Introductory Econometrics

Multiple Linear Regression Model (II)

Yaohan Chen

School of Big Data and Statistics, Anhui University

Spring, 2025

Small Sample Properties of OLS Estimator

- Recall the OLS estimator as follows

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}.$$

- OLS estimator is unbiased.

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{u}) = \boldsymbol{\beta} + \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{u} \\
\mathrm{E}\left(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right) &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\,\mathrm{E}\left(\boldsymbol{u} \mid \boldsymbol{X}\right) = \boldsymbol{\beta}
\end{aligned}$$

Small Sample Properties of OLS Estimator

- Variance of OLS estimator $\hat{\boldsymbol{\beta}}$

$$
\begin{aligned}
\operatorname{Var}\left(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right) &= \mathrm{E}\left[(\hat{\boldsymbol{\beta}} - \mathrm{E}(\hat{\boldsymbol{\beta}}))(\hat{\boldsymbol{\beta}} - \mathrm{E}(\hat{\boldsymbol{\beta}}))' \mid \boldsymbol{X}\right] \\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\mathrm{E}\left(\boldsymbol{u}\boldsymbol{u}'\right)\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\sigma^2\boldsymbol{I}_n\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}
\end{aligned}
$$

- Among all the linear unbiased estimator of $\boldsymbol{\beta}$, denoted by $\tilde{\boldsymbol{\beta}}$,

$$
\operatorname{Var}\left(\tilde{\boldsymbol{\beta}} \mid \boldsymbol{X}\right) - \operatorname{Var}\left(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}\right)
$$

  is semi-positive definite matrix (positive semi-definite matrix).

Large Sample Properties of OLS Estimator

- OLS estimator $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}$.

$$\begin{aligned} P\lim\hat{\boldsymbol{\beta}} &= \boldsymbol{\beta} + P\lim\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{u} \\ &= \boldsymbol{\beta} + \left(P\lim\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} P\lim\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{u}\right) \end{aligned}$$

- According to the law of large numbers, we have

$$\begin{aligned} P\lim\frac{1}{n}\boldsymbol{X}'\boldsymbol{X} &= P\lim\frac{1}{n}\sum \boldsymbol{X}_i\boldsymbol{X}_i' = \mathrm{E}\left(\boldsymbol{X}_i\boldsymbol{X}_i'\right) = \boldsymbol{Q} \\ P\lim\frac{1}{n}\boldsymbol{X}'\boldsymbol{u} &= P\lim\frac{1}{n}\sum \boldsymbol{X}_i' u_i = \mathrm{E}\left(\boldsymbol{X}_i' u_i\right) = \boldsymbol{0} \end{aligned}$$

where $\boldsymbol{X}_i = (1, X_{i1}, X_{i2}, \cdots, X_{ik})'$.

Large Sample Properties of OLS Estimator

- OLS estimator $\hat{\boldsymbol{\beta}}$ is asymptotically efficient. Besides, under some regular conditions, $\hat{\boldsymbol{\beta}}$ asymptotically follows a multivariate normal distribution

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\xrightarrow{d} N\left(0,\boldsymbol{Q}^{-1}\boldsymbol{V}\boldsymbol{Q}^{-1}\right).$$

where

$$\begin{aligned}\boldsymbol{Q} &\equiv \mathrm{E}\left(\boldsymbol{X}_i\boldsymbol{X}_i'\right)\\ \boldsymbol{V} &\equiv \mathrm{E}\left(\boldsymbol{X}_i\boldsymbol{X}_i'u_i^2\right)\end{aligned}$$

Hypothesis Testing

- With the classical assumptions **Assumption 1** to **Assumption 5**, for small sample we have

$$\hat{\boldsymbol{\beta}} \mid \boldsymbol{X} \sim N\left[\boldsymbol{\beta}, \sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right],$$

  and for large sample we can relax the **Assumption 5** associated with normality assumption and have

$$\hat{\boldsymbol{\beta}} \mid \boldsymbol{X} \overset{a}{\sim} N\left[\boldsymbol{\beta}, \sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right].$$

Hypothesis Testing

- Single linear restriction: t-test. For the single linear restriction, we can express the null hypothesis H_0 and the alternative hypothesis H_1 in matrix form as follows

$$H_0 : c'\boldsymbol{\beta} = r \text{ versus } H_1 : c'\boldsymbol{\beta} \neq r$$

  where c refers to a $(k+1)$ vector and r is a scalar.
- With **Assumption 5**,

$$c'\hat{\boldsymbol{\beta}} \mid \boldsymbol{X} \sim N\left(c'\boldsymbol{\beta}, \sigma^2 c' \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} c\right)$$

Hypothesis Testing

- Under H_0,

$$\frac{c'\hat{\boldsymbol{\beta}} - r}{\sqrt{\sigma^2 c' \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} c}} \sim N(0,1).$$

Specifically, if c is a vector with 1 in its jth place and 0 elsewhere, and $r = \beta_j$, then

$$\hat{\beta}_j \sim N\left(\beta_j, \sigma^2 c_{jj}\right)$$

where c_{jj} denotes the jth element on the diagonal of square matrix $\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$.

