[2025-1] 계량경제학 - Final

Jiho KWAK

June 8, 2025

1 Classical Assumption in DGP

A0. X_t is not random (not stochastic)

a) $\text{Cov}(X_1, X_2) \neq 0$: Multicollinearity $\rightarrow$ high R^2, low t

b) $\text{Cov}(X, \epsilon) \neq 0$: Endogeniety

- IV Estimation

A1. $\mathbb{E}[\epsilon_t] = 0$

A2. $\text{Var}(\epsilon_t) = \sigma^2$ (homoskedasticity)

a) $\text{Var}(\epsilon_t) = \sigma_t^2 \rightarrow$ unbiased but not efficient

- Apply Gauss-Markov Theorem with modified error (GLS)
- Re-estimate by employing HAC estimator

A3. $\text{Cov}(\epsilon_t, \epsilon_s) = 0$ for $t \neq s$ (no serial correlation)

a) $\text{Cov}(\epsilon_t, \epsilon_s) \neq 0$: unbiased but not efficient

- Apply Gauss-Markov Theorem with modified error (GLS)
- Re-estimate by employing HAC estimator

A4. $\epsilon_t \sim N(0, \sigma^2)$

If A1, A2, A3 ar satisfied, then the OLS estimator has the smallest variance among all linear and unbiased estimators for β^* (Best Linear Unbiased Estimator).

2 Heteroskedasticity

2.1 Effect of Dropping Homoskedasticity

- 기존 가정: $\text{Var}(\epsilon_t) = \sigma^2$
- $\text{Var}(\epsilon_t) = \sigma_t^2$

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{T-1} \\ \varepsilon_T \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_T^2 \end{pmatrix} \right) = \mathcal{N}(0, \Omega).$$

- Unbiasedness: $\mathbb{E}(\hat{\beta}) = \beta^*$

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \epsilon) \\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon \\ \mathbb{E}[\hat{\beta}] &= \mathbb{E}[\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon] \\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}[\epsilon] \quad (\because \text{A0}) \\ &= \beta \quad (\because \text{A1})\end{aligned}$$

- Variance: $\text{Var}(\hat{\beta})$

$$\begin{aligned}Var(\hat{\beta}) &= \text{Var}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \epsilon)) \\ &= Var(\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon) \\ &= Var((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\text{Var}(\epsilon)\mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}\mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1} \\ &\neq (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

- Any inference based on $\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$ is invalid. (not efficient)

2.2 White Test

$$y_t = \beta_0^* + \beta_1^* X_{t1} + \beta_2^* X_{t2} + \epsilon_t$$

- $\sigma_t^2 = h(\alpha_0^* + \alpha_1^* X_{t1} + \alpha_2^* X_{t2} + \alpha_3^* X_{t1}^2 + \alpha_4^* X_{t2}^2 + \alpha_5^* X_{t1} X_{t2})$
- H_0: homoskedasticity ($\alpha_1^* = \alpha_2^* = \cdots = \alpha_5^* = 0$)
  H_1: $\sim H_0$
- Steps:
  1. Base regression $\rightarrow$ get e_t
  2. Regress e_t^2 on 1, X_t, quadratic terms, interaction terms
  3. Get R^2 at 2nd step regression
  4. Test statistic: $T \cdot R^2 \sim \chi^2$(# of restrictions in H_0 i.e. 등호의 개수)

3 Serial Correlation

3.1 Effect of Dropping No Serial Correlation

- Serial correlation in ϵ_t

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_{T-1} \\ \varepsilon_T \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1,T-1} & \sigma_{1,T} \\ \sigma_{21} & \sigma^2 & \sigma_{23} & \cdots & \sigma_{2,T-1} & \sigma_{2,T} \\ \sigma_{31} & \sigma_{32} & \sigma^2 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \sigma_{T-2,T-1} & \vdots \\ \sigma_{T-1,1} & \sigma_{T-1,2} & \cdots & \ddots & \sigma^2 & \sigma_{T-1,T} \\ \sigma_{T,1} & \sigma_{T,2} & \cdots & \cdots & \sigma_{T,T-1} & \sigma^2 \end{pmatrix} \right) = \mathcal{N}(0, \Omega).$$

