Time Series Econometrics, 2ST111

Lecture 6. Linear Regression Models
Covariance Stationary Vector Processes

Yukai Yang

Department of Statistics, Uppsala University

Outline of Today's Lecture

Outline:

- Linear Regression Models
- Covariance Stationary Vector Processes (not 10.2-10.3)

The Content

- Repetition of econometrics content
- Introduce results we will rely on later: covariance matrices of estimators under heteroscedasticity, in AR models, etc
- Traditional assumptions not fulfilled, but what are the consequences?
- We focus on the linear regression model:

$$y_t = \mathbf{x}_t' \beta + u_t \tag{1}$$

- Note that much of the head-scratching in practice comes from the fact that the data is fundamentally different from what we usually deal with in econometrics courses
- By varying the assumptions for $\mathbf{x}_t$ and u_t, several different cases arise

The Basics of Linear Regression

- Given some data, $(y_1, \ldots, y_T)$, the OLS estimator of the $k \times 1$ parameter vector $\boldsymbol{\beta}$ in (1) is the minimizer of the residual sum of squares:

$$RSS = \sum_{t=1}^{T} (y_t - \mathbf{x}_t' \boldsymbol{\beta})^2 \tag{2}$$

- We write the estimator for $\boldsymbol{\beta}$ as

$$\mathbf{b} = \left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \left(\sum_{t=1}^{T} \mathbf{x}_t y_t \right) \tag{3}$$

given that $\left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1}$ exists.

- The existence of the inverse is equivalent to a full rank assumption, which in turn requires $T > k$

- We may write (1) with matrix notation:

$$\begin{aligned} y_1 &= \mathbf{x}_1'\boldsymbol{\beta} + u_1 \\ y_2 &= \mathbf{x}_2'\boldsymbol{\beta} + u_2 \\ &\vdots \\ y_T &= \mathbf{x}_T'\boldsymbol{\beta} + u_T \end{aligned} \qquad \text{(i)}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1'\boldsymbol{\beta} + u_1 \\ \mathbf{x}_2'\boldsymbol{\beta} + u_2 \\ \vdots \\ \mathbf{x}_T'\boldsymbol{\beta} + u_T \end{pmatrix} \qquad \text{(ii)}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_T' \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix} \qquad \text{(iii)}$$

$$\underbrace{\mathbf{y}}_{T\times 1} = \underbrace{\mathbf{X}}_{T\times k}\underbrace{\boldsymbol{\beta}}_{k\times 1} + \underbrace{\mathbf{u}}_{T\times 1} \qquad \text{(iv)}$$

- The OLS estimator can thus also be written as

$$\begin{aligned}
\mathbf{b} &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} \\
&= \left(\begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_T \end{pmatrix} \begin{pmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \\ \vdots \\ \mathbf{x}'_T \end{pmatrix}\right)^{-1} \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_T \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} \\
&= \left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}'_t\right)^{-1} \left(\sum_{t=1}^{T} \mathbf{x}_t y_t\right)
\end{aligned}$$

- The latter form is convenient from an asymptotic point of view, as we can divide by T in both the first term (i.e. the inverse) and the second term and using various LLN/CLT arguments

The Basics of Linear Regression

- The residual is $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\mathbf{b}$, which Hamilton refers to as the *sample* residual (as opposed to the *population* residual $\mathbf{u} = \mathbf{y} - \mathbf{X}\beta$)
- I will use residual to refer to the sample residual and error term to refer to the population residual
- The residuals are by construction orthogonal to $\mathbf{X}$:

$$\begin{aligned}
\hat{\mathbf{u}}'\mathbf{X} &= (\mathbf{y}' - \mathbf{b}'\mathbf{X}')\mathbf{X} \\
&= (\mathbf{y}' - \mathbf{y}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}')\mathbf{X} \\
&= \mathbf{y}'\mathbf{X} - \mathbf{y}'\mathbf{X}\underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}}_{=\mathbf{I}_k} \\
&= \mathbf{0}
\end{aligned}$$

- A useful representation of the OLS estimator is:

$$\mathbf{b} = \beta + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u} \tag{4}$$

Case 1: Classical Regression Assumptions

- The "classical" regression assumptions
- Assumptions:
  a) $\mathbf{x}_t$ is a vector of deterministic variables
  b) u_t is i.i.d. with mean 0 and variance σ^2
  c) u_t is Gaussian
- b) and c) is the same as saying u_t is iid $\sim N(0, \sigma^2)$

Case 1: Classical Regression Assumptions

- By using the useful form of $\mathbf{b}$ from (4), the expectation is easily seen to be

$$
\begin{aligned}
E(\mathbf{b}) &= E\left(\boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u}\right) \\
&= \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'E(\mathbf{u}) \\
&= \boldsymbol{\beta}
\end{aligned}
$$

The variance-covariance matrix is

$$
\begin{aligned}
E\left[(\mathbf{b}-\boldsymbol{\beta})(\mathbf{b}-\boldsymbol{\beta})'\right] &= E\left[\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u}\right)\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u}\right)'\right] \\
&= E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right] \\
&= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'E(\mathbf{u}\mathbf{u}')\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\
&= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\sigma^2\mathbf{I}_T\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\
&= \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}
\end{aligned}
$$

