ECON 61001: *Lecture 5*

Alastair R. Hall

The University of Manchester

Outline of this lecture

- Large sample behaviour of OLS with time series data
  - time series regression models
  - concepts and conditions
- OLS as projection
- Non-spherical errors
  - Properties of OLS
  - Generalized Least Squares

Time series variables

Two types of time series:

- *flow variable* - measured over an interval of time, for example monthly consumption expenditures;
- *stock variable* - measured at a moment in time, such as price or quantity of shares owned.

- Assume time series observed at *regularly spaced* intervals.
- Frequency at which time series is observed known as *sampling frequency*.

$$y_t = \beta_{0,1} + \beta_{0,2} y_{t-1} + \beta_{0,3} h_t + \beta_{0,4} h_{t-1} + u_t = x_t' \beta_0 + u_t$$

- contains lagged "y" and contemporaneous and lagged values of other variables (in this case "h").

Presence of lagged variables on the rhs of regression model $\Rightarrow$ difference between the number of observations on the variables and the number of observations used in the estimation.

If start with $T_\star$ observations and p is the longest lag on rhs then effective sample size is $T_\star - p$ with the first p observations being used for conditioning.

For ease of notation, assume effective sample runs $t = 1, 2, \ldots, T$ and $y_0, h_0', , \ldots, y_{-p+1}, h_{-p+1}'$ are available for conditioning.

The sampling framework for time series is fundamentally different from that in analysis of cross-section data. Based on stochastic process theory.

In stochastic process theory: time series v_t is viewed as evolving before we start observing it and continuing to evolve after we stop observing it, that is

$$\ldots, v_{-3}, v_{-2}, v_{-1}, v_0 \underbrace{v_1, v_2, v_3, \ldots v_T}_{sample}, v_{T+1}, v_{T+2}, \ldots.$$

The entire process $\{v_t\}_{t=-\infty}^{\infty}$ is known as a *realization* of v_t.

Key difference: sample *once* leading to a particular realization of the series.

Statistical background

So, as the sample size grows, we see more of one realization of the process.

Does this allow us to uncover the underlying probability distribution of v_t as $T \to \infty$?

Answer: Yes! Under certain conditions $\rightarrow$ stationarity and *weak dependence*.

We distinguish two forms of stationarity: strong- and weak-stationarity.

Strong stationarity

The time series $\{v_t\}_{t=-\infty}^{\infty}$ is said to be strongly stationary if the joint probability distribution function, $F(.)$, of any subset of $\{v_t\}$ satisfies:

$$F(v_{t_1}, v_{t_2}, \ldots, v_{t_n}) = F(v_{t_1+c}, v_{t_2+c}, \ldots, v_{t_n+c})$$

for any integer n and integer constant c.

- Also known as “strict”- stationarity

The time series $\{v_t\}_{t=-\infty}^{\infty}$ is said to be weakly stationary if for all t, s we have:

(i) $E[v_t] = \mu$; (independent of t)

(ii) $Var[v_t] = \Sigma$; (independent of t)

(iii) $Cov[v_t, v_s] = \Sigma_{t-s}$. (depend only on $t - s$)

- if v_t is scalar then $Cov[v_t, v_s]$ is $|t-s|^{th}$ autocovariance of v_t.

- if v_t is vector then $Cov[v_t, v_s]$ is $(t-s)^{th}$ autocovariance matrix of v_t:
  - diagonal elements are autocovariances of $v_{t,i}$.
  - off-diagonal elements are $Cov[v_{t,i}, v_{s,j}]$.
  - $Cov[v_t, v_s] = \{Cov[v_s, v_t]\}'$.

Weak dependence places restrictions on the memory of v_t that is, the relationship between v_t and v_{t-s} as $s \to \infty$.

If v_t is weakly stationary process then weak dependence implies that $Cov[v_t, v_{t-s}] \to 0$ as $s \to \infty$ and at a sufficiently fast rate.

Can derive WLLN and CLT for (strong or weak) stationary and weak dependent series (subject to “certain other conditions” that are taken to hold without statement)

Note these conditions are sufficient and not necessary - see lecture notes.

Weak Law of Large Numbers: *Let v_t be a stationary and weakly dependent time series with $E[v_t] = \mu$ then, subject to certain other conditions, it follows that*

$$T^{-1} \sum_{t=1}^{T} v_t \xrightarrow{p} \mu.$$

Central Limit Theorem: *Let v_t be a stationary and weakly dependent time series with $E[v_t] = \mu$ and $Cov[v_t, v_{t-j}] = \Gamma_j$ then, subject to certain other conditions, it follows that*

$$T^{-1/2} \sum_{t=1}^{T} (v_t - \mu) \xrightarrow{d} N(0, \Omega),$$

where $\Omega = \sum_{-\infty}^{\infty} \Gamma_i = \Gamma_0 + \sum_{i=1}^{\infty} \{\Gamma_i + \Gamma_i'\}$.

