Classical Linear Regression Model

- Normality Assumption
- Hypothesis Testing Under Normality
- Maximum Likelihood Estimator
- Generalized Least Squares
Normality Assumption

- Assumption 5
  \( \varepsilon |X \sim N(0, \sigma^2 I_n) \)

- Implications of Normality Assumption
  - \( (b - \beta) | X \sim N(0, \sigma^2 (X'X)^{-1}) \)
  - \( (b_k - \beta_k) | X \sim N(0, \sigma^2 ([X'X]^{-1})_{kk}) \)

\[
z_k = \frac{b_k - \beta_k}{\sqrt{\sigma^2 ([X'X]^{-1})_{kk}}} \sim N(0, 1)
\]
Hypothesis Testing under Normality

• Implications of Normality Assumption
  – Because $\varepsilon_i/\sigma \sim N(0,1)$,
    
    $\frac{(n-K)s^2}{\sigma^2} = \frac{e'e}{\sigma^2} = \left( \frac{\varepsilon}{\sigma} \right)'M \left( \frac{\varepsilon}{\sigma} \right) \sim \chi^2(\text{trace}(M))$

    where $M = I - X(X'X)^{-1}X'$ and $\text{trace}(M) = n-K$. 
Hypothesis Testing under Normality

• If $\sigma^2$ is not known, replace it with $s^2$. The standard error of the OLS estimator $\beta_k$ is

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

• Suppose A.1-5 hold. Under $H_0$: $\beta_k = \bar{\beta}_k$ the t-statistic defined as

$$t_k = \frac{b_k - \bar{\beta}_k}{SE(b_k)} = \frac{b_k - \bar{\beta}_k}{\sqrt{s^2[(X'X)^{-1}]_{kk}}} \sim t(n-K).$$
Hypothesis Testing under Normality

Proof of

\[ t_k = \frac{b_k - \bar{\beta}_k}{\text{SE}(b_k)} = \frac{b_k - \bar{\beta}_k}{\sqrt{s^2[(X'X)^{-1}]_{kk}}} \sim t(n-K) \]

\[ \frac{b_k - \bar{\beta}_k}{\sqrt{\sigma^2[(X'X)^{-1}]_{kk}}} \sim \frac{N(0,1)}{\sqrt{(n-K)s^2 / \sigma^2}} \sqrt{\frac{\chi^2(n-K)}{n-K}} = t(n-K) \]
Hypothesis Testing under Normality

• Testing Hypothesis about Individual Regression Coefficient, $H_0: \beta_k = \bar{\beta}_k$
  
  – If $\sigma^2$ is known, use $z_k \sim N(0,1)$.
  
  – If $\sigma^2$ is not known, use $t_k \sim t(n-K)$.

  Given a level of significance $\alpha$, 
  
  $\text{Prob}(-t_{\alpha/2}(n-K) < t < t_{\alpha/2}(n-K)) = 1-\alpha$

  
  $-t_{\alpha/2}(n-K) < \frac{b_k - \bar{\beta}_k}{\text{SE}(b_k)} < t_{\alpha/2}(n-K)$
Hypothesis Testing under Normality

- Confidence Interval

\[ b_k - SE(b_k) t_{\alpha/2} (n - K) < \beta_k < b_k + SE(b_k) t_{\alpha/2} (n - K) \]

\[ [b_k - SE(b_k) t_{\alpha/2} (n - K), b_k + SE(b_k) t_{\alpha/2} (n - K)] \]

- p-Value: \( p = \text{Prob}(t>|t_k|) \times 2 \)

\[ \text{Prob}(-|t_k| < t < |t_k|) = 1 - p \]

since \( \text{Prob}(t>|t_k|) = \text{Prob}(t<-|t_k|) \).

Accept \( H_0 \) if \( p > \alpha \). Reject otherwise.
Hypothesis Testing under Normality

- Linear Hypotheses \( H_0: R\beta = q \)

\[
\begin{bmatrix}
  r_{11} & r_{12} & \cdots & r_{1K} \\
  r_{21} & r_{22} & \cdots & r_{2K} \\
  \vdots & \vdots & \ddots & \vdots \\
  r_{J1} & r_{J2} & \cdots & r_{JK}
\end{bmatrix}
\begin{bmatrix}
  \beta_1 \\
  \beta_2 \\
  \vdots \\
  \beta_K
\end{bmatrix}
=
\begin{bmatrix}
  q_1 \\
  q_2 \\
  \vdots \\
  q_J
\end{bmatrix}
\]
Hypothesis Testing under Normality

