11. Further Issues in Using OLS with TS Data

- With TS, including lags of the dependent variable often allow us to fit much better the variation in y
- Exact distribution theory is rarely available in TS applications, because we often have y_{t-1} as one of the regressors and so violate $E(u_t|X) = 0$.
- LST can be used to show consistency and asymptotic normality as long as $E(x_t' u_t) = 0$
- However, LLN and CLTs for TS are more complex than those for i.i.d. samples.
- We need to formalize some notions of dependence for TS.

Definition: $\{x_t\}$ is (strictly) stationary if the joint distribution of $(x_{t_1}, x_{t_2}, \cdots, x_{t_m})$ is the same as that of $(x_{t_1+h}, x_{t_2+h}, \cdots, x_{t_m+h})$ for all subsets $\{t_1, t_2, \cdots, t_m\} \in T$ and $\forall h \in \mathbb{Z}$.

- This says that all finite dimensional distributions of the TS are invariant to translation.
- Clearly, i.i.d. sequences are stationary
- Stationarity imposes no restrictions on dependence, e.g.

$$\{x_t\} = \begin{cases} (0,0,0,\cdots) & \text{w.p. } 1/2 \\ (1,1,1,\cdots) & \text{w.p. } 1/2 \end{cases}$$

Definition: $\{x_t\}$ is nonstationary if it is NOT stationary

- Clearly, lots of ways to be nonstationary, eg trending data, or distributions that vary over time

Definition: $\{x_t\}$ is covariance (aka weakly or second-order) stationary if

(i) $E(x_t^2) = \gamma(0) < \infty$

(ii) $E(x_t) = \mu$ (a constant)

(iii) $\forall (t, h) \;\; E(x_t x_{t+h}) = \gamma(h)$ (doesn't depend on t)

Rk:

Covariance stationary is NOT a special case of strictly stationary (strictly stationary processes need not have any moments–e.g. a sequence of i.i.d. Pareto r.v's)

There are many notions of weak dependence for TS that allow us to derive LLNs and CLTs. Most require some serious probability theory just to state them.

a. A simple notion is *asymptotically uncorrelated*, i.e. $corr(x_t, x_{t+h}) \to 0$ as $h \uparrow \infty$. Although easy to describe, this isn't useful for nonlinear functions.

b. A TS is *m-dependent* if x_t and x_{t+h} are independent for $|h| > m$. Easy to describe and a good model for thinking about TS.

c. Other notions include *ergodic*, α- (or β-) *mixing*, *martingales*, *mixingales*

- We can construct LLN and CLTs for each of these notions. For example: If $\{x_t\}$ is stationary and ergodic, and $E|x_1| < \infty$, then

$$plim\,\frac{1}{n}\sum_{1}^{n} x_t = E(x_1)$$

Examples of weakly dependent processes

(0) White noise: $\{e_t\}$ where $Ee_t = 0$ and $Ee_s e_t = \sigma_e^2 \delta_{st}$

- $E(e_t) = 0$
- $V(e_t) = \sigma_e^2 < \infty$
- $cov(e_s, e_t) = 0 \quad s \neq t$

White noise is the building block for a rich collection of processes. If $\{e_t\}$ is an i.i.d. sequence, we'll call it *i.i.d. white noise.* Q: Why "white noise"?

(1) MA(1) process: $x_t = e_t + \alpha_1 e_{t-1}$ $\{e_t\}$ is white noise

- $E(x_t) = E(e_t) + \alpha_1 E(e_{t-1}) = 0$
- $V(x_t) = \sigma_e^2 + \alpha_1^2 \sigma_e^2$
- $cov(x_t, x_{t-1}) = \alpha_1 \sigma_e^2$
- $cov(x_t, x_{t-h}) = 0 \quad |h| > 1$

Rks:

- MA stands for *moving average*
- If $\{e_t\}$ is i.i.d. white noise, then $\{x_t\}$ is m-dependent
- Using the lag operator, we can write

$$x_t = \alpha(L) e_t \qquad \alpha(L) = 1 + \alpha_1 L$$

(2) MA (q) process: $x_t = \alpha(L)e_t \quad \alpha(L) = 1 + \alpha_1 L + \cdots + \alpha_q L^q$