Ω is known as the long run variance of v_t.

Form of $\Omega = lim_{T \to \infty} \Omega_T$ comes from:

$$\Omega_T = Var\left[T^{-1/2} \sum_{t=1}^{T} (v_t - \mu) \right] = T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} Cov[v_t, v_s]$$

$$= T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \Gamma_{t-s}, \text{ using stationarity,}$$

$$= \Gamma_0 + \sum_{i=1}^{T-1} \left(\frac{T-i}{T} \right) (\Gamma_i + \Gamma_{-i}).$$

So $lim_{T \to \infty} \Omega_T = \Gamma_0 + \sum_{i=1}^{\infty} \{\Gamma_i + \Gamma_i'\}$

Finite sample properties of OLS

Recall model with stochastic regressors.

Assumptions:

- *SR1*: true model is: $y = X\beta_0 + u$.
- *SR2*: X is stochastic.
- *SR3*: X is rank k with probability 1.

- *SR4*: $E[u|X] = 0$.
- *SR5*: $Var[u|X] = \sigma_0^2 I_T$.

- *SR6*: $u|X \sim$ Normal.

Argued OLS unbiased via:

$$\begin{aligned} E[\hat{\beta}_T] &= \beta_0 + E_X\left[E_{u|X}\left[(X'X)^{-1}X'u\right]\right] \\ &= \beta_0 + E_X\left[(X'X)^{-1}X'E_{u|X}[u]\right] = \beta_0 \end{aligned}$$

Assumption *SR4* with time series

Nature of condition $E[u|X] = 0$ in time series.
Whether it holds depends on "degree of exogeneity" of variables in x_t:

- x_t is said to be contemporaneously exogenous if $E[u_t|x_t] = 0$
  $\Rightarrow$ u_t and x_t are uncorrelated.

- x_t is said to be strictly exogenous if $E[u_t|\{x_t\}_{t=1}^T] = 0 \Rightarrow$ $\{u_t\}_{t=1}^T$ and $\{x_t\}_{t=1}^T$ are uncorrelated.

Since $E[u_t|\{x_t\}_{t=1}^T] = E[u_t|X]$, $E[u|X] = 0$ only holds if x_t is strictly exogenous.

Note: Assumption *SR4* must fail if x_t contains lagged values of y, see Lecture Notes. In these cases, in general, $E[\hat{\beta}_T] \neq \beta_0$.

Assuming $x_t = (1, x'_{2,t})'$ and $x_{2,t}$ is function of (vector) h_t and y_{t-i}, h_{t-j} (for $i, j > 0$).

- **Assumption TS1**: $y_t = x'_t\beta_0 + u_t$, $t = 1, 2, \ldots T$
- **Assumption TS2**: For (y_t, h'_t) is a weakly stationary, weakly dependent time series.

- **Assumption TS3**: $E[x_t x'_t] = Q$, a finite, positive definite matrix.

- **Assumption TS4**: $E[u_t \,|\, x_t] = 0$ for all $t = 1, 2, \ldots, T$.
- **Assumption TS5**: $Var[u_t \,|\, x_t] = \sigma_0^2$ for all $t = 1, 2, \ldots, T$.
- **Assumption TS6**: For all $t \neq s$, $E[u_t u_s \,|\, x_t, x_s] = 0$

Large sample analysis of OLS

Under these conditions can use essentially same arguments as for cross-section data to deduce the following results:

Theorem *If Assumptions* TS1 - TS4 *hold then* $\hat{\beta}_T$ *is a consistent estimator for* β_0.

Theorem *If Assumptions* TS1 - TS6 *hold then:*

$$T^{1/2}(\hat{\beta}_T - \beta_0) \xrightarrow{d} N\left(0, \sigma_0^2 Q^{-1}\right).$$

All large sample inference procedures described in Lecture 4 go through.

Dynamic completeness

Regression models can be interpreted as a statement about the conditional mean of y_t given an information set.

In cross-section data, this information set contains information about the i^{th} sampling unit: $E[y_i|x_i] = x_i'\beta_0$.

