Microeconometrics

The Linear Model

by Vanessa Berenguer-Rico

University of Oxford

Michaelmas Term 2016

Outline

The Linear Model

- The Simple Regression Model (review)
- Multiple Regression under Classical Assumptions
- Relaxing Classical Assumptions

The Simple Regression Model

"The simple regression model can be used to study the relationship between two variables [...] It has limitations as a general tool for empirical analysis. Nevertheless it is sometimes appropriate as an empirical tool. (Moreover) Learning how to interpret the simple regression model is good practice for studying multiple regression," Wooldridge (2013) p. 20

The Simple Regression Model

Examples (Stock and Watson, 2012, p. 149)

- A state implements tough new penalties on drunk drivers: What is the effect on highway fatalities?
- A school district cuts the size of its elementary school classes: What is the effect on its students' standardized test scores?
- You successfully complete one more year of college classes: What is the effect on your future earnings?

The Simple Regression Model

$$y_i = \beta_1 + \beta_2 x_i + u_i,$$

- y_i and x_i are observable random scalars
- u_i is the unobservable random disturbance or error
- β_1 and β_2 are the parameters (constants) we would like to estimate

The Simple Regression Model: OLS

- The OLS objective function

$$\min_{b \in \mathbb{R}^2} \sum_{i=1}^{n} u_i^2 = \min_{b \in \mathbb{R}^2} \sum_{i=1}^{n} (y_i - b_1 - b_2 x_i)^2 = L$$

- System of Normal Equations: First Order Conditions

$$\begin{aligned} \frac{\partial L}{\partial b_1} &= -2 \sum_{i=1}^{n} (y_i - b_1 - b_2 x_i) = 0 \\ \frac{\partial L}{\partial b_2} &= -2 \sum_{i=1}^{n} (y_i - b_1 - b_2 x_i)\, x_i = 0 \end{aligned}$$

The Simple Regression Model: OLS

- The OLS solution

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{n^{-1}\sum_{i=1}^{n} y_i x_i - \bar{y}\bar{x}}{n^{-1}\sum_{i=1}^{n} x_i^2 - \bar{x}^2}$$

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

The Simple Regression Model: OLS

- Example: Wage and Education (Wooldridge, 2013, p. 31)

$$\widehat{wage}_i = -0.90 + 0.54 educ_i$$

- Interpreting estimates (caution!)

The Simple Regression Model: OLS

OLS algebraic facts: Let $\hat{u}_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i$

1. $$\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i = 0$$

2. $$\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i x_i = 0$$

Recall system of normal equations

The Simple Regression Model: OLS

OLS algebraic facts: Let $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$

3.

$$\frac{1}{n}\sum_{i=1}^{n} \hat{y}_i = \bar{y}$$

4.

$$SST = SSE + SSR$$

where

$$SST \equiv \sum_{i=1}^{n} (y_i - \bar{y})^2 \quad SSE \equiv \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \quad SSR \equiv \sum_{i=1}^{n} u_i^2$$

The Simple Regression Model: OLS

OLS algebraic facts:

5. Goodness-of-Fit: Coefficient of Determination:

$$R^2 \equiv \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

6.

$$0 \leq R^2 \leq 1$$

7.

$$\sqrt{R^2} = r_{yx} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

The Simple Regression Model: OLS

OLS algebraic facts:

8.

$$\hat{\beta}_2 = \beta_2 + \frac{n^{-1}\sum_{i=1}^{n}(x_i - \bar{x})\,u_i}{n^{-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

9.

$$\hat{\beta}_1 = \beta_1 + \bar{u} - \left(\hat{\beta}_2 - \beta_2\right)\bar{x}$$

(Some) Classical Assumptions

A1.

$$E\left[u_i|x_i\right] = 0$$

Note: A1 implies $Cov\left(x_i, u_i\right) = 0$ (no the other way around) and $E\left[u_i\right] = 0$

A2. $\left(y_i, x_i\right)$ are independent and identically distributed (i.i.d.) across observations

A.3. Existence of moments: $0 < E\left(y_i^4\right) < \infty$ and $0 < E\left(x_i^4\right) < \infty$ (finite kurtosis)

A.4. Homoskedasticity: $var\left[u_i|x_i\right] = \sigma_u^2$

The Simple Regression Model: OLS Properties

OLS Properties under (Some) Classical Assumptions

1. Expectations:

$$E\left[\hat{\beta}_1\right] = \beta_1 \quad \text{and} \quad E\left[\hat{\beta}_2\right] = \beta_2$$

OLS is unbiased!

2. Conditional Variances

$$Var\left(\hat{\beta}_2 | x_i\right) = \frac{\sigma_u^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \quad \text{and} \quad Var\left(\hat{\beta}_1 | x_i\right) = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}\sigma_u^2$$

Simple Regression: Gauss-Markov Theorem

Gauss-Markov Conditions

(a)

$$E\left[u_i|x_1, ..., x_n\right] = 0$$

(b)

$$var\left[u_i|x_1, ..., x_n\right] = \sigma_u^2, \;\; 0 < \sigma_u^2 < \infty$$

(c)

$$E\left[u_i u_j|x_1, ..., x_n\right] = 0 \;\;\; i \neq j$$

Simple Regression: Gauss-Markov Theorem

Gauss-Markov Theorem for $\hat{\beta}_2$

Under condition (a), (b) and (c), the OLS estimator $\hat{\beta}_2$ is **BLUE** (**B**est (most efficient) **L**inear conditionally **U**nbiased **E**stimator)

(See Stock and Watson, p. 218)

Simple Regression: Gauss-Markov Theorem

Linear Conditionally Unbiased Estimators

- Linearity:

$$\tilde{\beta}_1 = \sum_{i=1}^{n} a_i Y_i$$

where the weights a_i can depend on x_i but not on y_i

- Conditionally Unbiased:

$$E\left[\tilde{\beta}_1 | x_1, ..., x_n\right] = \beta_1$$

The Multiple Regression Model

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + ... + \beta_K x_{Ki} + u_i$$

Example:

$$wage_i = \beta_1 + \beta_2 educ_i + \beta_3 \text{exper}_i + u_i$$

Multiple Regression: OLS

- The OLS objective function:

$$L = \min_{b \in \mathbb{R}^K} \sum_{i=1}^{n} u_i^2$$

- In this case:

$$L = \min_{b \in \mathbb{R}^K} \sum_{i=1}^{n} (y_i - b_1 x_{1i} - b_2 x_{2i} - b_3 x_{3i} - ... - b_K x_{Ki})^2$$

Multiple Regression: OLS

System of Normal Equations: First Order Conditions

$$\begin{aligned}
\frac{\partial L}{\partial b_1} &= -2\sum_{i=1}^{n}\left(y_i - b_1 x_{1i} - b_2 x_{2i} - b_3 x_{3i} - \ldots - b_K x_{Ki}\right) x_{1i} = 0 \\
\frac{\partial L}{\partial b_2} &= -2\sum_{i=1}^{n}\left(y_i - b_1 x_{1i} - b_2 x_{2i} - b_3 x_{3i} - \ldots - b_K x_{Ki}\right) x_{2i} = 0 \\
&\vdots \\
\frac{\partial L}{\partial b_K} &= -2\sum_{i=1}^{n}\left(y_i - b_1 x_{1i} - b_2 x_{2i} - b_3 x_{3i} - \ldots - b_K x_{Ki}\right) x_{Ki} = 0
\end{aligned}$$

Multiple Regression: OLS

Too Long: Matrix Notation!

$$Y = X\beta + U$$

where

$$\begin{aligned} Y &= \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{K1} \\ x_{12} & x_{22} & \cdots & x_{K2} \\ x_{13} & x_{23} & \cdots & x_{K3} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{Kn} \end{pmatrix}, U = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{pmatrix} \\ \beta &= \begin{pmatrix} \beta_1 & \beta_2 & \cdots & \beta_K \end{pmatrix}' \end{aligned}$$

Multiple Regression: OLS

- The OLS objective function: $U(b) = Y - Xb$

$$
\begin{aligned}
L &= \min_{b \in \mathbb{R}^k} U(b)' U(b) \\
&= \min_{b \in \mathbb{R}^k} \left[(Y - Xb)' (Y - Xb) \right] \\
&= \min_{b \in \mathbb{R}^k} \left[Y'Y - Y'Xb - b'X'Y + b'X'Xb \right] \\
&= \min_{b \in \mathbb{R}^k} \left[Y'Y - 2Y'Xb + b'X'Xb \right]
\end{aligned}
$$

Multiple Regression: OLS

- **System of Normal Equations**: First Order Conditions

$$\frac{\partial L}{\partial b} = -2X'Y + 2X'Xb = 0$$

- OLS estimator

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

Multiple Regression: OLS

Note that OLS can also be written as

$$\begin{aligned} \hat{\beta} &= \left(X'X\right)^{-1}\left(X'Y\right) \\ &= \left(n^{-1}\sum_{i=1}^{n} X_i'X_i\right)^{-1}\left(n^{-1}\sum_{i=1}^{n} X_i'Y_i\right) \quad [\textit{Wooldridge}] \\ &= \left(n^{-1}\sum_{i=1}^{n} \tilde{X}_i\tilde{X}_i'\right)^{-1}\left(n^{-1}\sum_{i=1}^{n} \tilde{X}_iY_i\right) \quad [\textit{Hayashi}] \end{aligned}$$

where X_i is the ith row in X, $\tilde{X}_i = X_i'$, and Y_i the ith element in Y.