• Let $m = Rb - q$, where $b$ is the unrestricted least squares estimator of $\beta$.
  - $E(m|X) = E(Rb - q|X) = R\beta - q = 0$
  - $\text{Var}(m|X) = \text{Var}(Rb - q|X) = R\text{Var}(b|X)R' = \sigma^2R(X'X)^{-1}R'$

• Wald Principle
  
  $W = m'\text{Var}(m|X)^{-1}m = (Rb-q)'[\sigma^2R(X'X)^{-1}R']^{-1}(Rb-q)$
  
  $\sim \chi^2(J)$, where $J$ is the number of restrictions

• Define $F = (W/J)/(s^2/\sigma^2)$
  
  $= (Rb-q)'[s^2R(X'X)^{-1}R']^{-1}(Rb-q)/J$
Hypothesis Testing under Normality

• Suppose A.1-5 holds. Under $H_0: \mathbf{R}\beta = \mathbf{q}$, where $\mathbf{R}$ is $J \times K$ with $\text{rank}(\mathbf{R}) = J$, the F-statistic defined as

\[
F = \frac{(\mathbf{Rb} - \mathbf{q})' [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1} (\mathbf{Rb} - \mathbf{q}) / J}{s^2}
\]

is distributed as $F(J,n-K)$, the $F$ distribution with $J$ and $n-K$ degrees of freedom.
Discussions

• Residuals $e = y - Xb \sim N(0, \sigma^2 M)$ if $\sigma^2$ is known and $M = I - X(X'X)^{-1}X'$

• If $\sigma^2$ is unknown and estimated by $s^2$,

$$\frac{e_i}{\sqrt{s^2[1 - x_i'(X'X)^{-1}x_i]}} \sim t(n - K)$$

$i = 1, 2, ..., n$
Discussions

• Wald Principle vs. Likelihood Principle: By comparing the restricted (R) and unrestricted (UR) least squares, the F-statistic is shown

\[
F = \frac{(SSR_R - SSR_{UR}) / J}{SSR_{UR} / (n - K)} = \frac{(R_{UR}^2 - R_R^2) / J}{(1 - R_{UR}^2) / (n - K)}
\]
Discussions

- Testing $R^2 = 0$:
  Equivalently, $H_0$: $R\beta = q$, where

\[
\begin{bmatrix}
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1 \\
\end{bmatrix}
\begin{bmatrix}
\beta_1 \\
\beta_2 \\
\vdots \\
\beta_K \\
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
\vdots \\
0 \\
\end{bmatrix}
\]

$J = K-1$, and $\beta_1$ is the unrestricted constant term. The F-statistic follows $F(K-1,n-K)$. 
Discussions

• Testing $\beta_k = 0$: Equivalently, $H_0: \mathbf{R}\beta = \mathbf{q}$, where

$$
\begin{bmatrix}
\beta_1 \\
\beta_2 \\
\vdots \\
\beta_K
\end{bmatrix}
= 0
$$

$$
F(1, n-K) = b_k[\text{Est Var}(\mathbf{b})]^{-1}_{kk} b_k
$$

t-ratio: $t(n-K) = \frac{b_k}{\text{SE}(b_k)}$
Discussions

• t vs. F:
  – $t^2(n-K) = F(1,n-K)$ under $H_0: R\beta = q$ when $J=1$
  – For $J > 1$, the F test is preferred to multiple t tests

• Durbin-Watson Test Statistic for Time Series Model:

$$DW = \frac{\sum_{i=2}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

  – The conditional distribution, and hence the critical values, of DW depends on X…
Maximum Likelihood

• Assumption 1, 2, 4, and 5 imply
  \( y|X \sim N(X\beta, \sigma^2 I_n) \)

• The conditional density or likelihood of \( y \) given \( X \) is

\[
f(y \mid X; \beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left[ -\frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta) \right]
\]
Maximum Likelihood

- Likelihood Function
  \[ L(\beta, \sigma^2) = f(y|X; \beta, \sigma^2) \]

- Log Likelihood Function
  \[
  \log L(\beta, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta)
  \]
Maximum Likelihood

- ML estimator of \((\beta, \sigma^2)\)
  \[= \operatorname{argmax}_{(\beta, \gamma)} \log L(\beta, \gamma)\), where we set \(\gamma = \sigma^2\)

\[
\frac{\partial \log L(\beta, \gamma)}{\partial \beta} = -\frac{1}{2\gamma} \frac{\partial \text{SSR} (\beta)}{\partial \beta} = 0
\]

\[
\frac{\partial \log L(\beta, \gamma)}{\partial \gamma} = -\frac{n}{2\gamma} + \frac{1}{2\gamma^2} \text{SSR} (\beta) = 0
\]
Maximum Likelihood