- Unbiasedness: $\mathbb{E}(\hat{\beta}) = \beta^*$
- Variance: $\text{Var}(\hat{\beta}) \neq \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$
- Any inference based on $\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$ is invalid. (not efficient)

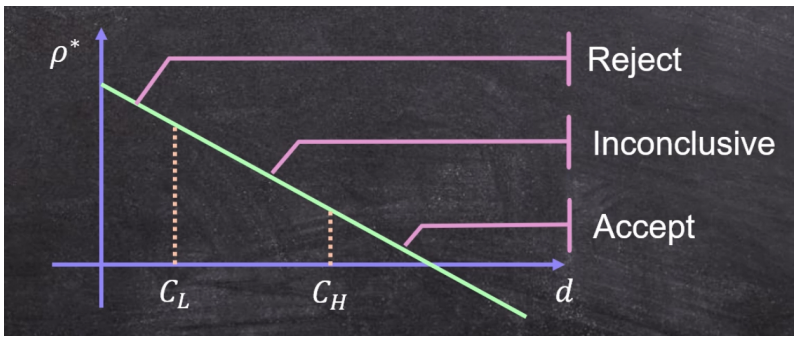
3.2 Durbin-Watson Test

- $\epsilon_t = \rho^* \epsilon_{t-1} + \eta_t$: AR(1) Process
- H_0: $\rho^* = 0$

  H_1: $\rho^* > 0$
- Steps:
  1. Base regression $\rightarrow$ get e_t
  2. Test Statistic: $d = \dfrac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \geq 0$

     d is approximately equal to $2(1-\hat{\rho})$, therefore $d = 2$ indicates no correlation.
  3. If d is close to 0, reject the null. (serially correlated!)

     If d is far away from 0, accept the null.



3.3 GLS

- White variance: $\sigma^2\mathbf{I}$
- If $\text{Var}(\epsilon) = \mathbf{\Omega} \neq \sigma^2\mathbf{I}$, we have to employ GLS instead of OLS.
- New DGP: $\mathbf{\Omega}^{-1/2}\mathbf{y} = \mathbf{\Omega}^{-1/2}\mathbf{X}\beta + \mathbf{\Omega}^{-1/2}\epsilon$

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\beta + \tilde{\epsilon}, \quad \text{where } \text{Var}(\tilde{\epsilon}) = \mathbf{I} = \mathbf{\Omega}^{-1/2}\mathbf{\Omega}\mathbf{\Omega}^{-1/2}$$

  - No heteroskedasticity
  - No serial correlation

- Very clean error term → Gauss-Markov Theorem can be applied.
- GLS estimator:

$$\begin{aligned}\tilde{\beta} &= (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{y}} \\ &= \left(\mathbf{X}^\top \boldsymbol{\Omega}^{-1/2\top} \boldsymbol{\Omega}^{-1/2} \mathbf{X}\right)^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1/2\top} \boldsymbol{\Omega}^{-1/2} \mathbf{y} \\ &= (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{y}\end{aligned}$$

- Variance of GLS estimator:

$$\begin{aligned}\mathrm{Var}(\tilde{\beta}) &= \mathrm{Var}\big((\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{y}}\big) \\ &= \mathrm{Var}\Big(\left(\mathbf{X}^\top \boldsymbol{\Omega}^{-1/2\top} \boldsymbol{\Omega}^{-1/2} \mathbf{X}\right)^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1/2\top} \boldsymbol{\Omega}^{-1/2} \varepsilon\Big) \\ &= (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathrm{Var}(\varepsilon) \boldsymbol{\Omega}^{-1} \mathbf{X} (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \\ &= (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} \boldsymbol{\Omega}^{-1} \mathbf{X} (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \\ &= (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1}\end{aligned}$$

  cf. Variance of OLS estimator: $\mathrm{Var}(\hat{\beta}) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega} \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1}$

- Goal: $\mathrm{Var}(\hat{\beta}) - \mathrm{Var}(\tilde{\beta})$ is positive definite.

