EC708 Discussion 1
Linear Models and Asymptotic Theory

Yan Liu

Department of Economics
Boston University

January 29, 2021

Outline

1. Linear Models
2. Convergence
3. Consistency & Laws of Large Numbers
4. Asymptotic Normality & Central Limit Theory

Contents are mainly based on *Asymptotic Theory for Econometricians* (White, 2002).

Linear Models

Data Generating Processes

$$y_t = x_t'\beta + u_t, \quad t = 1, \ldots, T$$

where

- we have T observations on y_t and $x_t = (x_{t1}, \ldots, x_{tk})'$;
- y_t is the outcome variable (or dependent variable);
- x_t is a $k \times 1$ vector of independent variables (or covariates, regressors);
- u_t is unobserved;
- $\beta \in \mathbb{R}^k$ is an unknown parameter we are interested in.

Linear Models

Exogeneity

We need to make assumptions on u_t to learn about β from $\{(y_t, x_t)\}_{t=1}^T$:

- Strong exogeneity: $E[u|X] = 0, u = (u_1, \ldots, u_T)', X = (x_1, \ldots, x_T)'$.
  - leads to unbiasedness of the OLS estimator;
  - too strong to be justified in many applications especially in time series context. E.g. it rules out lagged dependent variables:

$$\underbrace{y_t}_{\substack{\text{output}\\\text{growth}}} = \underbrace{z_t'\beta}_{\substack{\text{macro}\\\text{variables}}} + \underbrace{y_{t-1}\delta}_{\substack{\text{output growth in}\\\text{previous quarter}}} + \underbrace{u_t}_{\substack{\text{productivity}\\\text{shock}}}$$

- Weak exogeneity: $E[u_t x_t] = 0$.
  - Under $E[u_t] = 0$, u_t and x_t are uncorrelated.

Weak exogeneity provides identification of β:

$$E[x_t(y_t - x_t'\beta)] = 0 \;\Rightarrow\; E[x_t y_t] = E[x_t x_t']\beta.$$

A natural estimator is to use sample analogues:

$$\begin{aligned}\hat{\beta} &= \left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} x_t y_t \\ &= \arg\min_{\beta} \sum_{t=1}^{T} (y_t - x_t'\beta)^2 \;\Rightarrow\; \text{OLS estimator.}\end{aligned}$$

Frisch-Waugh-Lovell (FWL):

We can partition $x_t = (d_t', w_t')'$. For example, in wage gender gap analysis,

$$\underset{\text{wage}}{y_t} = \underset{\substack{\text{gender}\\\text{indicator}}}{d_t'\beta_1} + \underset{\text{controls}}{w_t'\beta_2} + u_t.$$

Define the partialling-out operator

$$\check{v}_t = v_t - w_t'\hat{\gamma}_{vw}, \quad \hat{\gamma}_{vw} = \arg\min_b \sum_{t=1}^{T} (v_t - w_t' b)^2.$$

Then,

$$\hat{\beta}_1 = \arg\min_b \sum_{t=1}^{T} (\check{y}_t - \check{d}_t' b)^2 = \left(\frac{1}{T} \sum_{t=1}^{T} \check{d}_t \check{d}_t' \right)^{-1} \frac{1}{T} \sum_{t=1}^{T} \check{d}_t \check{y}_t.$$

We use asymptotic approximations to analyze the properties of $\hat{\beta}$:

- Assumptions on sampling:
  $\{(y_t, x_t)\}_{t=1}^T$ satisfies regularity conditions on heterogeneity and dependence, e.g. i.i.d. (independent & identically distributed).
- We often write $\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V)$.
- In finite samples, we care about (when β is scalar)

$$\left.\begin{array}{l} P(\hat{\beta} > \beta + c)(\text{“overshooting”}) \\ P(\hat{\beta} < \beta - c)(\text{“undershooting”}) \end{array}\right. \approx 1 - F(\sqrt{T}c)$$

where F is the CDF of $N(0, V)$.

Convergence

Modes of Convergence

Let $\{Z_t : t = 1, 2, \dots\}$ be a sequence of random variables.