- So x_t is a weighted moving average of current and up to q lags of the white noise sequence

$$x_t = e_t + \alpha_1 e_{t-1} + \cdots + \alpha_q e_{t-q}$$

- Every MA (q) process is covariance stationary
- More generally, linear combinations of (covariance) stationary processes are (covariance) stationary
- (Almost) every covariance stationary process is an MA(∞) process (the exceptions are series with *linearly deterministic* components, i.e. series that have contributions that are random across different realizations but perfectly predictable from a given sample path)

(3) AR(1) process: $x_t = \rho_1 x_{t-1} + e_t$

- To show that it is stationary, write it as an MA(∞) process.
- Using the lag operator

$$\rho(L)x_t = e_t \qquad \rho(L) = 1 - \rho_1 L$$

- Does it make any sense to write

$$x_t = \frac{1}{\rho(L)} e_t \text{ ?}$$

- Consider

$$\frac{1}{1 - \rho_1 L} = 1 + \rho_1 L + \rho_1^2 L^2 + \cdots$$

  The RHS looks like it might make sense if $|\rho_1| < 1$. We can formalize this intuition (because $\mathcal{L}^2$space is complete)
- AR(1) process is (covariance) stationary iff $|\rho_1| < 1$

(4) AR(p) process: $\rho(L)x_t = e_t \quad \rho(L) = 1 - \rho_1 L - \cdots - \rho_p L^p$

- A weighted average of current and p lags of x_t is white noise

$$x_t = \rho_1 x_{t-1} + \cdots + \rho_p x_{t-p} + e_t$$

- To see if is stationary, factor the autoregressive polynomial

$$\rho(L) = \prod_{i=1}^{p}(1 - \theta_i L)$$

where (in general) θ_i is complex-valued.

- if $|\theta_i| < 1$ $(\forall i)$ we can invert $\rho(L)$ factor by factor
- If $z \in \mathbb{C}$, $|z|$ is its *modulus*, i.e. *sqrt*$(\mathrm{Re}^2(z) + \mathrm{Im}^2(z))$
- If $(1 - \theta_i z_i) = 0$, z_i is called a *root* of the polynomial $\rho(z)$
- $|\theta_i| < 1 \Leftrightarrow |z_i| > 1$, so we say "$x_t$ is covariance stationary iff roots of $\rho(z)$ lie strictly outside the unit circle"

(5) ARMA(p,q) process: $\rho(L)x_t = \alpha(L)e_t$

- Here p refers to the order of the autoregressive polynomial and q to the order of the moving average polynomial:

$$\rho(L) = 1 - \rho_1 L - \cdots - \rho_p L^p$$

$$\alpha(L) = 1 + \alpha_1 L + \cdots + \alpha_q L^q$$

- x_t is covariance stationary iff roots of $\rho(z)$ lie strictly outside the unit circle
- Note that if roots of $\alpha(z)$ lie strictly outside the unit circle, called the *invertibility condition*, then we can also write an MA(q) process as an AR(∞) process

LS Properties of OLS with TS data ("GM" Assumptions)

- TS.1$'$: $\{(y_t, x_t)\}$ is stationary and weakly dependent (so that LLN and CLTs apply to sample averages), and

$$y_t = x_t\beta + u_t$$

- TS.2$'$: X has full column rank (w.p.1). Moreover, the eigenvalues of $X'X/n$ are bounded below at a value that is strictly above zero

- Under TS.1$'$-TS.2$'$, we can write (w.p.1)

$$\widehat{\beta} = \beta + \left(\frac{1}{n}\sum x_t' x_t\right)^{-1}\left(\frac{1}{n}\sum x_t' u_t\right)$$

- TS.3$'$: $E(x_t' u_t) = 0$
  - A stronger assumption is $E(u_t | x_t) = 0$. We don't need it to show that $\widehat{\beta}$ converges to β, but in practice we would add nonlinear terms to the regression function to try to get the BLP and conditional mean as close as possible.
- By TS.1$'$,

$$plim \frac{1}{n} \sum x_t' u_t = 0$$

- By TS.2$'$, either $plim \frac{1}{n} \sum x_t' x_t$ is a pos. def. matrix or its eigenvalues stay strictly positive, so the eigenvalues of its inverse stay strictly bounded from above. $\therefore$

$$plim \left(\frac{1}{n} \sum x_t' x_t \right)^{-1} \left(\frac{1}{n} \sum x_t' u_t \right) = 0$$

and $\widehat{\beta}$ is a consistent estimator of β.