  Proof. Lemma: If $A - B$ is positive definite, $\dfrac{1}{B} - \dfrac{1}{A}$ is also positive definite.

$$\begin{aligned}\mathrm{Var}(\tilde{\beta})^{-1} \;-\; \mathrm{Var}(\hat{\beta})^{-1} &= (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X}) \;-\; (\mathbf{X}^\top \mathbf{X}) (\mathbf{X}^\top \boldsymbol{\Omega} \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{X}) \\ &= (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} \boldsymbol{\Omega}^{-1} \mathbf{X}) \;-\; (\mathbf{X}^\top \mathbf{X}) (\mathbf{X}^\top \boldsymbol{\Omega} \mathbf{X})^{-1} (\mathbf{X}^\top \boldsymbol{\Omega} \mathbf{X}) (\mathbf{X}^\top \boldsymbol{\Omega} \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{X})\end{aligned}$$

  Letting $\boldsymbol{V} = \mathbf{X}^\top \boldsymbol{\Omega}^{-1} - (\mathbf{X}^\top \mathbf{X})(\mathbf{X}^\top \boldsymbol{\Omega} \mathbf{X})^{-1} \mathbf{X}^\top$,

$$\mathrm{Var}(\tilde{\beta})^{-1} \;-\; \mathrm{Var}(\hat{\beta})^{-1} = \boldsymbol{V} \boldsymbol{\Omega} \boldsymbol{V}^\top$$

  ■

3.4 Cochrane-Orcutt Method

- Original DGP: $\mathbf{y} = \mathbf{X}\beta + \epsilon$
- Backward one period: $\rho L\mathbf{y} = \rho L\mathbf{X}\beta + \rho L\epsilon$ (L: lag operator)
- New DGP: original - backward one period

$$\begin{aligned}\mathbf{y} - \rho L\mathbf{y} &= (\mathbf{X} - \rho L\mathbf{X})\beta + \epsilon - \rho L\epsilon \\ \tilde{\mathbf{y}} &= \tilde{\mathbf{X}}\beta + \eta_t\end{aligned}$$

  - No heteroskedasticity
  - No serial correlation

- Very clean error term → Gauss-Markov Theorem can be applied.

3.5 HAC (Heteroskedasticity and Autocorrelation Consistent) Estimator

- Newey-West Estimator (one of them)

$$\begin{aligned}\hat{\beta} - \beta^* &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \epsilon \\ \sqrt{T}(\hat{\beta} - \beta^*) &= \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{X}_t \mathbf{X}_t^\top \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{X}_t \epsilon_t \\ &\xrightarrow{d} \boldsymbol{Q}^{-1} N(0, \boldsymbol{V}) = N(0, \boldsymbol{Q}^{-1} \boldsymbol{V} \boldsymbol{Q}^{-1})\end{aligned}$$

- $\boldsymbol{Q} = \mathbb{E}(\mathbf{X}_t \mathbf{X}_t^\top)$ is unknown: $\hat{\boldsymbol{Q}} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{X}_t \mathbf{X}_t^\top$
- $\boldsymbol{V} = \mathbb{E}(\epsilon_t^2 \mathbf{X}_t \mathbf{X}_t^\top)$ unknown:

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{X}_t \epsilon_t\right) &= \frac{1}{T} \mathrm{Var}\left(\sum_{t=1}^{T} \mathbf{X}_t \epsilon_t\right) \\ &= \frac{1}{T} \sum_{t=1}^{T} \mathrm{Var}(\mathbf{X}_t \epsilon_t) + \frac{1}{T} \sum_{t}^{T} \sum_{s, s \neq t}^{T} \mathrm{Cov}(\mathbf{X}_t \epsilon_t, \mathbf{X}_s \epsilon_s)\end{aligned}$$

$$\hat{\boldsymbol{V}} = \frac{1}{T} \sum_{t=1}^{T} e_t^2 \mathbf{X}_t \mathbf{X}_t^\top + \frac{1}{T} \sum_{j=1}^{H-1} w_j \sum_{t=j+1}^{T} (e_t^2 \mathbf{X}_{t-j} \mathbf{X}_t^\top + \epsilon_t^2 \mathbf{X}_t \mathbf{X}_{t-j}^2)$$

where $w_j = 1 - \frac{j}{H}$ (Barlett's Kernel)

and H is truncation lag (weight function) → contribution decreases as time goes by (long-term, short-term memory인지에 따라 설정)

4 Multicollinearity

4.1 Correlogram, VIF

- Rule of thumb (correlogram): If the absolute of correlation coefficient $|c| > 0.7$, severe multicollinearity may be present.
- VIF
  1. Run the OLS regression for each X variable.

$$X_{1i} = \alpha_1 + \alpha_2 X_{2i} + \cdots + \alpha_k X_{ki} + \nu_i$$

  2. Calculate the VIF for $\hat{\beta}_i$

$$VIF(\hat{\beta}_i) = \frac{1}{1 - R_i^2}$$

  3. Rule of thumb: If $VIF(\hat{\beta}_i) > 5$, then severe multicollinearity may be present.

5 Endogeneity

5.1 Instrumental Variable Estimation (IV)

- Setup: $y_t = \mathbf{X}_t^\top \beta + \epsilon_t$ where $\mathbb{E}[\mathbf{X}_t \epsilon_t] \neq 0$
- Suppose

  1) $\exists \mathbf{W}_t (l \times 1)$ with one in the 1st element.
  2) $\mathbb{E}[\mathbf{W}_t \epsilon_t] = 0$
  3) $\mathbb{E}[\mathbf{W}_t \mathbf{X}_t] \neq 0$ and invertible.

- Case 1. $l = k$
  - Since $\mathbb{E}[\mathbf{W}_t \epsilon_t] = 0$,

$$\frac{1}{T}\sum_{t=1}^{T} \mathbf{W}_t(\mathbf{y}_t - \mathbf{X}_t^\top \beta) = 0$$
$$\Leftrightarrow \frac{1}{T}\sum_{t=1}^{T} \mathbf{W}_t \mathbf{y}_t = \frac{1}{T}\sum_{t=1}^{T} \mathbf{W}_t \mathbf{X}_t^\top \beta$$

  - Therefore, we can get this estimator

$$\hat{\beta}_{IV} = \left(\sum_{t=1}^{T} \mathbf{W}_t \mathbf{X}_t\right)^{-1} \sum_{t=1}^{T} \mathbf{W}_t \mathbf{y}_t$$

  - IV Estimator with $\mathbf{W}$: $\hat{\beta} = (\mathbf{W}^\top \mathbf{X})^{-1} \mathbf{W}^\top \mathbf{y} \rightarrow$ unbiased

$$\begin{aligned}\mathbb{E}[\hat{\beta}] &= \mathbb{E}[(\mathbf{W}^\top \mathbf{X})^{-1} \mathbf{W}^\top \mathbf{y}] \\ &= \mathbb{E}[(\mathbf{W}^\top \mathbf{X})^{-1} \mathbf{W}^\top (\mathbf{X}\beta + \epsilon)] \\ &= \mathbb{E}[(\mathbf{W}^\top \mathbf{X})^{-1} \mathbf{W}^\top \mathbf{X}\beta + (\mathbf{W}^\top \mathbf{X})^{-1} \mathbf{W}^\top \epsilon] = \beta\end{aligned}$$

    Note: $\hat{\beta} = (\mathbf{W}^\top \mathbf{W})^{-1} \mathbf{W}^\top \mathbf{y}$ is unbiased.

