Advanced Econometrics

Professor: Julie Shi

Timekeeper: Rui Zhou

Fall 2023

Contents

8 Time Series Analysis

Time series data are *not* random samples (for example, GDPs in adjacent years are not independent, nor are they drawn from identical distributions). Indeed, a time series should be understood as a *single* occurrence of a random event.

Standard models/methods may not apply/work.

$$y_t = \beta x_t + u_t$$

- The random sampling assumption fails, almost for sure.
- There might be endogeneity problem: $\mathrm{E}[u_t | x_t] \neq 0$.
- The disturbances might be serially correlated.

Autoregression: $\mathrm{Cov}[y_t, y_s] \neq 0, \text{ for } t \neq s$.

Serial correlation: $\mathrm{Cov}[\varepsilon_t, \varepsilon_s] \neq 0, \text{ for } t \neq s$.

8.1 Stationarity and Ergodicity

Unconditional mean:

$$\mu_t = \mathrm{E}[y_t] = \operatorname{plim}_{l \to \infty} \frac{1}{l} \sum_{i=1}^{l} y_t$$

Unconditional jth autocovariance ($j = 0, \pm 1, \pm 2, \cdots$):

$$\gamma_{jt} = \mathrm{E}\left[(y_t - \mu_t)\left(y_{t-j} - \mu_{t-j}\right)\right] = \operatorname{plim}_{l \to \infty} \frac{1}{l} \sum_{i=1}^{l} \left[y_t^{(i)} - \mu_t\right]\left[y_{t-j}^{(i)} - \mu_{t-j}\right]$$

Covariance-stationary or weakly stationary: μ_t and γ_{jt} do not depend on date t:

$$\begin{aligned} \mu_t &= \mu, \forall t \\ \gamma_{jt} &= \gamma_j, \forall t, j \end{aligned}$$

White noise

$$\begin{cases} \mathrm{E}[\varepsilon_t] = 0 \\ \mathrm{E}[\varepsilon_t^2] = \mathrm{Var}[\varepsilon_t] = \sigma^2 \\ \mathrm{E}[\varepsilon_t \varepsilon_\tau] = \mathrm{Cov}[\varepsilon_t, \varepsilon_\tau] = 0, \forall t \neq \tau \end{cases}$$

Gaussian white noise

$$\varepsilon_t \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$$

8.2 Moving Average Process

MA(1) process:

$$y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}, \text{ with } \varepsilon_t \sim \text{w.n.}\left(\sigma^2\right)$$

MA(1) is necessarily stationary. Moments are

$$\mathrm{E}\left[y_t\right] = \mu$$

$$\begin{cases} \gamma_0 = \sigma^2 + \theta^2\sigma^2 = \left(1+\theta^2\right)\sigma^2 \\ \gamma_1 = \mathrm{E}\left[\left(y_t - \mu\right)\left(y_{t-1} - \mu\right)\right] = \mathrm{E}\left[\left(\varepsilon_t + \theta\varepsilon_{t-1}\right)\left(\varepsilon_{t-1} + \theta\varepsilon_{t-2}\right)\right] = \theta\mathrm{E}\left[\varepsilon_{t-1}^2\right] = \theta\sigma^2 \\ \gamma_j = 0, \forall j \geq 2 \end{cases}$$

$$\begin{cases} \rho_0 = 1 \\ \rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta}{1+\theta^2} \\ \rho_j = 0, \forall j \geq 2 \end{cases}$$

MA(∞) process

$$y_t = \mu + \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} = \mu + \theta_0\varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots, \text{ with } \varepsilon_t \sim \text{w.n.}\left(\sigma^2\right)$$

MA(∞) is not necessarily stationary. A sufficient condition for stationary: θ's are absolutely summable, i.e., $\sum_{j=0}^{\infty} \left|\theta_j\right| < \infty$.

