

OLS Estimation - Assumptions

- CLM Assumptions
- (A1) DGP: $\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}$ is correctly specified.
- $(\mathbf{A2}) \; \mathrm{E}[\boldsymbol{\epsilon} \,|\, \mathbf{X}] = 0$
- (A3) $\operatorname{Var}[\boldsymbol{\epsilon} \,|\, \mathbf{X}] = \sigma^2 \, \mathbf{I}_{\mathrm{T}}$
- (A4) **X** has full column rank $-\operatorname{rank}(\mathbf{X}) = k$ -, where $T \ge k$.

Q: What happens when (A1) is not correctly specified?

• In this lecture, we look at (A1), always in the context of linearity. Are we omitting a relevant regressor? Are we including an irrelevant variable? What happens when we impose restrictions in the DGP?

Specification Errors: Omitted Variables

• Omitting relevant variables: Suppose the correct model is $\mathbf{y} = \mathbf{X}_1 \mathbf{\beta}_1 + \mathbf{X}_2 \mathbf{\beta}_2 + \mathbf{\epsilon}$ -i.e., with two sets of variables. But, we compute OLS omitting \mathbf{X}_2 . That is, <= the "short regression."

Some easily proved results:

 $\mathbf{y} = \mathbf{X}_1 \mathbf{\beta}_1 + \mathbf{\varepsilon}$

(1) $E[\mathbf{b}_1 | \mathbf{X}] = E[(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1' \mathbf{y}] = \mathbf{\beta}_1 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\mathbf{\beta}_2 \neq \mathbf{\beta}_1$. So, unless $X_1'X_2 = 0$, b_1 is *biased*. The bias can be huge. It can reverse the sign of a price coefficient in a "demand equation."

(2) $\operatorname{Var}[\mathbf{b}_1 | \mathbf{X}] \leq \operatorname{Var}[\mathbf{b}_{1,2} | \mathbf{X}]$. (The latter is the northwest submatrix of the full covariance matrix.) The proof uses **M**, the residual maker. We get a smaller variance when we omit \mathbf{X}_2 .

Specification Errors: Omitted Variables

• We get a smaller variance when we omit \mathbf{X}_2 .

<u>Interpretation</u>: Omitting \mathbf{X}_2 amounts to using extra information –i.e., $\beta_2 = 0$. Even if the information is wrong, it reduces the variance.

(3) MSE

 \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.

<u>Note</u>: Suppose $X_1'X_2 = 0$. Then the bias goes away. Interpretation, the information is not "right," it is irrelevant. \mathbf{b}_1 is the same as $\mathbf{b}_{1,2}$.

Omitted Variables Example: Gasoline Demand We have a linear model for the demand for gasoline: G = PG β₁ + Y β₂ + ε, Q: What happens when you wrongly exclude Income (Y)? E[b₁|X] = β₁ + Cov[Price, Income] β₂ In time series data, β₁ < 0, β₂ > 0 (usually) Cov[Price, Income] > 0 in time series data. ⇒ The short regression will overestimate the price coefficient. In a simple regression of G (demand) on a constant and PG, the Price Coefficient (β₁) should be negative.



Estimation of a 'Demand' Equation (Greene): Multiple Regression - Theory Works.

Ordinary	least squares regres	sion .		
LHS=G	Mean	=	226.09444	l
	Standard deviation	=	50.59182	2
	Number of observs.	=	36	5
Model size	Parameters	=	3	3
	Degrees of freedom	=	33	3
Residuals	Sum of squares	=	1472.79834	L
	Standard error of e	=	6.68059)
Fit	R-squared	=	.98356	5
	Adjusted R-squared	=	.98256	5
Model test	F[2, 33] (prob)	= 98	87.1(.0000)	
Variable Co	pefficient Standard	Error	t-ratio	P[T >t]
Constant	-79.7535*** 8.67	255	-9.196	.0000
Υ	.03692*** .00	132	28.022	.0000
	1 = 1	024	0 010	0000

• <u>Note</u>: Income is helping us to identify a demand equation –i.e., with a negative slope for the price variable.





Other Models

• Looking ahead to nonlinear models: neither of the preceding results extend beyond the linear regression model.