Multiple Regression: Classical Assumptions

Greene (2012) p. 56 and p. 92

A1. Linear in parameters
A2. Full rank
A3. Regression (Conditional Expectation)
A.4. Homoscedasticity and nonautocorrelation
A.5. Fixed or Random regressors
A.6. Normality

Multiple Regression: Classical Assumptions

A1. Linear in parameters: The model specifies a linear relationship between y and $x_1, x_2, ..., x_K$:

$$Y = X\beta + U$$

A2. Full rank: There is no exact linear relationship among any of the regressors in the model:

$$X \text{ is an } n \times K \text{ matrix with rank } K$$

This assumption is also known as the identification condition.

A3. Regression (Conditional Expectation): The regressors do not carry useful information for prediction of u_i:

$$E\left[u_i | X\right] = 0$$

Multiple Regression: Classical Assumptions

A.4. **Homoscedasticity and nonautocorrelation**: Each disturbance, u_i, has the same finite variance, σ_U^2, and is uncorrelated with every other disturbance, u_j:

$$\begin{aligned} Var\left[u_i|X\right] &= \sigma_U^2 \quad \text{for all } i = 1, ..., n \\ Cov\left[u_i u_j|X\right] &= 0 \quad \text{for all } i \neq j \end{aligned}$$

In matrix form,

$$E\left[UU'|X\right] = \sigma_U^2 I$$

Multiple Regression: Classical Assumptions

A.5. **Fixed or Random regressors**: The data in $(x_{1i}, ..., x_{iK})$ may be a mixture of constants and random variables:

$$X \text{ may be fixed or random}$$

A.6. **Normality**: The disturbances are normally distributed:

$$U|X \sim N\left[0, \sigma_U^2 I\right]$$

OLS: Finite Sample Properties under Classical Assumptions

Expectation:

$$
\begin{aligned}
E\left[\hat{\beta}\right] &= E\left[E\left[\hat{\beta}|X\right]\right] \\
&= E\left[E\left[\beta + \left(X'X\right)^{-1}\left(X'U\right)|X\right]\right] \\
&= E\left[\beta + \left(X'X\right)^{-1}X'E\left[U|X\right]\right] \\
&= \beta
\end{aligned}
$$

Variance:

$$
\begin{aligned}
Var\left[\hat{\beta}|X\right] &= E\left[\left(\hat{\beta}-\beta\right)\left(\hat{\beta}-\beta\right)'|X\right] \\
&= E\left[\left(X'X\right)^{-1}X'UU'X\left(X'X\right)^{-1}|X\right] \\
&= \left(X'X\right)^{-1}X'E\left[UU'|X\right]X\left(X'X\right)^{-1} \\
&= \sigma_U^2\left(X'X\right)^{-1}
\end{aligned}
$$

Multiple Regression: Gauss-Markov

Gauss-Markov Theorem for $\hat{\beta}$

Under the classical assumptions A1-A5, the OLS estimator $\hat{\beta}$ is **BLUE** (**B**est (most efficient) **L**inear conditionally **U**nbiased **E**stimator)

Multiple Regression: The Normality Assumption

- The Gauss-Markov theorem is derived without using the normality assumption, A6
- Normality is useful to perform exact inference in finite samples since, under A1-A6,

$$\hat{\beta}|X \sim N\left(\beta, \sigma_U^2 \left(X'X\right)^{-1}\right),$$

i.e. $\hat{\beta}|X$ is multivariate normal

Multiple Regression: The Normality Assumption

- Hence,

$$\frac{\hat{\beta}_k - \beta_k}{sd\left(\hat{\beta}_k\right)} \sim N\left(0, 1\right) \quad \text{and} \quad \frac{\hat{\beta}_k - \beta_k}{se\left(\hat{\beta}_k\right)} \sim t_{n-K},$$

where

$$sd\left(\hat{\beta}_k\right) = \sqrt{\sigma_U^2 \left(X'X\right)_{kk}^{-1}} \quad \text{and} \quad se\left(\hat{\beta}_k\right) = \sqrt{\hat{\sigma}_U^2 \left(X'X\right)_{kk}^{-1}}$$

and

$$\hat{\sigma}_U^2 = \frac{\hat{U}'\hat{U}}{n-K}$$

Multiple Regression: The Normality Assumption

Testing a hypothesis about a coefficient
The Three Ingredients

- The hypothesis:

$$\begin{aligned} H_o &: \beta_k = \beta_k^o \\ H_a &: \beta_k \neq \beta_k^o \end{aligned}$$

- The statistic

$$t = \frac{\hat{\beta}_k - \beta_k^o}{se\left(\hat{\beta}_k\right)}$$

- Decision Rule

$$t \sim t_{n-K} : RH_o \quad \text{if} \quad |t| > t_{\alpha, n-K}$$

Multiple Regression: The Normality Assumption

Example: Wooldridge (2013) p. 121

$$\begin{aligned} \widehat{\text{col}\,GPA} &= \underset{(0.33)}{1.39} + \underset{(0.094)}{0.412}hsGPA + \underset{(0.011)}{0.015}ACT - \underset{(0.026)}{0.083}skipped \\ n &= 141, \;\; R^2 = 0.234 \end{aligned}$$

where:

- col GPA denotes college grade point average
- *hsGPA* denotes high school GPA
- *ACT* achievement test score (to measure skills and knowledge)
- *skipped* average number of lectures missed per week

Multiple Regression: The Normality Assumption

Statistical Significance Test

- Hypothesis

$$H_o : \beta_k = 0 \quad \text{vs} \quad H_a : \beta_k \neq 0$$

- t-statistic

$$\widehat{\text{col}\,GPA} = \underset{(0.33)}{1.39} + \underset{(4.38)}{0.412} hsGPA + \underset{(1.36)}{0.015} ACT - \underset{(-3.19)}{0.083}\, skipped$$

- Decision Rule:

$$t_{(\alpha/2=.1/2);137} = 1.64;\ t_{(\alpha/2=.05/2);137} = 1.96;\ t_{(\alpha/2=.01/2);137} = 2.57$$

Multiple Regression: The Normality Assumption

Other types of hypothesis

- One coefficient is one:

$$
\begin{aligned}
H_o &: \beta_2 = 1 \\
H_a &: \beta_2 \neq 1
\end{aligned}
$$

- Two coefficients are equal:

$$
\begin{aligned}
H_o &: \beta_2 = \beta_3 \\
H_a &: \beta_2 \neq \beta_3
\end{aligned}
$$

- A set of coefficients add up to one:

$$
\begin{aligned}
H_o &: \beta_2 + \beta_3 + \beta_4 = 1 \\
H_a &: \beta_2 + \beta_3 + \beta_4 \neq 1
\end{aligned}
$$

Multiple Regression: The Normality Assumption

Other types of hypothesis:

- More than one hypothesis

$$
\begin{aligned}
H_o &: \beta_1 + \beta_2 = 1 \\
H_a &: \beta_1 + \beta_2 \neq 1
\end{aligned}
$$

and

$$
\begin{aligned}
H_o &: \beta_3 = \beta_4 \\
H_a &: \beta_3 \neq \beta_4
\end{aligned}
$$

Testing General Linear Restrictions
(Greene, p. 153)

- Extended Form

$$\begin{aligned} r_{11}\beta_1 + r_{12}\beta_2 + ... + r_{1k}\beta_K &= q_1 \\ r_{21}\beta_1 + r_{22}\beta_2 + ... + r_{2k}\beta_K &= q_2 \\ &\dots \\ r_{J1}\beta_1 + r_{J2}\beta_2 + ... + r_{Jk}\beta_K &= q_J \end{aligned}$$

- Matrix Form

$$R\beta = q$$

Testing General Linear Restrictions: Examples
(Greene, p. 153)

- One coefficient is zero: $\beta_j = 0$

$$R = \begin{bmatrix} 0 & 0 & ... & 0 & 1 & 0 & ... & 0 \end{bmatrix} \quad \text{and} \quad q = 0$$

- Two coefficients are equal: $\beta_i = \beta_j$

$$R = \begin{bmatrix} 0 & 0 & 1 & ... & -1 & ... & 0 \end{bmatrix} \quad \text{and} \quad q = 0$$

- A set of coefficients add up to one: $\beta_2 + \beta_3 + \beta_4 = 1$

$$R = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & ... & 0 \end{bmatrix} \quad \text{and} \quad q = 1$$

Multiple Regression: The Normality Assumption

Wald Test
(Greene, p. 157)

- Hypothesis

$$H_o : R\beta - q = 0 \quad \text{vs} \quad H_a : R\beta - q \neq 0$$

- Wald statistic (based on the discrepancy $m = R\hat{\beta} - q$):

$$\begin{aligned} W &= m' \left[var\left[m|X\right]\right]^{-1} m \\ &= \left(R\hat{\beta} - q\right)' \left[\sigma_U^2 R \left(X'X\right)^{-1} R'\right]^{-1} \left(R\hat{\beta} - q\right) \end{aligned}$$

- Under the null hypothesis (by the normality assumption)

$$W \sim \chi_J^2$$

Multiple Regression: The Normality Assumption

- Wald statistic is not feasible since σ_U^2 is unknown. We could estimate it in practice and use $\hat{\sigma}_U^2$.
- Let

$$
\begin{aligned}
F &= \frac{W}{J}\frac{\sigma_U^2}{\hat{\sigma}_U^2} \\
&= \frac{\left(R\hat{\beta}-q\right)'\left[R\sigma_U^2\left(X'X\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}-q\right)/J}{\left[\left(n-K\right)\hat{\sigma}_U^2/\sigma_U^2\right]/\left(n-K\right)} \\
&= \frac{\left(R\hat{\beta}-q\right)'\left[R\hat{\sigma}_U^2\left(X'X\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}-q\right)}{J}
\end{aligned}
$$

- Therefore, under the null hypothesis (by the normality assumption)

$$F \sim F_{J,n-K}$$

- Decision Rule: Reject the null if $F > F_{J,n-K}$

Multiple Regression: The Normality Assumption

Example

- The model

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

- The hypothesis

$$\begin{aligned} H_o &: \beta_2 = 0 \\ H_a &: \beta_2 \neq 0 \end{aligned}$$

- In this case, it can be written as

$$\begin{aligned} H_o &: R\beta = q \\ H_a &: R\beta \neq q \end{aligned}$$

where $\beta = \begin{pmatrix} \beta_1 & \beta_2 & \beta_3 \end{pmatrix}'$ is the 3×1 vector of parameters, $R = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$ is a 1×3 vector of zeros and 1 at position 2 and $q = 0$.