$$\begin{aligned}\mathbb{E}[\hat{\beta}] &= \mathbb{E}[(\mathbf{W}^\top \mathbf{W})^{-1} \mathbf{W}^\top \mathbf{y}] \\ &= \mathbb{E}[(\mathbf{W}^\top \mathbf{W})^{-1} \mathbf{W}^\top (\mathbf{X}\beta + \epsilon)] \\ &= \mathbb{E}[(\mathbf{W}^\top \mathbf{W})^{-1} \mathbf{W}^\top \mathbf{X}\beta + (\mathbf{W}^\top \mathbf{X})^{-1} \mathbf{W}^\top \epsilon] \neq \beta\end{aligned}$$

  - 2SLS: regress $\mathbf{y}$ on $\hat{\mathbf{X}} = \mathbf{P}_\mathbf{W}\mathbf{X}$ ($\mathbf{P}_\mathbf{W} = \mathbf{W}(\mathbf{W}^\top \mathbf{W})^{-1}\mathbf{W}^\top$)
    1. Setup: $y_t = \mathbf{X}_t^\top \beta + \epsilon_t$ where $\mathbb{E}[\mathbf{X}_t \epsilon_t] \neq 0$
    2. Regression: $\mathbf{X}_t = \mathbf{W}_t^\top \alpha + \eta_t \rightarrow \hat{\mathbf{X}}_t = \mathbf{W}_t^\top \hat{\alpha}$
    3. Regression: $\mathbf{y}_t = \hat{\mathbf{X}}_t^\top \beta + \epsilon_t \rightarrow \hat{\mathbf{Y}}_t = \hat{\mathbf{X}}_t^\top \hat{\beta}$

5.2 System of Structural Equation

- Check the endogeneity: $\mathrm{Cov}(\epsilon_t, X_{t1}) = \mathbb{E}(\epsilon_t X_{t1}) = 0$?
- Supply function: $y_t = \alpha_0 + \alpha_1 X_{t1} + \eta_t$

$$X_{t1} = -\frac{\alpha_0}{\alpha_1} + \frac{1}{\alpha_1} y_t - \eta_t = \gamma_0 + \gamma_1 y_t + \xi_t$$

Therefore, $\mathbb{E}(\epsilon_t X_{t1}) = \mathbb{E}(\epsilon_t(\gamma_0 + \gamma_1 y_t + \xi_t))$
$= \gamma \mathbb{E}(\epsilon_t y_t) \neq 0$

6 Asymptotic Normality

6.1 Asymptotic Assumptions

A0. X_t is not random - violated

A1. $\mathbb{E}[\epsilon_t \mid \mathbf{X}] = 0$

A2. $\mathbb{E}(\epsilon_t^2 \mid \mathbf{X}) = \sigma^2$

A3. $\mathbb{E}(\epsilon_t \epsilon_s \mid \mathbf{X}) = 0$ where $t \neq s$

A4. $\epsilon \mid \mathbf{X} \sim N(0, \sigma^2)$

- Conditional expectation is brought in.
- Unbiasedness: $\mathbb{E}(\hat{\beta}) = \beta^*$

$$\begin{aligned}
\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \epsilon) \\
&= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon \\
\mathbb{E}[\hat{\beta} \mid \mathbf{X}] &= \mathbb{E}[\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon \mid \mathbf{X}] \\
&= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}[\epsilon \mid \mathbf{X}] \\
&= \beta \quad (\because \text{A1}) \\
\mathbb{E}(\hat{\beta}) &= \mathbb{E}(\mathbb{E}(\hat{\beta} \mid \mathbf{X})) = \mathbb{E}(\beta) = \beta
\end{aligned}$$

- Asymptotic assumptions are valid when T is large enough.

