Introductory Econometrics

Multiple Linear Regression Model (III)

Yaohan Chen

School of Big Data and Statistics, Anhui University

Spring, 2026

Constrained Least Squares

- We can rewrite the OLS procedure in matrix form

$$\min_{\boldsymbol{\beta}} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$$

- We have matrix derivatives

$$\begin{aligned} \nabla_x (Ax - b) &= A' \\ \nabla_x (Ax - b)'(Ax - b) &= 2A'(Ax - b) \end{aligned}$$

- By taking first order derivative of $(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ and set it equal to $\mathbf{0}$, we have

$$2\boldsymbol{X}'\left(\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Y}\right) = \mathbf{0}.$$

Constrained Least Squares

- Now we consider the constrained least squares expressed in matrix form,

$$\min_{\boldsymbol{\beta}} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \quad \text{s.t.} \quad R\boldsymbol{\beta} = r.$$

- Establishing Lagrangian

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{\lambda}'(R\boldsymbol{\beta} - r)$$

Constrained Least Squares

- Taking first derivatives of $\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ and setting $\partial\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda})/\partial\boldsymbol{\beta} = \mathbf{0}$, $\partial\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda})/\partial\boldsymbol{\lambda} = \mathbf{0}$,

$$\begin{aligned} \frac{\partial\mathcal{L}(\hat{\boldsymbol{\beta}}_*, \boldsymbol{\lambda})}{\boldsymbol{\partial\beta}} &= -2\boldsymbol{X}'\boldsymbol{Y} + 2\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}}_* + R'\boldsymbol{\lambda} = \mathbf{0} && (\dagger) \\ \frac{\partial\mathcal{L}(\hat{\boldsymbol{\beta}}_*, \boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}} &= R\hat{\boldsymbol{\beta}}_* - r = 0 && (\ddagger) \end{aligned}$$

- Equation (†) implies that

$$\hat{\boldsymbol{\beta}}_* = \hat{\boldsymbol{\beta}} - \frac{1}{2}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} R'\boldsymbol{\lambda}.$$

Constrained Least Squares

- By substituting $\hat{\boldsymbol{\beta}}_*$ in (‡) we obtain

$$R\hat{\boldsymbol{\beta}} - \frac{1}{2}\left[R\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}R'\right]\boldsymbol{\lambda} = r,$$

and we can solve λ from it as follows

$$\boldsymbol{\lambda} = 2\left[R\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}R'\right]^{-1}(R\hat{\boldsymbol{\beta}} - r).$$

- By substituting $\boldsymbol{\lambda}$ in $\hat{\boldsymbol{\beta}}_*$, we obtain the solution to the constrained least squares,

$$\hat{\boldsymbol{\beta}}_* = \hat{\boldsymbol{\beta}} - \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}R'\left[R\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}R'\right]^{-1}(R\hat{\boldsymbol{\beta}} - r).$$

Implications from Constrained Least Squares

- Let RSS_U denote the residual sum of squares associated with unconstrained least squares, and RSS_R denote the the residual sum of squares associated with constrained least squares. Then
$$\text{RSS}_\text{U} \leqslant \text{RSS}_\text{R} .$$
- RSS does not increase as the number of explained variables increase. Thus, RSS is a non-increasing function in k.
- In connection with the hypothesis testing,
$$F_n = \frac{(\text{RSS}_\text{R} - \text{RSS}_\text{U})/q}{\text{RSS}_\text{U}/(n-k-1)} \sim F(q, n-k-1) .$$

Implications from Constrained Least Squares

- For this scenario, we consider imposing constraints such that
$$\beta_j = 0, \quad j = 1, 2, \cdots, k.$$
Equivalently,
$$\boldsymbol{Y} = \iota\beta_0 + \boldsymbol{u}$$
where $\iota = (1, \cdots, 1)'$. For this constrained least squares, $\mathrm{ESS}_{\mathrm{R}} = 0$.
- We can derive that (when $q = k$)
$$F_n = \frac{\mathrm{ESS}_{\mathrm{U}}/q}{\mathrm{RSS}_{\mathrm{U}}/(n-k-1)} \quad \text{or} \quad F_n = \frac{R^2_{\mathrm{U}}/q}{(1-R^2_{\mathrm{U}})/(n-k-1)}.$$