Based on MA processes, it is straightforward to calculate the expected marginal effect of current innovation on future endogenous variable. Impulse response function (IRF):

$$IRF(j) = \frac{\partial \mathrm{E}_t\left[y_{t+j}\right]}{\partial \varepsilon_t} = \theta_j, \forall j = 0, 1, \cdots$$

8.3 Autoregressive (AR) Process

AR(1) process:

$$y_t = c + \phi y_{t-1} + \varepsilon_t, \text{ with } \varepsilon_t \sim \text{w.n.}\left(\sigma^2\right)$$

AR(1) is not necessarily stationary. Stationary condition: $|\phi| < 1$. If yes, moments are

$$\mathrm{E}\left[y_t\right] = c + \phi \mathrm{E}\left[y_{t-1}\right] \Longrightarrow \mathrm{E}\left[y_t\right] = \frac{c}{1-\phi}$$

$$\begin{cases} \gamma_0 = \mathrm{Var}\left[y_t\right] = \phi^2 \mathrm{Var}\left[y_{t-1}\right] + \sigma^2 = \phi^2 \gamma_0 + \sigma^2 \Longrightarrow \gamma_0 = \dfrac{\sigma^2}{1-\phi^2} \\ \gamma_1 = \mathrm{E}\left[\left(y_t - \mu\right)\left(y_{t-1} - \mu\right)\right] = \mathrm{E}\left[\left(\phi\left(y_{t-1} - \mu\right) + \varepsilon_t\right)\left(y_{t-1} - \mu\right)\right] = \phi \gamma_0 \\ \rho_1 = \dfrac{\gamma_1}{\gamma_0} = \dfrac{\phi \gamma_0}{\gamma_0} = \phi \end{cases}$$

$$\begin{cases} \gamma_2 = \mathrm{E}\left[\left(y_t - \mu\right)\left(y_{t-2} - \mu\right)\right] = \mathrm{E}\left[\left(\phi\left(y_{t-1} - \mu\right) + \varepsilon_t\right)\left(y_{t-2} - \mu\right)\right] = \phi \gamma_1 = \phi^2 \gamma_0 \\ \rho_2 = \dfrac{\gamma_2}{\gamma_0} = \dfrac{\phi^2 \gamma_0}{\gamma_0} = \phi^2 \end{cases}$$

Let $\mu = \mathrm{E}\left[y_t\right] = \dfrac{c}{1-\phi}$,

$$y_t = c + \phi y_{t-1} + \varepsilon_t \Longrightarrow y_t - \mu = \phi\left(y_{t-1} - \mu\right) + \varepsilon_t$$

AR(1) is a special case of AR(p) process:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \text{ with } \varepsilon_t \sim \text{w.n.}\left(\sigma^2\right)$$

AR(p) is not necessarily stationary. Stationary condition: all roots of the pth-order equation $1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$ lie outside unit circle, i.e., all have a modulus strictly greater than 1. If yes, moments are

$$\begin{cases} \mathrm{E}\left[y_t\right] = c/\left(1 - \phi_1 - \phi_2 - \cdots - \phi_p\right) \\ \gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \cdots + \phi_p \gamma_p + \sigma^2 \\ \gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \cdots + \phi_p \gamma_{j-p}, \forall j = 1, 2, \cdots \end{cases}$$

Unconditional and conditional moments of a stationary AR(1):

$$y_t = c + \phi y_{t-1} + \varepsilon_t, \text{ with } \varepsilon_t \sim \text{w.n.}\left(\sigma^2\right), |\phi| < 1$$

Unconditional moments:

$$\begin{cases} \mathrm{E}\left[y_t\right] = \dfrac{c}{1-\phi} \\ \gamma_j = \dfrac{\phi^j}{1-\phi^2} \sigma^2, \forall j = 0, 1, \cdots \\ \rho_j = \phi^j, \forall j = 0, 1, \cdots \end{cases}$$

