LECTURE 16
TIME SERIES TOPICS

Definitions

Econometricians often work with data that come in the form of a *time series* or *stochastic process*: a collection of observations on the same variable or vector of variables, indexed by the date of measurement. The data are typically collected at equally spaced intervals (daily, weekly, monthly, etc.), and indexed by $t = 1, \dots, T$:

$$\{Y_t : t = 1, \dots, T\}.$$

The index t represents time, and Y_t is the value of the process at time t. The observed sample is typically assumed to be a segment of a process that extends into the infinite past and will go on indefinitely:

$$\{\dots, Y_{-1}, Y_0, Y_1, Y_2, \dots, Y_T, Y_{T+1}, \dots\}.$$

In the iid or *cross-sectional* case, the sample consists of n independent draws from the same distribution. Since each Y_i is drawn from the same distribution, as the sample size increases, the sample average $n^{-1}\sum_{i=1}^{n} Y_i$ reveals the population average $\mathrm{E}[Y_i]$, which is the same for all i. The time-series setting differs from the iid framework in that each element of the trajectory $Y_1, Y_2, \dots, Y_T$ is observed only once. To learn about the underlying population model, we must address two concerns. First, do different observations share a common structure? The mean $\mathrm{E}[Y_t] = \mu_t$ may vary with t. Second, do measurements made at different dates effectively sample from the population distribution? When observations are highly correlated, they do not. The following definitions formalize these requirements.

Definition 1. *The process $\{Y_t\}$ is strictly stationary if $(Y_{t_1}, \dots, Y_{t_k}) \overset{d}{=} (Y_{t_1+h}, \dots, Y_{t_k+h})$ for all $k \geq 1$, all integers h, and all $t_1, \dots, t_k$.*

Definition 2. *Suppose $\mathrm{Var}(Y_t) < \infty$ for all t. The process $\{Y_t\}$ is (weakly or covariance) stationary if*

$$\begin{aligned}
\mathrm{E}[Y_t] &= \mu \textit{ for all } t, \\
\mathrm{Cov}(Y_t, Y_{t-k}) &= \mathrm{Cov}(Y_s, Y_{s-k}) \\
&= \gamma(k) \textit{ for all } s \textit{ and } k.
\end{aligned}$$

The function $\gamma(k)$ is called the *autocovariance function.* The stationarity assumption implies that the covariance $\mathrm{Cov}(Y_t, Y_s) = \gamma(t-s)$, and therefore, in the scalar case,

$$\begin{aligned}
\gamma(k) &= \mathrm{Cov}(Y_t, Y_{t-k}) \\
&= \mathrm{Cov}(Y_{t-k}, Y_t) \\
&= \gamma(-k).
\end{aligned}$$

Similarly, in the vector case, we obtain that

$$\gamma(k) = \gamma(-k)^{\top}.$$

For a stationary process, $\mathrm{Var}(Y_t) = \gamma(0)$. The *autocorrelation function* is $\rho(k) = \gamma(k)/\gamma(0)$.

Definition 3. *(Informal.) The process $\{Y_t\}$ is ergodic if $\gamma(k) \to 0$ as $k \to \infty$.*

This is an informal characterization. Formally, ergodicity for a strictly stationary process is defined in terms of mixing conditions, for example, α-mixing or strong mixing; see Chapter 3 of Hamilton (1994, *Time Series Analysis*) or Chapter 14 of Davidson (1994, *Stochastic Limit Theory*). For covariance stationary processes, a sufficient condition for the LLN below is absolute summability of the autocovariance function: $\sum_{k=0}^{\infty} |\gamma(k)| < \infty$.

Stationarity ensures that observations at different times share a common underlying model. Ergodicity ensures that distant observations are nearly uncorrelated.

LLNs for a covariance stationary process

The LLN for a *strictly stationary* and *ergodic* sequence $\{Y_t : t \geq 1\}$ says that if $\mathrm{E}[Y_t] = \mu$ (the mean is finite), then, as $T \to \infty$,

$$T^{-1} \sum_{t=1}^{T} Y_t \to_p \mu.$$

The following weaker result has a direct proof. Let $\{Y_t : t \geq 1\}$ be a covariance stationary sequence of random variables, and assume that $\sum_{k=1}^{\infty} |\gamma(k)| < \infty$. Then, as $T \to \infty$,

$$T^{-1} \sum_{t=1}^{T} Y_t \to_p \mathrm{E}[Y_t].$$

The condition $\sum_{k=1}^{\infty} |\gamma(k)| < \infty$ above says that the autocovariance function of Y_t has to be absolutely summable. For this to hold, it must be true that $\gamma(k) \to 0$ as $k \to \infty$ (Y_t is ergodic). However, unlike the first result, this version requires the existence of second moments.