Hypothesis Testing

- Ton construct a valid statistic under H_0, we use the sample estimation $\hat{\sigma}^2$ as the proxy for σ^2. Under H_0,

$$t = \frac{\hat{\beta}_j - \beta_j}{S_{\hat{\beta}_j}} = \frac{\hat{\beta}_j - \beta_j}{\sqrt{c_{jj}\hat{\sigma}^2}} \sim t(n-k-1),$$

where

$$\hat{\sigma}^2 = \frac{\boldsymbol{e}'\boldsymbol{e}}{n-k-1}.$$

Hypothesis Testing

- Multiple linear restrictions: F-test. We consider testing q linear restrictions on $\boldsymbol{\beta}$,

$$H_0 : R\boldsymbol{\beta} = r \text{ versus } H_1 : R\boldsymbol{\beta} \neq r$$

where R is a known matrix of order $q \times (k+1)$ with $q < k+1$ and r is a known $q \times 1$ vector. We assume $\operatorname{rank}(R) = q$.

Example. If

$$R = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} = [\mathbf{0} \quad \boldsymbol{I}_k] \quad r = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \quad q = k$$

Hypothesis Testing

Example (Cont'd). This is equivalent to testing

$$H_0 \;:\; \beta_1 = \ldots = \beta_k = 0.$$
$$H_1 \;:\; \exists\, \beta_j \neq 0 \quad (j = 1, 2, \cdots, k).$$

Hypothesis Testing

- In general, under H_0

$$F_n \equiv \frac{1}{q}(R\hat{\boldsymbol{\beta}}-r)' \left[\hat{\sigma}^2 R \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} R'\right]^{-1} (R\hat{\boldsymbol{\beta}}-r) \sim F(q, n-k-1).$$

- Equivalently, when $H_0: \ \beta_1 = \ldots = \beta_k = 0$, $q = k$:

$$F_n = \frac{\text{ESS}/q}{\text{RSS}/(n-k-1)} \quad \text{or} \quad F_n = \frac{R^2/q}{(1-R^2)/(n-k-1)}.$$

- We reject H_0 if $F_n > F_\alpha\,(q, n-k-1)$, where $F_\alpha\,(q, n-k-1)$ refers to the threshold for a given level of significance.

Hypothesis Testing

- t-test and F-test. Suppose that we are still interested in testing $H_0 : \beta_j = \beta_{j0}$ (for instance, $\beta_{j0} = 0$ or any other value you want test).

- In this case, $q = 1$, $r = \beta_{j0}$, and R is row vector with 1 in its jth place and 0 elsewhere. Then under H_0, $R\hat{\boldsymbol{\beta}} - r = \hat{\beta}_j - \beta_j$ and $R\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}R' = \left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{jj}$. The test static is

$$F_n = \left\{ \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 \left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right]_{jj}}} \right\}^2 \sim F(1, n-k-1) \text{ under } H_0.$$

- Note the expression inside the curly bracket is just the t-statistic.

Prediction

- Given the fitted model (or sample regression function)

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$$

and explanatory variable $\boldsymbol{X}_0 = (1, X_{01}, X_{02}, , \cdots , X_{0k})$, we can express the prediction for Y_0 as follows

$$\hat{Y}_0 = \boldsymbol{X}_0\hat{\boldsymbol{\beta}}$$

- For a given $\boldsymbol{X}_0$

$$\mathrm{E}\left(\hat{Y}_0\right) = \mathrm{E}\left(\boldsymbol{X}_0\hat{\boldsymbol{\beta}}\right) = \boldsymbol{X}_0 \, \mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{X}_0\boldsymbol{\beta} = \mathrm{E}\left(Y_0\right)$$

Prediction

- For a given $\boldsymbol{X}_0$

$$\begin{aligned}
\operatorname{Var}\left(\hat{Y}_0\right) &= \mathrm{E}\left(\boldsymbol{X}_0\hat{\boldsymbol{\beta}} - \boldsymbol{X}_0\boldsymbol{\beta}\right)^2 \\
&= \mathrm{E}\left[\boldsymbol{X}_0(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{X}_0'\right] \\
&= \boldsymbol{X}_0 \operatorname{Var}\left(\hat{\boldsymbol{\beta}}\right) \boldsymbol{X}_0' \\
&= \sigma^2 \boldsymbol{X}_0 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}_0'
\end{aligned}$$

- If we further have normality assumption, then

$$\hat{Y}_0 \sim N\left(\boldsymbol{X}_0\boldsymbol{\beta}, \sigma^2 \boldsymbol{X}_0 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}_0'\right)$$

Prediction

- Confidence interval for $\mathrm{E}(Y_0)$.
  - Note that for given $\boldsymbol{X}$ and $\boldsymbol{X}_0$

$$\frac{\hat{Y}_0 - \mathrm{E}(Y_0)}{\hat{\sigma}\sqrt{\boldsymbol{X_0}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}_0'}} \sim t(n-k-1)$$

