# EC708 Discussion 1
# Linear Models and Asymptotic Theory

Yan Liu

Department of Economics
Boston University

January 21, 2022

# 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).

# Table of Contents


1. Linear Models
2. Convergence
3. Consistency
4. Asymptotic Normality

# 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.

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

- **Strong exogeneity**: $E[u|X] = 0$, where $u = (u_1, \dots, u_T)', X = (x_1, \dots, x_T)'$.
  - 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:

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

- **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 $\beta$:

$$E[x_t(y_t - x_t'\beta)] = 0 \;\Rightarrow\; \beta = (E[x_t x_t'])^{-1} E[x_t y_t].$$

A natural estimator is to use sample analogues:

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

We can always interpret $\beta$ as the least squares or projection parameter in the population:

$$\beta = \underset{b \in \mathbb{R}^k}{\arg\min}\, E[(y_t - x_t' b)^2]$$

and $\hat{\beta}$ as the least squares or projection parameter in the sample:

$$\hat{\beta} = \underset{b \in \mathbb{R}^k}{\arg\min}\, \frac{1}{T} \sum_{t=1}^{T} (y_t - x_t' b)^2 \;\Rightarrow\; \text{OLS estimator.}$$

By FOC, we obtain the decomposition identity

$$y_t = x_t'\beta + \varepsilon_t, \quad E[\varepsilon_t x_t] = 0.$$

However, without the linearity assumption on the true DGP, $\beta$ is not necessarily a parameter of a structural or causal economic model.

**Frisch-Waugh-Lovell (FWL):**

We can partition $x_t = (\underset{1\times k_1}{d_t'}, \underset{1\times k_2}{w_t'})'$. E.g. in wage gender gap analysis,

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

For a random variable $v_t$, define the partialling-out operator w.r.t. $w_t$:

$$\check{v}_t = v_t - w_t'\hat{\gamma}_{vw}, \quad \hat{\gamma}_{vw} = \arg\min_{b\in\mathbb{R}^{k_2}} \sum_{t=1}^{T} (v_t - w_t'b)^2.$$

Then,

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

# Linear Models

## Inference

- In finite samples, we care about (when $\beta$ 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 $N(0, V)$.

- We use asymptotic approximations: $\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V)$.
- 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).

# Table of Contents


1 Linear Models

2 Convergence

3 Consistency

4 Asymptotic Normality

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 $Z_T \xrightarrow{p}$ ( or $\xrightarrow{a.s.}$)$c$. Let $g$ be a function continuous at point $c$. Then $g(Z_T) \xrightarrow{p}$ ( or $\xrightarrow{a.s.}$)$g(c)$.

**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.

# Convergence



- $Z_T = O_{a.s.}(T^\lambda)$ 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 $T^{-\lambda} Z_T \stackrel{p}{\to} 0$.

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

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

# Table of Contents

1 Linear Models

2 Convergence

3 Consistency

4 Asymptotic Normality

In the linear model, under

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

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

# Laws of Large Numbers

## General Form

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

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

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

**Kolmogrov's LLN (IID data)**

Let $\{Z_t\}$ be a sequence of i.i.d. random variables. Then $\bar{Z}_T \stackrel{a.s.}{\to} \mu$ 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 $\bar{Z}_T - \bar{\mu}_T \stackrel{a.s.}{\to} 0$.

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

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

**Stationarity** (time-series counterpart of identical distribution):

- $\{Z_t\}$ 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 $s$.
- Stationarity means that the distribution is constant over time.

**Ergodicity** (time-series counterpart of independence):

- Stationarity alone is not sufficient for the LLN. E.g. $Z_t = Z$ for some random variable $Z$, then $\bar{Z}_T$ will be inconsistent for $E[Z_t]$.
- $\{Z_t\}$ is ergodic if $\{Z_t\}$ is stationary and for every set $A, B$ of real sequences,

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

i.e. $(Z_1, Z_2, \dots)$ and $(Z_{t+1}, Z_{t+2}, \dots)$ are asymptotically independent on average.

**Ergodic theorem**

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

- If $Z_t$ is stationary ergodic and $X_t = \phi(Z_t, Z_{t-1}, Z_{t-2}, \dots)$ is a random vector, then $X_t$ is stationary ergodic.

# Table of Contents


1. Linear Models
2. Convergence
3. Consistency
4. Asymptotic Normality

# Asymptotic Normality

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 random 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 random vectors such as $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t u_t$. $\Rightarrow$ Cramér-Wold device

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.

# 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

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

**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 $E|Z_t|^{2+\delta}$.

**Martingale difference sequence:**

- $E[u_t x_t] = 0$ can be justified by the theory of rational expectations:

$$E[u_t | X_t, X_{t-1}, \ldots ; u_{t-1}, u_{t-2}, \ldots] = 0 \quad \text{(unforecastability)}$$

- We call $Z_t$ a martingale difference sequence (MDS) if $E[Z_t | Z_{t-1}, Z_{t-2}, \ldots] = 0$.
- More rigorously, we can write the conditional expectation as $E[Z_t | \mathcal{F}_{t-1}] = 0$ where $\mathcal{F}_{t-1}$ is a $\sigma$-algebra (information set) generated by the infinite history $(Z_{t-1}, Z_{t-2}, \ldots)$.
  - $\mathcal{F}_{t-1} \subset \mathcal{F}_t$: information accumulates over time.
  - $\mathcal{F}_t$ can contain variables other than $Z_t$.

**MDS CLT:**

Let $Z_t$ be a strictly stationary and ergodic martingale difference sequence such that $\sigma^2 = \mathsf{Var}(Z_t) < \infty$. Then

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

**Example:** Estimation of autoregressive models

Let $y_t$ be a stationary and ergodic $AR(p)$ process

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} + u_t,$$

where $u_t$ is a MDS. Let $x_t = (1, y_{t-1}, \ldots, y_{t-p})'$. Since $x_t$ is part of $\mathcal{F}_{t-1}$, $E[x_t u_t|\mathcal{F}_{t-1}] = x_t E[u_t|\mathcal{F}_{t-1}] = 0$, i.e. $x_t u_t$ is a MDS.