In time series data, the relevant information set is not only the h_t but the history of both y and h that is,

$$\mathcal{I}_t \;=\; \{\, h_t, y_{t-1}, h_{t-1}, y_{t-2}, h_{t-2}, \ldots, y_1, h_1 \,\}\,.$$

Therefore if regression model specifies conditional mean then we are really stating that we believe:

$$E[y_t \,|\, \mathcal{I}_t] \;=\; x_t'\beta_0,$$

- if this statement is true then model is said to be *dynamically complete*.

It is convenient to introduce this property as a specific assumption.

- **Assumption TS7**: $E[y_t \,|\, \mathcal{I}_t] \;=\; x_t'\beta_0$.

Note:

- Assumption *TS7* $\Rightarrow$ Assumptions *TS4* and *TS6*, see Lecture Notes

Suppose wish to predict y_t given x_t using $\tilde{y}_t = c(x_t)$.

Issue: choice of $c(\cdot)$.

If choose $c(\cdot)$ to minimize

$$MSE(\tilde{y}_t) = E[y_t - \tilde{y}_t]^2$$

$\rightarrow c_o(x_t) = E[y_t|x_t]$.

If $c_o(\cdot)$ unknown then might restrict to class of linear forecasts $y_t^{lp} = \alpha' x_t$ but then what should α be?

OLS as projection

Choice that minimizes MSE (over class of linear forecasts) is one associated with the linear projection of y_t on x_t which has the property

$$E[(y_t - \alpha' x_t)x_t] = 0.$$

$$\rightarrow \alpha \;=\; \{E[x_t x_t']\}^{-1}\, E[x_t y_t] \sim \text{population analogue to } \hat{\beta}_T.$$

So OLS can be justified as estimator of weights in linear projection of y_t on x_t - but this does not justify using estimators to learn about how x_t affects y_t, for this we need to impose assumptions about the relationship between the variables.

Non-spherical errors

Have developed large sample framework for inference in cross-section and time series data.

Do assumptions always hold? No! In next part of course, we consider consequences of violations of assumptions about second moments of error term.

Terminology (using fixed regressor model to simplify):

- $Var[u] = \sigma_0^2 I_T$ then u is said to have a spherical distribution.
- $Var[u] \neq \sigma_0^2 I_T$ then u is said to have a non-spherical distribution.

See Lecture notes Section 4.1 for origins of these terms.

Specifically, we consider two examples of non-spherical errors:

- Cross-section data with $Var[u_i|x_i] = \sigma_i^2$ - heteroscedasticity

- Time series data with
  - $Var[u_t|x_t] = \sigma_t^2$ - heteroscedasticity
  - $Cov[u_t, u_s|x_t, x_s] \neq 0$ - serial correlation

OLS with non-spherical errors

Recall that our model is:

$$y = X\beta_0 + u$$

where

- *CA1*: true model is: $y = X\beta_0 + u$.

- *CA2*: X is fixed in repeated samples.
- *CA3*: X is rank k.
- *CA4*: $E[u] = 0$

- *CA5-NS* $Var[u] = \Sigma$ where Σ is a $T \times T$ positive definite matrix.
- *CA6*: $u \sim$ Normal.

- $E[\hat{\beta}_T] = \beta_0$ as impose Assumptions *CA1-CA4*.

- $Var[\hat{\beta}_T] = (X'X)^{-1}X'\,E[uu']\,X(X'X)^{-1}$ and so

$$Var[\hat{\beta}_T] = (X'X)^{-1}X'\Sigma X(X'X)^{-1} \neq \sigma_0^2(X'X)^{-1}.$$

- $\hat{\beta}_T \sim N\left(\beta_0, (X'X)^{-1}X'\Sigma X(X'X)^{-1}\right)$.

So inference procedures from Lectures 2, 3 and 4 are not valid because are based on wrong formula for $Var[\hat{\beta}_T]$.

Conditions of Gauss-Markov Theorem do not hold and so this result cannot be used to justify OLS is BLUE.

Can we characterize the BLUE estimator? Yes!

It is the The Generalized Least Squares (GLS) estimator of β_0 given by

$$\hat{\beta}_{GLS} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$$

If Assumptions *CA1-CA4*, *CA5-VS* and *CA6* hold then

$$\hat{\beta}_{GLS} \sim N\left(\beta_0, (X'\Sigma^{-1}X)^{-1}\right).$$

But need Σ - what happens if Σ is unknown?

Further reading

- Notes: Sections 2.12, 3.3-3.5, 4.1, 4.2.
- Greene:
  - time series data Section 20.1
  - conditions for limit theorems Section 20.2 and Section 20.4 (but more detail than in the course)
  - OLS with non-spherical errors Sections 9.1 and 9.2 (finite sample part)