Multiple Regression: The Normality Assumption

- Therefore, $R\hat{\beta} - q = \hat{\beta}_j$ and since there is only one restriction, $J = 1$. Hence

$$
\begin{aligned}
F &= \frac{\left(R\hat{\beta} - q\right)' \left[R\hat{\sigma}^2_U \left(X'X\right)^{-1} R'\right]^{-1} \left(R\hat{\beta} - q\right)}{J} \\
&= \frac{\hat{\beta}_j \left[R\hat{\sigma}^2_U \left(X'X\right)^{-1} R'\right]^{-1} \hat{\beta}_j}{1}
\end{aligned}
$$

Multiple Regression: The Normality Assumption

- What is $R\hat{\sigma}_U^2 \left(X'X\right)^{-1} R'$?
- Recall that $\widehat{AsyVar}\left(\hat{\beta}\right) = \hat{\sigma}_U^2 \left(X'X\right)^{-1}$, which is

$$\begin{pmatrix} \widehat{AsyVar}\left(\hat{\beta}_1\right) & \widehat{AsyCov}\left(\hat{\beta}_1, \hat{\beta}_2\right) & \widehat{AsyCov}\left(\hat{\beta}_1, \hat{\beta}_3\right) \\ \widehat{AsyCov}\left(\hat{\beta}_1, \hat{\beta}_2\right) & \widehat{AsyVar}\left(\hat{\beta}_2\right) & \widehat{AsyCov}\left(\hat{\beta}_2, \hat{\beta}_3\right) \\ \widehat{AsyCov}\left(\hat{\beta}_1, \hat{\beta}_3\right) & \widehat{AsyCov}\left(\hat{\beta}_2, \hat{\beta}_3\right) & \widehat{AsyVar}\left(\hat{\beta}_3\right) \end{pmatrix}$$

- Hence: $R\hat{\sigma}_U^2 \left(X'X\right)^{-1} R' = \widehat{AsyVar}\left(\hat{\beta}_2\right)$

Multiple Regression: The Normality Assumption

- Therefore,

$$
\begin{aligned}
F &= \frac{\left(R\hat{\beta}-q\right)'\left[R\hat{\sigma}_U^2\left(X'X\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}-q\right)}{J} \\
&= \left(\frac{\hat{\beta}_2}{\widehat{AsySD}\left(\hat{\beta}_2\right)}\right)^2 \\
&= \left(\frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}_U^2\left(X'X\right)_{22}^{-1}}}\right)^2 \\
&= t^2.
\end{aligned}
$$

- Recall, under normality: $F \sim F_{J,n-K}$. In this case, $F \sim F_{1,n-3}$. Decision Rule: Reject the null if $F > F_{\alpha;1,n-3}$

Multiple Regression: The Normality Assumption

For a linear restriction:

$$H_o : a_1\beta_1 + a_2\beta_2 + ... + a_K\beta_K = q$$

with a_i $i = 1, ..., K$ and q known, the t-statistic is

$$t = \frac{a_1\hat{\beta}_1 + a_2\hat{\beta}_2 + ... + a_K\hat{\beta}_K - q}{\widehat{AsySD}\left(a_1\hat{\beta}_1 + a_2\hat{\beta}_2 + ... + a_K\hat{\beta}_K\right)} \sim t_{n-K}$$

Example: For

$$H_o : \beta_1 + \beta_2 = 1$$

the statistic is

$$t = \frac{\hat{\beta}_1 + \hat{\beta}_2 - 1}{\widehat{AsySD}\left(\hat{\beta}_1 + \hat{\beta}_2\right)} \sim t_{n-K}$$

Decision Rule: Reject the null if $|t| > t_{\alpha/2;n-K}$

Multiple Regression: Asymptotics

- Unbiasedness, the Gauss-Markov theorem, and the distributions described above are all finite sample properties (Greene p.103 for a review)
- Next, we will study the properties of the OLS estimator and related statistics when the sample size grows ($n \to \infty$)
- Normality is not essential for the asymptotic analysis

Multiple Regression: Asymptotics

Consistency
(Greene p. 103 and Wooldridge (2010), p. 56)

C1. Linear in parameters

$$Y = X\beta + U$$

C2. Independence: (X_i, u_i) $i = 1, ..., n$ is a sequence of independent observations and $E[U|X] = 0$

C3. Rank condition:

$$Q = p \lim_{n \to \infty} \frac{X'X}{n}$$

has (full) rank K.

Multiple Regression: Asymptotics

Consistency
(Greene p. 103 and Wooldridge (2010), p. 56)

Theorem: Under Assumptions C1-C3,

$$\hat{\beta} \xrightarrow{p} \beta.$$

Multiple Regression: Asymptotics

Proof: Given C1, the OLS estimator $\hat{\beta}$ can be written as

$$\hat{\beta} = \beta + \left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'U}{n}\right).$$

By C3 and the Slutsky theorem,

$$\left(\frac{X'X}{n}\right)^{-1} \xrightarrow{p} Q^{-1}.$$

By C2 and the Law of Large Numbers,

$$\left(\frac{X'U}{n}\right) \xrightarrow{p} 0.$$

Multiple Regression: Asymptotics

Asymptotic Normality
(Greene p. 105 and Wooldridge (2010), p. 59)

AN1. Linear in parameters

$$Y = X\beta + U$$

AN2. Independence: (X_i, u_i) $i = 1, ..., n$ is a sequence of independent observations and $E[U|X] = 0$
AN3. Rank condition:

$$Q = \underset{n\to\infty}{p\lim} \frac{X'X}{n}$$

has (full) rank K
AN4. Homoskedasticity and No Autocorrelation:

$$E\left[UU'|X\right] = \sigma_U^2 I$$

Multiple Regression: Asymptotics

Asymptotic Normality
(Greene p. 105 and Wooldridge (2010), p. 59)

Theorem: Under Assumptions AN1-AN4,

$$\sqrt{n}\left(\hat{\beta}-\beta\right) \xrightarrow{d} N\left(0, \sigma_U^2 Q^{-1}\right).$$

Multiple Regression: Asymptotics

$AsyVar\left(\hat{\beta}_n\right)$ Estimation
Greene p. 107

By Theorem 1,

$$\sqrt{n}\left(\hat{\beta}-\beta\right) \overset{a}{\sim} N\left(0, \sigma_U^2 Q^{-1}\right)$$

and

$$\hat{\beta} \overset{a}{\sim} N\left(\beta, \frac{\sigma_U^2}{n} Q^{-1}\right).$$

See Greene, p. 1124.

Multiple Regression: Asymptotics

$AsyVar\left(\hat{\beta}_n\right)$ Estimation

Greene p. 107

Therefore

$$AsyVar\left(\hat{\beta}_n\right) = \frac{\sigma_U^2}{n} Q^{-1}.$$

This can be consistently estimated (see Greene p. 107) by

$$\widehat{AsyVar}\left(\hat{\beta}_n\right) = \hat{\sigma}_U^2 \left(X'X\right)^{-1},$$

where remember

$$\hat{\sigma}_U^2 = \frac{\hat{U}'\hat{U}}{n-K}.$$

Multiple Regression: Asymptotic Inference

"We know that normality plays no role in the unbiasedness of OLS, nor does it affect the conclusion that OLS is the best linear unbiased estimator under the Gauss-Markov assumptions. But exact inference based on t and F statistics requires [Normality of the error term]. Does this mean that, in our analysis (...), we must abandon the t statistic for determining which variables are statistically significant? Fortunately, the answer to this question is no,"
Wooldridge (2013) p. 167

Multiple Regression: Asymptotic Inference

- **Exact inference based on t and F: ONLY under the Normality Assumption**
- What can we do if Normality is not assumed?
- Asymptotic (large-sample) Tests via Central Limit Theorem (Wooldridge (2013) p. 165 and Greene p. 167)

Multiple Regression: Asymptotic Inference

Statistical Significance Test

Recall, the t-statistic for

$$\begin{aligned} H_o &: \beta_k = \beta_k^o \\ H_o &: \beta_k \neq \beta_k^o \end{aligned}$$

is

$$t = \frac{\hat{\beta}_k - \beta_k^o}{se\left(\hat{\beta}_k\right)}$$

where

$$se\left(\hat{\beta}_k\right) = \sqrt{\hat{\sigma}_U^2 \left(X'X\right)_{kk}^{-1}} \quad \text{and} \quad \hat{\sigma}_U^2 = \frac{\hat{U}'\hat{U}}{n-K}$$

Statistical Significance Test

The t-statistic can be rewritten as

$$t = \frac{\sqrt{n}\left(\hat{\beta}_k - \beta_k^o\right)}{\sqrt{\hat{\sigma}_U^2 \left(X'X/n\right)_{kk}^{-1}}},$$

hence under AD1-AD4, if $\hat{\sigma}_U^2$ is a consistent estimator of σ_U^2, under the null,

$$t = \frac{\sqrt{n}\left(\hat{\beta}_k - \beta_k^o\right)}{\sqrt{\hat{\sigma}_U^2 \left(X'X/n\right)_{kk}^{-1}}} \sim N\left(0,1\right).$$