- $Z_T \overset{a.s.}{\to} c$: Z_T converges almost surely to c if

$$P\{\omega : \lim_{T\to\infty} Z_T(\omega) = c\} = 1.$$

- $Z_T \overset{p}{\to} c$: Z_T converges in probability to c if for any $\varepsilon > 0$,

$$P\{\omega : |Z_T(\omega) - c| > \varepsilon\} \to 0 \text{ as } T \to \infty.$$

- $Z_T \overset{d}{\to} Z$: Z_T converges in distribution to Z if

$$F_{Z_T}(z) \to F_Z(z) \text{ for every continuity point } z \text{ of } F_Z.$$

Continuous mapping theorem:

Let $\{Z_T\}$ be a sequence of random variables such that $Z_T \xrightarrow{p} c$. Let g be a function continuous at point c. Then $g(Z_T) \xrightarrow{p} g(c)$.

Slutsky's theorem:

Let $Z_T \xrightarrow{d} Z$ and $Y_T \xrightarrow{p} c$. Then

- $Z_T + Y_T \xrightarrow{d} Z + c$;
- $Z_T Y_T \xrightarrow{d} cZ$;
- $Y_T^{-1} Z_T \xrightarrow{d} c^{-1} Z$ provided Y_T^{-1} and c^{-1} exist.

Both theorems hold when Z_T, Y_T, and g are scalar or vectorial.

- $Z_T = O_{a.s.}(T^\lambda)$ means for some $\Delta < \infty$ and $T^* < \infty$, $P(|T^{-\lambda} Z_T| < \Delta \text{ for all } T > T^*) = 1$.
- $Z_T = O_p(T^\lambda)$ means for every $\varepsilon > 0$ there exist finite $\Delta_\varepsilon > 0$ and $T_\varepsilon \in \mathbb{N}$ such that $P(|T^{-\lambda} Z_T| < \Delta_\varepsilon) > 1 - \varepsilon$ for all $T > T_\varepsilon$.
- $Z_T = o_{a.s.}(T^\lambda)$ means $T^{-\lambda} Z_T \stackrel{a.s.}{\to} 0$.
- $Z_T = o_p(T^\lambda)$ means $T^{-\lambda} Z_T \stackrel{p}{\to} 0$.

Convergence

Big O and little o notation

In particular,

- $Z_T = O_p(1)$: Z_T is bounded with probability approaching 1.
  - If $Z_T \xrightarrow{d} Z$, then $Z_T = O_p(1)$.
- $Z_T = o_p(1) \iff Z_T \xrightarrow{p} 0$.

Product rule:

If $\underset{k\times l}{A_T} = O_p(1)$ and $\underset{l\times 1}{b_T} = o_p(1)$ (component wise), then

$$\underset{k\times 1}{A_T b_T} = o_p(1).$$

In the linear model, under

- weak exogeneity,
- no perfect multicollinearity,
- restrictions on dependence, heterogeneity & moments of $\{(y_t, x_t)\}_{t=1}^{T}$,

$$\hat{\beta} = \left(\frac{1}{T} \sum_{t=1}^{T} x_t x_t' \right)^{-1} \frac{1}{T} \sum_{t=1}^{T} x_t y_t \overset{a.s.}{\rightarrow} \beta.$$

- We implicitly assume the model is correctly specified in the sense that the true DGP is a linear structure for some value of β.

Laws of Large Numbers

General Form

Given restrictions on the dependence, heterogeneity & moments of a sequence of random variables $\{Z_t\}$,

$$\bar{Z}_T - \bar{\mu}_T \overset{a.s.}{\to} 0,$$

where $\bar{Z}_T = \frac{1}{T}\sum_{t=1}^{T} Z_t$ and $\bar{\mu}_T = E(\bar{Z}_T)$.

- $\{Z_t\}$ is IID (independent & identically distributed)
- $\{Z_t\}$ is INID (independent & not identically distributed)
- $\{Z_t\}$ is dependent & identically distributed
- $\{Z_t\}$ is dependent & heterogeneously distributed

Kolmogrov's LLN (IID data)

Let $\{Z_t\}$ be a sequence of i.i.d. random variables. Then $\bar{Z}_T \overset{a.s.}{\to} \mu$ if and only if $E|Z_t| < \infty$ and $E(Z_t) = \mu$.

