Lecture 7: LLN and CLTs for Time Series

Raul Riva

FGV EPGE

October, 2025

Intro

Intro

- We are frequently interested in regressing y_t on x_t, x_{t-1}, etc;
- We can do that with OLS and be less restrictive than MLE;
- But if we want to make inference in a flexible way, we need to develop asymptotic theory;
- Standard LLN and CLT (Lindberg-Lévy, Lingberg-Feller, etc) will not apply. Why?

Intro

- We are frequently interested in regressing y_t on x_t, x_{t-1}, etc;
- We can do that with OLS and be less restrictive than MLE;
- But if we want to make inference in a flexible way, we need to develop asymptotic theory;
- Standard LLN and CLT (Lindberg-Lévy, Lingberg-Feller, etc) will not apply. Why?
- We will sketch some proofs and give references for further reading;
- I want you to focus on the main ideas, not the technical details;

Law of Large Numbers

Law of Large Numbers

- We will first develop a Law of Large Numbers for covariance-stationary time series;
- Assume that $\{y_t\}$ has mean μ and autocovariance function γ_h;
- As usual, assume $\sum_{h=-\infty}^{h=\infty} |\gamma_h| < \infty$;
- We will focus on the properties of the sample mean:

$$\bar{y}_T = \frac{1}{T} \sum_{t=1}^{T} y_t$$

Law of Large Numbers

- We will first develop a Law of Large Numbers for covariance-stationary time series;
- Assume that $\{y_t\}$ has mean μ and autocovariance function γ_h;
- As usual, assume $\sum_{h=-\infty}^{h=\infty} |\gamma_h| < \infty$;
- We will focus on the properties of the sample mean:

$$\bar{y}_T = \frac{1}{T}\sum_{t=1}^{T} y_t$$

- We notice that it is an unbiased estimator of μ:

$$\mathbb{E}[\bar{y}_T] = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[y_t] = \mu$$

The Variance of the Sample Mean

- If $\gamma_h = 0$ for $h \neq 0$, then $\mathbb{E}\left(\bar{y}_T - \mu\right)^2 = \frac{\gamma_0}{T}$;
- This is the result we would get if the y_t were i.i.d.

The Variance of the Sample Mean

- If $\gamma_h = 0$ for $h \neq 0$, then $\mathbb{E}\left(\bar{y}_T - \mu\right)^2 = \frac{\gamma_0}{T}$;
- This is the result we would get if the y_t were i.i.d.

Let's see the general case:

- To make computations simple, consider $\mathbf{Y}_T = (y_1 - \mu, \dots, y_T - \mu)'$.
- Consider a $T \times 1$ vector of ones $\mathbf{1}_T$;
- Then we can write: $\bar{y}_T - \mu = \frac{1}{T}\mathbf{1}_T'\mathbf{Y}_T$

The Variance of the Sample Mean

- If $\gamma_h = 0$ for $h \neq 0$, then $\mathbb{E}\left(\bar{y}_T - \mu\right)^2 = \frac{\gamma_0}{T}$;
- This is the result we would get if the y_t were i.i.d.

Let's see the general case:

- To make computations simple, consider $\mathbf{Y}_T = (y_1 - \mu, \dots, y_T - \mu)'$.
- Consider a $T \times 1$ vector of ones $\mathbf{1}_T$;
- Then we can write: $\bar{y}_T - \mu = \frac{1}{T}\mathbf{1}_T'\mathbf{Y}_T$
- If $\mathbf{V}_T$ is the $T \times T$ covariance matrix of $\mathbf{Y}_T$, then:

$$\mathbb{E}\left(\bar{y}_T - \mu\right)^2 = \frac{1}{T^2}\mathbf{1}_T'\mathbf{V}_T\mathbf{1}_T$$

- This is just the summation of all elements of $\mathbf{V}_T$ divided by T^2;

The Variance of the Sample Mean

- $\mathbf{Y}_T$ has mean zero and covariance matrix $\mathbf{V}_T$ given by:

$$\mathbf{V}_T = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{T-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{T-1} & \gamma_{T-2} & \gamma_{T-3} & \cdots & \gamma_0 \end{pmatrix}$$