Case 1: Classical Regression Assumptions

- Without the normality assumption (but assuming a) and b)), the Gauss-Markov theorem establishes optimality of **b** within the class of unbiased and linear estimators
- If we also assume normality, **b** is *exactly* Gaussian
- In general, if $\mathbf{z} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, then $\boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{z} \sim N(\boldsymbol{\mu}, \boldsymbol{\Lambda}\boldsymbol{\Sigma}\boldsymbol{\Lambda}')$
- So $\mathbf{b} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \sim N\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right)$
- **Important**: t and F tests are exact due to this

Case 2 and 3

- Case 2: a) $\mathbf{x}_t$ stochastic and independent of u_s, all t, s; b) $u_t \sim i.i.d. N(0, \sigma^2)$
  - The results are the same (except that $\mathbf{b}|\mathbf{X}$ is Gaussian and not $\mathbf{b}$)
- Case 3: a) $\mathbf{x}_t$ stochastic and independent of u_s for all t, s; b) u_t non-Gaussian but i.i.d. with mean zero, variance σ^2, and $E(u_t^4) = \mu_4 < \infty$
  - There are additional technical assumptions in c)-e), which assume $\mathbf{x}_t$ to be 'well-behaved', allowing for the use of previous results for martingale difference sequences
  - Still unbiased, but only asymptotically Gaussian
  - t and F tests are inexact (asymptotically valid)

Case 4: Stationary autoregression with independent errors

Assumption (8.4)

The regression model is

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t \tag{5}$$

with

- *roots of* $(1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p) = 0$ *outside of the unit circle,*
- $\{\epsilon_t\}$ *an i.i.d. sequence with* $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = \sigma^2$, *and* $E(\epsilon_t^4) = \mu_4 < \infty$.

Case 4: Autoregression

- The autoregression can be written as a typical regression model:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t \tag{6}$$

  where $\mathbf{x}_t' = \begin{pmatrix} 1 & y_{t-1} & \cdots & y_{t-p} \end{pmatrix}$ and $u_t = \epsilon_t$.
- **However**, it cannot satisfy the independence assumption found in the previous cases:

Assumption (8.2(a), 8.3(a))

$\mathbf{x}_t$ *stochastic and independent of* u_s *for all* t, s.

Case 4: Autoregression

- The assumption states that $\mathbf{x}_t$ and u_{t-1} are independent. But:

$$\mathbf{x}_t = \begin{pmatrix} 1 \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{pmatrix} \qquad E(\mathbf{x}_t u_{t-1}) = E \begin{pmatrix} u_{t-1} \\ y_{t-1}u_{t-1} \\ y_{t-2}u_{t-1} \\ \vdots \\ y_{t-p}u_{t-1} \end{pmatrix}$$

- Note that $E(y_{t-1}u_{t-1}) = E(u_{t-1}^2) = \sigma^2$
- Since $E(\mathbf{x}_t u_{t-1}) \neq \mathbf{0}$, $\mathbf{x}_t$ and u_{t-1} are dependent and Assumption 8.2(a) and 8.3(a) cannot hold
- Thus, the previous cases considered are insufficient in the case of an autoregression

Case 4: Autoregression

- Consequence: $\mathbf{b}_T$ is not unbiased

$$E(\mathbf{b}) = \beta + E\left[\left(T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1} T^{-1}\sum_{t=1}^{T}\mathbf{x}_t u_t\right] \tag{7}$$

- What about asymptotics? Consider the first part:

$$\sum_{t=1}^{T}\frac{\mathbf{x}_t\mathbf{x}_t'}{T} = \sum_{t=1}^{T}\frac{1}{T}\begin{pmatrix}1\\ y_{t-1}\\ \vdots\\ y_{t-p}\end{pmatrix}\begin{pmatrix}1 & y_{t-1} & \cdots & y_{t-p}\end{pmatrix}$$

$$= \begin{pmatrix} 1 & T^{-1}\sum y_{t-1} & \cdots & T^{-1}\sum y_{t-p} \\ T^{-1}\sum y_{t-1} & T^{-1}\sum y_{t-1}^2 & \cdots & T^{-1}\sum y_{t-1}y_{t-p} \\ \vdots & \vdots & \ddots & \vdots \\ T^{-1}\sum y_{t-p} & T^{-1}\sum y_{t-p}y_{t-1} & \cdots & T^{-1}\sum y_{t-p}^2 \end{pmatrix}$$

Case 4: Autoregression

- From Prop 7.5, we know that $T^{-1}\sum_{t=1}^{T} y_{t-j} \xrightarrow{p} E(y_t) = \mu$ and that

$$T^{-1}\sum_{t=1}^{T} y_{t-i}y_{t-j} \xrightarrow{p} E(y_{t-i}y_{t-j}) = \gamma_{|i-j|} + \mu^2.$$

- This gives us:

$$\mathbf{Q} = \begin{pmatrix} 1 & \mu & \mu & \cdots & \mu \\ \mu & \gamma_0 + \mu^2 & \gamma_1 + \mu^2 & \cdots & \gamma_{p-1} + \mu^2 \\ \mu & \gamma_1 + \mu^2 & \gamma_0 + \mu^2 & \cdots & \gamma_{p-2} + \mu^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu & \gamma_{p-1} + \mu^2 & \gamma_{p-2} + \mu^2 & \cdots & \gamma_0 + \mu^2 \end{pmatrix}$$