Therefore, the decision rule is, for $\alpha = 0.05$,

$$RH_o \quad \text{if} \quad |t| > 1.96$$

Multiple Regression: Asymptotic Inference

Wald Test for Linear Hypothesis
Greene p. 169

Recall, the hypotheses are:

$$\begin{aligned} H_o &: R\beta - q = 0 \\ H_a &: R\beta - q \neq 0 \end{aligned}$$

The (large sample) Wald statistic is

$$W = \left(R\hat{\beta} - q\right)' \left[R\hat{\sigma}_U^2 \left(X'X\right)^{-1} R'\right]^{-1} \left(R\hat{\beta} - q\right) = JF.$$

If AD1-AD4 hold and $\hat{\sigma}_U^2$ is a consistent estimator of σ_U^2, then under the null,

$$W \xrightarrow{d} \chi_J^2.$$

Multiple Regression: Asymptotic Inference

Wage Equation

The specified model:

$$wage = \beta_1 + \beta_2 educ + \beta_3 tenure + u$$

The estimated model:

$$\widehat{wage} = \underset{(0.64)}{-2.22} + \underset{(0.04)}{0.56} educ + \underset{(0.01)}{0.18} tenure$$

Multiple Regression: Asymptotic Inference

Wage Equation

The estimated model:

$$\widehat{wage} = \underset{(0.64)}{-2.22162} + \underset{(0.04)}{0.56914}educ + \underset{(0.01)}{0.18958}tenure$$

Variance/Covariance matrix of $\hat{\beta}$

$$\widehat{AsyVar}\left(\hat{\beta}\right) = \hat{\sigma}_U^2 \left(X'X\right)^{-1} = \begin{pmatrix} 0.40979 & -0.03018 & -0.00243 \\ -0.03018 & 0.00238 & 0.00005 \\ -0.00243 & 0.00005 & 0.00034 \end{pmatrix}$$

Multiple Regression: Asymptotic Inference

Wage Equation

Consider the hypothesis

$$
\begin{aligned}
H_o &: \beta_2 = \beta_3 \\
H_a &: \beta_2 \neq \beta_3
\end{aligned}
$$

It can be written as

$$
\begin{aligned}
H_o &: \beta_2 - \beta_3 = 0 \\
H_a &: \beta_2 - \beta_3 = 0
\end{aligned}
$$

Multiple Regression: Asymptotic Inference

Hypothesis:

$$\begin{aligned} H_o &: \beta_2 - \beta_3 = 0 \\ H_a &: \beta_2 - \beta_3 = 0 \end{aligned}$$

Test Statistic

$$F = \frac{\left(R\hat{\beta} - q\right)' \left[R\hat{\sigma}_U^2 \left(X'X\right)^{-1} R'\right]^{-1} \left(R\hat{\beta} - q\right)}{J}$$

where

$$J = 1; \quad R = \left[\begin{array}{ccc} 0 & 1 & -1 \end{array}\right] \quad \text{and } q = 0$$

Multiple Regression: Asymptotic Inference

Therefore

$$
\begin{aligned}
F &= \frac{\left(R\hat{\beta}-q\right)'\left[R\hat{\sigma}_U^2\left(X'X\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}-q\right)}{J} \\
&= \frac{\left(\hat{\beta}_2-\hat{\beta}_3\right)^2}{\widehat{AsyVar}\left(\hat{\beta}_2\right)+\widehat{AsyVar}\left(\hat{\beta}_3\right)-2\widehat{AsyCov}\left(\hat{\beta}_2,\hat{\beta}_3\right)} \\
&= \left(\frac{\hat{\beta}_2-\hat{\beta}_3}{\sqrt{\widehat{AsyVar}\left(\hat{\beta}_2-\hat{\beta}_3\right)}}\right)^2 \\
&= t^2 \\
&= (7.415)^2 \\
&= 54.9822
\end{aligned}
$$

Multiple Regression: Asymptotic Inference

Decision Rule:

$$F \sim F_{J,n-K} = F_{1,523} \quad \text{and} \quad t \sim t_{n-K} = t_{523}$$

Assuming a 5% level of significance (α):

$$F_{\alpha;1,523} = 3.84 \quad \text{and} \quad t_{\alpha/2;523} = 1.96$$

We reject the null hypothesis.

Outline

Linear Model: Relaxing Classical Assumptions

- Heteroscedasticity: White test and White standard errors
- Measurement Error and Omitted Variables Problems: IV

Heteroscedasticity

Heteroscedasticity

Recall: Wage Equation
(Wooldridge data)

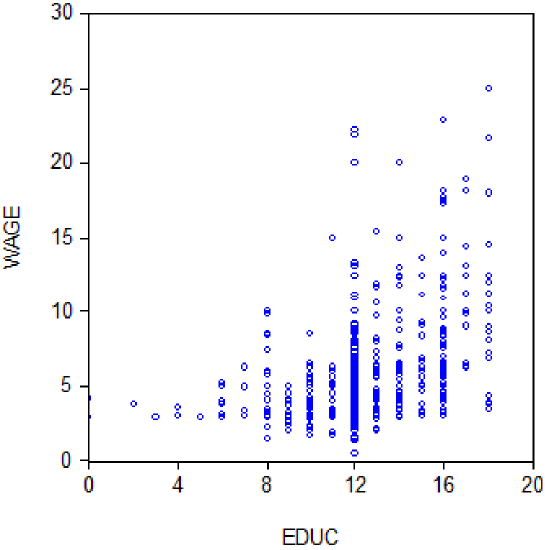
The estimated model:

$$\widehat{wage} = \underset{(0.64)}{-2.22162} + \underset{(0.04)}{0.56914} educ + \underset{(0.01)}{0.18958} tenure$$

Variance/Covariance matrix of $\hat{\beta}$

$$\widehat{AsyVar}\left(\hat{\beta}\right) = \hat{\sigma}_U^2 \left(X'X\right)^{-1} = \begin{pmatrix} 0.40979 & -0.03018 & -0.00243 \\ -0.03018 & 0.00238 & 0.00005 \\ -0.00243 & 0.00005 & 0.00034 \end{pmatrix}$$

Heteroscedasticity



Heteroscedasticity

- **Heteroscedasticity Example**: Stock and Watson, p. 199
- The Economic Value of a Year of Education

$$Earning = \beta_1 + \beta_2 YearsEducation$$

- Data: Hourly earnings and Years of education for 29- to 30-year olds in the US in 2008
- *"The spread of the distribution of earnings increases with the years of education. While some workers with many years of education have low-paying jobs, very few workers with low levels of education have high-paying jobs. (...) The variance of the residuals in the regression of Equation [above] depends on the regressor"*

Recall: Wage Equation
(Wooldridge data)

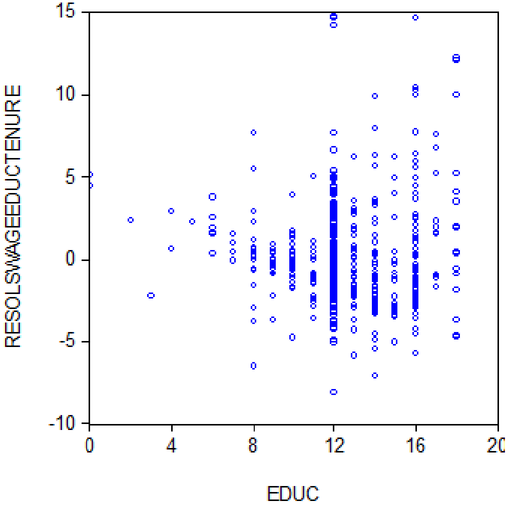
The estimated model:

$$\widehat{wage} = \underset{(0.64)}{-2.22162} + \underset{(0.04)}{0.56914}educ + \underset{(0.01)}{0.18958}tenure$$

Variance/Covariance matrix of $\hat{\beta}$

$$\widehat{AsyVar}\left(\hat{\beta}\right) = \hat{\sigma}_U^2 \left(X'X\right)^{-1} = \begin{pmatrix} 0.40979 & -0.03018 & -0.00243 \\ -0.03018 & 0.00238 & 0.00005 \\ -0.00243 & 0.00005 & 0.00034 \end{pmatrix}$$

Heteroscedasticity



Heteroscedasticity

Recall: Classical Assumptions

A1. Linear in parameters
A2. Full rank
A3. Regression (Conditional Expectation)
A.4. **Homoscedasticity** and nonautocorrelation
A.5. Fixed or Random regressors
A.6. Normality

Heteroscedasticity

A.4. **Homoscedasticity** and nonautocorrelation: **Each disturbance**, u_i, **has the same finite variance,** σ_U^2, and is uncorrelated with every other disturbance, u_j:

$$\begin{aligned} Var\left[u_i|X\right] &= \sigma_U^2 \quad \text{for all } i = 1, ..., n \\ Cov\left[u_i u_j|X\right] &= 0 \quad \text{for all } i \neq j \end{aligned}$$

In matrix form,

$$E\left[UU'|X\right] = \sigma_U^2 I$$

Heteroscedasticity

Generalized Linear Regression Model
Greene, p. 297

The Generalized Linear Regression model is

$$\begin{aligned} Y &= X\beta + U \\ E\left[U|X\right] &= 0 \\ E\left[UU'|X\right] &= \sigma_U^2 \Omega = \Sigma \end{aligned}$$

where Ω is a positive definite symmetric matrix.