Conditional moments:

$$\begin{aligned}
y_{t+j} &= c + \phi y_{t+j-1} + \varepsilon_{t+j} \\
&= c + \phi \left(c + \phi y_{t+j-2} + \varepsilon_{t+j-1}\right) + \varepsilon_{t+j} \\
&= (1+\phi) c + \phi^2 \left(c + \phi y_{t+j-3} + \varepsilon_{t+j-2}\right) + \phi \varepsilon_{t+j-1} + \varepsilon_{t+j} \\
&= \left(1 + \phi + \cdots + \phi^{j-1}\right) c + \phi^j y_t + \left(\phi^{j-1} \varepsilon_{t+1} + \cdots + \varepsilon_{t+j}\right) \\
\Longrightarrow \mathrm{E}_t \left[y_{t+j}\right] &= \mathrm{E}_t \left[\left(1 + \phi + \cdots + \phi^{j-1}\right) c + \phi^j y_t\right] \\
&= \frac{1-\phi^j}{1-\phi} c + \phi^j y_t \\
&= \left(1 - \phi^j\right) \mathrm{E}_t \left[y_t\right] + \phi^j y_t, \forall j = 0, 1, \cdots \\
\mathrm{Var} \left[y_{t+j}\right] &= \mathrm{Var} \left[\phi^{j-1} \varepsilon_{t+1} + \cdots + \varepsilon_{t+j}\right] \\
&= \left(\phi^{2(j-1)} + \phi^{2(j-2)} + \cdots + \phi^{2 \cdot 1} + 1\right) \sigma^2 \\
&= \frac{1-\phi^{2j}}{1-\phi^2} \sigma^2 \\
&= \left(1 - \phi^{2j}\right) \gamma_0, \forall j = 0, 1, \cdots
\end{aligned}$$

Remarks:

- $\mathrm{E}\left[y_{t+j}\right]$ is a weighted average of $\mathrm{E}_t\left[y_t\right]$ and y_t, converging to $\mathrm{E}_t\left[y_t\right]$ in the long run.
- $\mathrm{Var}_t\left[y_{t+j}\right]$ rises in j, converging to γ_0 in the long run.

8.4 Autoregressive Moving Average (ARMA) Process

ARMA(p, q) process:

$$\begin{aligned}
y_t =& c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \\
& + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}, \\
& \text{with } \varepsilon_t \sim \text{w.n.} \left(\sigma^2\right)
\end{aligned}$$

ARMA(p, q) is not necessarily stationary. Its stationarity relies entirely on AR coefficients $\left(\phi_1, \phi_2, \cdots, \phi_p\right)$, not on MA coefficients $\left(\theta_1, \theta_2, \cdots, \theta_q\right)$; indeed, ARMA$(p, q)$ and the corresponding AR(p) share the same stationarity condition: all roots of the pth-order equation $1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$ lie outside unit circle, i.e., all have a modulus strictly greater than 1. If yes, moments are

$$\begin{cases}
\mathrm{E}\left[y_t\right] = c / \left(1 - \phi_1 - \phi_2 - \cdots - \phi_p\right) \\
\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \cdots + \phi_p \gamma_{j-p}, \forall j > q \\
\gamma_j \text{ has no analytical expression for } j = 0, 1, \cdots q
\end{cases}$$

8.5 Lag Polynomial Representation

Lag operator is defined as

$$
\begin{aligned}
&Ly_t = y_{t-1}, L^2 y_t = y_{t-2}, \cdots, L^p y_t = y_{t-p} \\
&L\beta = \beta, L\left(\beta y_t\right) = \beta L y_t, L\left(y_{1t} \pm y_{2t}\right) = Ly_{1t} \pm Ly_{2t}, L^p L^q y_t = L^{p+q} y_t
\end{aligned}
$$

MA(q) in a lag polynomial representation:

$$
\begin{aligned}
&y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \\
\Longrightarrow &y_t = \mu + \theta\left(L\right)\varepsilon_t, \text{ where } \theta\left(L\right) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q
\end{aligned}
$$

AR(p) in a lag polynomial representation:

$$
\begin{aligned}
&y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \\
\Longrightarrow &\phi\left(L\right) = c + \varepsilon_t, \text{ where } \phi\left(L\right) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p
\end{aligned}
$$

ARMA(p, q) in a lag polynomial representation:

$$
\begin{aligned}
y_t =& c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} \\
&+ \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}, \\
\Longrightarrow \phi\left(L\right) y_t =& c + \theta\left(L\right)\varepsilon_t
\end{aligned}
$$