Proof. First, by Markov's inequality,

$$\Pr\left(\left|T^{-1} \sum_{t=1}^{T} Y_t - \mathrm{E}[Y_t]\right| > \varepsilon\right) \leq \frac{\mathrm{E}\left|T^{-1} \sum_{t=1}^{T} Y_t - \mathrm{E}[Y_t]\right|^2}{\varepsilon^2}.$$

We need to show that the numerator on the right-hand side of the above expression converges to zero as $T \to \infty$.

$$\begin{aligned}
\mathrm{E}\left|T^{-1} \sum_{t=1}^{T} Y_t - \mathrm{E}[Y_t]\right|^2 &= \mathrm{E}\left|T^{-1} \sum_{t=1}^{T} (Y_t - \mathrm{E}[Y_t])\right|^2 \\
&= T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathrm{E}\left[(Y_t - \mathrm{E}[Y_t])(Y_s - \mathrm{E}[Y_s])\right] \\
&= T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} \gamma(t-s) \\
&= T^{-2} \left(T\gamma(0) + 2 \sum_{j=1}^{T-1} (T-j)\gamma(j)\right).
\end{aligned}$$

To verify the last equality, observe that $\sum_{t=1}^{T} \sum_{s=1}^{T} \gamma(t-s)$ is given by the sum of all elements in the following $T \times T$ matrix.

$$\begin{pmatrix}
\gamma(0) & \gamma(1) & \dots & \gamma(T-2) & \gamma(T-1) \\
\gamma(1) & \gamma(0) & \dots & \gamma(T-3) & \gamma(T-2) \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\gamma(T-2) & \gamma(T-3) & \dots & \gamma(0) & \gamma(1) \\
\gamma(T-1) & \gamma(T-2) & \dots & \gamma(1) & \gamma(0)
\end{pmatrix}$$

Since $(T-j)/T < 1$ for $1 \le j \le T$,

$$
\begin{aligned}
\mathrm{E}\left|T^{-1}\sum_{t=1}^{T} Y_t - \mathrm{E}[Y_t]\right|^2 &= T^{-1}\left(\gamma(0) + 2\sum_{j=1}^{T-1}\frac{T-j}{T}\gamma(j)\right) \\
&\le T^{-1}\left(\gamma(0) + 2\sum_{j=1}^{T-1}\frac{T-j}{T}\left|\gamma(j)\right|\right) \\
&\le T^{-1}\left(\gamma(0) + 2\sum_{j=1}^{T-1}\left|\gamma(j)\right|\right) \\
&\le T^{-1}\left(\gamma(0) + 2\sum_{j=1}^{\infty}\left|\gamma(j)\right|\right) \\
&\to 0,
\end{aligned}
$$

since, by assumption, $\sum_{j=1}^{\infty}|\gamma(j)| < \infty$.

The result can be extended to sequences of random p-vectors. In this case, $\gamma(j)$ is a $p \times p$ matrix. Its element (r,s) is given by $\mathrm{Cov}(Y_{r,t}, Y_{s,t-j})$. In the vector case, the condition becomes: for all $1 \le r, s \le p$, $\sum_{j=0}^{\infty}|\mathrm{Cov}(Y_{r,t}, Y_{s,t-j})| < \infty$. The long-run variance-covariance matrix is

$$
\gamma(0) + \sum_{j=1}^{\infty}\gamma(j) + \sum_{j=1}^{\infty}\gamma(j)^{\top}.
$$

Examples

White noise

Suppose that

$$
\begin{aligned}
\mathrm{E}[U_t] &= 0 \text{ for all } t, \\
\mathrm{E}[U_t^2] &= \sigma^2 \text{ for all } t, \\
\mathrm{E}[U_t U_s] &= 0 \text{ for all } t \ne s. \qquad (1)
\end{aligned}
$$

Such a process is called *white noise*. Sometimes condition (1) is replaced by a stronger one, which says that U_s and U_t are independent for all $s \ne t$. Such a process is called *independent white noise*. A white-noise process is stationary and ergodic.