The Variance of the Sample Mean

- $\mathbf{Y}_T$ has mean zero and covariance matrix $\mathbf{V}_T$ given by:

$$\mathbf{V}_T = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{T-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{T-1} & \gamma_{T-2} & \gamma_{T-3} & \cdots & \gamma_0 \end{pmatrix}$$

- The sum of all elements in $\mathbf{V}_T$ is:

$$\sum_{i=1}^{T} \sum_{j=1}^{T} \gamma_{|i-j|} = T\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \ldots + 2\gamma_{T-1}$$

- Therefore:

$$\mathbb{E}\left(\bar{y}_T - \mu\right)^2 = \frac{1}{T^2}\left[T\gamma_0 + 2(T-1)\gamma_1 + \ldots + 2\gamma_{T-1}\right] = \frac{1}{T^2} \sum_{h=-(T-1)}^{T-1} (T - |h|)\gamma_h$$

Absolute Summability Helps a Lot

- If $\sum\limits_{h=-\infty}^{\infty} |\gamma_h| < \infty$, then:

$$\begin{aligned}
\lim_{T\to\infty} \mathbb{E}(\bar{y}_T - \mu)^2 &= \lim_{T\to\infty} \frac{1}{T} \sum_{h=-(T-1)}^{T-1} \left(1 - \frac{|h|}{T}\right) \gamma_h \\
&\leq \lim_{T\to\infty} \frac{1}{T} \underbrace{\sum_{h=-(T-1)}^{T-1} \left(1 - \frac{|h|}{T}\right) |\gamma_h|}_{\text{finite as } T \text{ grows}} \\
&= 0
\end{aligned}$$

- In fact, by Chebyshev's inequality, we have that $\bar{y}_T \xrightarrow{p} \mu$;
- This is the Weak Law of Large Numbers for covariance-stationary time series;
- $\sum\limits_{h=-\infty}^{\infty} |\gamma_h| < \infty$ = "the process can be time-dependent *but* not *too* dependent";

The Limiting Variance of the Sample Mean

- The previous slide suggests another limiting result;
- A conjecture: is it true that $\lim_{T\to\infty}\left(T\cdot\mathbb{E}\left(\bar{y}_T-\mu\right)^2\right)=\sum_{h=-\infty}^{\infty}\gamma_h$? Yes? No? Maybe?

The Limiting Variance of the Sample Mean

- The previous slide suggests another limiting result;
- A conjecture: is it true that $\lim_{T\to\infty}\left(T\cdot\mathbb{E}\left(\bar{y}_T-\mu\right)^2\right)=\sum_{h=-\infty}^{\infty}\gamma_h$? Yes? No? Maybe?
- The answer is *yes*. And the proof is actually nice;
- In proper EPGE style, let $\epsilon>0$;
- Notice that absolute summability implies that $\sum_{h=q}^{\infty}|\gamma_h|$ is very small for large q;
- We can find q such that $\sum_{h=q+1}^{\infty}|\gamma_h|<\epsilon/4$;

The Limiting Variance of the Sample Mean

Now we limit the following difference:

$$\begin{aligned}\left|\sum_{h=-\infty}^{\infty} \gamma_h - T \cdot \mathbb{E}\left(\bar{y}_T - \mu\right)^2\right| &= \left|(\gamma_0 + 2\gamma_1 + \ldots) - \left[\gamma_0 + 2\left(1 - \frac{1}{T}\right)\gamma_1 + \ldots + 2\left(1 - \frac{T-1}{T}\right)\gamma_{T-1}\right]\right| \\ &\leq \sum_{j=1}^{q} \frac{2j}{T}|\gamma_j| + \sum_{j=q+1}^{\infty} 2|\gamma_j| \\ &\leq \sum_{j=1}^{q} \frac{2j}{T}|\gamma_j| + \epsilon/2\end{aligned}$$

- But the first term can be made smaller than $\epsilon/2$ for large T;
- The whole expression is smaller than ϵ for large T. Hence:

$$\boxed{\lim_{T\to\infty}\left(T \cdot \mathbb{E}\left(\bar{y}_T - \mu\right)^2\right) = \sum_{h=-\infty}^{\infty} \gamma_h}$$