- Define the matrix $V_n = V(\sum x_t' u_t)/n$. We can write

$$\sqrt{n}\, V_n^{-1/2}\left(\frac{1}{n}\sum x_t' u_t\right) = \sqrt{n}\, V_n^{-1/2}\left(\frac{1}{n}\sum x_t' x_t\right)(\widehat{\beta} - \beta)$$

- Recall the CLT says, in the scalar case,

$$\left(\frac{\overline{Y}_n - E(\overline{Y}_n)}{\sqrt{Var(\overline{Y}_n)}}\right) \sim^a N(0,1)$$

- Applying the vector analog we have

$$\sqrt{n}\, V_n^{-1/2}\left(\frac{1}{n}\sum x_t' u_t\right) = \sqrt{n}\, V_n^{-1/2}\left(\frac{1}{n}\sum x_t' x_t\right)(\widehat{\beta} - \beta) \sim^a N(0, I_K)$$

so, using my abuse of notation

$$\widehat{\beta} \sim^a N\left(\beta, n^{-1}\left(\frac{1}{n}\sum x_t' x_t\right)^{-1} V_n \left(\frac{1}{n}\sum x_t' x_t\right)^{-1}\right)$$

Notice that

$$n^{-1}\left(\frac{1}{n}\sum x_t'x_t\right)^{-1}V_n\left(\frac{1}{n}\sum x_t'x_t\right)^{-1} = n\left(\sum x_t'x_t\right)^{-1}V_n\left(\sum x_t'x_t\right)^{-1}$$

So what does V_n look like? Because it has zero mean,

$$\begin{aligned} V\left(\sum_{t=1}^{T} x_t'u_t\right) &= E\left(\sum_{t=1}^{T} x_t'u_t\right)\left(\sum_{s=1}^{T} x_s'u_s\right)' \\ &= E\left(\sum_{t=1}^{T}\sum_{s=1}^{T} u_t u_s x_t' x_s\right) \\ &= E\left(\sum_{t=1}^{T} u_t^2 x_t' x_t\right) + 2E\left(\sum_{t=1}^{T}\sum_{s>t}^{T} u_t u_s x_t' x_s\right) \end{aligned}$$

The simplest case mimics the CLM

- TS.4$'$: $var(u_t|x_t) = \sigma^2$
- TS.5$'$: $E(u_t u_s|x_t, x_s) = 0$
- Under TS.1$'$-TS.5$'$:

$$V_n = \frac{\sigma^2}{n} E\left(\sum_{t=1}^{T} x_t' x_t\right)$$

and

$$plim\left(\frac{1}{n}\sum x_t' x_t\right)^{-1} V_n = \sigma^2$$

$$\therefore\ \hat{\beta} \sim^a N\left(\beta, \sigma^2 \left(\sum x_t' x_t\right)^{-1}\right)$$

Illustrative examples

1. AR(1) model for y

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

where $x_t = y_{t-1}$ and $\{u_t\}$ is an i.i.d. white noise sequence

- We can have $E(u_t|x_t) = 0$ (or even the stronger property $E(u_t|x_t, x_{t-1}, \cdots) = 0$, but it makes no sense to assume $E(u_t|x_{t+1}, x_t, x_{t-1}, \cdots) = 0$ as $u_t = x_{t+1} - \beta_0 - \beta_1 x_t$!
- As long as $|\beta_1| < 1$, we can write y_t as an MA(∞) process, so $\{(y_t, x_t)\}$ is weakly dependent. Therefore the OLS estimator will be consistent and asymptotically normal.
- If $\beta_1 = 1$ (random walk with drift), then OLS estimator will be consistent, but not asymptotically normal. This suggests that the small sample distribution will be poorly approximated by usual LST if β_1 is close to 1