Moving average models

Suppose that

$$
Y_t = U_t + \theta_1 U_{t-1} + \ldots + \theta_q U_{t-q},
$$

where U_t is a white-noise process. Such a process is called a *moving average of order* q, denoted MA(q). A moving average process is covariance stationary. To see this, observe that $\mathrm{E}[Y_t] = 0$ for all t (a nonzero mean can be obtained if we include a nonzero constant on the right-hand side of the above equation). Next,

$$
\begin{aligned}
\gamma(0) &= \sigma^2 \left(1 + \theta_1^2 + \ldots + \theta_q^2\right), \\
\gamma(1) &= \mathrm{E}[Y_t Y_{t-1}] \\
&= \mathrm{E}\left[(U_t + \theta_1 U_{t-1} + \ldots + \theta_q U_{t-q})(U_{t-1} + \theta_1 U_{t-2} + \ldots + \theta_q U_{t-1-q})\right] \\
&= \theta_1 \mathrm{E}[U_{t-1}^2] + \theta_2\theta_1 \mathrm{E}[U_{t-2}^2] + \ldots + \theta_q\theta_{q-1} \mathrm{E}[U_{t-q}^2] \\
&= \sigma^2 (\theta_1 + \theta_2\theta_1 + \ldots + \theta_q\theta_{q-1}). \\
&\vdots \\
\gamma(q) &= \sigma^2 \theta_q, \\
\gamma(k) &= 0 \text{ for } k > q.
\end{aligned}
$$

Hence, MA(q) is covariance stationary and ergodic.

Autoregressive models

Suppose that

$$Y_t = \beta_1 Y_{t-1} + \ldots + \beta_p Y_{t-p} + U_t,$$

where U_t is a white-noise process. Such a process is called an *autoregression of order* p, denoted AR(p). A nonzero mean can be accommodated by including an intercept.

Consider the case of AR(1). Write

$$
\begin{aligned}
Y_t &= \beta Y_{t-1} + U_t \\
&= \beta^2 Y_{t-2} + \beta U_{t-1} + U_t \\
&= \beta^t Y_0 + \sum_{j=0}^{t-1} \beta^j U_{t-j}.
\end{aligned}
$$

Assume that $|\beta| < 1$. Then, $\lim_{j\to\infty} \beta^j = 0$, and therefore

$$Y_t = \sum_{j=0}^{\infty} \beta^j U_{t-j}.$$

This is called the MA(∞) representation of an AR(1) process. We have

$$
\begin{aligned}
\gamma(0) &= \sigma^2 \sum_{j=0}^{\infty} \beta^{2j} \\
&= \frac{\sigma^2}{1-\beta^2}, \\
\gamma(1) &= \sigma^2 \sum_{j=0}^{\infty} \beta^j \beta^{j+1} \\
&= \beta\sigma^2 \sum_{j=0}^{\infty} \beta^{2j} \\
&= \frac{\beta\sigma^2}{1-\beta^2}. \\
\gamma(k) &= \sigma^2 \sum_{j=0}^{\infty} \beta^j \beta^{j+k} \\
&= \frac{\beta^k \sigma^2}{1-\beta^2}.
\end{aligned}
$$

The autocovariances $\gamma(j)$ do not depend on t, so AR(1) is weakly stationary. It is also ergodic since $\lim_{j\to\infty}\gamma(j)=0$. The long-run variance of the AR(1) process is given by

$$\begin{aligned}
&\frac{\sigma^2}{1-\beta^2}+2\sum_{j=1}^{\infty}\frac{\beta^j\sigma^2}{1-\beta^2}\\
&=\frac{\sigma^2}{1-\beta^2}\left(1+2\beta\sum_{j=0}^{\infty}\beta^j\right)\\
&=\frac{\sigma^2}{1-\beta^2}\left(1+2\frac{\beta}{1-\beta}\right)\\
&=\frac{\sigma^2}{(1-\beta)^2}.
\end{aligned}$$