Generalized Linear Regression Model

(Conditional) Variance/Covariance Matrix of the error term

$$
\begin{aligned}
E\left[UU'|X\right] &= E\left[\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}\begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix}|X\right] \\
&= E\left[\begin{pmatrix} u_1^2 & u_1u_2 & \cdots & u_1u_n \\ u_2u_1 & u_2^2 & \cdots & u_2u_n \\ \vdots & \vdots & \ddots & \\ u_nu_1 & u_nu_2 & \cdots & u_n^2 \end{pmatrix}|X\right]
\end{aligned}
$$

Heteroscedasticity

$$\begin{aligned}
E\left[UU'|X\right] &= E\left[\begin{pmatrix} u_1^2 & u_1u_2 & \cdots & u_1u_n \\ u_2u_1 & u_2^2 & \cdots & u_2u_n \\ \vdots & \vdots & \ddots & \\ u_nu_1 & u_nu_2 & \cdots & u_n^2 \end{pmatrix} |X\right] \\
&= \begin{pmatrix} E\left[u_1^2|X\right] & E\left[u_1u_2|X\right] & \cdots & E\left[u_1u_n|X\right] \\ E\left[u_2u_1|X\right] & E\left[u_2^2|X\right] & \cdots & E\left[u_2u_n|X\right] \\ \vdots & \vdots & \ddots & \\ E\left[u_nu_1|X\right] & E\left[u_nu_2|X\right] & \cdots & E\left[u_n^2|X\right] \end{pmatrix} \\
&= \sigma_U^2\Omega = \Sigma
\end{aligned}$$

where Ω is a positive definite symmetric matrix.

Heteroscedasticity

Generalized Linear Regression Model
Greene, p. 297

In AD3 above we assumed $\Omega = I$, that is,

$$\begin{aligned} Y &= X\beta + U \\ E\left[U|X\right] &= 0 \\ E\left[UU'|X\right] &= \sigma_U^2 I. \end{aligned}$$

The generalized linear regression model allows Ω to be different from the identity matrix.

Heteroscedasticity

- Heteroscedasticity:

$$\sigma_U^2 \Omega = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

- Typically (conditional) Heteroscedasticity:

$$\sigma_U^2 \Omega = \begin{bmatrix} \sigma_1^2(X) & 0 & \cdots & 0 \\ 0 & \sigma_2^2(X) & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \sigma_n^2(X) \end{bmatrix}$$

Heteroscedasticity

- Heteroscedasticity tends to be a more relevant problem when dealing with cross-sectional data
- Time Series, on the other hand, typically deals with autocorrelation and nonstationarities in the error term
- Definition (Stock and Watson, p. 198): The error term u_i given X is **homoscedastic** if the variance of the conditional distribution of u_i given X is constant for $i = 1, ..., n$ and in particular does not depend on X. Otherwise, the error term is **heteroscedastic**.

Generalized Linear Regression Model
Greene, p. 297

The Generalized Linear Regression model is

$$\begin{aligned} Y &= X\beta + U \\ E\left[U|X\right] &= 0 \\ E\left[UU'|X\right] &= \sigma_U^2\Omega = \Sigma \end{aligned}$$

where Ω is a positive definite matrix.

Heteroscedasticity

Finite Sample Consequences for OLS
Greene, p. 298

- In the Generalized model, we relax A4 allowing for heteroscedasticity and autocorrelation
- What are the **finite sample** consequences for OLS?

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

Heteroscedasticity

Finite Sample Consequences for OLS
Greene, p. 299

- Expectation:

$$E\left[\hat{\beta}\right] = E\left[E\left[\hat{\beta}|X\right]\right] = E\left[E\left[\beta + \left(X'X\right)^{-1}\left(X'U\right)|X\right]\right] = \beta$$

- Variance

$$\begin{aligned} Var\left[\hat{\beta}|X\right] &= E\left[\left(\hat{\beta}-\beta\right)\left(\hat{\beta}-\beta\right)'|X\right] \\ &= E\left[\left(X'X\right)^{-1}X'UU'X\left(X'X\right)^{-1}|X\right] \\ &= \left(X'X\right)^{-1}X'\left(\sigma_U^2\Omega\right)X\left(X'X\right)^{-1} \end{aligned}$$

- Normality

$$\hat{\beta}|X \sim N\left[\beta, \sigma_U^2\left(X'X\right)^{-1}X'\Omega X\left(X'X\right)^{-1}\right]$$

Heteroscedasticity

Finite Sample Consequences for OLS
Greene, p. 299

- OLS estimator is **unbiased** in the generalized regression model described above
- The conditional variance is not $\sigma_U^2 \left(X'X\right)^{-1}$ but $\sigma_U^2 \left(X'X\right)^{-1} X'\Omega X \left(X'X\right)^{-1}$
- Statistical inference based on $\hat{\sigma}_U^2 \left(X'X\right)^{-1}$ may be misleading
- $\hat{\sigma}_U^2$ biased

Large Sample Consequences for OLS
Greene, p. 299

- In the Generalized model, we relax A4 allowing for heteroscedasticity and autocorrelation
- What are the **large sample** consequences for OLS?

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

- Consistency and Asymptotic Normality affected???

Multiple Regression: Asymptotics

Recall: Consistency

(Greene p. 103 and Wooldridge (2010), p. 56)

C1. **Linear in parameters**

$$Y = X\beta + U$$

C2. **Independence**: (X_i, u_i) $i = 1, ..., n$ is a sequence of independent observations and $E[U|X] = 0$

C3. **Rank condition:**

$$Q = p\lim_{n\to\infty} \frac{X'X}{n}$$

has (full) rank k.

Heteroscedasticity

Large Sample Consequences for OLS
Greene, p. 299

- Consistency of the OLS estimator can be derived without stating an assumption about the variance covariance matrix of the errors. Hence, OLS will remain consistent in the Generalized model.
- This argument uses the strong LLN, i.e., almost sure convergence. We could also use a mean square error argument to prove consistency. In this case, the following limit should exist:

$$Q^* = p \lim_{n \to \infty} \frac{X'\Omega X}{n}.$$

(see Greene p. 300)

Multiple Regression

Recall: Asymptotic Normality
(Greene p. 105 and Wooldridge (2010), p. 59)

AN1. Linear in parameters

$$Y = X\beta + U$$

AN2. Independence: (X_i, u_i) $i = 1, ..., n$ is a sequence of independent observations and $E[U|X] = 0$
AN3. Rank condition:

$$Q = p\lim_{n\to\infty} \frac{X'X}{n}$$

has (full) rank K
AN4. Homoskedasticity and no autocorrelation:

$$E\left[UU'|X\right] = \sigma_U^2 I$$

Heteroscedasticity

Large Sample Consequences for OLS

- Condition AN4, imposes homoscedasticity and no autocorrelation to obtain (recall):

Theorem: Under Assumptions AN1-AN4,

$$\sqrt{n}\left(\hat{\beta}-\beta\right) \xrightarrow{d} N\left(0, \sigma_U^2 Q^{-1}\right).$$

- We should expect some impact of relaxing assumption AN4

Heteroscedasticity

Generalized Model: OLS Asymptotic Normality
(Greene p. 301)

GAN1. Linear in parameters

$$Y = X\beta + U \quad \text{and} \quad E\left[UU'|X\right] = \sigma_U^2\Omega = \Sigma$$

GAN2. Independence: (X_i, u_i) $i = 1, ..., n$ is a sequence of independent observations and $E[U|X] = 0$

GAN3. Rank condition:

$$Q = p\lim_{n\to\infty}\frac{X'X}{n}$$

has (full) rank K

GAN4. Heteroskedasticy and/or Autocorrelation:
$E\left[UU'|X\right] = \sigma_U^2\Omega = \Sigma$ and

$$Q^* = p\lim_{n\to\infty}\frac{X'\Omega X}{n}.$$

exists.

Large Sample Consequences for OLS

Under Assumptions GAN1-GAN4,

$$\sqrt{n}\left(\hat{\beta}-\beta\right) \xrightarrow{d} N\left(0, \sigma_U^2 Q^{-1} Q^* Q^{-1}\right),$$

where $Q^* = p\lim_{n\to\infty} X'\Omega X/n$.
Hence,

$$\hat{\beta} \overset{a}{\sim} N\left(\beta, \frac{\sigma_U^2}{n} Q^{-1} Q^* Q^{-1}\right).$$

See Greene, p. 1124.

Heteroscedasticity

Consequences for OLS
Greene p.299

- $\hat{\beta}$ is unbiased and consistent
- But
$$AsyVar\left(\hat{\beta}\right) = \frac{\sigma_U^2}{n} Q^{-1} Q^{*} Q^{-1}$$
- OLS is no longer BLUE
- Standard inference based on t and F tests is not appropriate

Heteroscedasticity

Recall: Wage Equation

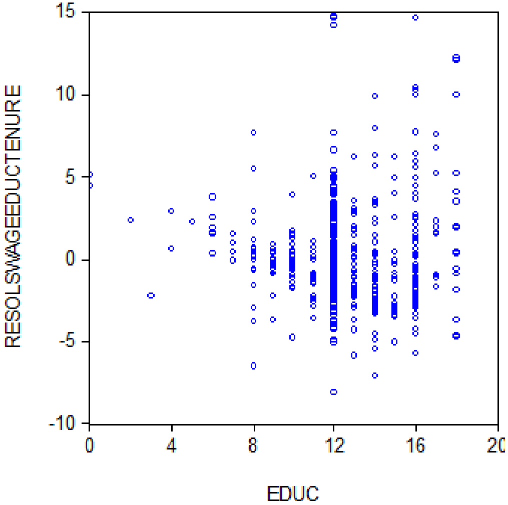
The estimated model:

$$\widehat{wage} = \underset{(0.64)}{-2.22162} + \underset{(0.04)}{0.56914}educ + \underset{(0.01)}{0.18958}tenure$$

Variance/Covariance matrix of $\hat{\beta}$

$$\widehat{AsyVar}\left(\hat{\beta}\right) = \hat{\sigma}_U^2 \left(X'X\right)^{-1} = \begin{pmatrix} 0.40979 & -0.03018 & -0.00243 \\ -0.03018 & 0.00238 & 0.00005 \\ -0.00243 & 0.00005 & 0.00034 \end{pmatrix}$$

Heteroscedasticity



Testing for Heteroscedasticity

How to asses statistically the presence of Heteroscedasticty?