6.2 Asymptotic Approach

- As $T \to \infty$, $\hat{\beta} \xrightarrow{p} \beta^*$ (consistent)

$$\begin{aligned}
\hat{\beta} &= \beta^* + \underbrace{\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t\mathbf{X}_t^\top\right)^{-1}}_{\xrightarrow{p}\mathbb{E}(\mathbf{X}_t\mathbf{X}_t^\top)=\boldsymbol{Q}} \underbrace{\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t\epsilon_t}_{\xrightarrow{p}\mathbb{E}(\mathbf{X}_t\epsilon_t)=0} \\
&\xrightarrow{p} \beta^*
\end{aligned}$$

As $T \to \infty$,

$$\begin{aligned}
\sqrt{T}(\hat{\beta} - \beta^*) &= \underbrace{\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t\mathbf{X}_t^\top\right)^{-1}}_{\xrightarrow{p}\mathbb{E}(\mathbf{X}_t\mathbf{X}_t^\top)=\boldsymbol{Q} \text{ by LLN}} \underbrace{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{X}_t\epsilon_t}_{\xrightarrow{d}N(0,\sigma^2\boldsymbol{Q}) \text{ by CLT}} \\
&\xrightarrow{d} \boldsymbol{Q}^{-1}N(0, \sigma^2\boldsymbol{Q}) = N(0, \sigma^2\boldsymbol{Q}^{-1})
\end{aligned}$$

- (IV Estimation) As $T \to \infty$, $\hat{\beta}_{IV} \xrightarrow{p} \beta^*$ (consistent)

$$\hat{\beta}_{IV} = \beta^* + \underbrace{\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{W}_t\mathbf{X}_t^\top\right)^{-1}}_{\xrightarrow{p}\mathbb{E}(\mathbf{W}_t\mathbf{X}_t^\top)=\boldsymbol{Q}_{\mathbf{WX}}} \underbrace{\frac{1}{T}\sum_{t=1}^{T}\mathbf{W}_t\epsilon_t}_{\xrightarrow{p}\mathbb{E}(\mathbf{W}_t\epsilon_t)=0 \text{ by assumption}}$$
$$\xrightarrow{p} \beta^*$$

  As $T \to \infty$,

$$\sqrt{T}(\hat{\beta} - \beta^*) = \underbrace{\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{W}_t\mathbf{X}_t^\top\right)^{-1}}_{\xrightarrow{p}\mathbb{E}(\mathbf{W}_t\mathbf{X}_t^\top)=\boldsymbol{Q}_{\mathbf{WX}} \text{ by LLN}} \underbrace{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{W}_t\epsilon_t}_{\xrightarrow{d}N(0,\sigma^2\boldsymbol{Q}_{\mathbf{WW}}) \text{ by CLT}}$$
$$\xrightarrow{d} \boldsymbol{Q}_{\mathbf{WX}}^{-1}N(0,\sigma^2\boldsymbol{Q}_{\mathbf{WW}}) = N(0,\sigma^2\boldsymbol{Q}_{\mathbf{WX}}^{-1}\boldsymbol{Q}_{\mathbf{WW}}\boldsymbol{Q}_{\mathbf{XW}}^{-1})$$

- (IV estimation when $l > k$) As $T \to \infty$, $\hat{\beta}_{IV} = (\mathbf{X}^\top\mathbf{P}_\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{P}_\mathbf{W}\mathbf{y} \xrightarrow{p} \beta^*$ (consistent)

$$\hat{\beta}_{IV} = \beta^* + \Bigg[\underbrace{\left(\frac{1}{T}\mathbf{X}^\top\mathbf{W}\right)}_{\xrightarrow{p}\mathbf{Q}_{\mathbf{XW}}}\underbrace{\left(\frac{1}{T}\mathbf{W}^\top\mathbf{W}\right)^{-1}}_{\xrightarrow{p}\mathbf{Q}_{\mathbf{WW}}}\underbrace{\left(\frac{1}{T}\mathbf{W}^\top\mathbf{X}\right)}_{\xrightarrow{p}\mathbf{Q}_{\mathbf{WX}}}\Bigg]^{-1}\left(\frac{1}{T}\mathbf{X}^\top\mathbf{W}\right)\left(\frac{1}{T}\mathbf{W}^\top\mathbf{W}\right)^{-1} \underbrace{\frac{1}{T}\sum_{t=1}^{T}\mathbf{W}_t\epsilon_t}_{\xrightarrow{p}\mathbb{E}(\mathbf{W}_t\epsilon_t)=0 \text{ by assumption}}$$
$$\xrightarrow{p} \beta^*$$