Markov's LLN (INID data)

Let $\{Z_t\}$ be a sequence of independent random variables such that $E|Z_t|^{1+\delta} < M < \infty$ for some $\delta > 0$ and all $t > 0$. Then $\bar{Z}_T - \bar{\mu}_T \overset{a.s.}{\to} 0$.

Remark: $E|Z_t|^{1+\delta} < M < \infty$ implies Markov's condition

$$\sum_{t=1}^{\infty} E|Z_t - \mu_t|^{1+\delta}/t^{1+\delta} < \infty \text{ where } \mu_t = E(Z_t).$$

- $\{Z_t\}$ is stationary if the joint distribution of $(Z_{t_1}, Z_{t_2}, \ldots, Z_{t_m})$ is the same as that of $(Z_{t_1+s}, Z_{t_2+s}, \ldots, Z_{t_m+s})$ for any $(t_1, \ldots, t_m)$ and s.
- $\{Z_t\}$ is ergodic if $\{Z_t\}$ is stationary and for every set A, B of real sequences,

$$
\begin{aligned}
\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} & P\{(Z_1, Z_2, \ldots) \in A \text{ and } (Z_{t+1}, Z_{t+2}, \ldots) \in B\} \\
&= P\{(Z_1, Z_2, \ldots) \in A\} P\{(Z_1, Z_2, \ldots) \in B\}.
\end{aligned}
$$

($(Z_1, Z_2, \ldots)$ and $(Z_{t+1}, Z_{t+2}, \ldots)$ are independent on average in the limit.)

Ergodic theorem

Let $\{Z_t\}$ be a stationary ergodic scalar sequence with $E|Z_t| < \infty$. Then $\bar{Z}_T \overset{a.s.}{\to} \mu = E(Z_t)$.

$E(u_t x_t) = 0$ can be justified by the theory of rational expectations:

$$E(u_t | X_t, X_{t-1}, \dots; u_{t-1}, u_{t-2}, \dots) = 0.$$

- Let $\mathcal{F}_t$ be a σ-algebra generated by all information available at time t such that $\mathcal{F}_{t-1} \subset \mathcal{F}_t$ for all t. E.g. $\mathcal{F}_t = \sigma(\dots, Z_{t-1}, Z_t)$.
- Let Z_t be adapted to $\mathcal{F}_t$ so that Z_t is measurable with respect to $\mathcal{F}_t$.
- $\{Z_t, \mathcal{F}_t\}$ is a martingale difference process if $E(Z_t | \mathcal{F}_{t-1}) = 0$.

Chow's LLN

Let $\{Z_t, \mathcal{F}_t\}$ be a martingale difference sequence such that $E|Z_t|^{2\delta} < M < \infty$ for some $\delta \geq 1$ and all t. Then $\bar{Z}_T \stackrel{a.s.}{\to} 0$.

Asymptotic Normality

Overview

In the linear model, under

- weak exogeneity,
- no perfect multicollinearity,
- restrictions on dependence, heterogeneity & moments of $\{(y_t, x_t)\}_{t=1}^T$,

$$V_T^{-1/2}\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, I)$$

where

$$V_T = Q_T^{-1}\Sigma_T Q_T^{-1}, \quad Q_T = E\left(\frac{1}{T}\sum_{t=1}^T x_t x_t'\right), \quad \Sigma_T = \mathsf{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T x_t u_t\right).$$

Central Limit Theory

General Form

Given restrictions on the dependence, heterogeneity & moments of a scalar sequence $\{Z_t\}$,

$$\sqrt{T}(\bar{Z}_T - \bar{\mu}_T)/\bar{\sigma}_T \xrightarrow{d} N(0,1),$$

where $\bar{Z}_T = \frac{1}{T}\sum_{t=1}^{T} Z_t$, $\bar{\mu}_T = E(\bar{Z}_T)$, and $\bar{\sigma}_T^2 = \mathsf{Var}(\sqrt{T}\bar{Z}_T)$.

However, we usually need the asymptotic normality of vectors such as $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t u_t$.

Let $\{Z_t\}$ be a sequence of $k \times 1$ random vectors. Suppose that for any $b \in \mathbb{R}^k$ such that $\|b\| = b'b = 1$,

$$b' Z_T \xrightarrow{d} b' Z,$$

where Z is a $k \times 1$ random vector with distribution function F. Then,

$$Z_T \xrightarrow{d} Z.$$

Hence, it is only necessary to study CLT for sequences of scalars.

