10. Regression with Time Series Data

Definition: A *time series* $\{x_t\}$ is a collection of random variables indexed by time.

Rks:

- We shall deal only with discrete time series, i.e. where $t \in \mathbb{N}$ or perhaps $t \in \mathbb{Z}$. Although samples will involve only a finite number of observations, it is useful to think of them as being part of an infinite (or doubly infinite) collection.
- We shall not deal with continuous time models, where $t \in [0, T]$
- A time series (TS) is also called a *stochastic process*.

- With random sampling, it is useful to think of our sample as a sequence of draws from an underlying distribution.
- With time series, it is useful to think of an entire sample as being drawn at once as a point $\omega \in \Omega$. We sometimes write the collection of random variables as $\{x_t(\omega)\}$. Fixing ω, and letting t vary we get a realized *trajectory/time path/history/ensemble*.
- Probability statements for TS describe what would happen across different draws of ω. So one of the questions that has to be addressed is whether we can learn ensemble averages can be estimated from a single realization of a trajectory.

Application

- Distributed Lag Models

$$y_t = \beta_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t \qquad t = 1, \cdots, n$$

  where $\{u_t\}$ is an i.i.d sequence. Notice that here the effect of z_t on the dependent variable does not all occur in period t, unless $\delta_1 = \delta_2 = 0$ (a static model).

- Suppose z is increased by one unit at time t, but its values in all other periods are held constant. In period t, we get the "impact" response δ_0. But y_{t+1} is expected to increase by δ_1, and y_{t+2} by δ_2. Values of y_{t+s} for $s \geq 2$ are unaffected.

- If z is increased permanently by one unit at time t, we get an initial response of δ_0 and a "long run" multiplier of $\delta_0 + \delta_1 + \delta_2$

Lag operator

When dealing with (discrete) time series, it is useful to introduce the *Lag* (or *Backshift*) operator

$$L(z_t) = z_{t-1}$$

- Strictly speaking, the lag operator maps an entire time series into an entire time series, but it's convenient to represent it by what it does to a typical component.
- We often drop the brackets and write just Lz_t
- The lag operator is linear: $L(a_1 z_{1t} + a_2 z_{2t}) = a_1 z_{1t-1} + a_2 z_{2t-1}$
- Define $L^2 z_t = z_{t-2}$, $L^{-1} z_t = z_{t+1}$, $L^0 z_t = z_t$ and in general $L^n z_t = z_{t-n}$, so we can make polynomials in the lag operator.

- With the lag operator, we can write the distributed lag model as

  $$y_t = \beta_0 + \delta(L)z_t + u_t \quad \text{where } \delta(L) = \delta_0 L^0 + \delta_1 L^1 + \delta_2 L^2$$

  The impact multiplier is δ_0 and the long run multiplier is $\delta(1)$

- Notice that for observation $t = 1$, we need the observations for the explanatory variable at time $t = 0$ and $t = -1$. We'll assume that these two presample observations are available.

- In practice, introducing dynamics, such as lags of z_t into the regression may force use to "drop" the first two observations from the sample.

Finite Sample Properties of OLS under Classical Assumptions

- If we don't have a random sample, because the explanatory variables are drawn from a time series, it is still possible to construct a finite sample theory for the OLS estimators that may be applicable. In matrix terms, the assumptions are exactly the same as in the CLM, but now we have to be a bit more careful to verify them.
- S1: $y = X\beta + u \quad (\Leftrightarrow y_t = x_t\beta + u_t \quad \forall t)$
  - x_t could contain lags of the variables $z_{1t}, z_{2t},$ etc
- S2: X has full column rank

Under S1-S2, the OLS estimator is unique and given by

$$\widehat{\beta} = (X'X)^{-1}X'y = \beta + Lu$$

- S3: $E(u|X) = 0 \quad (\Leftrightarrow E(u_t|x_1, x_2, \cdots x_n) = 0 \quad \forall t)$
  - If $E(u_t|x_t) = 0$ we say x_t is *contemporaneously exogenous*. This is enough for a large sample theory, but not to obtain finite sample results such as no bias.
  - $E(u_t|x_1, x_2, \cdots x_n) = 0 \quad \forall t$ says the regressors are *strictly exogenous*

Under S1-S3

$$E(\widehat{\beta}|X) = E(\beta + Lu|X) = \beta + LE(u|X) = \beta$$

$$\Rightarrow E(\widehat{\beta}) = E\left[E(\widehat{\beta}|X)\right] = \beta$$

- S4: $E(uu'|X) = \sigma^2 I_n$
  - S4a: $E(u_t^2|x_1, x_2, \cdots x_n) = \sigma^2 \quad \forall t$
  - S4b: $E(u_s u_t|x_1, x_2, \cdots x_n) = 0 \quad \forall s \neq t$

Under S1-S4

$$V(\widehat{\beta}|X) = LV(u|X)L' = \sigma^2 (X'X)^{-1}$$

- Under S1-S4, $\widehat{\beta}$ is the Gauss-Markov estimator, i.e. it is BLUE.