Testing Structural Change

- Suppose we have following two regressions

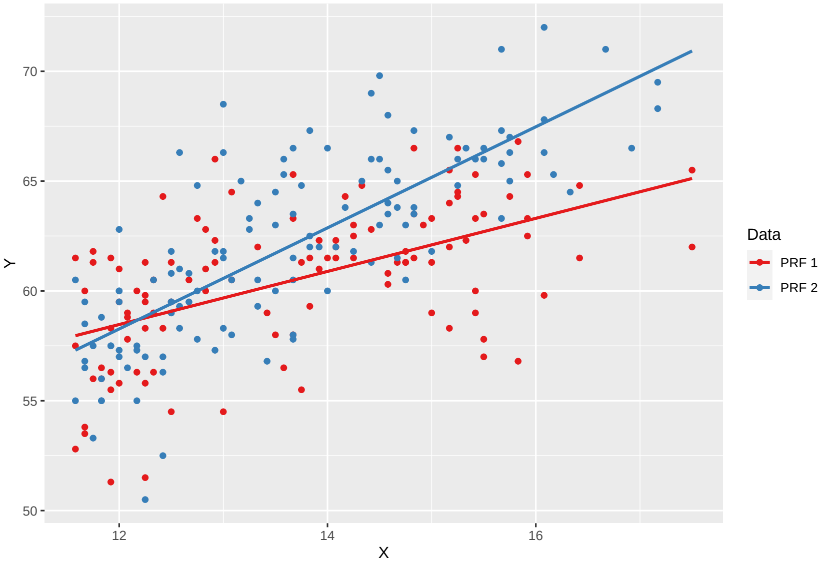
$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + u_1 \\ Y &= \alpha_0 + \alpha_1 X_1 + \cdots + \alpha_k X_k + u_2 \end{aligned}$$

- We can stack these two regressions in matrix form

$$\begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{X}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix} + \begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \end{pmatrix}$$

and separately

$$\begin{aligned} \underset{n_1 \times 1}{\boldsymbol{Y}_1} &= \underset{n_1 \times (k+1)}{\boldsymbol{X}_1} \underset{(k+1) \times 1}{\boldsymbol{\beta}} + \underset{n_1 \times 1}{\boldsymbol{u}_1} \\ \underset{n_2 \times 1}{\boldsymbol{Y}_2} &= \underset{n_2 \times (k+1)}{\boldsymbol{X}_2} \underset{(k+1) \times 1}{\boldsymbol{\alpha}} + \underset{n_2 \times 1}{\boldsymbol{u}_2} \end{aligned}$$


70
65
60
55
50
Y
12
14
16
X
Data
PRF 1
PRF 2

Testing Structural Change

- We want to test whether $\boldsymbol{\beta} = \boldsymbol{\alpha}$. We can use $\boldsymbol{\beta} = \boldsymbol{\alpha}$ as the hull hypothesis H_0.
- Without $\boldsymbol{\beta} = \boldsymbol{\alpha}$ restriction, OLS suggests that

$$\underbrace{\begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix}}_{\boldsymbol{Y}} = \begin{pmatrix} \boldsymbol{X}_1 \hat{\boldsymbol{\beta}} \\ \boldsymbol{X}_2 \hat{\boldsymbol{\alpha}} \end{pmatrix} + \underbrace{\begin{pmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \end{pmatrix}}_{\boldsymbol{e}}.$$

  Therefore, $\boldsymbol{e}'\boldsymbol{e} = \boldsymbol{e}_1'\boldsymbol{e}_1 + \boldsymbol{e}_2'\boldsymbol{e}_2$, and $\boldsymbol{e}'\boldsymbol{e}$ is RSS_U, $\boldsymbol{e}_1'\boldsymbol{e}_1$ is RSS_1, $\boldsymbol{e}_2'\boldsymbol{e}_2$ is RSS_2.

Testing Structural Change

- By imposing $\boldsymbol{\beta} = \boldsymbol{\alpha}$ restriction ($(k+1)$ restrictions), we obtain

  $$\begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \end{pmatrix}$$

  and

  $$\begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \hat{\boldsymbol{\beta}}_* + \boldsymbol{e}_*,$$

  where $\hat{\boldsymbol{\beta}}_*$ refers to the restricted least squares solution and $\boldsymbol{e}_*$ refers to the corresponding residual. $\boldsymbol{e}_*'\boldsymbol{e}_*$ denotes the RSS_R.