• Suppose Assumptions 1-5 hold. Then the ML estimator of $\beta$ is the OLS estimator $b$ and ML estimator of $\gamma$ or $\sigma^2$ is

$$\frac{SSR}{n} = \frac{e'e}{n} = \frac{n - K}{n} s^2$$
Maximum Likelihood

- Maximum Likelihood Principle
  - Let $\theta = (\beta, \gamma)$
  - Score: $s(\theta) = \partial \log L(\theta) / \partial \theta$
  - Information Matrix: $I(\theta) = E(s(\theta)s(\theta)'|X)$
  - Information Matrix Equality:

$$I(\theta) = -E\left[ \frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'} \right] = \begin{bmatrix} \frac{1}{\sigma^2} X'X & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}$$
Maximum Likelihood

• Maximum Likelihood Principle
  – Cramer-Rao Bound: $I(\theta)^{-1}$
    That is, for an unbiased estimator of $\theta$ with a finite variance-covariance matrix:

$$\text{Var}(\hat{\theta}) \geq I(\theta)^{-1} = \begin{bmatrix} \sigma^2 (X'X)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix}$$
Maximum Likelihood

• Under Assumptions 1-5, the ML or OLS estimator $b$ of $\beta$ with variance $\sigma^2(X'X)^{-1}$ attains the Cramer-Rao bound.

• ML estimator of $\sigma^2$ is biased, so the Cramer-Rao bound does not apply.

• OLS estimator of $\sigma^2$, $s^2 = e'e/(n-K)$ with $E(s^2|X) = \sigma^2$ and $\text{Var}(s^2|X) = 2\sigma^4/(n-K)$, does not attain the Cramer-Rao bound $2\sigma^4/n$. 
Discussions

- Concentrated Log Likelihood Function

\[ \log L_c(\beta) = \log L(\beta, \text{SSR}(\beta)/n) \]

\[ = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\text{SSR}(\beta)/n) - \frac{n}{2} \]

\[ = -\frac{n}{2} \left[ \log(2\pi/n) + 1 \right] - \frac{n}{2} \log(\text{SSR}(\beta)) \]

- Therefore, \( \arg\max_\beta \log L(\beta) = \arg\min_\beta \text{SSR}(\beta) \)
Discussions

• Hypothesis Testing $H_0: R\beta = q$
  
  – Likelihood Ratio Test
  \[
  \lambda = \frac{L_{UR}}{L_R} = \left( \frac{SSR_R}{SSR_{UR}} \right)^{n/2}
  \]
  
  – F Test as a Likelihood Ratio Test
  \[
  F = \frac{(SSR_R - SSR_{UR})/J}{SSR_{UR}/(n-K)}
  = \frac{n-K}{J} \left( \frac{SSR_R}{SSR_{UR}} - 1 \right) = \frac{n-K}{J} \left( \frac{\lambda^{2/n}}{n} - 1 \right)
  \]
Discussions

• Quasi-Maximum Likelihood
  – Without normality (Assumption 5), there is no guarantee that ML estimator of $\beta$ is OLS or that the OLS estimator $b$ achieves the Cramer-Rao bound.
  – However, $b$ is a quasi- (or pseudo-) maximum likelihood estimator, an estimator that maximizes a misspecified (normal) likelihood function.
Generalized Least Squares

• Assumption 4 Revisited:
  \( E(\mathbf{e}'\mathbf{e}|\mathbf{X}) = \text{Var}(\mathbf{e}|\mathbf{X}) = \sigma^2 \mathbf{I}_n \)

• Assumption 4 Relaxed (Assumption 4’):
  \( E(\mathbf{e}'\mathbf{e}|\mathbf{X}) = \text{Var}(\mathbf{e}|\mathbf{X}) = \sigma^2 \mathbf{V}(\mathbf{X}) \), with nonsingular and known \( \mathbf{V}(\mathbf{X}) \).
  – OLS estimator of \( \mathbf{b} \), \( \mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \), is not efficient although it is still unbiased.
  – t-test and F-test are no longer valid.
Generalized Least Squares

