Classical Linear Regression Model

- Finite Sample Properties of OLS
- Restricted Least Squares
- Specification Errors
  - Omitted Variables
  - Irreverent Variables

Finite Sample Properties of OLS

- Finite Sample Properties of **b**
  $\mathbf{b} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$
  - Under A.1-3, $E(\mathbf{b}|\mathbf{X}) = \boldsymbol{\beta}$
  - Under A.1-4, $Var(\mathbf{b}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$
  - Under A.1-4, the OLS estimator is **efficient** in the class of linear unbiased estimators (Gauss-Markov Theorem).

Finite Sample Properties of OLS

- Proof of Gauss-Markov Theorem
  - $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, $\mathrm{Var}(\mathbf{b}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$
  - Let $\mathbf{d}=\mathbf{A}\mathbf{y}$ be another linear unbiased estimator of $\boldsymbol{\beta}$, where $\mathbf{A}= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'+\mathbf{C}$ and $\mathbf{C}\neq\mathbf{0}$, $\mathbf{C}\mathbf{X}=\mathbf{0}$
    - $E(\mathbf{d}|\mathbf{X}) = E[((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'+\mathbf{C})(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})|\mathbf{X}] = \boldsymbol{\beta}$
    - $\mathrm{Var}(\mathbf{d}|\mathbf{X}) = \mathrm{Var}(\mathbf{A}\mathbf{y}|\mathbf{X})$
      $= \sigma^2((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'+\mathbf{C})((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'+\mathbf{C})'$
      $= \sigma^2((\mathbf{X}'\mathbf{X})^{-1}+\mathbf{C}\mathbf{C}')$
      $= \mathrm{Var}(\mathbf{b}|\mathbf{X})+\sigma^2\mathbf{C}\mathbf{C}'$
    - Therefore, $\mathrm{Var}(\mathbf{d}|\mathbf{X}) \geq \mathrm{Var}(\mathbf{b}|\mathbf{X})$

Finite Sample Properties of OLS

- OLS estimator is BLUE. Assumption 2 (exogeneity) plays an important role to establish these results:
  - $\mathbf{b}$ is linear in $\mathbf{y}$ and $\boldsymbol{\varepsilon}$.
  - $\mathbf{b}$ is unbiased estimator of $\boldsymbol{\beta}$:
    $E(\mathbf{b}) = E(E(\mathbf{b}|X)) = \boldsymbol{\beta}$
  - $\mathbf{b}$ is efficient or best:
    $Var(\mathbf{b}) = E(Var(\mathbf{b}|X))$ is the minimum variance-covariance matrix

Finite Sample Properties of OLS

- The relationship between s^2 and σ^2
  - $s^2 = \mathbf{e}'\mathbf{e}/(n-K)$, $\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$
- Finite Sample Properties of s^2
  - Under A.1-4, $E(s^2|\mathbf{X}) = \sigma^2$ (and hence $E(s^2) = \sigma^2$), provided $n > K$.

Finite Sample Properties of OLS

- Proof of $E(s^2|\mathbf{X})=\sigma^2$: s^2 is an unbiased estimator of σ^2 (Recall: $s^2=\mathbf{e}'\mathbf{e}/(n\text{-}K)$, $\mathbf{e}=\mathbf{M}\boldsymbol{\varepsilon}$, $\mathbf{M}=\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$)
  - $E(\mathbf{e}'\mathbf{e}|\mathbf{X})=E(\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}|\mathbf{X})= E(\text{trace}(\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon})|\mathbf{X})$
    $= E(\text{trace}(\mathbf{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')|\mathbf{X})=\text{trace}(\mathbf{M}E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}))$
    $= \text{trace}(\mathbf{M}\sigma^2\mathbf{I})=\sigma^2\text{trace}(\mathbf{M})=\sigma^2(n\text{-}K)$
  - Therefore, $E(s^2|\mathbf{X})=E(\mathbf{e}'\mathbf{e}/(n\text{-}K)|\mathbf{X})=\sigma^2$

Finite Sample Properties of OLS

- Estimate of Var($\mathbf{b}|\mathbf{X}$) = $s^2(\mathbf{X}'\mathbf{X})^{-1}$
- Standard Errors

$$\mathrm{SE}(b_k) = \sqrt{s^2[(\mathbf{X}'\mathbf{X})^{-1}]_{kk}}$$

Restricted Least Squares

- $\mathbf{b}^* = \text{argmin}_{\beta}\ SSR(\boldsymbol{\beta}) = (\mathbf{y}\text{-}\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}\text{-}\mathbf{X}\boldsymbol{\beta})$
  s.t. $\mathbf{R}\boldsymbol{\beta} = \mathbf{q}$
- $(\mathbf{b}^*,\boldsymbol{\lambda}^*) = \text{argmin}_{(\beta,\lambda)}\ SSR^*(\boldsymbol{\beta},\boldsymbol{\lambda})$
  $= (\mathbf{y}\text{-}\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}\text{-}\mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\lambda}'(\mathbf{R}\boldsymbol{\beta} - \mathbf{q})$
  $\mathbf{R}$: $J \times K$ restriction matrix
  $\mathbf{q}$: $J \times 1$ vector of restricted values
  $\boldsymbol{\lambda}$: $J \times 1$ vector of Langrangian multiplier