Collecting the Results

So we showed that:

1. $\bar{y}_T \xrightarrow{p} \mu$ (Weak LLN);
2. $\lim\limits_{T\to\infty} \left(T \cdot \mathbb{E}\left(\bar{y}_T - \mu\right)^2\right) = \sum\limits_{h=-\infty}^{\infty} \gamma_h$;

Collecting the Results

So we showed that:

1. $\bar{y}_T \xrightarrow{p} \mu$ (Weak LLN);
2. $\lim\limits_{T\to\infty} \left(T \cdot \mathbb{E}\left(\bar{y}_T - \mu\right)^2\right) = \sum\limits_{h=-\infty}^{\infty} \gamma_h$;

- This implies that estimating means will be feasible and simple;
- This result is also hinting that the right “notion” of variance is $\sum\limits_{h=-\infty}^{\infty} |\gamma_h|$;
- We call this term the *Long-Run Variance* of the process;
- This is tricky to estimate: infinite parameters;

Questions?

A CLT for Martingale Difference Sequences

A CLT for Martingale Difference Sequences

- Independence is always the same thing, but dependence comes in all shapes and forms!
- There is no such a thing as "**the** CLT for time series";
- Different setups will require different asymptotic theory;
- We will cover useful results that appear here and there;
- An amazing reference for econometricians is Davidson's book (*Stochastic Limit Theory*);

A CLT for Martingale Difference Sequences

- A sequence $\{y_t\}$ is a Martingale Difference Sequence (MDS) with respect to the information set $\mathcal{F}_t$ if:
  1. y_t is known given $\mathcal{F}_t$;
  2. $\mathbb{E}[|y_t|] < \infty$;
  3. $\mathbb{E}[y_t|\mathcal{F}_{t-1}] = 0$ a.s.;

A CLT for Martingale Difference Sequences

- A sequence $\{y_t\}$ is a Martingale Difference Sequence (MDS) with respect to the information set $\mathcal{F}_t$ if:
  1. y_t is known given $\mathcal{F}_t$;
  2. $\mathbb{E}[|y_t|] < \infty$;
  3. $\mathbb{E}[y_t|\mathcal{F}_{t-1}] = 0$ a.s.;
- If $\{y_t\}$ is an MDS with respect to $\mathcal{F}_t$, it has mean zero and $\gamma_h = 0$ for $h \neq 0$;
- Still an uncorrelated sequence over time, but this is much weaker than independence;

A CLT for Martingale Difference Sequences

- A sequence $\{y_t\}$ is a Martingale Difference Sequence (MDS) with respect to the information set $\mathcal{F}_t$ if:
  1. y_t is known given $\mathcal{F}_t$;
  2. $\mathbb{E}[|y_t|] < \infty$;
  3. $\mathbb{E}[y_t|\mathcal{F}_{t-1}] = 0$ a.s.;
- If $\{y_t\}$ is an MDS with respect to $\mathcal{F}_t$, it has mean zero and $\gamma_h = 0$ for $h \neq 0$;
- Still an uncorrelated sequence over time, but this is much weaker than independence;

Example: $y_t = e_t e_{t-1}$, where e_t is i.i.d. with mean zero and $\mathcal{F}_t =$ "the entire path of $e_s \leq t$";

- We have that $\mathbb{E}[y_t|\mathcal{F}_{t-1}] = e_{t-1}\mathbb{E}[e_t|\mathcal{F}_{t-1}] = 0$.
- But we have $Cov(y_t^2, y_{t-1}^2) = \sigma^4 \left(\mathbb{E}[e_t^4] - \sigma^4\right) > 0 \implies$ dependence;

A CLT for Martingale Difference Sequences

Theorem (CLT for MDS - Proposition 7.8 from Hamilton's book)