• Since $V = V(X)$ is known, $V^{-1} = C' C$
• Let $y^* = Cy$, $X^* = CX$, $\varepsilon^* = C\varepsilon$
• $y = X\beta + \varepsilon \Rightarrow y^* = X^*\beta + \varepsilon^*$
  - Checking A.2: $E(\varepsilon^*|X^*) = E(\varepsilon^*|X) = 0$
  - Checking A.4: $E(\varepsilon^*\varepsilon^*'|X^*) = E(\varepsilon^*\varepsilon^*'|X) = \sigma^2CVC' = \sigma^2I_n$
• GLS: OLS for the transformed model $y^* = X^*\beta + \varepsilon^*$
Generalized Least Squares

- \( \mathbf{b}_{\text{GLS}} = (\mathbf{X}'\mathbf{X}^*)^{-1}\mathbf{X}'\mathbf{y}^* = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \)
- \( \text{Var}(\mathbf{b}_{\text{GLS}}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X}^*)^{-1} = \sigma^2 (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \)
- If \( \mathbf{V} = \mathbf{V}(\mathbf{X}) = \text{Var}(\mathbf{e}|\mathbf{X})/\sigma^2 \) is known,
  - \( \mathbf{b}_{\text{GLS}} = (\mathbf{X}'[\text{Var}(\mathbf{e}|\mathbf{X})]^{-1}\mathbf{X})^{-1}\mathbf{X}'[\text{Var}(\mathbf{e}|\mathbf{X})]^{-1}\mathbf{y} \)
  - \( \text{Var}(\mathbf{b}_{\text{GLS}}|\mathbf{X}) = (\mathbf{X}'[\text{Var}(\mathbf{e}|\mathbf{X})]^{-1}\mathbf{X})^{-1} \)
  - GLS estimator \( \mathbf{b}_{\text{GLS}} \) of \( \mathbf{\beta} \) is BLUE.
Generalized Least Squares

- Under Assumption 1-3, $E(b_{GLS}|X) = \beta$.
- Under Assumption 1-3, and 4’, $\text{Var}(b_{GLS}|X) = \sigma^2 (X'V(X)^{-1}X)^{-1}$
- Under Assumption 1-3, and 4’, the GLS estimator is efficient in that the conditional variance of any unbiased estimator that is linear in $y$ is greater than or equal to $[\text{Var}(b_{GLS}|X)]$. 
Discussions

• Weighted Least Squares (WLS)
  – Assumption 4”: \( V(X) \) is a diagonal matrix, or
    \[
    E(\varepsilon_i^2|X) = \text{Var}(\varepsilon_i|X) = \sigma^2 v_i(X) \]

    Then
    \[
    y_i^* = \frac{y_i}{\sqrt{v_i(X)}}, \quad x_i^* = \frac{x_i}{\sqrt{v_i(X)}}
    \]
    \((i = 1, 2, \ldots, n)\)
  – WLS is a special case of GLS.
Discussions

• If $V = V(X)$ is not known, we can estimate its functional form from the sample. This approach is called the **Feasible GLS**. $V$ becomes a random variable, then very little is known about the distribution and finite sample properties of the GLS estimator.
Example

• Cobb-Douglas Cost Function for Electricity Generation (Christensen and Greene [1976])
• Data: Greene’s Table F4.3
  – Id = Observation, 123 + 35 holding companies
  – Year = 1970 for all observations
  – Cost = Total cost,
  – Q = Total output,
  – Pl = Wage rate,
  – Sl = Cost share for labor,
  – Pk = Capital price index,
  – Sk = Cost share for capital,
  – Pf = Fuel price,
  – Sf = Cost share for fuel
Example

- **Cobb-Douglas Cost Function for Electricity Generation (Christensen and Greene [1976])**
  
  \[ \ln(\text{Cost}) = \beta_1 + \beta_2 \ln(PL) + \beta_3 \ln(PK) + \beta_4 \ln(PF) + \beta_5 \ln(Q) + \frac{1}{2} \beta_6 \ln(Q)^2 + \beta_7 \ln(Q) \ln(PL) + \beta_8 \ln(Q) \ln(PK) + \beta_9 \ln(Q) \ln(PF) + \varepsilon \]

- **Linear Homogeneity in Prices:**
  
  - \( \beta_2 + \beta_3 + \beta_4 = 1, \beta_7 + \beta_8 + \beta_9 = 0 \)

- **Imposing Restrictions:**
  
  - \( \ln(\text{Cost}/PF) = \beta_1 + \beta_2 \ln(PL/PF) + \beta_3 \ln(PK/PF) + \beta_5 \ln(Q) + \frac{1}{2} \beta_6 \ln(Q)^2 + \beta_7 \ln(Q) \ln(PL/PF) + \beta_8 \ln(Q) \ln(PK/PF) + \varepsilon \)