- The second term is an MDS and separately the limits are:

$$\left(\frac{\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'}{T}\right)^{-1} \xrightarrow{p} \mathbf{Q}^{-1}, \qquad \left(\frac{\sum_{t=1}^{T} \mathbf{x}_t u_t}{\sqrt{T}}\right) \xrightarrow{d} N(\mathbf{0}, \sigma^2\mathbf{Q})$$

Case 4: Autoregression

- Thus,

$$\mathbf{b}_T - \boldsymbol{\beta} \xrightarrow{p} \mathbf{0} \tag{8}$$

$$\sqrt{T}(\mathbf{b}_T - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \sigma^2 \mathbf{Q}^{-1}) \tag{9}$$

- In an autoregression, the OLS estimator is biased but consistent
- t and F tests are asymptotically valid
- Example: AR(1)

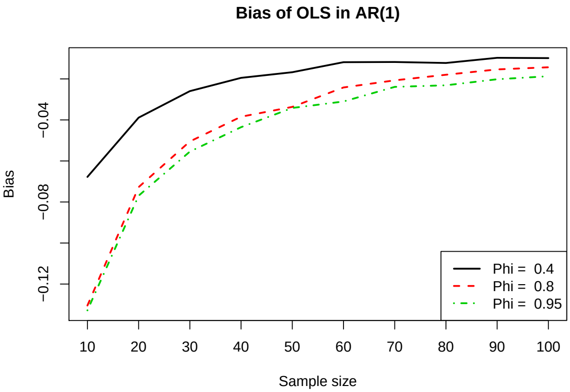
$$y_t = \phi y_{t-1} + \epsilon_t$$

  where the process is stationary ($|\phi| < 1$).
- The matrix $\mathbf{Q}$ is just a scalar, $E(y_{t-1}^2) = \gamma_0 = \sigma^2/(1-\phi^2)$
- The asymptotic distribution result above thus implies that

$$\sqrt{T}(\hat{\phi}_T - \phi) \xrightarrow{d} N\left(0, \sigma^2 \frac{1-\phi^2}{\sigma^2}\right) = N(0, 1-\phi^2)$$

Case 4: Autoregression

- How large is the bias?
- Model: $y_t = \phi y_{t-1} + \epsilon_t$, with $T = 10, 20, \ldots, 100$ and $\phi = 0.4, 0.8, 0.95$
- Bias: $\hat{\phi} - \phi$



Case 5: Errors Gaussian with Known Var-Cov Matrix

- Further relaxation of assumptions:
  (a) $\mathbf{x}_t$ stochastic
  (b) $\mathbf{u}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{V})$
  (c) $\mathbf{V}$ is a known positive definite matrix
- Heteroskedasticity: $\mathbf{V}$ diagonal, but $\mathbf{V} \neq \sigma^2\mathbf{I}_T$
- Autocorrelation: $\mathbf{V}$ non-diagonal
- OLS still unbiased:

$$E\left((\mathbf{b}-\boldsymbol{\beta})|\mathbf{X}\right) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}|\mathbf{X}) = \mathbf{0}$$

and since $E(Y) = E_X(E(Y|X))$:

$$E\left((\mathbf{b}-\boldsymbol{\beta})\right) = E_{\mathbf{X}}\left\{E[(\mathbf{b}-\boldsymbol{\beta})|\mathbf{X}]\right\} = E_{\mathbf{X}}(\mathbf{0}) = \mathbf{0}$$

- For the estimator:

$$\mathbf{b}|\mathbf{X} \sim N\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right)$$

- OLS inefficient, use GLS (Section 8.3)

Case 6: Errors Uncorr but with General Heteroskedasticity

- Unknown and general heteroskedasticity:

Assumption (8.6)

a) $\mathbf{x}_t$ *stochastic, including perhaps lags of* y

b) $\mathbf{x}_t u_t$ *is an MDS*

c) $E(u_t^2 \mathbf{x}_t \mathbf{x}_t') = \mathbf{\Omega}_t$ *(positive definite) and*

   (i) $\sum_{t=1}^{T} \frac{\mathbf{\Omega}_t}{T} \to \mathbf{\Omega}$

   (ii) $\sum_{t=1}^{T} \frac{u_t^2 \mathbf{x}_t \mathbf{x}_t'}{T} \xrightarrow{p} \mathbf{\Omega}$

d)-e) $\mathbf{x}_t$ *and* u_t *well-behaved such that certain asymptotic results apply*

- Example: let $\mathbf{x}_t = x_t$ and suppose that $E(x_t^2) = \mu_2$ and $E(x_t^4) = \mu_4$. Suppose the heteroskedasticity is of the form

$$E(u_t^2 | x_t) = a + b x_t^2$$

Case 6: Errors Uncorr but with General Heteroskedasticity

- In this case,

$$\begin{aligned}\mathbf{\Omega}_t &= E(u_t^2 x_t^2) = E_x[E(u_t^2|x_t^2)x_t^2] = E_x\left[\left(a + bx_t^2\right)x_t^2\right] \\ &= a\mu_2 + b\mu_4\end{aligned}$$

  so $\mathbf{\Omega}_t = \mathbf{\Omega}$ for all t and (i) is satisfied.