  As $T \to \infty$,

$$\sqrt{T}(\hat{\beta}_{IV} - \beta^*) = \Bigg[\underbrace{\left(\frac{1}{T}\mathbf{X}^\top\mathbf{W}\right)}_{\xrightarrow{p}\mathbf{Q}_{\mathbf{XW}}}\underbrace{\left(\frac{1}{T}\mathbf{W}^\top\mathbf{W}\right)^{-1}}_{\xrightarrow{p}\mathbf{Q}_{\mathbf{WW}}}\underbrace{\left(\frac{1}{T}\mathbf{W}^\top\mathbf{X}\right)}_{\xrightarrow{p}\mathbf{Q}_{\mathbf{WX}}}\Bigg]^{-1}\left(\frac{1}{T}\mathbf{X}^\top\mathbf{W}\right)\left(\frac{1}{T}\mathbf{W}^\top\mathbf{W}\right)^{-1} \underbrace{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{W}_t\epsilon_t}_{\xrightarrow{d}N(0,\sigma^2\mathbf{Q}_{\mathbf{WW}})}$$
$$\xrightarrow{d} N(0,\sigma^2\boldsymbol{Q}_{\mathbf{WX}}^{-1}\boldsymbol{Q}_{\mathbf{WW}}\boldsymbol{Q}_{\mathbf{XW}}^{-1})$$

7 Practical

7.1 Model Selection

- Adjusted R^2

$$R^2_{adj} = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

  Adjusted R^2 increases only if the significantly explanatory variable is added.

- AIC: the smaller the better!

$$AIC = 2k - 2\ln(\hat{L})$$

7.2 Multicollinearity

- Effect of near multicollinearity: $\text{Var}(\hat{\beta}_1) \uparrow$
- 그러나 모든 t값이 유의미하지 않지는 않음: inflate 된 variance 하에서도 유의미하다면 원래는 더 유의미함

7.3 Chow Test

- Used to test for the presence of a structural break at a period which can be assumed to be **known**.
- Restricted OLS (ROLS)

$$\min_{\beta} = (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta) \;\; \text{s.t. } R\beta - r = 0$$

Thus,

$$\begin{aligned} L &= (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta) + (\mathbf{R}\beta - r)^\top \lambda \\ \frac{\partial L}{\partial \beta} &= -2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X}\beta + \mathbf{R}^\top \lambda = 0 \\ \frac{\partial L}{\partial \lambda} &= \mathbf{R}\beta - r = 0 \end{aligned}$$

The solution is

$$\begin{aligned} \tilde{\beta} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \; - \; (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \left[\mathbf{R} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top\right]^{-1} \left[\mathbf{R} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} - r\right] \\ &= \hat{\beta} - \underbrace{(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \left[\mathbf{R} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top\right]^{-1} \left[\mathbf{R}\hat{\beta} - r\right]}_{\text{The correction term given new information}} \end{aligned}$$

- Expected value of $\tilde{\beta}$

$$\begin{aligned} \mathbb{E}(\tilde{\beta}) &= \mathbb{E}(\hat{\beta}) - \mathbb{E}((\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \left[\mathbf{R} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top\right]^{-1} \left[\mathbf{R}\hat{\beta} - r\right]) \\ &= \beta \end{aligned}$$

Variance of $\tilde{\beta}$

$$\text{Var}(\tilde{\beta}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} - \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \left[\mathbf{R} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top\right]^{-1} \mathbf{R} (\mathbf{X}^\top \mathbf{X})^{-1}$$

We can show that $\sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \left[\mathbf{R} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top\right]^{-1} \mathbf{R} (\mathbf{X}^\top \mathbf{X})^{-1}$ is p.s.d. Thus $\tilde{\beta}$ is unbiased and more efficient than $\hat{\beta}$ if the restriction is true.