  - Given the level of significance α

$$\hat{Y}_0 - t_{\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{\mathbf{X}_0\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}_0'} < \mathrm{E}(Y_0) < \hat{Y}_0 + t_{\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{\boldsymbol{X}_0\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}_0'}$$

Prediction

- We can also use $\hat{Y}_0$ as the prediction for Y_0. The prediction error for Y_0 is $e_0 = Y_0 - \hat{Y}_0$.
- For given $\boldsymbol{X}_0$

$$
\begin{aligned}
\mathrm{E}\left(e_0\right) &= \mathrm{E}\left(\boldsymbol{X}_0\boldsymbol{\beta} + u_0 - \boldsymbol{X}_0\hat{\boldsymbol{\beta}}\right) \\
&= \mathrm{E}\left(u_0\right) - \boldsymbol{X}_0\,\mathrm{E}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \\
&= \mathrm{E}\left(u_0\right) = 0
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{Var}\left(e_0\right) &= \mathrm{E}\left[u_0 - \boldsymbol{X}_0\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{u}\right]^2 \\
&= \sigma^2\left(1 + \boldsymbol{X}_0\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}_0'\right)
\end{aligned}
$$

Prediction

- Confidence interval for Y_0.
  - For given $\boldsymbol{X}$ and $\boldsymbol{X}_0$

$$t = \frac{\hat{Y}_0 - Y_0}{\hat{\sigma}_{e_0}} \sim t(n-k-1)$$

where

$$\hat{\sigma}_{e_0}^2 = \hat{\sigma}^2 \left[1 + \boldsymbol{X}_0 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}_0'\right]$$

  - Given the level of significance α

$$\hat{Y}_0 - t_{\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{1 + \boldsymbol{X}_0 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}_0'} < Y_0 < \hat{Y}_0 + t_{\frac{\alpha}{2}} \times \hat{\sigma}\sqrt{1 + \boldsymbol{X}_0 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}_0'}$$

Linearization

- How to linearize a non-linear function $f(x)$. Log-linearization is one technique we commonly adopt in economics analysis.
- Recall the Cobb-Douglas function

$$Q = AK^{\alpha}L^{\beta}.$$

By taking logs on both sides of the equation, we have

$$\ln Q = \ln A + \alpha \ln K + \beta \ln L.$$

Linearization

- CES function

$$Q = A\left(\delta_1 K^{-\rho} + \delta_2 L^{-\rho}\right)^{-\frac{1}{\rho}} e^u \quad (\delta_1 + \delta_2 = 1)$$

- Define
$$f(K, L) = A\left(\delta_1 K^{-\rho} + \delta_2 L^{-\rho}\right)^{-\frac{1}{\rho}}$$

- Elasticity of substitution,

$$\mathrm{EIS}_{LK} = -\frac{d\ln(L/K)}{d\ln(f_L/f_K)}.$$

Linearization

- Recall that

$$\begin{aligned} f_K &= A\left(\delta_1 K^{-\rho}+\delta_2 L^{-\rho}\right)^{-\frac{1}{\rho}-1}\delta_1 K^{-\rho-1} \\ f_L &= A\left(\delta_1 K^{-\rho}+\delta_2 L^{-\rho}\right)^{-\frac{1}{\rho}-1}\delta_2 L^{-\rho-1} \end{aligned}$$

and

$$\begin{aligned} \ln\left(f_L/f_K\right) &= \ln\left(\delta_2/\delta_1\right)-\left(\rho+1\right)\ln\left(L/K\right) \\ d\ln\left(f_L/f_K\right) &= -\left(\rho+1\right)\ln\left(L/K\right) \end{aligned}$$

Therefore,

$$\text{EIS}_{LK}=\frac{1}{1+\rho}.$$

Linearization

- By taking log of Q, we obtain

$$\ln Q = \ln A - \frac{1}{\rho} \ln \left(\delta_1 K^{-\rho} + \delta_2 L^{-\rho} \right) + u$$

- By expanding $\ln \left(\delta_1 K^{-\rho} + \delta_2 L^{-\rho} \right)$ at $\rho = 0$ using Taylor series to the second order, we obtain

$$\ln Y = \ln A + \delta_1 \ln K + \delta_2 \ln L - \frac{1}{2} \rho \delta_1 \delta_2 \left[\ln \left(\frac{K}{L} \right) \right]^2 .$$

Nonlinear least squares

- We can summarize the multivariate linear regression model in a more generic form,

$$f\left(X_1, X_2, \cdots, X_k, \beta_0, \beta_1, \cdots, \beta_k\right) + u.$$

- Given the sample $\{\boldsymbol{X}_i, Y_i\}_{i=1}^n$, we seek $\hat{\boldsymbol{\beta}}$ such that

$$Q\left(\hat{\boldsymbol{\beta}}\right) = \sum_{i=1}^{n}\left[Y_i - f\left(\boldsymbol{X}_i, \hat{\boldsymbol{\beta}}\right)\right]^2$$

  is minimized.

- Numerical methods are needed for obtaining $\hat{\boldsymbol{\beta}}$.