For a log polynomial, say $\phi\left(L\right)$, if all roots of $\phi\left(z\right) = 0$ lie outside unit circle, then there exists a lag polynomial with absolutely summable coefficients, denoted by $\phi\left(L\right)^{-1}$, such that $\phi\left(L\right)^{-1}\phi\left(L\right) y_t = y_t$. Example:

$$
\begin{aligned}
&\phi\left(L\right) = 1 - \phi L, |\phi| < 1 \\
\Longrightarrow &\phi\left(L\right)^{-1} = 1 + \phi L + \phi^2 L^2 + \cdots
\end{aligned}
$$

A stationary AR(p) has a stationary MA(∞) representation:

$$
\begin{aligned}
&\phi\left(L\right) y_t = c + \varepsilon_t \\
\Longrightarrow &y_t = \phi\left(L\right)^{-1} c + \phi\left(L\right)^{-1}\varepsilon_t
\end{aligned}
$$

Take AR(1) for example:

$$
\begin{aligned}
&y_t = c + \phi y_{t-1} + \varepsilon_t \iff \left(1 - \phi L\right) y_t = c + \varepsilon_t \\
\Longrightarrow &y_t = \left(1 - \phi L\right)^{-1} c + \left(1 - \phi L\right)^{-1}\varepsilon_t \\
&\quad = \left(1 + \phi L + \phi^2 L^2 + \cdots\right) c + \left(1 + \phi L + \phi^2 L^2 + \cdots\right)\varepsilon_t \\
&\quad = \left(1 + \phi + \phi^2 + \cdots\right) c + \left(1 + \phi L + \phi^2 L^2 + \cdots\right)\varepsilon_t \\
&\quad = \frac{c}{1 - \phi} + \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2}\cdots
\end{aligned}
$$

Take MA(1) as another example:

$$y_t = \mu + \varepsilon_t - \theta\varepsilon_{t-1}, |\theta| < 1$$

Note that here we rewrite the term $\theta\varepsilon_{t-1}$ as a negative one, which convenient our computation without distorting its mathematical meaning.

$$\begin{aligned}
&y_t = \mu + \varepsilon_t - \theta\varepsilon_{t-1}, |\theta| < 1 \\
\Longrightarrow &y_t - \mu = (1 - \theta L)\,\varepsilon_t \\
\Longrightarrow &(1 - \theta L)^{-1}(y_t - \mu) = \varepsilon_t \\
\Longrightarrow &\left(1 + \theta L + \theta^2 L^2 + \cdots\right) y_t - \frac{\mu}{1-\theta} = \varepsilon_t \\
\Longrightarrow &y_t = \frac{\mu}{1-\theta} + \varepsilon_t - \theta y_{t-1} - \theta^2 y_{t-2} - \cdots
\end{aligned}$$

which is an AR(∞).

8.6 Estimation of ARMA: MLEs

Case 1: Gaussian AR(1)

$$y_t = c + \phi y_{t-1} + \varepsilon_t, \varepsilon_t \overset{\text{i.i.d.}}{\sim} N\left(0, \sigma^2\right)$$

This is a Markov process, the probability density conditioning on all histories is equivalent to conditioning on current history only. In order to obtain conditional distributions, we focus the "random" part. In AR(1), we turn to $\varepsilon_t = y_t - c - \phi y_{t-1}$, which follows a distribution of $N\left(0, \sigma^2\right)$.

$$\begin{aligned}
f\left(y_2 | y_1\right) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\left(y_2 - c - \phi y_1\right)^2}{2\sigma^2}\right] \\
f\left(y_3 | y_1, y_2\right) = f\left(y_3 | y_2\right) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\left(y_3 - c - \phi y_2\right)^2}{2\sigma^2}\right] \\
&\cdots \\
f\left(y_T | y_1, y_2, \cdots, y_T\right) = f\left(y_T | y_{T-1}\right) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\left(y_T - c - \phi y_{T-1}\right)^2}{2\sigma^2}\right]
\end{aligned}$$