Let $\{y_t\}$ *be a scalar MDS with respect to* $\mathcal{F}_t$ *such that:*

- $\mathbb{E}[y_t^2] = \sigma_t^2 > 0$ *such that* $\frac{1}{T}\sum_{t=1}^{T} \sigma_t^2 \to \sigma^2 > 0$;
- $\mathbb{E}[|Y_t|^r] < \infty$ *for some* $r > 2$ *and all* t;
- $\frac{1}{T}\sum\limits_{t=1}^{T} Y_t^2 \xrightarrow{p} \sigma^2$;

Then: $\sqrt{T} \cdot \bar{y}_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_t \xrightarrow{d} \mathcal{N}(0, \sigma^2)$

- This result generalizes to vectors and to triangular arrays;
- The proof is not trivial, but it is not too hard either;
- It will also use tricks involving the convergence of Fourier transforms;
- This result will come in handy when we study the OLS estimator;

Questions?

CLTs for Weakly Dependent Time Series

CLTs for Weakly Dependent Time Series

- Consider a covariance-stationary process $\{y_t\}$ with mean μ and ACF $\{\gamma_h\}$;
- The whole challenge to get a CLT in this case is controlling dependence;

CLTs for Weakly Dependent Time Series

- Consider a covariance-stationary process $\{y_t\}$ with mean μ and ACF $\{\gamma_h\}$;
- The whole challenge to get a CLT in this case is controlling dependence;
- A typical way of doing that is to use *mixing conditions*;
- These are technical conditions on how fast dependence fades away as $h \to \infty$;
- There are several types of mixing conditions: α-mixing, β-mixing, ϕ-mixing, etc;
- Things get *super* complicated *very* quickly;

CLTs for Weakly Dependent Time Series

- Consider a covariance-stationary process $\{y_t\}$ with mean μ and ACF $\{\gamma_h\}$;
- The whole challenge to get a CLT in this case is controlling dependence;
- A typical way of doing that is to use *mixing conditions*;
- These are technical conditions on how fast dependence fades away as $h \to \infty$;
- There are several types of mixing conditions: α-mixing, β-mixing, ϕ-mixing, etc;
- Things get *super* complicated *very* quickly;
- Typical trade-off: stronger mixing condition $\rightarrow$ weaker moment conditions and vice-versa;

CLTs for Weakly Dependent Time Series

- Consider a covariance-stationary process $\{y_t\}$ with mean μ and ACF $\{\gamma_h\}$;
- The whole challenge to get a CLT in this case is controlling dependence;
- A typical way of doing that is to use *mixing conditions*;
- These are technical conditions on how fast dependence fades away as $h \to \infty$;
- There are several types of mixing conditions: α-mixing, β-mixing, ϕ-mixing, etc;
- Things get *super* complicated *very* quickly;
- Typical trade-off: stronger mixing condition $\to$ weaker moment conditions and vice-versa;
- But the "outcome" of these CLTs is roughly the same:
  $\sqrt{T} \cdot (\bar{y}_T - \mu) \xrightarrow{d} \mathcal{N}\left(0, \sum_{h=-\infty}^{\infty} \gamma_h\right)$;
- We will cover two results but many more exist.

IID Innovations

Theorem (Theorem 7.11 from Hamilton's book)

Let $y_t = \mu + \sum_{j=0}^{\infty} \psi_j e_{t-j}$, *where* $\{e_t\}$ *is i.i.d. with mean zero and finite variance. Assume that* $\sum_{j=0}^{\infty} |\psi_j| < \infty$. *Then:*

$$\sqrt{T} \cdot (\bar{y}_T - \mu) \xrightarrow{d} \mathcal{N}\left(0, \sum_{h=-\infty}^{\infty} \gamma_h\right)$$

Correlated Innovations and Strong Mixing

Correlated Innovations and Strong Mixing

- Mixing is implicit in the assumption that $\{e_t\}$ is i.i.d.;
- The result generalizes for vectors;
- You can also prove it for MDS innovation but other conditions are needed;
- See Phillips and Solo (Annals of Statistics, 1992) for a complete treatment;

Strong Mixing

- We need a way to **quantify how fast dependence fades with time**;
- For two events (A,B), define the discrepancy

$$\alpha(A,B) \;=\; \left|\, \mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B) \,\right|$$

which is zero if A and B are independent and positive otherwise;