- Test of the restriction
  - H_0: $\mathbf{R}\beta^* = r$
  - H_1: $\mathbf{R}\beta^* \neq r$
- Example:

$$y_t = \beta_0 + \beta_1 D_t + \beta_2 Z_t + \beta_3 (Z_t D_t) + \beta_4 X_t + \epsilon_t$$

where D_t is a dummy variable (Male = 1).

  - Male version: $y_t = \beta_0 + \beta_1 + \beta_2 Z_t + \beta_3 Z_t + \beta_4 X_t + \epsilon_t$
  - Female version: $y_t = \beta_0 + \beta_2 Z_t + \beta_4 X_t + \epsilon_t$
  - H_0: $\beta_1 = \beta_3 = 0$ v.s. H_1: $\sim H_0$
  - If $\beta_1 = 0$, there is no gender wage gap in average.
  - If $\beta_3 = 0$, there is no gender wage gap by education level.
- Wald Statistics (q: 제약조건 수)

$$W_0 = \frac{(USS_R - USS_U)/q}{USS_U/(T-k)} = \frac{(\tilde{e}^\top \tilde{e} - e^\top e)/q}{e^\top e/(T-k)} \sim F(q, T-k)$$

- Chow Test: structural break 전후 dummy 혹은 두 가지 모델로

$$W_0 = \frac{(USS_R - USS_U)/q}{USS_U/(T-2k)} = \frac{(\tilde{e}^\top \tilde{e} - e^\top e)/q}{e^\top e/(T-2k)} \sim F(q=k, T-2k)$$

7.4 Conditional Expectation

- Properties of Conditional Expectation
  1. $\mathbb{E}(X \mid X) = X$
  2. $\mathbb{E}(XY \mid X) = X \cdot \mathbb{E}(Y \mid X)$
  3. $\mathbb{E}(Y + Z \mid X) = \mathbb{E}(Y \mid X) + \mathbb{E}(Z \mid X)$
  4. If $X \perp\!\!\!\perp Y$, then $\mathbb{E}(Y \mid X) = \mathbb{E}(Y)$
  5. If $\mathbb{E}(Y \mid X) = \mathbb{E}(Y)$, then $\mathrm{Cov}(X, Y) = 0$
  6. $\mathbb{E}(\mathbb{E}(Y \mid X)) = \mathbb{E}(Y)$

7.5 GLS Estimator

- No serial correlation & $\mathrm{Var}(\epsilon) = \begin{cases} \sigma_1^2 & t = 1, 2, \cdots, T_1 \\ \sigma_2^2 & t = T_1 + 1, \cdots, T \end{cases}$
  - Then, $\Omega = \begin{pmatrix} \sigma_1^2 I_{T_1} & 0 \\ 0 & \sigma_2^2 I_{T-T_1} \end{pmatrix}$,
  - where $\hat{\sigma}_1^2 = \frac{1}{T_1 - k} \sum_{t=1}^{T_1} e_t^2$, $\hat{\sigma}_2^2 = \frac{1}{T - T_1 - k} \sum_{t=T_1+1}^{T} e_t^2$
- No heteroskedasticity & $\epsilon_t = \rho \epsilon_{t-1} + \eta_t$
  - Then $\Omega = \sigma^2 \begin{pmatrix} 1 & \rho^* & (\rho^*)^2 & \cdots & (\rho^*)^{T-1} \\ \rho^* & 1 & \rho^* & \cdots & (\rho^*)^{T-2} \\ (\rho^*)^2 & \rho^* & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \rho^* \\ (\rho^*)^{T-1} & (\rho^*)^{T-2} & \cdots & \rho^* & 1 \end{pmatrix} \neq \sigma^2 I_T.$
  - cf. Cochrane-Orcutt

7.6 Example

- F-Test: Is this model meaningful?
- R-square: Is the explanatory variable effective?
- White Test: χ^2 검정 (If $T \cdot R^2 > C$, reject H_0: heteroskedasticity)
- Durbin-Watson
- VIF
- IV: endogeneity check