- By the LLN we have (ii)

$$\sum_{t=1}^{T} \frac{u_t^2 x_t^2}{T} \xrightarrow{p} E(u_t^2 x_t^2) = \mathbf{\Omega}$$

- Case 6 thus allows for fairly general types of conditional heteroskedasticity

- By Assumption 8.6 (and Proposition 7.9):

$$\left(\sum_{t=1}^{T} \frac{\mathbf{x}_t \mathbf{x}_t'}{T}\right)^{-1} \xrightarrow{p} \mathbf{Q}^{-1}$$

$$\sum_{t=1}^{T} \frac{\mathbf{x}_t u_t}{\sqrt{T}} \xrightarrow{d} N(\mathbf{0}, \mathbf{\Omega})$$

- Hence

$$\sqrt{T}(\mathbf{b} - \beta) \xrightarrow{d} N(\mathbf{0}, \mathbf{Q}^{-1}\mathbf{\Omega}\mathbf{Q}^{-1})$$

- White (1980) proposed: use $\hat{\mathbf{Q}}_T = T^{-1}\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'$ and $\hat{\mathbf{\Omega}}_T = T^{-1}\sum_{t=1}^{T} \hat{u}_t^2 \mathbf{x}_t \mathbf{x}_t'$
- Then (Proposition 8.3):

$$\hat{\mathbf{Q}}_T^{-1}\hat{\mathbf{\Omega}}_T\hat{\mathbf{Q}}_T^{-1} \xrightarrow{p} \mathbf{Q}^{-1}\mathbf{\Omega}\mathbf{Q}^{-1}$$

Case 6: Errors Uncorr but with General Heteroskedasticity

- We can treat $\mathbf{b}_T$ as if

$$\mathbf{b}_T \approx N\left(\boldsymbol{\beta}, \frac{\hat{\mathbf{Q}}_T^{-1}\hat{\boldsymbol{\Omega}}_T\hat{\mathbf{Q}}_T^{-1}}{T}\right)$$

- As it turns out, the expression for the variance is quite nice:

$$\frac{\hat{\mathbf{Q}}_T^{-1}\hat{\boldsymbol{\Omega}}_T\hat{\mathbf{Q}}_T^{-1}}{T} = (\mathbf{X}'_T\mathbf{X}_T)^{-1}\left(\sum_{t=1}^{T}\hat{u}_t^2\mathbf{x}_t\mathbf{x}_t'\right)(\mathbf{X}'_T\mathbf{X}_T)^{-1}$$

- Consistent, even when an unknown form of heteroskedasticity is present
- Also known as the sandwich estimator

- Note here that a model with autocorrelation may have it either in the variable itself, or in the error term:

$$
\begin{aligned}
\phi(L) y_t &= \epsilon_t & y_t &= \epsilon_t \\
\epsilon_t &= u_t & \phi(L)\epsilon_t &= u_t \\
u_t &\sim iid(0, \sigma^2) & u_t &\sim iid(0, \sigma^2)
\end{aligned}
$$

- However, for zero-mean processes the two models are identical:

$$
\begin{aligned}
y_t &= [\phi(L)]^{-1} u_t \\
u_t &\sim iid(0, \sigma^2)
\end{aligned}
$$

- With regressors in the equations as well, things get more complicated. You may find the details in last part of the chapter

Introduction to Vector Autoregressions

- Stochastic vector processes is a straight-forward generalization of univariate processes
- Most of the previous results are (principally) the same, only need to adapt them to a multivariate context
- Chapter 10 is quite technical, with more of a mathematical focus on vector time series
- Chapter 11 is also technical (it's still Hamilton), but focuses more on empirical issues and intepretations

Introduction to Vector Autoregressions

- Previously, we considered univariate stochastic processes such as an autoregression:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \tag{10}$$

where we assumed

$$E(\epsilon) = 0, \quad E(\epsilon_t \epsilon_\tau) = \begin{cases} \sigma^2, & \text{for } t = \tau, \\ 0, & \text{otherwise.} \end{cases}$$

- The generalization to a vector process is made by replacing the scalar y_t by the $n \times 1$ vector

$$\mathbf{y}_t = \begin{pmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{n,t} \end{pmatrix}$$

- The vector equivalent of (10) is thus:

$$\mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \mathbf{\Phi}_2 \mathbf{y}_{t-2} + \cdots + \mathbf{\Phi}_p \mathbf{y}_{t-p} + \varepsilon_t \tag{11}$$

  where $\mathbf{y}_t$, $\mathbf{c}$ and ε are all $n \times 1$ and each $\mathbf{\Phi}_j$ is $n \times n$.
- Alternatively, the process can be formulated using a lag polynomial:

$$(\mathbf{I}_n - \mathbf{\Phi}_1 L - \mathbf{\Phi}_2 L^2 - \cdots - \mathbf{\Phi}_p L^p)\mathbf{y}_t = \mathbf{c} + \varepsilon_t$$

  where

$$E(\varepsilon_t) = \mathbf{0},$$
$$E(\varepsilon_t \varepsilon_\tau') = \begin{cases} \mathbf{\Omega}, & \text{for } t = \tau, \\ \mathbf{0}, & \text{otherwise.} \end{cases}$$