2. Efficient Market Hypothesis (EMH)

Suppose we believe returns are unpredictable from their own past history. We can formalize this as saying

$$E(y_t | y_{t-1}, y_{t-2}, \cdots) = E(y_t)$$

- A strategy would be to build a model

  $$y_t = \beta_0 + x_t \beta_1 + u_t$$

  where $x_t = (y_{t-1}, y_{t-2}, \cdots, y_{t-p})$
- Under EMH, this AR(p) model has all its roots outside the unit circle (they are at infinity), so if $\{u_t\}$ is i.i.d., then $\{(y_t, x_t)\}$ is weakly dependent.
- OLS estimator $\widehat{\beta}_1$ will be consistent and asymptotically normal. We can use LST to test $H_0 : \beta_1 = 0$

3. Expectations Augmented Phillips Curve

$$\text{PC}(\ast): \ \pi_t - \pi_t^e = \beta_1(u_t - u_t^n) + e_t$$

where

π_t = inflation

π_t^e = expected inflation

u_t = unemployment rate (*not* the disturbance!)

u_t^n = natural unemployment rate

e_t = disturbance

Suppose we don't see π_t^e, u_t^n

- Strategy 1: Assume

$$\pi_t^e = \pi_{t-1} \qquad \text{(static expectations)}$$

$$u_t^n = \mu \qquad \text{(a constant)}$$

Then PC($*$) becomes

$$\text{PC}(**) \quad \Delta\pi_t = \beta_0 + \beta_1 u_t + e_t \qquad \Delta = 1 - L$$

Rk: Suppose e_t is an AR(1) process, so $\rho(L)e_t = \varepsilon_t$ and $\{\varepsilon_t\}$ is white noice. Multiply both sides of PC($**$) by $\rho(L)$ to get

$$\Delta\pi_t = \widetilde{\beta}_0 + \widetilde{\beta}_1 u_t + \widetilde{\beta}_2 \Delta\pi_{t-1} + \widetilde{\beta}_3 u_{t-1} + \varepsilon_t$$

so dependence in the errors looks like dynamics in the PRF.

- Strategy 2: Assume

$$\pi_t^e = \pi_{t-1} + \theta_1 \Delta\pi_{t-1} + \theta_2 \Delta\pi_{t-2}$$

$$u_t^n = \gamma_1 u_t + \gamma_2 u_{t-1} + \gamma_3 u_{t-3}$$

Then PC($*$) becomes

$$\text{PC}(***) \quad \Delta\pi_t = \beta_0 + \beta_1 u_t + \beta_2 u_{t-1} + \beta_3 u_{t-3} + \beta_4 \Delta\pi_{t-1} + \beta_5 \Delta\pi_{t-2} + e_t$$

Rk: Notice that using a proxy for unobserved expectations or for a smoothed series (natural unemployment, permanent income, etc.) often introduces dynamics (lags) into the PRF

- We can extend the LST for weakly dependent processes to include seasonals and time trends as regressors. But highly persistent series lead to a very different theory.

Highly persistent series

Consider the model

$$(\ast) \quad y_t = y_{t-1} + e_t$$

- If $(\forall t)$
  - (i) $E|y_t| < \infty$
  - (ii) $E(e_t|y_{t-1}, y_{t-2}, \cdots)$

  Then $\{y_t\}$ is called a *martingale* (w.r.t. $\sigma(y_{t-1}, y_{t-2}, \cdots)$) and $\{e_t\}$ is called a *martingale difference sequence*. There is no requirement that $\{e_t\}$ be i.i.d.

- If we have $\{e_t\}$ i.i.d. white noise, then $\{y_t\}$ is called a *random walk*

- Recursive substitution in $(\ast)$ gives

$$y_t = y_0 + e_t + e_{t-1} + \cdots e_1$$

Therefore

$$(\forall t)\ E(y_t) = E(y_0) \quad \text{but} \quad var(y_t) = \sigma_e^2 t$$

so the random walk model is not (covariance) stationary. In fact it's variance "blows up".