- There are several statistical procedures to test for Heteroscedasticity
- See for instance: Wooldridge1, p. 265
- We will consider **White's test (1980)**
- "A Heteroskedastic-Covariance Matrix Estimator and a Direct Test for Heteroscedasticity," Econometrica, 50, 483-499

Testing for Heteroscedasticity: White's Test

White (1980) Test: Example: $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$

- **Step 1**: Estimate the model by OLS and compute the OLS residuals $\hat{u}_i = \hat{\beta}_1 - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i}$
- **Step 2**: Run the following regression:

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 x_{2i} + \alpha_3 x_{3i} + \gamma_2 x_{2i}^2 + \gamma_3 x_{Ki}^2 + \gamma_1 x_{2i} x_{3i} \qquad (1)$$

- **Step 3**: Test the null (homoscedasticity) hypothesis

$$H_o : \alpha_2 = \alpha_3 = \gamma_1 = \gamma_2 = \gamma_3 = 0$$

$$\text{Test Statistic: } nR^2 \sim \chi^2_{p-1}$$

where n, R^2 and p are the sample size, coefficient of determination and number of parameters of (1), respectively

Heteroscedasticity

Recall: Wage Equation

Step 1: Estimate the model

$$\widehat{wage}_i = \underset{(0.64)}{-2.22162} + \underset{(0.04)}{0.56914} educ_i + \underset{(0.01)}{0.18958} tenure_i$$

and obtain residuals:

$$\hat{u}_i = wage_i + 2.22162 - 0.56914 educ_i - 0.18958 tenure_i$$

Heteroscedasticity

Step 2: Auxiliary Regression:

$$\begin{aligned}\hat{u}_i^2 &= \underset{(11.39610)}{24.84743} - \underset{(1.765585)}{4.407294} educ - \underset{(0.660221)}{0.465109} tenure \\ &\quad - \underset{(0.012984)}{0.013566} tenure^2 + \underset{(0.071008)}{0.213336} educ^2 \\ &\quad + \underset{(0.045025)}{0.129456} educ * tenure\end{aligned}$$

with $n = 526$, $R^2 = 0.106035$.

Step 3: Test the null hypothesis of homoscedasticity

$$nR^2 = 55.77452 > \chi^2_{\alpha=0.05;5} \implies RH_o$$

Heteroscedasticity

Recall: Wage Equation

$$\widehat{wage} = \underset{(0.64)}{-2.22162} + \underset{(0.04)}{0.56914} educ + \underset{(0.01)}{0.18958} tenure$$

Recall: Heteroscedasticity consequences for OLS

Unbiased, Consistent but NOT Efficient

+

$\sigma_U^2 Q^{-1}$ **vs** $\sigma_U^2 Q^{-1} Q^* Q^{-1}$

Heteroscedasticity

What can we do in practice to account for heteroscedasticity?

- There exist several alternative ways to do so
- In this course, we will consider White (1980) proposal:
- Use OLS estimator with robust standard errors (HAC)
- "A Heteroskedastic-Covariance Matrix Estimator and a Direct Test for Heteroscedasticity," Econometrica, 50, 483-499

Heteroscedasticity

Some Proposed Solutions: HAC standard errors
(Greene, p. 302)

- Principle: Use OLS estimator with HAC standard errors.
- Recall, under heteroscedasticity and autocorrelation

$$AsyVar\left(\hat{\beta}\right) = \frac{\sigma_U^2}{n} Q^{-1} Q^{*} Q^{-1}$$

- HAC standard errors are based on the sample analog

Heteroscedasticity

Some Proposed Solutions: HAC standard errors

- Heteroscedasticity case (Greene, p. 313): White (1980) estimator

$$\widehat{AsyVar}\left(\hat{\beta}\right) = \frac{1}{n}\left[\frac{1}{n}X'X\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2 X_i X_i'\right]\left[\frac{1}{n}X'X\right]^{-1}$$

Heteroscedasticity

Wage Equation - OLS "standard" standard errors

$$\widehat{wage} = \underset{(0.64)}{-2.22162} + \underset{(0.04)}{0.56914} educ + \underset{(0.01)}{0.18958} tenure$$

Wage Equation - OLS "White" standard errors

$$\widehat{wage} = \underset{(0.72)}{-2.22162} + \underset{(0.05)}{0.56914} educ + \underset{(0.02)}{0.18958} tenure$$

Endogenity

Endogeneity: OLS Bias
Greene, p. 259

Examples:
1. Measurement errors
2. Omitted Variables
3. Simultaneous Equations
etc....

OLS under Measurement Errors

Wooldridge (2013) p.307 and Greene p. 137

"There are a number of cases in which observed data are imperfect measures of their theoretical counterparts in the regression model. Examples include income, education, ability, health, the interest rate, output, capital, and so on."

Greene, p. 137

Measurement Errors

OLS under Measurement Errors

Wooldridge (2013) p.307 and Greene p. 137

It is convenient to distinguish between:

- Measurement Error in the Dependent Variable
- Measurement Error in an Explanatory Variable

Measurement Errors

OLS under Measurement Errors in the Dependent Variable
Wooldridge (2013) p.307 and Greene p. 137

Let

$$y^* = \beta_1 + \beta_2 x_2 + ... + \beta_k x_k + u$$

and assume that the Gauss-Markov assumptions are satisfied. In addition, let y be the observable measure of y^* so that

$$y = y^* + e$$

where e is the measurement error.

Measurement Errors

OLS under Measurement Errors in the Dependent Variable
Wooldridge (2013) p.307 and Greene p. 137

In that case, since $y^* = y - e$, the model can be written as

$$y = \beta_1 + \beta_2 x_2 + ... + \beta_k x_k + u + e$$

Since we have assumed the Gauss-Markov conditions, the properties of the OLS estimator will depend on the properties of e.

Measurement Errors

OLS under Measurement Errors in the Dependent Variable
Wooldridge (2013) p.307 and Greene p. 137

If $e_i \sim i.i.d\left(0, \sigma_e^2\right)$ and is independent of the explanatory variables x_j, then:

- the OLS estimator is unbiased and consistent
- the usual OLS inference procedures (t, F) are valid
- $V\left(u+e\right) = V\left(u\right) + V\left(e\right) > V\left(u\right)$; this results in larger variances of the OLS estimators

Measurement Errors

OLS under Measurement Errors in an Explanatory Variable
Wooldridge (2013) p.310 and Greene p. 139

Let

$$y = \beta_1 + \beta_2 x_2^* + u$$

and the Gauss-Markov conditions are satisfied.
Assume that x_2^* is not observed and but we observe

$$x_2 = x_2^* + v$$

where v is a measurement error.

Measurement Errors

OLS under Measurement Errors in an Explanatory Variable
Wooldridge (2013) p.310 and Greene p. 139

The properties of the OLS estimator which uses x_2 instead of x_2^* depend on the properties of the measurement error v
The following two assumptions will be maintained in the following:
(a) $E[v] = 0$
(b) u is uncorrelated with x_2^* and x_2

Measurement Errors

OLS under Measurement Errors in an Explanatory Variable
Wooldridge (2013) p.310 and Greene p. 139

Case 1:

$$Cov\left(x_2, v\right) = 0$$

In this case,

$$y = \beta_1 + \beta_2 x_2 + \left(u - \beta_2 v\right)$$

Measurement Errors

OLS under Measurement Errors in an Explanatory Variable
Wooldridge (2013) p.310 and Greene p. 139

$$y = \beta_1 + \beta_2 x_2 + (u - \beta_2 v)$$

- Note that: $E[u - \beta_2 v] = 0$ and $Cov(x_2, u - \beta_2 v) = 0$
- Therefore, OLS with x_2 instead of x_2^* is consistent
- On the other hand,
  $Var[u - \beta_2 v] = Var[u] + \beta_2^2 Var[v] > Var[u]$ if $\beta_2 \neq 0$

Measurement Errors

OLS under Measurement Errors in an Explanatory Variable
Wooldridge (2013) p.311 and Greene p. 139

Case 2: Classical errors-in-variables

$$Cov\left(x_2^*, v\right) = 0$$

In this case,

$$\begin{aligned} Cov\left(x_2, v\right) &= E\left[x_2 v\right] \\ &= E\left[\left(x_2^* + v\right) v\right] \\ &= E\left[x_2^* v\right] + E\left[v^2\right] \\ &= V\left[v\right] \end{aligned}$$

Measurement Errors

OLS under Measurement Errors in an Explanatory Variable
Wooldridge (2013) p.311 and Greene p. 139

$$y = \beta_1 + \beta_2 x_2 + (u - \beta_2 v)$$

- Note that
$$E\left[u - \beta_2 v\right] = 0$$
- but
$$Cov\left(x_2, u - \beta_2 v\right) = -\beta_2 Cov\left(x_2, v\right) = -\beta_2 V\left[v\right]$$

Measurement Errors

OLS under Measurement Errors in an Explanatory Variable
Wooldridge (2013) p.311 and Greene p. 139

Recall

$$\hat{\beta}_2 = \beta_2 + \frac{n^{-1}\sum_{i=1}^{n}\left(x_{2i}-\bar{x}_2\right)\left(u_i-\beta_2 v_i\right)}{n^{-1}\sum_{i=1}^{n}\left(x_{2i}-\bar{x}_2\right)^2}$$

Therefore, $\hat{\beta}_2$ is inconsistent. In particular,

$$\begin{aligned} p\lim\left(\hat{\beta}_2\right) &= \beta_2 + \frac{Cov\left(x_1, u-\beta_2 v\right)}{Var\left(x_2\right)} \\ &= \beta_2 - \frac{\beta_2 V\left[v\right]}{V\left[x_2^*\right]+V\left[v\right]} \\ &= \beta_2\left(\frac{V\left[x_2^*\right]}{V\left[x_2^*\right]+V\left[v\right]}\right) \end{aligned}$$

Measurement Errors

OLS under Measurement Errors in an Explanatory Variable
Wooldridge (2013) p.311 and Greene p. 139

Notice that

$$0 < \frac{V[x_2^*]}{V[x_2^*] + V[v]} < 1$$

therefore

$$p\lim\left(\hat{\beta}_2\right) = \beta_2 \left(\frac{V[x_2^*]}{V[x_2^*] + V[v]}\right)$$

is always closer to zero than β_2 is. This is known as the **attenuation bias**.