Wold decomposition

The MA(∞) representation is central to time-series analysis, owing to the Wold decomposition theorem. If $\{Y_t\}$ is a covariance stationary and ergodic process, then there exist a white-noise sequence $\{U_t\}$ and a sequence of constants $\{\theta_j\}$ such that

$$Y_t=\mu+\sum_{j=0}^{\infty}\theta_j U_{t-j},\ \text{where}$$

$$\sum_{j=0}^{\infty}\theta_j^2<\infty.$$

Such a process is called a *linear process.*

CLT for linear processes

Let Y_t be a linear process:

$$Y_t=\sum_{j=0}^{\infty}\theta_j U_{t-j},$$

where $\{U_t\}$ are iid with $\mathrm{E}[U_t]=0$ and $\mathrm{E}[U_t^2]=\sigma^2<\infty$, the sequence $\{\theta_j\}$ is absolutely summable and $\sum_{j=0}^{\infty}\theta_j\neq 0$. Then,

$$T^{-1/2}\sum_{t=1}^{T}Y_t\to_d N\left(0,\sum_{j=-\infty}^{\infty}\gamma(j)\right),$$

where $\gamma(j)$ is the j-th autocovariance of Y_t, and, therefore, $\sum_{j=-\infty}^{\infty}\gamma(j)$ is the long-run variance of Y_t. In the scalar case, the long-run variance equals

$$\sigma^2\left(\sum_{j=0}^{\infty}\theta_j\right)^2.$$

Time-series regression

Suppose that the econometrician observes $\{(Y_t,X_t):t=1,\ldots,T\}$, where

- $Y_t = X_t^\top \beta + U_t$.
- $\beta \in \mathbb{R}^k$ is a vector of unknown regression coefficients.
- $\{X_t\}$ is strictly stationary and ergodic with finite second moments, and $\mathrm{E}[X_t X_t^\top]$ is positive definite.
- $\{U_t X_t\}$ is a linear process satisfying the conditions of the CLT stated above.

Consider the OLS estimator of β:

$$T^{1/2}\left(\widehat{\beta}_T - \beta\right) = \left(T^{-1}\sum_{t=1}^{T} X_t X_t^\top\right)^{-1} T^{-1/2}\sum_{t=1}^{T} U_t X_t.$$

By the LLN,

$$\left(T^{-1}\sum_{t=1}^{T} X_t X_t^\top\right)^{-1} \to_p \left(\mathrm{E}[X_t X_t^\top]\right)^{-1}.$$

The CLT implies that

$$T^{-1/2}\sum_{t=1}^{T} U_t X_t \to_d N\left(0,\Omega\right),$$

where Ω is the long-run variance-covariance matrix of $U_t X_t$:

$$\Omega = \mathrm{E}[U_t^2 X_t X_t^\top] + \sum_{j=1}^{\infty} \mathrm{E}[U_t U_{t-j}\left(X_t X_{t-j}^\top + X_{t-j} X_t^\top\right)].$$

Therefore,

$$T^{1/2}\left(\widehat{\beta}_T - \beta\right) \to_d N(0,V),$$

where

$$V = \left(\mathrm{E}[X_t X_t^\top]\right)^{-1}\Omega\left(\mathrm{E}[X_t X_t^\top]\right)^{-1}.$$

Here, U_t may be heteroskedastic and serially correlated. The result is similar to the iid case, except for the expression for Ω, which now includes autocovariance terms that account for serial dependence. The asymptotic variance matrix V can be consistently estimated by

$$\widehat{V}_T = \left(T^{-1}\sum_{t=1}^{T} X_t X_t^\top\right)^{-1} \widehat{\Omega}_T \left(T^{-1}\sum_{t=1}^{T} X_t X_t^\top\right)^{-1},$$

where $\widehat{\Omega}_T$ is the heteroskedasticity and autocorrelation consistent (HAC) estimator of Ω of Newey and West (1987, Econometrica):

$$\widehat{\Omega}_T = T^{-1}\sum_{t=1}^{T} \widehat{U}_t^2 X_t X_t^\top + T^{-1}\sum_{l=1}^{L}\sum_{t=l+1}^{T}\left(1-\frac{l}{L+1}\right)\widehat{U}_t\widehat{U}_{t-l}\left(X_t X_{t-l}^\top + X_{t-l} X_t^\top\right),$$

where L is called the truncation parameter. The weights $1 - l/(L+1)$ are called Bartlett (or triangular) kernel weights; they decline linearly with the lag l and reach zero at $l = L+1$. This downweighting is necessary because sample autocovariances at long lags are estimated imprecisely, and including them without attenuation would prevent $\widehat{\Omega}_T$ from being positive semidefinite. The HAC estimator is consistent provided $L \to \infty$ and $L/T \to 0$ as $T \to \infty$.