- S5: $u|X \sim N$

Under S1-S5

$$\widehat{\beta}|X \sim N(\beta, \sigma^2(X'X)^{-1})$$

- So all the test statistics derived under S1-S5 for random samples, also hold in a time series setting.
- Under S1-S5, $\widehat{\beta}$ achieves the Cramer-Rao lower bound, and so it is the MVUE, i.e. it has minimum variance in the class of all unbiased estimators

So what's the big deal about time series?

- With random samples, i.e. $\{(x_t, u_t)\}$ an i.i.d. sequence,

$$E(u_t|x_t) = 0 \Leftrightarrow E(u|X) = 0$$

$$E(u_t^2|x_t) = \sigma^2 \Leftrightarrow E(u^2|X) = \sigma^2 I_n$$

  so it was enough to verify the left-hand-side conditions which involved only disturbances and regressors for each observation. With time series (dependent sample), we have to verify the right-hand-side equalities directly.
- Why would LHS equalities be OK but not RHS?
  - x_t could respond to y_{t-1} (or *be* y_{t-1}), then $E(u_t|x_{t+1}) \neq 0$
  - disturbances could be serially correlated (a violation of S4b): $E(u_s u_t|X) \neq 0 \quad \forall s \neq t$

Some Functional Form issues

With time series data, we may want coefficients to vary over time.

$$y_t = x_t\beta_t + u_t$$

1. Structural Breaks

$$\beta_t = \begin{cases} \beta^* & t \in T^* \\ \beta & t \notin T^* \end{cases}$$

where T^* denotes a subsample. For example, T^* could be the observations

- before an event occured (say the introduction of same change in regime–NAFTA, Sarbanes-Oxley)
- during a specific episode (a war, or a monetary targeting regime, an "event" such as a corporate name change).

2. Seasonals

$$\beta_t = \beta + D_t\delta$$

where

$$D_t = \begin{bmatrix} D_{1t} & D_{2t} & \cdots & D_{S-1,t} \end{bmatrix}$$

and $D_{jt} = 1$ if observation t belongs to season j, 0 otherwise.

- Modelling returns, we often allow the intercept to vary with the season. For example, in monthly returns, allow for a "January effect".
- Trading volume has a time-of-day seasonal. It’s highest at the open and just before the close.

Time trends

Linear

$$y_t = \beta_0 + \beta_1 t + u_t \qquad \therefore\ E(y_t) = \beta_0 + \beta_1 t$$

or log-linear

$$\ln y_t = \beta_0 + \beta_1 t + u_t \qquad \therefore\ E(y_t) \propto \exp(\beta_1 t)$$

Rk: We can also extend to higher order time polynomials

- Should you include time trends as regressors? Consider

$$y_t = X\beta_1 + \beta_2 t + u_t$$

- From the FW theorem, we know that if we include t as a regressor, only the detrended part of X is used to estimate β_1
- Because many time series display a trend, t serves as a proxy for many left out variables. Not including a time trend makes $\widehat{\beta}_1$ highly susceptible to the usual left-out variable bias.

Goodness of fit with time series data

- If y displays a trend, then it's not unusual to get very high values for R^2 (say $.98$) whereas in cross-section data, a value of $R^2 = 0.2$. Are time-series regressions more informative?

- A suggestion: Recall

$$R^2 = \frac{\hat{y}' A \hat{y}}{y' A y} \quad \text{where } A = I - n^{-1}\iota\iota'$$

With time series data, it makes some sense to report

$$\tilde{R}^2 = \frac{\hat{y}' \tilde{A} \hat{y}}{y' \tilde{A} y} = 1 - \frac{SSR}{y' \tilde{A} y}$$

where $\tilde{A}$ projects onto $Sp^{\perp}\{\iota, t, seasonals\}$. This is just the usual R^2 from the regression of $\tilde{A}y$ on (ιX)

More on Seasonality

- We know from the FW theorem, that if we run the regression

$$y_t = X\beta + D_S\delta + u_t$$

where D_S are seasonal dummies, that it's the same as running deviations from seasonal means of y on deviations from seasonal means of the variables in X.

- Statistical agencies often report *deseasonalized* data. When we run regressions using such data, it's almost as if they have taken out the seasonal means.

- However, the agencies' procedures are a bit more complicated because
  - they allow the seasonal component to evolve over time (it's not a constant). This means that the data are revised as seasonal factors are estimated more accurately with time.
  - they adjust the data for outliers as well (possibly due to events such as strikes or that are weather related)
- The adjustments are mechanical and done outside a model. This has some advantages, but in practice you should always deseasonalize yourself, whenever possible. Standard approaches often take out *too much* variation in the data.