Macro II: Stochastic Processes II

Professor Griffy

UAlbany

Spring 2021

Introduction

- Today: study a variety of stochastic processes that show up in macroeconomics.
- Then, discuss detrending data.

Stochastic Processes

- Conditional expectations and linear projections
- White noise
- AR(1)
- MA(1)
- ARMA(p,q)
- Detrending data

What is a Stochastic Process?

- Stochastic process is an infinite sequence of random variables $\{X_t\}_{t=-\infty}^{\infty}$
- j'th autocovariance $= \gamma_j = C(X_t, X_{t-j})$
- Strict stationarity: distribution of $(X_t, X_{t+j_1}, X_{t+j_2}, ...X_{t+j_n},)$ does not depend on t
- Covariance stationarity: $\bar{X}_t$ and $C(X_t, X_{t-j})$ do not depend on t

Defining a Conditional Density

- Work with random vector $\underline{x} = (X, Y) \sim F(x, y)$.
  - X and Y are random variables
  - x and y are realizations of the random variables
  - $F(x, y)$ is joint cumulative distribution
  - $f(x, y)$ is joint density function

Conditional Variables and Independence

- Conditional probability
  - when $\Pr(\underline{x} \in B) > 0$,

$$\Pr(\underline{x} \in A | \underline{x} \in B) = \Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

- Conditional distribution $F(y|x)$ (handles $\Pr(B) = 0$)
  - Marginal distribution: $F_X(x) = \Pr(X \leq x)$
  - $F(y|x)$ is $\Pr(Y \leq y)$ conditional on $X \leq x$

- Independence: The random variables X and Y are independent if

$$F(x, y) = F_X(x) F_Y(y)$$

  - If X and Y are independent, then

$$F(y|x) = F_Y(y)$$

    and

$$F(x|y) = F_X(x)$$

  - i.i.d means independent and identically distributed
- Conditional (mathematical, rational) expectation

$$E(Y|x) = \int_{-\infty}^{\infty} y dF(y|x) = \int_{-\infty}^{\infty} y f(y|x) dy.$$

Conditional Expectations as OLS

- Conditional expectaton: function that minimizes the mean squared forecast error

$$E\left(Y|\underline{X}\right) = \arg \min_{\{f(\cdot)\}} E\left(\left[Y - f\left(\underline{X}\right)\right]^2\right)$$

  where $\underline{X}$ is a vector
- Best linear predictor or linear projection: linear function that minimizes the mean squared forecast error

$$\hat{E}\left(Y|\underline{X}\right) = \arg \min_{\{\text{ linear } f(\cdot)\}} E\left(\left[Y - f\left(\underline{X}\right)\right]^2\right)$$

- $\hat{E}\left(Y|\underline{X}\right) = E\left(Y|\underline{X}\right)$ when $E\left(Y|\underline{X}\right)$ is linear
- OLS estimate since OLS gives best linear unbiased estimates

White noise

$$\{\varepsilon_t\}_{t=-\infty}^{\infty}$$

- $E(\varepsilon_t) = 0, \ \forall t$
- $V(\varepsilon_t) = \sigma_\varepsilon^2, \ \forall t$
- $C(\varepsilon_t, \varepsilon_{t-j}) = 0, \ \forall t, \ j \neq 0$

First-order autoregressive (AR(1)) process

$$x_t = \alpha + \phi x_{t-1} + \varepsilon_t$$

- ε_t is white noise and $|\phi| < 1$ as required by stationarity
- By recursive substitution under stationarity

$$\begin{aligned} x_t &= \alpha + \varepsilon_t + \phi\left[\alpha + \phi x_{t-2} + \varepsilon_{t-1}\right] \\ &= \frac{\alpha}{1-\phi} + \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}. \end{aligned}$$

- $E\left(x_t\right) = \alpha / \left(1 - \phi\right)$

Moments

- Facts

$$V(aX + bY) = a^2 V(X) + b^2 V(Y) + 2abC(X, Y)$$

$$C(aX + bY, cX + dY) = acV(X) + bdV(Y) + (ad + bc)\, C(X, Y)$$

- Since the value of x_t can be expressed as

$$x_t = \frac{\alpha}{1 - \phi} + \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j},$$

- The variance of x_t is

$$V(x_t) = \sum_{j=0}^{\infty} \left(\phi^j\right)^2 \sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2}{1 - \phi^2}.$$

Covariances

- Covariance

$$
\begin{aligned}
C\left(x_t, x_{t-1}\right) &= C\left(\alpha + \phi x_{t-1} + \varepsilon_t, x_{t-1}\right) \\
&= 0 + \phi V\left(X\right) + 0 = \phi \frac{\sigma_\varepsilon^2}{1-\phi^2}, \\
C\left(x_t, x_{t-k}\right) &= C\left(\phi^k x_{t-k} + \sum_{j=0}^{k-1} \phi^j \varepsilon_{t-j}, x_{t-k}\right) \\
&= \phi^k \frac{\sigma_\varepsilon^2}{1-\phi^2} = \phi^k V\left(x_t\right).
\end{aligned}
$$

- Expectation: If $\{\varepsilon_t\}$ is i.i.d. and $\alpha = 0$, $E\left(x_t \mid x_{t-k}\right) = \phi^k x_{t-k}$

AR(p)