Lindeberg-Lévy (IID data)

Let $\{Z_t\}$ be a sequence of i.i.d. random scalars with $\mu = E(Z_t)$ and $\sigma^2 = \mathsf{Var}(Z_t) < \infty$. If $\sigma^2 \neq 0$, then

$$\sqrt{T}(\bar{Z}_T - \mu)/\sigma \xrightarrow{d} N(0, 1).$$

Liapounov's CLT (INID data)

Let $\{Z_t\}$ be a sequence of independent random scalars with $E|Z_t - E(Z_t)|^{2+\delta} < \Delta < \infty$ for some $\delta > 0$ and all $t > 0$. If $\bar{\sigma}_T^2 > \delta' > 0$ for all T sufficiently large, then

$$\sqrt{T}(\bar{Z}_T - \bar{\mu}_T)/\bar{\sigma}_T \xrightarrow{d} N(0, 1).$$

Remark: We can obtain CLT by imposing a uniform bound on $E|Z_t|^{2+\delta}$.

CLT for martingale difference sequences:

Let $\{Z_t, \mathcal{F}_t\}$ be a martingale difference sequence such that $E|Z_t|^{2+\delta} < M < \infty$ for some $\delta > 0$ and all t. If $\bar{\sigma}_T^2 > \delta' > 0$ for all T sufficiently large and $\frac{1}{T}\sum_{t=1}^{T} Z_t^2 - \bar{\sigma}_T^2 \xrightarrow{p} 0$, then

$$\sqrt{T}\bar{Z}_T/\bar{\sigma}_T \xrightarrow{d} N(0,1).$$

With stationarity, we can allow $\{Z_t\}$ to only behave asymptotically like martingale difference processes.

- $\{Z_t, \mathcal{F}_t\}$ is an adapted mixingale if $E(Z_t^2) < \infty$ and there exist finite nonnegative sequences $\{c_t\}$ and $\{\gamma_m\}$ s.t. $\gamma_m \to 0$ as $m \to \infty$ and

$$\left(E(E(Z_t|\mathcal{F}_{t-m})^2)\right)^{1/2} \leq c_t \gamma_m.$$

  We say γ_m is of size $-a$ if $\gamma_m = O(m^{-a-\varepsilon})$ for some $\varepsilon > 0$.

Scott's CLT

Let $\{Z_t, \mathcal{F}_t\}$ be a stationary ergodic adapted mixingale of size -1. Then $\bar{\sigma}_T^2 = \mathsf{Var}(\sqrt{T}\bar{Z}_T) \to \bar{\sigma}^2 < \infty$ as $T \to \infty$ and if $\bar{\sigma}^2 > 0$, then

$$\sqrt{T}\bar{Z}_T/\bar{\sigma} \xrightarrow{d} N(0,1).$$

Central Limit Theory

Delta Method

If $\sqrt{T}(Z_T - c) \xrightarrow{d} N(0, \Sigma)$ and g is continuously differentiable at c, then

$$\sqrt{T}(g(Z_T) - g(c)) \xrightarrow{d} N\left(0, \frac{\partial g(c)}{\partial c'} \Sigma \left(\frac{\partial g(c)}{\partial c'}\right)'\right).$$

- Follows from a stochastic Taylor expansion and Slutsky's theorem:

$$\sqrt{T}(g(Z_T) - g(c)) = \frac{\partial g(\tilde{Z}_T)}{\partial c'} \sqrt{T}(Z_T - c)$$

  where $\tilde{Z}_T$ lies between Z_T and c so that $\tilde{Z}_T \xrightarrow{p} c$.

- If $c \in \mathbb{R}^k$ and $g : \mathbb{R}^k \to \mathbb{R}^r$, then

$$\frac{\partial g}{\partial c'} = \begin{pmatrix} \frac{\partial g_1}{\partial c_1} & \dots & \frac{\partial g_1}{\partial c_k} \\ \dots & \dots & \dots \\ \frac{\partial g_r}{\partial c_1} & \dots & \frac{\partial g_r}{\partial c_k} \end{pmatrix} \text{ is a } r \times k \text{ matrix.}$$