Testing Structural Change

- We can construct statistic for testing structural change. Under H_0 (corresponding to $\mathrm{RSS_R}$),

$$F_n = \frac{\left(\mathrm{RSS_R} - \mathrm{RSS_U}\right)/(k+1)}{\mathrm{RSS_U}\,/\,[n_1+n_2-2(k+1)]} \sim F\left(k+1, n_1+n_2-2(k+1)\right)$$

or

$$F_n = \frac{\left[\mathrm{RSS_R} - \left(\mathrm{RSS_1}+\mathrm{RSS_2}\right)\right]/(k+1)}{\left(\mathrm{RSS_1}+\mathrm{RSS_2}\right)/\,[n_1+n_2-2(k+1)]} \sim F\left(k+1, n_1+n_2-2(k+1)\right)$$

Testing Structural Change

- The testing procedure can be summarized as follows
  1. Splitting the sample into two parts, $n = n_1 + n_2$, and run OLS for the separate samples to obtain RSS_1 and RSS_2 respectively.
  2. Running OLS for the original sample assuming there is no structural change and obtain RSS_R.
  3. Calculating testing statistic under H_0 using RSS_R, RSS_1 and RSS_2. Comparing the calculated statistic with the threshold value.
- This testing procedure refers to the Chow test, introduced by Gregory Chow in 1960.

Large Sample Hypothesis Testing: Wald, LR and LM Tests

- Recall that under $H_0 : R\boldsymbol{\beta} = r$,

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

  This result holds for finite sample.

- When $n \to \infty$ and assuming regular conditions hold, under H_0 and use $\mathrm{RSS}_{\mathrm{U}} / n$ as the proxy for $\hat{\sigma}^2$,

$$W_n \equiv qF_n \xrightarrow{d} \chi^2(q).$$

  This refers to **Wald test**.

Large Sample Hypothesis Testing: Wald, LR and LM Tests

- Given the expression for W_n, we can derive

$$\frac{\mathrm{RSS_R} - \mathrm{RSS_U}}{\mathrm{RSS_U}/(n-k-1)} \xrightarrow{d} \chi^2(q).$$

- nR^2 test.

$$nR^2 \xrightarrow{d} \chi^2(q)$$

where q refers to the number of restrictions and R^2 refers to R^2 associated with following auxiliary regression

$$e_{\mathrm{R}} = \delta_0 + \delta_1 X_1 + \delta_2 X_2 + \cdots + \delta_k X_k + \varepsilon.$$

Large Sample Hypothesis Testing: Wald, LR and LM Tests

- To see nR^2 test statistic more clearly, suppose we impose q restrictions such that q elements in $(\beta_1, \cdots, \beta_k)$ are zeros.

$$\text{RSS}_\text{R} = \boldsymbol{e}_\text{R}' \boldsymbol{e}_\text{R}$$

- Note that (why ?)

$$\text{RSS}_\text{R} - \text{RSS}_\text{U} \quad = \quad \underbrace{\text{ESS}_\text{aux}}_{\text{ESS of auxiliary regression}}$$

Consequently,

$$qF_n = \frac{R^2}{(1 - R^2)/(n - k - 1)} \xrightarrow{d} \chi^2(q).$$

Large Sample Hypothesis Testing: Wald, LR and LM Tests

- It can be shown that given $qF_n \xrightarrow{d} \chi^2(q)$,

$$(n-k-1)R^2 \xrightarrow{d} \chi^2(q).$$

Therefore,

$$nR^2 = \frac{n}{n-k-1}(n-k-1)R^2 \xrightarrow{d} \chi^2(q).$$

Large Sample Hypothesis Testing: Wald, LR and LM Tests

- Maximum Likelihood Estimation relies on likelihood function $L\left(\boldsymbol{\beta}, \sigma^2\right)$.
- Unconstrained MLE:

$$\text{Max} : L\left(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2\right).$$

- Constrained MLE:

$$\text{Max} : L\left(\tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2\right) \quad \text{s.t.} \quad g(\boldsymbol{\beta}) = \mathbf{0}.$$

  Or through Lagrangian

$$L\left(\boldsymbol{\beta}, \sigma^2\right) - \boldsymbol{\lambda}' g(\boldsymbol{\beta}).$$

Large Sample Hypothesis Testing: Wald, LR and LM Tests

- The key idea of **Likelihood Ratio (LR) test** is to test $g(\boldsymbol{\beta}) = \mathbf{0}$ by comparing likelihood of unconstrained model and likelihood of constrained model.
- Suppose we have q restrictions.

$$\mathrm{LR}_n = -2\left[\ln L\left(\tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2\right) - \ln L\left(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2\right)\right] \xrightarrow{d} \chi^2(q).$$

Large Sample Hypothesis Testing: Wald, LR and LM Tests

- Lagrangian multiplier $\boldsymbol{\lambda}$ can be used for constructing test statistic. Under linear restrictions, we have the corresponding test (**Lagrangian Multiplier, LM test**) statistic under $H_0 : R\boldsymbol{\beta} = r$,

$$\text{LM}_n = \tilde{\sigma}^2 \tilde{\boldsymbol{\lambda}}' R \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} R' \tilde{\boldsymbol{\lambda}},$$

  where $\tilde{\sigma}^2$ and $\tilde{\boldsymbol{\lambda}}$ refers to the solutions to constrained MLE.