- Autoregressive function of p lagged $x's$

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + \varepsilon_t$$

- Defining $x_{t-j} = L^j x_t$, we can rewrite an AR(p) process as

$$\left(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p\right) x_t = \alpha + \varepsilon_t$$

- Stationarity condition: The roots of

$$1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p = 0$$

lie outside the unit circle ($|z| > 1$ when real)

AR(p)

- For AR(1), $\phi < 1$ yields stationarity
- If stationarity holds, we can rewrite x_t as a function of infinitely lagged $\varepsilon's$

First-order moving average (MA(1)) process

$$x_t = \alpha + \varepsilon_t + \theta \varepsilon_{t-1}$$

- ε_t is white noise
- $E(x_t) = \alpha$
- $V(x_t) = \left(1 + \theta^2\right) \sigma_\varepsilon^2$
- $C(x_t, x_{t-1}) = C(\varepsilon_t + \theta\varepsilon_{t-1}, \varepsilon_{t-1} + \theta\varepsilon_{t-2}) = \theta\sigma_\varepsilon^2$
- $C(x_t, x_{t-k}) = C(\varepsilon_t + \theta\varepsilon_{t-1}, \varepsilon_{t-k} + \theta\varepsilon_{t-k-1}) = 0, \; k > 1$

MA cont'd

- Rewrite with lag operator as

$$x_t - \alpha = (1 + \theta L)\varepsilon_t$$

- When the root of

$$1 + \theta z = 0$$

lie outside unit circle (when $|\theta| < 1$) x_t is said to be invertible

$$\begin{aligned} \varepsilon_t &= \frac{(x_t - \alpha)}{(1 + \theta L)} \\ &= \frac{-\alpha}{1 + \theta} + \sum_{j=0}^{\infty} (-\theta)^j x_{t-j}, \end{aligned}$$

- Express residual as infinite recursion of lagged $x\prime s$

MA(q)

- x_t is a function of q lagged residuals

$$x_t = \alpha + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

- Rewriting with lag operator yields

$$\left(1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q\right) \varepsilon_t = x_t - \alpha$$

- Invertibility condition is that the roots of

$$1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q = 0$$

  lie outside the unit circle ($|z| > 1$ when real)
- If the invertibility condition holds, we can write the ε_t as an infinite function of lagged $x's$

ARMA process

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p}$$
$$+ \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}.$$

- Stationarity condition
  - Depends entirely on autoregressive coefficients
  - The roots of

$$1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p = 0,$$

must lie outside the unit circle ($| \cdot | > 1$ when real)

EX

- Example: For AR(1)

$$1 - \phi_1 z = 0$$

implying

$$z = \frac{1}{\phi_1}$$

need

$$|z| = |\phi_1^{-1}| > 1 \quad \text{requiring } |\phi_1| < 1$$

Invertibility Condition

- The roots of

$$1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_p z^p = 0,$$

  must lie on or outside the unit circle ($|\cdot| > 1$ when real)
- Example: For MA(1)

$$1 - \theta_1 z = 0$$

  implying

$$z = \frac{1}{\theta_1}$$

  need

$$|z| = |\theta_1^{-1}| \geq 1 \quad \text{requiring } |\theta_1| \leq 1,$$

  where unity is included as a limit

Stationarity and invertibility

- Stationarity and invertibility imply
  - if ε_t is i.i.d., then ε_t is the innovation to x_t

$$\varepsilon_t = x_t - E\left(x_t \middle| x_{t-1}, x_{t-2}, ...\right).$$

  - knowledge of the entire sequences $\{\varepsilon_{t-j}\}_{j=0}^{\infty}$ and $\{x_{t-j}\}_{j=0}^{\infty}$ is equivalent

Detrending

- Most of our (business cycle) models have nothing to say about trends in the data.
- i.e., these models generally don't explain growth.
- Need to detrend to get an appropriate data series.
- $Y_t = X_t + z_t$
  - X_t is stationary
  - z_t is a trend
- Trend stationary
  - z_t is deterministic
  - example: $z_t = \alpha t$

Difference Stationary

- Difference stationary
  - z_t is a random walk with $\{\varepsilon_t\}$ a white noise process

$$z_t = z_0 + \sum_{j=1}^{t} \varepsilon_{t-j}$$

$$z_{t-1} = z_0 + \sum_{j=1}^{t-1} \varepsilon_{t-j}$$

$$z_t - z_{t-1} = \varepsilon_t$$

Three approaches

- Linear detrending: Regress data on time and take residuals
- Use Hodrick-Prescott filter to separate data into a trend component and residuals and take residuals
- First difference the data

HP Filter

- Most common approach: HP Filter.
- Idea: isolate low-frequency trends from high frequency cycles.
- Let $\{y_t\}_{t=1}^{\infty}$ be a given series, where $y_t = x_t + z_t$ as before.
- x_t is the trend component, z_t is cyclical.
- Let λ be a parameter to be specified later, and consider the problem

$$\min_{x_1, x_2, \ldots, x_T} \sum_{t=1}^{T} (y_t - x_t)^2 + \lambda \sum_{t=2}^{T-1} [(x_{t+1} - x_t) - (x_t - x_{t-1})]^2 \tag{1}$$

- What is going on here?
  - We are minimizing the cyclical component (first part), by moving the trend closer to the data.
  - But we are getting penalized (λ) for making the trend too closely reflect the data.

Next Time

- Discuss expectations in linear difference equations.
- Please turn your homework in by this evening.
- See my webpage for new homework (may not be up by class).