- Stationarity applies in the same way to vector processes
- Covariance stationary if $E(\mathbf{y}_t)$ and $E[(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_{t-j} - \boldsymbol{\mu})']$ are independent of t
- Take expectations of (11):

$$E(\mathbf{y}_t) = \mathbf{c} + \mathbf{\Phi}_1 E(\mathbf{y}_{t-1}) + \mathbf{\Phi}_2 E(\mathbf{y}_{t-2}) + \cdots + \mathbf{\Phi}_p E(\mathbf{y}_{t-p}) + E(\varepsilon_t)$$

- If $\mathbf{y}_t$ is covariance stationary, $E(\mathbf{y}_t) = E(\mathbf{y}_{t-1}) = \cdots = \boldsymbol{\mu}$ for all t:

$$\boldsymbol{\mu} = \mathbf{c} + (\mathbf{\Phi}_1 + \mathbf{\Phi}_2 + \cdots + \mathbf{\Phi}_p)\boldsymbol{\mu}$$
$$\boldsymbol{\mu} = (\mathbf{I} - \mathbf{\Phi}_1 - \mathbf{\Phi}_2 - \cdots - \mathbf{\Phi}_p)^{-1}\mathbf{c}$$

- Thus, this is the obvious multivariate extension of the unconditional mean for an AR(p) process:

$$E(y_t) = (1 - \phi_1 - \phi_2 - \cdots - \phi_p)^{-1} c$$
$$E(\mathbf{y}_t) = (\mathbf{I} - \mathbf{\Phi}_1 - \mathbf{\Phi}_2 - \cdots - \mathbf{\Phi}_p)^{-1}\mathbf{c}$$

- It is sometimes useful to write the process in deviations from the mean. Note that from the previous slide:

$$\mathbf{c} = \boldsymbol{\mu} - \boldsymbol{\Phi}_1\boldsymbol{\mu} - \boldsymbol{\Phi}_2\boldsymbol{\mu} - \cdots - \boldsymbol{\Phi}_p\boldsymbol{\mu}$$

- So (11) is

$$\begin{aligned}\mathbf{y}_t &= \boldsymbol{\mu} - \boldsymbol{\Phi}_1\boldsymbol{\mu} - \boldsymbol{\Phi}_2\boldsymbol{\mu} - \cdots - \boldsymbol{\Phi}_p\boldsymbol{\mu} \\ &\quad + \boldsymbol{\Phi}_1\mathbf{y}_{t-1} + \boldsymbol{\Phi}_2\mathbf{y}_{t-2} + \cdots + \boldsymbol{\Phi}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t \\ (\mathbf{y}_t - \boldsymbol{\mu}) &= \boldsymbol{\Phi}_1(\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\Phi}_2(\mathbf{y}_{t-2} - \boldsymbol{\mu}) + \cdots + \boldsymbol{\Phi}_p(\mathbf{y}_{t-p} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t\end{aligned}$$

- Just as for the AR process, we can write this as a process of order one (companion form):

$$\vec{Y}_t = \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \phi^1_{11} y_{1,t-1} + \phi^1_{21} y_{2,t-1} \\ \phi^2_{11} y_{1,t-1} + \phi^1_{21} y_{2,t-1} \end{pmatrix} + \vec{\varepsilon}_t$$

$$= \phi \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix}$$

$$\begin{pmatrix} \mathbf{y}_t - \boldsymbol{\mu} \\ \mathbf{y}_{t-1} - \boldsymbol{\mu} \\ \vdots \\ \mathbf{y}_{t-p} - \boldsymbol{\mu} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \cdots & \boldsymbol{\Phi}_{p-1} & \boldsymbol{\Phi}_p \\ \mathbf{I}_n & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_n & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{y}_{t-1} - \boldsymbol{\mu} \\ \mathbf{y}_{t-2} - \boldsymbol{\mu} \\ \vdots \\ \mathbf{y}_{t-p-1} - \boldsymbol{\mu} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}$$

- Or in short:

$$\boldsymbol{\xi}_t = \mathbf{F}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t$$

Introduction to Vector Autoregressions

- Just as for the AR process, the effects of previous shocks must eventually die out

$$
\begin{aligned}
\boldsymbol{\xi}_t &= \mathbf{F}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t \\
&= \mathbf{F}^2\boldsymbol{\xi}_{t-2} + \mathbf{F}\mathbf{v}_{t-1} + \mathbf{v}_t \\
&= \mathbf{F}^3\boldsymbol{\xi}_{t-3} + \mathbf{F}^2\mathbf{v}_{t-2} + \mathbf{F}\mathbf{v}_{t-1} + \mathbf{v}_t \\
&\vdots \\
&= \mathbf{F}^s\boldsymbol{\xi}_{t-s} + \mathbf{F}^{s-1}\mathbf{v}_{t-s+1} + \cdots + \mathbf{F}\mathbf{v}_{t-1} + \mathbf{v}_t
\end{aligned}
\tag{12}
$$

- By Proposition 10.1, the eigenvalues of $\mathbf{F}$ satisfy

$$
|\mathbf{I}_n\lambda^p - \boldsymbol{\Phi}_1\lambda^{p-1} - \boldsymbol{\Phi}_2\lambda^{p-2} - \cdots - \boldsymbol{\Phi}_p| = 0
$$

- Hence, covariance stationary if and only if all solutions λ satisfy $|\lambda| < 1$, i.e. *inside* the unit circle.