- It can shown that under H_0

$$\text{LM}_n = nR^2 \xrightarrow{d} \chi^2(q),$$

  R^2 is associated with the auxiliary regression.

Large Sample Hypothesis Testing: Wald, LR and LM Tests

- Under some regular conditions, it can be shown that

$$\begin{aligned} W_n &= n\left(\frac{\mathrm{RSS_R} - \mathrm{RSS_U}}{\mathrm{RSS_U}}\right) \\ \mathrm{LM}_n &= n\left(\frac{\mathrm{RSS_R} - \mathrm{RSS_U}}{\mathrm{RSS_R}}\right) \\ \mathrm{LR}_n &= n\ln\left(\frac{\mathrm{RSS_R}}{\mathrm{RSS_U}}\right) \end{aligned}$$

- Let $x = (\mathrm{RSS_R} - \mathrm{RSS_U})/\mathrm{RSS_U}$ and using the the inequality that for $x > 0$, $x/(1+x) < \ln(1+x) < x$, we have

$$\mathrm{LM}_n \leqslant \mathrm{LR}_n \leqslant W_n.$$

Dummy Variables

- Dummy variables in econometrics model.

$$D = \begin{cases} 1 & \text{Bachelor's degree or higher education} \\ 0 & \text{Without Bachelor's degree or highre education} \end{cases}$$

- We can also introduce dummy variable to indicate the gender differences.

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + u_i$$

where Y_i refers to the "wage income", X_i refers to the "years of employmen", and $D_i = 1$ for male, $D_i = 0$ for female.

Dummy Variables

- Dummy variables can be added as additional regressors or in combination with other variables.
- Dummy variables as additional regressors

$$\mathrm{E}\left(Y_i \mid X_i, D=0\right)=\beta_0+\beta_1 X_i$$

$$\mathrm{E}\left(Y_i \mid X_i, D=1\right)=\left(\beta_0+\beta_2\right)+\beta_1 X_i$$

Dummy Variables

- Dummy variables in combination with other variables

$$D = \begin{cases} 1 & \text{rural residents} \\ 0 & \text{urban residents} \end{cases}$$

and we want to check the relationship between consumption (C_i) and income (X_i) using regression

$$C_i = \beta_0 + \beta_1 X_i + \beta_2 D_i X_i + u_i$$

- D_i distinguishes the marginal effect of income on consumption.

$$\mathrm{E}\left(C_i \mid X, D = 1\right) = \beta_0 + \boxed{(\beta_1 + \beta_2)} X_i$$

$$\mathrm{E}\left(C_i \mid X, D = 0\right) = \beta_0 + \boxed{\beta_1} X_i$$

Dummy Variables

- Dummy variables should satisfy the rank condition.
- Suppose we have following seasonal dummies

$$D_{i1} = \begin{cases} 1 & \text{Spring} \\ 0 & \text{Other} \end{cases} ; D_{i2} = \begin{cases} 1 & \text{Summer} \\ 0 & \text{Other} \end{cases} ; D_{i3} = \begin{cases} 1 & \text{Autumn} \\ 0 & \text{Other} \end{cases}$$

and following regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \alpha_1 D_{i1} + \alpha_2 D_{i2} + \alpha_3 D_{i3} + u_i$$

Dummy Variables

- We cannot add additional dummy D_{i4} indicating Winter, since otherwise for the regression model in matrix form

$$\boldsymbol{Y} = (\boldsymbol{X}, \boldsymbol{D}) \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} + \boldsymbol{u}$$

  the first column of $\boldsymbol{X}$ can be represented by any linear combination of vectors of $\boldsymbol{D}$, therefore $(\boldsymbol{X}, \boldsymbol{D})$ is not full rank.

Dummy Variables

- Testing structural change using dummy variables.

$$Y_i = \beta_0 + \delta_0 D_i + \beta_1 X_{i1} + \delta_1 (D_i X_{i1}) + \cdots + \beta_k X_{ik} + \delta_k (D_i X_{ik})$$

$$D_i = \begin{cases} 1 & \{Y_i, X_{i1}, \cdots, X_{ik}\} \text{ from sample 1} \\ 0 & \{Y_i, X_{i1}, \cdots, X_{ik}\} \text{ from sample 2} \end{cases}$$

- Testing structural change(s) is equivalent to test

$$H_0 : \delta_0 = 0, \delta_1 = 0, \cdots, \delta_k = 0.$$