Restricted Least Squares

- Linear Restrictions: $\mathbf{R}\boldsymbol{\beta} = \mathbf{q}$

$$\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1K} \\ r_{21} & r_{22} & \cdots & r_{2K} \\ \vdots & \vdots & \vdots & \vdots \\ r_{J1} & r_{J2} & \cdots & r_{JK} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_J \end{bmatrix}$$

Algebra of Restricted Least Squares

- $\partial SSR^*(\boldsymbol{\beta},\lambda)/\partial\boldsymbol{\beta} = -2\mathbf{X}'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})+\mathbf{R}'\boldsymbol{\lambda} = \mathbf{0}$
  $\partial SSR^*(\boldsymbol{\beta},\lambda)/\partial\boldsymbol{\lambda} = \mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \mathbf{0}$
- $\boldsymbol{\lambda}^* = 2[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}-\mathbf{q}]$
  $\mathbf{b}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}-\frac{1}{2}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\boldsymbol{\lambda}^*$
  $= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}-\mathbf{q}]$
  $= \mathbf{b} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b}-\mathbf{q})$
- $\mathbf{e}^* = \mathbf{y} - \mathbf{X}\mathbf{b}^* = \mathbf{e} + \mathbf{X}(\mathbf{b}-\mathbf{b}^*)$
  $\mathbf{e}^{*\prime}\mathbf{e}^* = \mathbf{e}'\mathbf{e} + (\mathbf{b}-\mathbf{b}^*)'\mathbf{X}'\mathbf{X}(\mathbf{b}-\mathbf{b}^*) > \mathbf{e}'\mathbf{e}$
  $\mathbf{e}^{*\prime}\mathbf{e}^* - \mathbf{e}'\mathbf{e} = (\mathbf{R}\mathbf{b}-\mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b}-\mathbf{q})$
- $E(\mathbf{b}^*|\mathbf{X}) = \boldsymbol{\beta} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}[\mathbf{R}\boldsymbol{\beta}-\mathbf{q}]$
  $E(\mathbf{b}^*|\mathbf{X}) \neq \boldsymbol{\beta}$ unless $\mathbf{R}\boldsymbol{\beta}=\mathbf{q}$
- $Var(\mathbf{b}^*|\mathbf{X}) = Var(\mathbf{b}|\mathbf{X})-\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}$

Discussions

- Linear regression without constant term (or intercept), a special case of restricted least squares.
- Restricted least squares estimator is biased if the restriction is incorrectly specified.

Application: Specification Errors

- Omitting relevant variables: Suppose the correct model is $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$. That is, two sets of variables. Compute least squares omitting $\mathbf{X}_2$. Some easily proved results:
- Var[$\mathbf{b}_1$] is smaller than Var[$\mathbf{b}_{1.2}$]. (The latter is the northwest sub-matrix of the full covariance matrix. The proof uses the residual maker (again!). That is, you get a smaller variance when you omit $\mathbf{X}_2$. (One interpretation: Omitting $\mathbf{X}_2$ amounts to using extra information ($\boldsymbol{\beta}_2 = \mathbf{0}$). *Even if the information is wrong, it reduces the variance.*

Application: Specification Errors

- $E[\mathbf{b}_1] = \boldsymbol{\beta}_1 + (\mathbf{X_1}'\mathbf{X_1})^{-1}\mathbf{X_1}'\mathbf{X_2}\boldsymbol{\beta}_2 \neq \boldsymbol{\beta}_1$. So, $\mathbf{b}_1$ is **biased**.(!!!) The bias can be huge. Can reverse the sign of a price coefficient in a "demand equation."
- $\mathbf{b}_1$ may be more "precise."
  Precision = Mean squared error
  = variance + squared bias.
  Smaller variance but positive bias. If bias is small, may still favor the short regression.
- Suppose $\mathbf{X_1}'\mathbf{X_2} = \mathbf{0}$. Then the bias goes away. Interpretation, the information is not "right," it is irrelevant. $\mathbf{b}_1$ is the same as $\mathbf{b}_{1.2}$.

Application: Specification Errors

- Including superfluous variables: Just reverse the results.
- Including superfluous variables increases variance. (The cost of not using information.)
- Does not cause a bias, because if the variables in $\mathbf{X}_2$ are truly superfluous, then $\boldsymbol{\beta}_2 = \mathbf{0}$, so $E[\mathbf{b}_{1.2}] = \boldsymbol{\beta}_1$.

Example

- Linear Regression Model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
  $G = \beta_0 + \beta_1 PG + \beta_2 Y + \beta_3 PNC + \beta_4 PUC + \varepsilon$
  $\mathbf{y} = G$; $\mathbf{X} = [1\ PG\ Y\ PNC\ PUC]$
  Note: All variables are log transformed.
- Linear Restrictions: $\mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \mathbf{0}$
  $\beta_3 = \beta_4 = 0$

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$