- But, similarly, we have that

$$\begin{aligned}&|\mathbf{I}_n\lambda^p - \mathbf{\Phi}_1\lambda^{p-1} - \mathbf{\Phi}_2\lambda^{p-2} - \cdots - \mathbf{\Phi}_p| \\ &= \lambda^{np}|\mathbf{I}_n - \mathbf{\Phi}_1 z - \mathbf{\Phi}_2 z^2 - \cdots - \mathbf{\Phi}_p z^p|\end{aligned}$$

  where $z = \lambda^{-1}$.
- So, this is equivalent to

$$|\mathbf{I}_n - \mathbf{\Phi}_1 z - \mathbf{\Phi}_2 z^2 - \cdots - \mathbf{\Phi}_p z^p| = 0,$$

  so the process is stationary if all solutions z satisfy $|z| > 1$, i.e. *outside* of the unit circle.

- MA processes also exist in vector form, called VMA(q):

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t + \boldsymbol{\Theta}_1 \boldsymbol{\varepsilon}_{t-1} + \cdots + \boldsymbol{\Theta}_q \boldsymbol{\varepsilon}_{t-q}$$

- As in the univariate case, VMA(q) processes:
  - are always stationary
  - have $\boldsymbol{\Gamma}_j = \mathbf{0}$ if $|j| > q$
- Compare the autocovariances for q-th order processes:

$$\boldsymbol{\Gamma}_j = \begin{cases} \boldsymbol{\Theta}_j \boldsymbol{\Omega} + \boldsymbol{\Theta}_{j+1} \boldsymbol{\Omega} \boldsymbol{\Theta}_1' + \cdots + \boldsymbol{\Theta}_q \boldsymbol{\Omega} \boldsymbol{\Theta}_{q-j}', & j = 1, \ldots, q \\ \boldsymbol{\Omega} \boldsymbol{\Theta}_{-j}' + \boldsymbol{\Theta}_1 \boldsymbol{\Omega} \boldsymbol{\Theta}_{-j+1}' + \cdots + \boldsymbol{\Theta}_{q-j} \boldsymbol{\Omega} \boldsymbol{\Theta}_q', & j = -1, \ldots, -q \\ \mathbf{0}, & |j| > q \end{cases}$$

$$\gamma_j = \begin{cases} (\theta_j + \theta_{j+1}\theta_1 + \cdots + \theta_q \theta_{q-j})\sigma^2, & j = 1, \ldots, q \\ 0, & |j| > q \end{cases}$$

- For the VMA(∞) process, we change notation and write the model as:

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{\varepsilon}_{t-2} + \cdots$$

- If the process

$$(\mathbf{I}_n - \mathbf{\Phi}_1 L - \mathbf{\Phi}_2 L^2 - \cdots - \mathbf{\Phi}_p L^p)(\mathbf{y}_t - \boldsymbol{\mu}) = \boldsymbol{\varepsilon}_t$$

is stationary, it admits a moving average representation:

$$\begin{aligned} \mathbf{y}_t - \boldsymbol{\mu} &= \boldsymbol{\varepsilon}_t + \mathbf{\Psi}_1 \boldsymbol{\varepsilon}_{t-1} + \mathbf{\Psi}_2 \boldsymbol{\varepsilon}_{t-2} + \cdots \\ &= (\mathbf{I}_n + \mathbf{\Psi}_1 L + \mathbf{\Psi}_2 L^2 + \cdots)\boldsymbol{\varepsilon}_t \end{aligned}$$

- Hence,

$$(\mathbf{I}_n - \mathbf{\Phi}_1 L - \mathbf{\Phi}_2 L^2 - \cdots - \mathbf{\Phi}_p L^p)(\mathbf{I}_n + \mathbf{\Psi}_1 L + \mathbf{\Psi}_2 L^2 + \cdots) = \mathbf{I}_n$$

and restrictions on L^s coefficients yield

$$\begin{aligned} \mathbf{\Psi}_1 - \mathbf{\Phi}_1 &= \mathbf{0} \\ \mathbf{\Psi}_2 - \mathbf{\Phi}_1 \mathbf{\Psi}_1 - \mathbf{\Phi}_2 &= \mathbf{0} \\ \mathbf{\Psi}_3 - \mathbf{\Phi}_2 \mathbf{\Psi}_1 - \mathbf{\Phi}_1 \mathbf{\Psi}_2 - \mathbf{\Phi}_3 &= \mathbf{0} \end{aligned}$$

and so on.