Omitted Variables

OLS under Omitted Variables
Wooldridge (2013) p.491 and Greene p. 259

Example (Wooldridge (2013) p.491): Consider

$$\log(wage) = \beta_1 + \beta_2 educ + \beta_3 abil + e$$

where e is the error term and *abil* is not observed. If the following model

$$\log(wage) = \beta_1 + \beta_2 educ + w$$

where $w = \beta_3 abil + e$ is estimated by OLS and *educ* is correlated with *abil*, then $\hat{\beta}_2$ is biased and inconsistent.

Endogeneity

Endogeneity
Greene, p. 259

As we have seen in the above discussion, since

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

the assumption that X_i and u_i are uncorrelated is fundamental, specially for unbiasedness and consistency. **Endogeneity** problems translate in practice into situations in which the uncorrelatedness assumption does not hold.

Endogeneity: OLS Bias
Greene, p. 259

Examples:
1. Measurement errors
2. Omitted Variables
3. Simultaneous Equations
etc...

Endogeneity: IV Solution

IV Assumptions

IV Assumptions: Regressors X ($n \times K$)
Greene (2012) p. 263

AIV1. **Linear in parameters**
AIV2. **Full rank**
AIV3. **Endogeneity !!!**
AIV.4. **Homoscedasticity and nonautocorrelation**
Note: Normality: We won't use it. (Asymptotics)

IV Assumptions

AIV1. **Linear in parameters**: The model specifies a linear relationship between y and $x_1, x_2, ..., x_K$:

$$Y = X\beta + U$$

AIV2. **Full rank**: There is no exact linear relationship among any of the regressors in the model:

$$X \text{ is an } n \times K \text{ matrix with rank } K$$

This assumption is also known as the identification condition.

IV Assumptions

AIV3. **Endogeneity!!!**:. The regressors provide information about the disturbances

$$E\left[u_i | X_i\right] = \eta_i$$

Therefore, $E\left[u_i X_i\right] = \gamma_i$ and the regressors are no longer exogenous.

IV Assumptions

AIV4. **Homoscedasticity and nonautocorrelation**: Each disturbance, u_i, has the same finite variance, σ_U^2, and is uncorrelated with every other disturbance, u_j:

$$
\begin{aligned}
Var\left[u_i | X\right] &= \sigma_U^2 \quad \text{for all } i = 1, ..., n \\
Cov\left[u_i u_j | X\right] &= 0 \quad \text{for all } i \neq j
\end{aligned}
$$

In matrix form,

$$E\left[UU' | X\right] = \sigma_U^2 I$$

OLS Properties under IV Assumptions

OLS Properties under IV Assumptions
Greene p. 265

- OLS is biased:

$$
\begin{aligned}
E\left[\hat{\beta}_{OLS}\right] &= E\left[E\left[\hat{\beta}_{OLS}|X\right]\right] \\
&= E\left[\beta + \left(X'X\right)^{-1}X'\boldsymbol{\eta}\right] \\
&= \beta + E\left[\left(X'X\right)^{-1}X'\boldsymbol{\eta}\right] \\
&\neq \beta
\end{aligned}
$$

Therefore, the Gauss-Markov theorem does not hold.

OLS Properties under IV Assumptions

OLS Properties under IV Assumptions
Greene p. 265

- OLS is inconsistent:

$$
\begin{aligned}
\underset{n\to\infty}{p\lim}\hat{\beta}_{OLS} &= \underset{n\to\infty}{p\lim}\left(\beta + \left(X'X\right)^{-1}X'U\right) \\
&= \beta + \underset{n\to\infty}{p\lim}\left(\frac{X'X}{n}\right)^{-1}\underset{n\to\infty}{p\lim}\left(\frac{X'U}{n}\right) \\
&= \beta + Q_{XX}^{-1}\gamma \\
&\neq \beta
\end{aligned}
$$

IV Solution

The IV principle relies on the assumptions that an additional set of variables, Z, with the following two properties, is available:

1. Exogeneity: The instruments are uncorrelated with the error term

2. Relevance: The instruments are correlated with the endogenous independent variables

IV Solution

- Endogeneity in the simplest model:

$$y = \beta x + u \quad \text{where } Cov\left(x, u\right) \neq 0$$

- Instrument:

$$x = \alpha z + e$$

  - Exogeneity:

$$Cov\left(z, u\right) = 0 \quad \text{and} \quad Cov\left(z, e\right) = 0$$

  - Relevance:

$$\alpha \neq 0$$

IV Solution

- Suppose
$$Y = X\beta + U$$
where
$$X = [1 \;\; x_{1i} \;\; x_{2i} \;\; x_{3i}]$$
and x_{3i} is endogenous ($Cov\left(u_i, x_{3_i}\right) \neq 0$). We have an instrument for x_{3i}, say IV_i.

- Instruments matrix:
$$Z = [1 \;\; x_{1i} \;\; x_{2i} \;\; IV_i]$$

IV Solution

The IV principle relies on the assumptions that an additional set of variables, Z, with the following two properties, is available:

1. Exogeneity: The instruments are uncorrelated with the error term
2. Relevance: The instruments are correlated with the endogenous independent variables

Lets formalize this.

IV Assumptions

IV Assumptions: Regressors X ($n \times K$)
Greene (2012) p. 263

AIV1. **Linear in parameters**
AIV2. **Full rank**
AIV3. **Endogeneity !!!**
AIV.4. **Homoscedasticity and nonautocorrelation**
Note: Normality: We won't use it. (Asymptotics)

IV Assumptions

IV Assumptions: Instruments Z ($n \times L$)
for now $L = K$
Example

- Regressors: $K = 3$

$$X = [1 \;\; x_{1i} \;\; x_{2i} \;\; x_{3i}]$$

- Instruments matrix: $L = K$

$$Z = [1 \;\; x_{1i} \;\; x_{2i} \;\; IV_i]$$

IV Assumptions

AIV5. (X_i, Z_i, u_i) $i = 1, ..., n$ are i.i.d.
AIV6. $Q_{XX} = \underset{n\to\infty}{p\lim}\left(\frac{1}{n}X'X\right)$ is a finite, positive definite matrix
AIV7. $Q_{ZZ} = \underset{n\to\infty}{p\lim}\left(\frac{1}{n}Z'Z\right)$ is a finite, positive definite matrix with rank L
AIV8. $Q_{ZX} = \underset{n\to\infty}{p\lim}\left(\frac{1}{n}Z'X\right)$ is a finite, $L \times K$ matrix with rank K (relevance)
AIV9. $E\left[u_i|z_i\right] = 0$ (exogeneity)

The IV Estimator

The IV Estimator
Greene p. 265

The instrumental variables estimator of β is

$$\hat{\beta}_{IV} = \left(Z'X\right)^{-1} \left(Z'Y\right)$$

The IV Estimator: Large Sample Properties
Greene p. 266

The IV estimator can be expressed as

$$\begin{aligned}\hat{\beta}_{IV} &= \left(Z'X\right)^{-1}\left(Z'Y\right)\\ &= \left(Z'X\right)^{-1}\left(Z'\left(X\beta+U\right)\right)\\ &= \beta+\left(Z'X\right)^{-1}\left(Z'U\right)\end{aligned}$$

and hence

$$\sqrt{n}\left(\hat{\beta}_{IV}-\beta\right)=\left(\frac{Z'X}{n}\right)^{-1}\left(\frac{Z'U}{\sqrt{n}}\right)$$

The IV Estimator

The IV Estimator: Large Sample Properties
Greene p. 266

Theorem: Under assumptions AIV1-AIV9, if $L = K$, then

$$\sqrt{n}\left(\hat{\beta}_{IV} - \beta\right) \xrightarrow{d} N\left(0, \sigma_U^2 Q_{ZX}^{-1} Q_{ZZ} Q_{XZ}^{-1}\right)$$

Therefore,

$$\hat{\beta}_{IV} \overset{a}{\sim} N\left(\beta, \frac{\sigma_U^2}{n} Q_{ZX}^{-1} Q_{ZZ} Q_{XZ}^{-1}\right)$$

The IV Estimator

IV: Asymptotic Variance Estimation
Greene p. 266

Since,

$$\hat{\beta}_{IV} \overset{a}{\sim} N\left(\beta, \frac{\sigma_U^2}{n} Q_{ZX}^{-1} Q_{ZZ} Q_{XZ}^{-1}\right)$$

the asymptotic variance of $\hat{\beta}_{IV}$ is

$$AsyVar\left(\hat{\beta}_{IV}\right) = \frac{\sigma_U^2}{n} Q_{ZX}^{-1} Q_{ZZ} Q_{XZ}^{-1}$$

IV: Asymptotic Variance Estimation
Greene p. 266

$$AsyVar\left(\hat{\beta}_{IV}\right) = \frac{\sigma_U^2}{n} Q_{ZX}^{-1} Q_{ZZ} Q_{XZ}^{-1}$$

The sample analog is then a consistent estimator

$$\begin{aligned}
\widehat{AsyVar}\left(\hat{\beta}_{IV}\right) &= \frac{1}{n}\left\{\left(\frac{\hat{U}_{IV}'\hat{U}_{IV}}{n}\right)\left(\frac{Z'X}{n}\right)^{-1}\left(\frac{Z'Z}{n}\right)\left(\frac{X'Z}{n}\right)^{-1}\right\} \\
&= \hat{\sigma}^2_{\hat{U}_{IV}}\left(Z'X\right)^{-1}\left(Z'Z\right)\left(X'Z\right)^{-1}
\end{aligned}$$

IV: Normal Inference
Greene p. 266

Given

$$\hat{\beta}_{IV} \overset{a}{\sim} N\left(\beta, \frac{\sigma_U^2}{n} Q_{ZX}^{-1} Q_{ZZ} Q_{XZ}^{-1}\right)$$

and consistency of

$$\widehat{AsyVar}\left(\hat{\beta}_{IV}\right) = \hat{\sigma}^2_{\hat{U}_{IV}} \left(Z'X\right)^{-1} \left(Z'Z\right) \left(X'Z\right)^{-1}$$

inference can be carried out in the standard way.

