters simple, and that $|\phi| < 1$. If you estimate ϕ using OLS, you will obtain:

$$\begin{aligned}
\widehat{\phi} &= \left(\sum_{t=1}^{T} y_{t-1}^2\right)^{-1} \sum_{t=1}^{T} y_{t-1}y_t \\
&= \left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} y_{t-1}y_t
\end{aligned}$$

And we would like to know two things:

- Is $\widehat{\phi}$ consistent?
- What is the limiting distribution of $\sqrt{n}\left(\widehat{\phi} - \phi\right)$?

To answer both questions we will need to discover the asymptotic distribution of $\frac{1}{T}\sum\limits_{t=1}^{T} y_{t-1}^2$

4.1.1 Asymptotics for $\frac{1}{T}\sum\limits_{t=1}^{T} y_{t-1}^2$

We suppose that not only are the y_t stationary, but also the y_t^2 (Strong Stationarity plus the existence of fourth moments is sufficient, but not necessary) So, using the Ergodic Theorem,

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \longrightarrow_p E(y_t^2)$$

But in our case

$$E(y_t^2) = Var(y_t) = \frac{\sigma^2}{1-\phi^2}$$

4.1.2 Consistency

$\widehat{\phi} - \phi = \left(\frac{1}{T}\sum\limits_{t=1}^{T} y_{t-1}^2\right)^{-1} \frac{1}{T}\sum\limits_{t=1}^{T} y_{t-1}\varepsilon_t$

But $y_{t-1}\varepsilon_t$ is itself a stationary sequence, since:

$$\begin{aligned}
E(y_{t-1}\varepsilon_t) &= E(y_{t-1})E(\varepsilon_t) = 0 \\
Var(y_{t-1}\varepsilon_t) &= E(y_{t-1}^2\varepsilon_t^2) = E(y_{t-1}^2)E(\varepsilon_t^2) = \frac{\sigma^2}{1-\phi^2}\sigma^2 < \infty \\
Cov(y_{t-1}\varepsilon_t, y_{t-1-s}\varepsilon_{t-s}) &= E(y_{t-1}y_{t-1-s}\varepsilon_{t-s})E(\varepsilon_t) = 0
\end{aligned}$$

so again by LLN for stationary random variables:

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \longrightarrow_p E(y_{t-1}\varepsilon_t) = 0$$

so, since the denominator converges to something positive, the numerator converges to zero, and by Slutsky (since both converge to constants, so one of them can be identified to be a convergence in distribution) we have:

$$\widehat{\phi} - \phi \longrightarrow_p 0$$

and consistency follows.

4.1.3 Limiting distribution

$$\sqrt{T}\left(\widehat{\phi} - \phi\right) = \sqrt{T}\left(\sum_{t=1}^{T} y_{t-1}^2\right)^{-1} \sum_{t=1}^{T} y_{t-1}\varepsilon_t = \underbrace{\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2\right)^{-1}}_{A} \underbrace{\sqrt{T}\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t}_{B}$$

We already know the result for A, now, for B we will need to trust that it satisfies the necessary conditions (which are above this course level, but are satisfied by the AR(1) process). Then,

$$\sqrt{T}\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \longrightarrow_d N(0,V)$$

Now, we have to calculate this variance, which is:

$$\lim_{T\to\infty} Var\left(\sqrt{T}\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t\right)$$

so:

$$Var\left(\sqrt{T}\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t\right) = E\left(\left(\sqrt{T}\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t\right)^2\right)$$

however, the crossing terms are all of the kind $y_{t-1}\varepsilon_t, y_{t-1-s}\varepsilon_{t-s}$, and we know from above that they have zero mean (Covariance =0). So we only worry about the simple terms:

$$\begin{aligned}
&= \frac{1}{T}\sum_{t=1}^{T} E(y_{t-1}^2\varepsilon_t^2) \\
&= \frac{1}{T}\sum_{t=1}^{T} \frac{\sigma^2}{1-\phi^2}\sigma^2 \\
&= \frac{\sigma^4}{1-\phi^2}
\end{aligned}$$

so the limit when $T \to \infty$ is

$$V = \frac{\sigma^4}{1-\phi^2}$$

and using Slutsky,

$$\begin{aligned}
\sqrt{T}\left(\widehat{\phi}-\phi\right) &= \left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2\right)^{-1} \sqrt{T}\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \\
&\longrightarrow_d \left(\frac{\sigma^2}{1-\phi^2}\right)^{-1} N\left(0, \frac{\sigma^4}{1-\phi^2}\right) \\
&\sim N(0, 1-\phi^2)
\end{aligned}$$

4.1.4 Pivotal Statistics:

The OLS estimator is consistent, so

$$\begin{aligned} \widehat{\phi} &\longrightarrow {}_p\phi \\ &\Longrightarrow \left(\sqrt{1-\widehat{\phi}^2}\right)^{-1} \longrightarrow_p \left(\sqrt{1-\phi^2}\right)^{-1} \end{aligned}$$

by the Continuous Mapping Theorem

And now, we can do:

$$\frac{\sqrt{T}\left(\widehat{\phi}-\phi\right)}{\sqrt{1-\widehat{\phi}^2}} \longrightarrow_d \frac{1}{\sqrt{1-\phi^2}} N(0,1-\phi^2) \sim N(0,1)$$

and the test rejects the null hypothesis

$$H_0 : \widehat{\phi} = \phi_0$$

at the 5% confidence level if:

$$\left| \frac{\sqrt{T}\left(\widehat{\phi}-\phi\right)}{\sqrt{1-\widehat{\phi}^2}} \right| > 1.96$$

4.1.5 Comparison with known results

Just to show you that what you see is the same as the result you can find in the class notes:

$$\begin{aligned} \sqrt{T}\left(\widehat{\phi}-\phi\right) &\longrightarrow {}_dN(0,1-\phi^2) \\ &\sim N\left(0, \left(\frac{\sigma^2}{1-\phi^2}\right)^{-2} \frac{\sigma^4}{1-\phi^2}\right) \\ &\sim N\left(0, \left(\frac{\sigma^2}{1-\phi^2}\right)^{-1} \sigma^2\right) \\ &\sim N(0, \sigma^2 \left(E(y_{t-1}^2)\right)^{-1}) \end{aligned}$$

so, since we proved that $\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \longrightarrow_p E(y_t^2) = E(y_{t-1}^2)$

we have that:

$$\left(\sqrt{\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2}\right)^{-1} \sqrt{T}\left(\widehat{\phi}-\phi\right) \longrightarrow_d N(0,\sigma^2)$$

so if we allow the "asymptotic" treatment:

$$\left(\sqrt{\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2}\right)^{-1} \sqrt{T}\left(\widehat{\phi}-\phi\right) \overset{A}{\sim} N(0,\sigma^2)$$

$$\Longrightarrow \quad \sqrt{T}\left(\widehat{\phi}-\phi\right) \overset{A}{\sim} N\left(0,\sigma^2\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2\right)^{-1}\right)$$

which is equivalent to the

$$N(0,\sigma^2\left(\frac{X'X}{T}\right)^{-1})$$

that we are used to see.