- From the previous slide:

$$\mathbf{\Psi}_1 = \mathbf{\Phi}_1$$
$$\mathbf{\Psi}_2 = \mathbf{\Phi}_1\mathbf{\Psi}_1 + \mathbf{\Phi}_2$$
$$\mathbf{\Psi}_3 = \mathbf{\Phi}_2\mathbf{\Psi}_1 + \mathbf{\Phi}_1\mathbf{\Psi}_2 + \mathbf{\Phi}_3$$

- Two examples:

Example: VAR(1) ($p = 1$)

$$\begin{aligned}\mathbf{\Psi}_1 &= \mathbf{\Phi}_1 &&= \mathbf{\Phi}_1\\ \mathbf{\Psi}_2 &= \mathbf{\Phi}_1\mathbf{\Psi}_1 &&= \mathbf{\Phi}_1^2\\ \mathbf{\Psi}_3 &= \mathbf{\Phi}_1\mathbf{\Psi}_2 &&= \mathbf{\Phi}_1^3\end{aligned}$$

Example: VAR(2) ($p = 2$)

$$\begin{aligned}\mathbf{\Psi}_1 &= \mathbf{\Phi}_1 &&= \mathbf{\Phi}_1\\ \mathbf{\Psi}_2 &= \mathbf{\Phi}_1\mathbf{\Psi}_1 + \mathbf{\Phi}_2 &&= \mathbf{\Phi}_1^2 + \mathbf{\Phi}_2\\ \mathbf{\Psi}_3 &= \mathbf{\Phi}_1\mathbf{\Psi}_2 + \mathbf{\Phi}_2\mathbf{\Psi}_1 &&= \mathbf{\Phi}_1^3 + \mathbf{\Phi}_1\mathbf{\Phi}_2 + \mathbf{\Phi}_2\mathbf{\Phi}_1\end{aligned}$$

Introduction to Vector Autoregressions

- For the VMA(q) we saw that the autocovariances were different for negative and positive j. Why is this?
- The autocovariance is

$$\mathbf{\Gamma}_j = E\left[(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_{t-j} - \boldsymbol{\mu})'\right]$$

- In the univariate case $\gamma_j = \gamma_{-j}$, but $\mathbf{\Gamma}_j \neq \mathbf{\Gamma}_{-j}$.
- Instead, $\mathbf{\Gamma}_j = \mathbf{\Gamma}'_{-j}$
- The reason is that generally:

$$Cov(y_{k,t+j}, y_{l,t}) \neq Cov(y_{k,t}, y_{l,t+j})$$

- We have before noted that e.g. in an AR(1)

$$Cov(y_{t+1}, \epsilon_t) = \phi \neq Cov(y_t, \epsilon_{t+1}) = 0$$

Forecasting

- For forecasting, consider again (12) but for $t+s$:

$$\boldsymbol{\xi}_{t+s} = \mathbf{F}^s \boldsymbol{\xi}_t + \mathbf{F}^{s-1}\mathbf{v}_{t+1} + \cdots + \mathbf{F}\mathbf{v}_{t+s-1} + \mathbf{v}_{t+s} \tag{13}$$

- If we view $t+1, t+2, \ldots, t+s$ as the future, then this expresses the future $\boldsymbol{\xi}_{t+s}$ as a function of past (known) $(\mathbf{y}-\boldsymbol{\mu})$ and future errors ε
- If we want to forecast $\mathbf{y}_{t+s}$, then the first $n \times 1$ elements of $\boldsymbol{\xi}_{t+s}$ are what we want
- Recall equation [4.2.20] for Y_{t+s}:

$$\begin{aligned} Y_{t+s} - \mu = {} & f_{11}^{(s)}(Y_t - \mu) + f_{12}^{(s)}(Y_{t-1} - \mu) + \cdots + f_{1p}^{(s)}(Y_{t-p+1} - \mu) \\ & + \epsilon_{t+s} + \psi_1 \epsilon_{t+s-1} + \psi_2 \epsilon_{t+s-2} + \cdots + \psi_{s-1}\epsilon_{t+1} \end{aligned}$$

- Difference between the **F** matrices:

$$\mathbf{F}_{AR}^{s} = \underbrace{\begin{pmatrix} f_{11}^{(s)} & f_{12}^{(s)} & \cdots & f_{1p}^{(s)} \\ f_{21}^{(s)} & f_{22}^{(s)} & \cdots & f_{2p}^{(s)} \\ \vdots & \vdots & \ddots & \vdots \\ f_{p1}^{(s)} & f_{p2}^{(s)} & \cdots & f_{pp}^{(s)} \end{pmatrix}}_{p\times p}, \quad \mathbf{F}_{VAR}^{s} = \underbrace{\begin{pmatrix} \mathbf{F}_{11}^{(s)} & \mathbf{F}_{12}^{(s)} & \cdots & \mathbf{F}_{1p}^{(s)} \\ \mathbf{F}_{21}^{(s)} & \mathbf{F}_{22}^{(s)} & \cdots & \mathbf{F}_{2p}^{(s)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{F}_{p1}^{(s)} & \mathbf{F}_{p2}^{(s)} & \cdots & \mathbf{F}_{pp}^{(s)} \end{pmatrix}}_{np\times np}$$

- Hence, straight-forward to get an expression of only the first n rows of $\boldsymbol{\xi}_{t+s}$:

$$\begin{aligned} \mathbf{y}_{t+s} - \boldsymbol{\mu} = {} & \mathbf{F}_{11}^{(s)}(\mathbf{y}_t - \boldsymbol{\mu}) + \mathbf{F}_{12}^{(s)}(\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \cdots + \mathbf{F}_{1p}^{(s)}(\mathbf{y}_{t-p+1} - \boldsymbol{\mu}) \\ & + \boldsymbol{\varepsilon}_{t+s} + \boldsymbol{\Psi}_1 \boldsymbol{\varepsilon}_{t+s-1} + \boldsymbol{\Psi}_2 \boldsymbol{\varepsilon}_{t+s-2} + \cdots + \boldsymbol{\Psi}_{s-1} \boldsymbol{\varepsilon}_{t+1} \end{aligned} \tag{14}$$