Example

Example: Returns to Education
Wooldridge p. 504

$$\log(wage) = \beta_1 + \beta_2 educ + \beta_3 \exp er + \beta_4 \exp er^2 + ... + u$$

where u may be correlated with *educ* because *ability* is omitted.

Let *nearc*4 be a dummy variable equals 1 if individual i grew up near a four-year college and assume that it is correlated with *educ* but not with u.

Example

Example: Returns to Education
Wooldridge p. 504

OLS:

$$\log\left(wage\right) = \hat{\beta}_1 + \underset{(0.003)}{0.075} educ + \hat{\beta}_3 \exp er + \hat{\beta}_4 \exp er^2 + ... + u$$

IV:

$$\log\left(wage\right) = \tilde{\beta}_1 + \underset{(0.055)}{0.132} educ + \tilde{\beta}_3 \exp er + \tilde{\beta}_4 \exp er^2 + ... + u$$

Multiple Instruments: 2SLS

Multiple Instruments: 2SLS
Greene p. 270 and Wooldridge (2010) p. 96

- So far, we have assumed that the number of instruments in Z is equal to the number of variables (exogenous plus endogenous) in X.
- If Z contains more variables than X then $(Z'X)$ is $L \times K$ with rank $K < L$, therefore $(Z'X)$ is not invertible
- Since $E[Z'U] = 0$, every column of Z is uncorrelated with U as it is any linear combination of the columns of Z
- Which to choose?

Multiple Instruments: 2SLS

Multiple Instruments: 2SLS

Greene p. 270 and Wooldridge (2010) p. 96

- Which to choose?
- 2SLS Proposal:

$$\hat{X} = Z\left(Z'Z\right)^{-1} Z'X$$

- Why? It can be shown (see Wooldridge 2010 p. 103) that $\hat{X}$ delivers the most efficient estimator in the class of instrumental variables estimators using instruments linear in Z

Multiple Instruments: 2SLS

Multiple Instruments: 2SLS

Greene p. 270 and Wooldridge (2010) p. 96

- Interpretation:

$$\begin{aligned} Y &= X\beta + U \\ X &= Z\alpha + V \end{aligned}$$

- Then

$$\hat{\alpha} = \left(Z'Z\right)^{-1} Z'X$$

  and

$$\hat{X} = Z\hat{\alpha} = Z\left(Z'Z\right)^{-1} Z'X$$

- Hence,

$$\begin{aligned} \hat{\beta}_{2SLS} &= \left(\hat{X}'X\right)^{-1}\left(\hat{X}'Y\right) \\ &= \left(X'Z\left(Z'Z\right)^{-1} Z'X\right)^{-1}\left(X'Z\left(Z'Z\right)^{-1} Z'Y\right) \end{aligned}$$

Multiple Instruments: 2SLS

Multiple Instruments: 2SLS

Greene p. 270 and Wooldridge (2010) p. 96

- Recall: if $L = K$

$$\hat{\beta}_{IV} = \left(Z'X\right)^{-1}\left(Z'Y\right)$$

- 2SLS Proposal: use as instruments

$$\hat{X} = Z\left(Z'Z\right)^{-1}Z'X$$

- Hence,

$$\begin{aligned}\hat{\beta}_{2SLS} &= \left(\hat{X}'X\right)^{-1}\left(\hat{X}'Y\right) \\ &= \left(X'Z\left(Z'Z\right)^{-1}Z'X\right)^{-1}\left(X'Z\left(Z'Z\right)^{-1}Z'Y\right)\end{aligned}$$

Multiple Instruments: 2SLS

Asymptotic Normality of 2SLS
Wooldridge (2010) p. 101

Theorem: Under assumptions AIV1-AIV10,

$$\sqrt{n}\left(\hat{\beta}_{2SLS} - \beta\right) \xrightarrow{d} N\left(0, \sigma_U^2 \left(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX}\right)^{-1}\right)$$

Therefore,

$$\hat{\beta}_{2SLS} \overset{a}{\sim} N\left(\beta, \frac{\sigma_U^2}{n} \left(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX}\right)^{-1}\right)$$

Multiple Instruments: 2SLS

2SLS: Asymptotic Variance Estimation
Wooldridge (2010) p. 101

Since,

$$\hat{\beta}_{2SLS} \overset{a}{\sim} N\left(\beta, \frac{\sigma_U^2}{n}\left(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX}\right)^{-1}\right)$$

the asymptotic variance of $\hat{\beta}_{2SLS}$ is

$$AsyVar\left(\hat{\beta}_{2SLS}\right) = \frac{\sigma_U^2}{n}\left(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX}\right)^{-1}$$

Multiple Instruments: 2SLS

2SLS: Asymptotic Variance Estimation
Wooldridge (2010) p. 101

$$AsyVar\left(\hat{\beta}_{2SLS}\right) = \frac{\sigma_U^2}{n}\left(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX}\right)^{-1}$$

The sample analog is then a consistent estimator

$$\widehat{AsyVar}\left(\hat{\beta}_{2SLS}\right) = \hat{\sigma}^2_{\hat{U}_{2SLS}}\left(\left(X'Z\right)\left(Z'Z\right)^{-1}\left(Z'X\right)\right)^{-1}$$

where (IMPORTANT)

$$\hat{\sigma}^2_{\hat{U}_{2SLS}} = \frac{\hat{U}'_{2SLS}\hat{U}_{2SLS}}{n}$$

and

$$\hat{U}_{2SLS} = Y - X\hat{\beta}_{2SLS} \quad \left(\text{NO } Y - \hat{X}\hat{\beta}_{2SLS}\right)$$

Multiple Instruments: 2SLS

IV: Normal Inference
Wooldridge (2010) p. 101

Given

$$\hat{\beta}_{2SLS} \overset{a}{\sim} N\left(\beta, \frac{\sigma_U^2}{n}\left(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX}\right)^{-1}\right)$$

and consistency of

$$\widehat{AsyVar}\left(\hat{\beta}_{2SLS}\right) = \hat{\sigma}^2_{\hat{U}_{2SLS}}\left(\left(X'Z\right)\left(Z'Z\right)^{-1}\left(Z'X\right)\right)^{-1}$$

inference can be carried out in the standard way.

Application: Returns to Schooling

Griliches (1976): Wages of Very Young Men
(A summary from Hayashi, p. 236)

- The semi-log wage equation

$$LW = \alpha + \beta S + \gamma A + \delta' h + \varepsilon$$

  - LW is the log wage rate for the individual
  - S is schooling in years
  - A is a measure of ability
  - h is the vector of observable characteristics of the individual (experience, location dummies, etc...)
  - ε is the unobservable error term with zero mean

Application: Returns to Schooling

Griliches (1976): Wages of Very Young Men
(A summary from Hayashi, p. 236)

- The semi-log wage equation

$$LW = \alpha + \beta S + \gamma A + \delta' h + \varepsilon$$

- β measures the percentage increase in the wage rate the individual would receive for one more year of education (marginal return from investing in human capital)

- It is assumed that (S, A, h) are uncorrelated with ε

- "**Ability bias**": (Griliches) biases on the OLS estimate of β when A is not included or measured with error

Application: Returns to Schooling

Griliches (1976): Wages of Very Young Men
(A summary from Hayashi, p. 236)

- Griliches Data: Young Men's Cohort of the National Longitudinal Survey (NLS-Y)
- This database includes two measures of ability:
  - KWW: Knowledge of the World of Work
  - IQ score
- "**Ability bias**": (Griliches) biases on the OLS estimate of β when A is omitted or measured with error

Application: Returns to Schooling

Griliches (1976): Wages of Very Young Men
(A summary from Hayashi, p. 236)

- Griliches applies 2SLS to control for the bias in the wage equation
- Predetermined Regressors: $(1, S, h)$
- IQ as an approximation to ability subject to "classical" measurement error
- Instruments for IQ: age, age squared, KWW, mother's education, background variables (such as father's occupation)

Application: Returns to Schooling

Returns to Education for Working Women
Wooldridge data (W1 p.528; W2 p. 102)

- Data on 428 working, married women
- The equation of interest is

$$LW = \beta_1 + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + u$$

- We assume *exper* is exogenous, but we allow *educ* to be correlated with u.
- Instruments for *educ*: motheeduc, fatheduc, and huseduc

Application: Returns to Schooling

Returns to Education for Working Women
Wooldridge data (W1 p.528; W2 p. 102)

- OLS estimation:

$$\widehat{LW} = \underset{(-2.628)}{-\ 0.522} + \underset{(7.598)}{0.107} educ + \underset{(3.154)}{0.041} exper - \underset{(-2.062)}{0.0008} exper^2$$

- 2SLS:

$$\widehat{LW} = \underset{(-0.654)}{-\ 0.187} + \underset{(3.692)}{0.080} educ + \underset{(3.248)}{0.043} exper - \underset{(-2.177)}{0.0008} exper^2$$