The conditional joint density is:

$$f\left(y_2, y_3, \cdots, y_T | y_1\right) = \prod_{t=2}^{T} f\left(y_t | y_{t-1}\right) = \prod_{t=2}^{T} \left\{ \exp\left[-\frac{\left(y_t - c - \phi y_{t-1}\right)^2}{2\sigma^2}\right] \right\}$$

The conditional log likelihood function is:

$$\mathcal{L}\left(c, \phi, \sigma^2\right) = -\frac{T-1}{2} \ln 2\pi - \frac{T-1}{2} \ln \sigma^2 - \sum_{t=2}^{T} \frac{\left(y_t - c - \phi y_{t-1}\right)^2}{2\sigma^2}$$

Remarks:

- For AR processes, MLEs of coefficients are identical to OLS estimates; because OLS estimates are consistent, so are MLEs –this fact doesn't hinge on normality.
- For MA processes, MLEs have no analytical solution; MLEs must be found by numerical optimization.

8.7 Error Correction Model

ECM: It captures the idea that changes in a variable partially reflect adjustments of the variable to its deviations from a (long-run) equilibrium level/relationship.

Take AR(1) as an example.

$$\begin{aligned} y_t &= c + \phi y_{t-1} + \varepsilon_t \\ \mu = \mathrm{E}\left[y_t\right] &= \frac{c}{1-\phi} \\ \Longrightarrow \Delta y_t &= -\left(1-\phi\right)\left(y_{t-1} - \mu\right) + \varepsilon_t \end{aligned}$$

Example: An ADL model

$$y_t = \beta_0 + \beta_1 y_{t-1} + \gamma_0 x_t + \gamma_1 x_{t-1} + \varepsilon_t, |\beta_1| < 1$$

Long-run relationship (by assumption):

$$y^* = a + bx^*, \text{ where } y^* = \mathrm{E}\left[y_t\right], x^* = \mathrm{E}\left[x_t\right]$$

Easy to show that

$$a = \frac{\beta_0}{1-\beta_1}, b = \frac{\gamma_0 + \gamma_1}{1-\beta_1}$$

ECM representation:

$$\Delta y_t = \gamma_0 \Delta x_t - \left(1-\beta_1\right)\left(y_{t-1} - a - bx_{t-1}\right) + \varepsilon_t$$

8.8 Multivariate Processes

8.8.1 Vector Autoregressions (VAR): Stationarity

VAR(p) process:

$$\vec{y}_t = \vec{c} + \Phi_1 \vec{y}_{t-1} + \Phi_2 \vec{y}_{t-2} + \cdots + \Phi_p \vec{y}_{t-p} + \vec{\varepsilon}_t, \text{ with } \vec{\varepsilon} \sim \text{w.n.}\,(\Sigma)$$

Lag polynomial representation of VAR(p):

$$\Phi\,(L)\,\vec{y}_t = \vec{c} + \vec{\varepsilon}_t, \text{ where } \Phi\,(L) = \mathbf{I}_n - \Phi_1 L - \Phi_2 L^2 - \cdots - \Phi_p L^p$$

VAR(p) is not necessarily stationary. Stationarity condition: all roots of the npth-order equation $\det\left[\Phi\,(z)\right] = \det\left(\mathbf{I}_n - \Phi_1 z - \Phi_2 z^2 - \cdots - \Phi_p z^p\right) = 0$ lie outside unit circle, i.e., all have a modulus strictly greater than 1.

VAR(1) process:

$$\vec{y}_t = \vec{c} + \Phi_1 \vec{y}_{t-1} + \vec{\varepsilon}_t, \text{ with } \varepsilon_t \sim \text{w.n.}\left(\sigma^2\right)$$

Why VAR?

- Better fit; and can do a better forecasting job than early macro models.
- Solutions to DSGEs can have a VAR representation.

VAR: Estimation

- The idea is pretty much the same as in the univariate case.
- For coefficients, conditional MLE is equivalent to OLS estimates equation by equation.