Strong Mixing

- We need a way to **quantify how fast dependence fades with time**;
- For two events (A,B), define the discrepancy

$$\alpha(A,B) \; = \; \Big| \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) \Big|$$

  which is zero if A and B are independent and positive otherwise;
- Consider two information sets where $\sigma(.)$ informally denotes "all events generated by":
  - $\mathcal{F}^{t}_{-\infty} = \sigma(\dots, Y_{t-1}, Y_t)$ is the **past up to (t)**;
  - $\mathcal{F}^{\infty}_{t} = \sigma(Y_t, Y_{t+1}, \dots)$ is the **future from (t)**;

Strong Mixing

- We need a way to **quantify how fast dependence fades with time**;
- For two events (A,B), define the discrepancy

$$\alpha(A,B) \;=\; \Big|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)\Big|$$

  which is zero if A and B are independent and positive otherwise;
- Consider two information sets where $\sigma(.)$ informally denotes "all events generated by":
  - $\mathcal{F}^{t}_{-\infty} = \sigma(\dots, Y_{t-1}, Y_t)$ is the **past up to (t)**;
  - $\mathcal{F}^{\infty}_{t} = \sigma(Y_t, Y_{t+1}, \dots)$ is the **future from (t)**;
- Define $\alpha(l) \equiv \sup\limits_{A \in \mathcal{F}^{t-l}_{-\infty}, B \in \mathcal{F}^{\infty}_{t}} \alpha(A,B)$;
- We say that a process is **strong-mixing** if $\alpha(l) \to 0$ as $l \to \infty$;
- The faster $\alpha(l)$ goes to zero, the weaker the dependence;

A CLT for Strong-Mixing Processes

Theorem (Theorem 14.15 from Hansen's book)

Let y_t be a strictly stationary process with mixing coefficients $\alpha(l)$. Assume that:

1. $\mathbb{E}[y_t] = 0$;
2. $\mathbb{E}[|y_t|^r] < \infty$ *for some* $r > 2$;
3. $\sum_{l=1}^{\infty} \alpha(l)^{\frac{r-2}{r}} < \infty$;

Then:

$$\sqrt{T} \cdot \bar{y}_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_t \xrightarrow{d} \mathcal{N}\left(0, \sum_{h=-\infty}^{\infty} \gamma_h\right)$$

- Notice that condition (3) is a bound on how fast it must mix;
- The processes we will work with in this class will satisfy the mixing condition;
- Checking these conditions in practice is not trivial. Hansen's theorem 14.26;

Questions?

Results for Time Series Regression

Results for Time Series Regression

- Now we finally consider a linear regression model:

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + u_t$$

  where $\mathbb{E}[u_t \mathbf{x}_t] = 0$;

- We assume that $\mathbf{x}_t$ is $K \times 1$ vector containing the intercept;
- $\mathbf{x}_t$ might contain lags of y_t as well;
- We assume that $(y_t, \mathbf{x}_t)$ is strictly stationary and ergodic;

Results for Time Series Regression

- Now we finally consider a linear regression model:

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + u_t$$

  where $\mathbb{E}[u_t \mathbf{x}_t] = 0$;
- We assume that $\mathbf{x}_t$ is $K \times 1$ vector containing the intercept;
- $\mathbf{x}_t$ might contain lags of y_t as well;
- We assume that $(y_t, \mathbf{x}_t)$ is strictly stationary and ergodic;
- In such cases, $\boldsymbol{\beta}$ is identified:

$$\boldsymbol{\beta} = \mathbb{E}[\mathbf{x}_t \mathbf{x}_t']^{-1} \mathbb{E}[\mathbf{x}_t y_t]$$

- We implicitly assume that there is no multicollinearity and finite second moments;

The OLS Estimator

The OLS estimator is given by:

$$\hat{\boldsymbol{\beta}}_T = \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t y_t\right) = \boldsymbol{\beta} + \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t u_t\right)$$

- Ergodicity will ensure consistency;

The OLS Estimator

The OLS estimator is given by:

$$\hat{\beta}_T = \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t y_t\right) = \beta + \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t u_t\right)$$