- For a stationary AR(1) model, we have

$$E(y_{t+h}|y_t) = \rho_1^h y_t$$

which goes to zero as the horizon increases. But for the random walk (or martingale) model, the dependence never goes away. If we set $y_0 = E(y_0)$, we can show that

$$corr(y_t, y_{t+h}) = \sqrt{\frac{t}{t+h}}$$

so for large t the correlation among observations of any fixed distance can be arbitrarily large (we don't get weak dependence)

- Many economic data display what looks like "unit root behaviour" (aka look I(1)) because the have a root in their AR representation that looks like it is on the unit circle. For such models, we can generalize eq ($*$) to allow e_t to be a weakly dependent (I(0) which includes ARMA(p,q)) process). Examples include the exchange rate, interest rate, inflation.

- Some highly persistent processes display "trends", eg

$$y_t = \beta_0 + y_{t-1} + e_t$$

which is called a random walk with drift. Using recursive substitution, we see that we can write this as

$$y_t = \beta_0 t + y_0 + e_t + e_{t-1} + \cdots e_1$$

Transformations on I(1) processes

- By definition, y_t is I(1) iff Δy_t is I(0) (i.e. weakly dependent). We can define an entire hierarchy I(2), I(3), etc.
- Therefore, if y_t looks I(1), we can use its first difference to get a series that is weakly dependent. For example, suppose

$$y_t = \beta_0 + \rho_1 y_{t-1} + \rho_2 y_{t-2} + x_t\delta + e_t$$

and we suspect a unit root in $\rho(L)$ (i.e. $1 - \rho_1 - \rho_2 = 0$). Then we could impose a unit root and rewrite the model as

$$\Delta y_t = \beta_0 + \beta_1 \Delta y_{t-1} + x_t\delta + e_t$$

where $\{\Delta y_t\}$ is now a weakly dependent process.

- If $\{x_t\}$ is also I(1), there is the possibility that $\exists\gamma$ such that $y_t - x_t\gamma$ is stationary. If so, γ is called a *cointegrating vector* and differencing would be wrong. See 18.4.

Deciding whether a series is I(1)

- Formal tests are in Ch 18.
- Even if process is stationary, standard LST provides a poor approximation to the sampling distribution in finites samples if roots are close to the unit circle.
- Run the regression

$$y_t = \beta_0 + \rho_1 y_{t-1} + \cdots + \rho_p y_{t-p} + e_t$$

- To see if is stationary, and factor the AR polynomial $\widehat{\rho}(L) = \prod_{i=1}^{p}(1 - \widehat{\theta}_i L)$. If $max|\widehat{\theta}_i| > .9$ then decide I(1).
- If data display a trend, include t in the regression (to allow for trend stationarity); otherwise $max|\theta_i|$ is biased upwards.

Dynamically Complete Models

- Consider the model

$$y_t = \beta_0 + \beta_1 z_t + u_t$$

  where $\{(y_t, z_t)\}$ are time series.

- As long as $E(u_t|z_t) = 0$ (or even just $E(z_t' u_t)$), OLS will usually be a consistent estimator (whether the data display weak dependence or are highly persistent). But the disturbances will be serially correlated.

- What we would like is the property

$$(DC) \;\; E(u_t|z_t, u_{t-1}, z_{t-1}, \cdots) = E(u_t|z_t, y_{t-1}, z_{t-1}, \cdots) = 0$$

- Q: How to get this?
- A: Add lags of y_t, z_t to the regression!

- This leads to an augmented model

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 y_{t-1} + \beta_3 z_{t-1} + u_t$$

- If we add enough lags (and nonlinear terms) we can *guarantee* that the model is dynamically complete, i.e. satisfies (*DC*)
- A dynamically complete model also guarantees

$$\text{TS.5}' \quad E(u_s u_t | x_s, x_t) = 0 \qquad \forall s \neq t$$

- NOTE: Including only lags but not nonlinear terms won't get you TS.5$'$, but will guarantee $E(u_s u_t) = 0$