Introductory Econometrics

Remained Discussions

Yaohan Chen

School of Big Data and Statistics, Anhui University

Spring, 2025

Model Specification and Model Comparison

All models are wrong, but some are useful.

George E. P. Box

- Model misspecfication mainly refers to
  - Selecting variables incorrectly.
  - Specifying model functional form incorrectly.
- Model misspecification may induce an increase in estimator variance and is closely related to bias and inconsistency issues.
- To compare and select specific model(s), we can use various information criteria. AIC, BIC, and DIC,...

Serial Correlation and Time Series Data

- For linear regression model with observations indexed by t

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \ldots + \beta_k X_{tk} + u_t,$$

if for $t \neq s$, $\operatorname{Cov}(u_t, u_s) = \operatorname{E}(u_t u_s) \neq 0$, then obviously

$$\begin{aligned} \operatorname{Cov}(\boldsymbol{u}) &= \operatorname{E}(\boldsymbol{u}\boldsymbol{u}') = \begin{pmatrix} \sigma^2 & \cdots & \operatorname{E}(u_1 u_n) \\ \vdots & \ddots & \vdots \\ \operatorname{E}(u_n u_1) & \cdots & \sigma^2 \end{pmatrix} \\ &= \sigma^2 \boldsymbol{\Omega} \neq \sigma^2 \boldsymbol{I} \end{aligned}$$

- $\operatorname{E}(u_t u_s) \neq 0$ is referred to as the **serial correlation**, which violates the classical assumptions again.

Testing Serial Correlation

- Durbin-Watson Test

$$\text{D.W.} = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}.$$

- LM Test. Consider auxiliary regression

$$e_t = \beta_0 + \beta_1 X_{t1} + \cdots + \beta_k X_{tk} + \rho_1 e_{t-1} + \cdots + \rho_p e_{t-p} + \varepsilon_t$$

then under the null $H_0 : \rho_1 = \rho_2 = \cdots = \rho_p = 0$

$$\text{LM}_n = (n-p)R^2 \xrightarrow{d} \chi^2(p).$$

R^2 refers the R^2 of the auxiliary regression.

Remedies for Serial Correlation

- GLS if we know or we can estimate correlation coefficient properly.
- For the scenario with serial correlation of order 1,

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma_\varepsilon^2)$$

then

$$\mathrm{Cov}(\boldsymbol{u}\boldsymbol{u}') = \frac{\sigma_\varepsilon^2}{1-\rho^2} \begin{pmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho & 1 & \cdots & \rho^{n-2} \\ \cdots & \cdots & \cdots & \cdots \\ \rho^{n-1} & \rho^{n-2} & \cdots & 1 \end{pmatrix}$$

Remedies for Serial Correlation

- Difference method. Suppose

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + \varepsilon_t,$$

  then we can transform the original model by taking difference

$$\begin{aligned} & Y_t - \rho_1 Y_{t-1} - \cdots - \rho_p Y_{t-p} \\ = \; & \beta_0 \left(1 - \rho_1 - \cdots - \rho_p\right) + \beta_1 \left(X_{t1} - \rho_1 X_{t-1,1} - \cdots - \rho_p X_{t-p,1}\right) \\ + & \cdots + \beta_k \left(X_{tk} - \rho_1 X_{t-1,k} - \cdots - \rho_p X_{t-p,k}\right) + \varepsilon_t. \end{aligned}$$

- Since $\varepsilon_t \overset{i.i.d.}{\sim} (0, \sigma_\varepsilon^2)$, the transformed model satisfy the no serial-correlation assumption.