- Ergodicity will ensure consistency;
- Inference is a more complicated matter;
- Depending on the assumptions we make on $\{u_t\}$, we will get different results;
- The defining feature is whether $\mathbf{x}_t u_t$ is uncorrelated over time or not;

The OLS Estimator

The OLS estimator is given by:

$$\hat{\beta}_T = \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t y_t\right) = \beta + \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t u_t\right)$$

- Ergodicity will ensure consistency;
- Inference is a more complicated matter;
- Depending on the assumptions we make on $\{u_t\}$, we will get different results;
- The defining feature is whether $\mathbf{x}_t u_t$ is uncorrelated over time or not;
- In either case, assume that $\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right) \to \mathbf{Q}$, where $\mathbf{Q}$ is positive definite;
- We will get two different limiting results depending on the assumptions we use...

Uncorrelated Innovations

- Let $\mathcal{F}_t$ denote the information set up to time t;
- Assume that $\mathbf{x}_t$ is *known* at time $t-1$;
- Example: $\mathbf{x}_t = (1, y_{t-1}, y_{t-2})$, as would be the case in an AR(2) model;
- Assume that u_t is an MDS with respect to $\mathcal{F}_t$;
- Then $\mathbf{x}_t u_t$ is also an MDS;

Uncorrelated Innovations

- Let $\mathscr{F}_t$ denote the information set up to time t;
- Assume that $\mathbf{x}_t$ is *known* at time $t-1$;
- Example: $\mathbf{x}_t = (1, y_{t-1}, y_{t-2})$, as would be the case in an AR(2) model;
- Assume that u_t is an MDS with respect to $\mathscr{F}_t$;
- Then $\mathbf{x}_t u_t$ is also an MDS;

If both y_t and $\mathbf{x}_t$ have finite fourth moments (see Theorem 14.35 from Hansen's book), then

$$\boxed{\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}\left(0, \mathbf{Q}^{-1}\boldsymbol{\Sigma}\mathbf{Q}^{-1}\right)}$$

where $\Sigma = \mathbb{E}[\mathbf{x}_t\mathbf{x}_t' u_t^2]$.

Correlated Innovations

- It might be the case that u_t displays time-dependence;
- This will not invalidate consistency, but it will affect inference;

Correlated Innovations

- It might be the case that u_t displays time-dependence;
- This will not invalidate consistency, but it will affect inference;
- Assume that, for some $r > 4$, we have $\mathbb{E}[|y_t|^r] < \infty$, $\mathbb{E}[\|\mathbf{x}_t\|^r] < \infty$, and the mixing coefficients $\alpha(l)$ of the process $(y_t, \mathbf{x}_t)$ satisfy $\sum_{l=1}^{\infty} \alpha(l)^{\frac{r-4}{r}} < \infty$;

Then we have that

$$\boxed{\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}\left(0, \mathbf{Q}^{-1}\boldsymbol{\Omega}\mathbf{Q}^{-1}\right)}$$

where $\Omega = \sum_{h=-\infty}^{\infty} \mathbb{E}[\mathbf{x}_t \mathbf{x}_{t-h}' u_t u_{t-h}]$

- Notice this is the same long-run variance we saw before, but in vector form;

Questions?

How to Estimate the Covariance Matrix?

No Time-Dependence

When there is no time-dependence, we can estimate Σ with the sample analogue:

$$\hat{\mathbf{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \hat{u}_t^2$$

where $\hat{u}_t = y_t - \mathbf{x}_t' \hat{\boldsymbol{\beta}}_T$;

- This estimator is robust to heteroskedasticity but **not** to autocorrelation;
- This is the same White estimator we saw in the cross-sectional case;
- Standard errors for coefficients are given by the square roots of diagonal elements;
- Standard t-tests and Wald tests are valid;

Handling Time-Dependence

When there is time-dependence, things are more complicated:

- There is an infinite number of parameters to be estimated: $\Omega = \sum_{h=-\infty}^{\infty} \mathbb{E}[\mathbf{x}_t \mathbf{x}'_{t-h} u_t u_{t-h}]$;
- But we do know that these autocovariances *must* fade away as $|h|$ grows...