- The forecast of $\mathbf{y}_{t+s}$ is therefore:

$$\hat{\mathbf{y}}_{t+s|t} = \boldsymbol{\mu} + \mathbf{F}_{11}^{(s)}(\mathbf{y}_t - \boldsymbol{\mu}) + \mathbf{F}_{12}^{(s)}(\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \cdots + \mathbf{F}_{1p}^{(s)}(\mathbf{y}_{t-p+1} - \boldsymbol{\mu}) \tag{15}$$

$H_0: \tilde{\mu} = 0$
$H_1: \tilde{\mu} \neq 0$

$$T \cdot E(\bar{Y}_T - \mu)(\bar{Y}_T - \mu)' \to \sum \Gamma_i$$

$$T \cdot E(\bar{Y}_T - \mu)(\bar{Y}_T - \mu)' \approx \sum \Gamma_i$$

$$\Rightarrow Var(\bar{Y}) = E(\ )(\ ) = \frac{1}{T} \sum \Gamma_i$$

$$\Rightarrow \bar{Y} - \mu \sim N(0, \tfrac{1}{T} \sum \Gamma_i)$$

$$\Rightarrow \frac{\bar{Y} - \tilde{\mu}}{\sqrt{\frac{1}{T} \sum \Gamma_i}} \mathrel{\hat{\sim}} N(0,1)$$

$$\Rightarrow \left(\frac{\bar{Y} - \tilde{\mu}}{\sqrt{\frac{1}{T} \sum \Gamma_i}} \right)^2 \mathrel{\hat{\sim}} \chi^2_1$$

The Sample Mean of a Vector Process

- Proposition 10.5 says that for a covariance stationary process $\mathbf{y}_t$, with expectation $\boldsymbol{\mu}$ and $E[(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_{t-j} - \boldsymbol{\mu})'] = \boldsymbol{\Gamma}_j$ absolutely summable, it follows that $\bar{\mathbf{y}}_T \xrightarrow{p} \boldsymbol{\mu}$ and

$$\mathbf{S} = \lim_{T\to\infty} \left\{ T \cdot E[(\bar{\mathbf{y}}_T - \boldsymbol{\mu})(\bar{\mathbf{y}}_T - \boldsymbol{\mu})'] \right\} = \sum_{v=-\infty}^{\infty} \boldsymbol{\Gamma}_v$$

- If we assume a VMA(q), $\boldsymbol{\Gamma}_j = \mathbf{0}$ for all $|j| > q$. We can estimate $\boldsymbol{\Gamma}_v$, where $v = 0, 1, \ldots, q$, by

$$\hat{\boldsymbol{\Gamma}}_v = T^{-1} \sum_{t=v+1}^{T} (\mathbf{y}_t - \bar{\mathbf{y}}_T)(\mathbf{y}_{t-v} - \bar{\mathbf{y}}_T)'$$

which is consistent as long as $\mathbf{y}_t$ is ergodic for second moments.

- $\mathbf{S}$ can then be consistently estimated by

$$\hat{\mathbf{S}} = \hat{\boldsymbol{\Gamma}}_0 + \sum_{v=1}^{q} (\hat{\boldsymbol{\Gamma}}_v + \hat{\boldsymbol{\Gamma}}_v')$$

The Sample Mean of a Vector Process

- $\hat{\mathbf{S}}$ is not guaranteed to be positive semi-definite
- An adjusted estimator was proposed by Newey and West, and it is known as the Newey and West estimator:

$$\tilde{\mathbf{S}} = \hat{\mathbf{\Gamma}}_0 + \sum_{v=1}^{q} \left(1 - \frac{v}{q+1}\right) (\hat{\mathbf{\Gamma}}_v + \hat{\mathbf{\Gamma}}_v')$$

- Note: with the VMA(q), we say that the autocovariances are zero for $v > q$, but even if $E[(\mathbf{y}_t - \boldsymbol{\mu})(\mathbf{y}_s - \boldsymbol{\mu})']$ is non-zero for all t and s (e.g. VAR), $\tilde{\mathbf{S}}$ will still be consistent if q, the threshold for non-zero autocovariances, is allowed to increase alongside T
- In particular, if $q \to \infty$ and $T \to \infty$ such that

$$\frac{q}{T^{1/4}} \to 0,$$

then $\tilde{\mathbf{S}} \xrightarrow{p} \mathbf{S}$.

Final remarks

Some things to note:

- For an AR, we have only one dimension for its size; p, the number of lags
- For a VAR, we write VAR(p) and the cross-sectional dimension n is usually omitted
- Three types of possible error-term covariance:
  - Between equations, same time point
  - Within equations, over time
  - Between equations, over time

$$E(\varepsilon_t \varepsilon_\tau) = \begin{cases} \mathbf{\Omega}, & t = \tau \\ \mathbf{0}, & t \neq \tau \end{cases}$$

- With intercepts, n variables and p lags, the number of parameters to be estimated is $n(np + 1)$. Thus, an 8-variable VAR with four lags needs to estimate 264 parameters.

To be continued! Thank you!