Handling Time-Dependence

When there is time-dependence, things are more complicated:

- There is an infinite number of parameters to be estimated: $\Omega = \sum_{h=-\infty}^{\infty} \mathbb{E}[\mathbf{x}_t \mathbf{x}'_{t-h} u_t u_{t-h}]$;
- But we do know that these autocovariances *must* fade away as $|h|$ grows...
- Idea: estimate only a finite number of autocovariances and assume the rest are zero;
- But how many lags should we consider?
- How to ensure that the estimator is positive definite? Negative variances are not good...

Handling Time-Dependence

Let's rewrite Ω as:

$$\begin{aligned}\Omega &= \sum_{h=-\infty}^{\infty} \mathbb{E}[\mathbf{x}_t \mathbf{x}_{t-h}' u_t u_{t-h}] \\ &= \Gamma_0 + \sum_{h=1}^{\infty} (\Gamma_h + \Gamma_h')\end{aligned}$$

where $\Gamma_h \equiv \mathbb{E}[\mathbf{x}_t \mathbf{x}_{t-h}' u_t u_{t-h}]$. Notice that $\Gamma_h = \Gamma_{-h}'$;

Handling Time-Dependence

Let's rewrite Ω as:

$$\Omega = \sum_{h=-\infty}^{\infty} \mathbb{E}[\mathbf{x}_t \mathbf{x}'_{t-h} u_t u_{t-h}]$$
$$= \Gamma_0 + \sum_{h=1}^{\infty} (\Gamma_h + \Gamma'_h)$$

where $\Gamma_h \equiv \mathbb{E}[\mathbf{x}_t \mathbf{x}'_{t-h} u_t u_{t-h}]$. Notice that $\Gamma_h = \Gamma'_{-h}$;

- The sample estimator of Γ_h is $\hat{\Gamma}_h \equiv \frac{1}{T} \sum_{t=h+1}^{T} \mathbf{x}_t \mathbf{x}'_{t-h} \hat{u}_t \hat{u}_{t-h}$;
- If we pick a truncation lag q, we could try estimating Ω with:

$$\hat{\Omega} = \hat{\Gamma}_0 + \sum_{h=1}^{q} \left(\hat{\Gamma}_h + \hat{\Gamma}'_h\right)$$

The Newey-West Estimator

- The main issue with this approach is that $\hat{\Omega}$ might not be positive definite;

The Newey-West Estimator

- The main issue with this approach is that $\hat{\Omega}$ might not be positive definite;
- Newey and West (Econometrica, 1987) had an ingenious idea: use the Bartlett kernel!
- Define the weights $w_h = 1 - \frac{h}{q+1}$ for $h = 0, 1, \ldots, q$;
- The Newey-West estimator is given by:

$$\hat{\Omega}_{NW} = \hat{\Gamma}_0 + \sum_{h=1}^{q} w_h \left(\hat{\Gamma}_h + \hat{\Gamma}_h' \right)$$

- This dude is guaranteed to be positive semi-definite for a given q!
- Sometimes, this is estimator is also called the *HAC estimator* (Heteroskedasticity and Autocorrelation Consistent);

The Bandwidth Choice

- The choice of q (also called the *bandwidth*) is important:
    - Low q: you might ignore the tails;
    - High q: estimation gets noisier and noisier and you have to estimate more and more parameters...
- Theory tells us that q should increase with T but not too fast;
- Hansen (1992) showed that if q grows no faster than $T^{1/3}$, we get consistency;
- Andrews (1991) showed that $q \propto T^{1/3}$ minimizes asymptotic mean squared error under some conditions;
- In practice: if your main results depend a lot on the choice of q, that is not a good sign;
    - Be transparent about q and stick to the same value throughout the paper;
    - Different statistical packages use different values for q. Just be transparent;
    - Rule of thumb: q should be "much smaller" than T;

Questions?

The End

References

- Chapter 7 from Hamilton's book for LLN and CLT for weakly dependent time series;
- Chapter 14 from Hansen's book collects several interesting results;
- Davidson's book (*Stochastic Limit Theory*) is the definitive treatment – very dark magic!;