"Omitting relevant variables from a model is always costly. (No exceptions.) The benign result above almost never carries over to more involved nonlinear models." (Greene)

Specification and Functional Form: Non-linearity

• In the context of OLS estimation, we can introduce some nonlinearities: quadratic, cubic and interaction effects can be easily estimated by OLS. For example:

 $\boldsymbol{y} = \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 \, \mathbf{X}_2 + \boldsymbol{\beta}_3 \, \mathbf{X}_2^2 + \boldsymbol{\beta}_4 \, \mathbf{X}_2 \, \mathbf{X}_3 + \boldsymbol{\varepsilon}$

• Partial effects , $\partial y / \partial X_2$, (and standard errors) can be different. In the above model

$$\partial \mathbf{y} / \partial \mathbf{X}_2 = \beta_2 + 2 \beta_3 \mathbf{X}_2 + \beta_4 \mathbf{X}_3 \neq \beta_2$$

Note: Recall that in a simple linear model:

 $\mathbf{y} = \mathbf{\beta}_1 + \mathbf{\beta}_2 \mathbf{X}_2 + \mathbf{\beta}_3 \mathbf{X}_3 + \mathbf{\varepsilon}$ the partial effect is equal to the β_i coefficient: $\partial \mathbf{y}/\partial \mathbf{X}_2 = \beta_2$.

Specification and Functional Form: Non-linearity

• The estimator of partial effects and their variances are different from b_i and $Var[b_i | \mathbf{X}]$ in the presence of non-linearities

Example: Quadratic Effect

Population

Estimators

 $y = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 z + \varepsilon$ $\hat{y} = b_1 + b_2 x + b_3 x^2 + b_4 z$ $\delta_x = \frac{\partial E[y \mid x, z]}{\partial x} = \beta_2 + 2\beta_3 x$ $\hat{\delta}_x = b_2 + 2b_3 x$ Estimators $\hat{y} = b_1 + b_2 x + b_3 x^2 + b_4 z$

Estimator of the variance of $\hat{\delta}_x$

$$Est.Var[\hat{\delta}_{x}] = Var[b_{2}] + 4x^{2}Var[b_{3}] + 4xCov[b_{2}, b_{3}]$$

Note: Now, the partial effect and the variance are a function of the data! Usually, an average is used in the estimation.





Ordinary	least squares regression						
LHS=LOGY	Mean	2	=	-1.1574	6		
	Standard deviation		=	.49149			
	Number of o	oservs.	=	2732	2		
Model size	Parameters		=		4		
	Degrees of	freedom	=	2731	8		
Residuals	Sum of squa	res	=	6540.4598	8		
	Standard er	ror of e	=	.4893	1		
Fit	R-squared		=	.0089	6		
	Adjusted R-	squared	=	.0088	5		
Model test +	F[3, 2731	8] (prob)	=	82.4(.0000) 		
Variable (Coefficient	Standard	Erro	r b/St.Er.	P[Z >z]	Mean of X	
Constant	-1.22592***	.01	 605	-76.376	.0000		
AGE	.00227***	.00	036	6.240	.0000	43.5272	
FEMALE	.21239***	.02	363	8.987	.0000	.47881	
AGE_FEM	00620***	.00	052	-11.819	.0000	21.2960	

OLS Subject to Restrictions

• Restrictions: Theory imposes certain restrictions on parameters.

Examples:

(1) Dropping variables from the equation. That is, certain coefficients in **b** forced to equal 0. (Is variable x_3 =*size* significant?")

(2) Adding up conditions: Sums of certain coefficients must equal fixed values. Adding up conditions in demand systems. Constant returns to scale in production functions ($\alpha + \beta = 1$ in a Cobb-Douglas production function).

(3) Equality restrictions: Certain coefficients must equal other coefficients. Using real vs. nominal variables in equations.

• Usual formulation with J linear restrictions (**R** is Jxk and **q** is Jx1):

 $\operatorname{Min}_{\mathbf{b}} \left\{ \mathrm{S}(x_i, \theta) = \sum_{i=1}^{T} e_i^2 = e'e = (\mathbf{y} - \mathbf{X}\mathbf{b})' (\mathbf{y} - \mathbf{X}\mathbf{b}) \right\} \text{ s.t. } \mathbf{R}\mathbf{b} = \mathbf{q}$

Restricted Least Squares

• In practice, restrictions can usually be imposed by solving them out. (1) Dropping variables -i.e., force a coefficient to equal zero. <u>Problem</u>: $\min_b \sum_{i=1}^n (y_i - b_1 x_{i1} - b_2 x_{i2} - b_3 x_{i3})^2$ s.t. $b_3 = 0$ $\min_b \sum_{i=1}^n (y_i - b_1 x_{i1} - b_2 x_{i2})^2$ (2) Adding up. Do least squares subject to $b_1 + b_2 + b_3 = 1$. Then, $b_3 = 1 - b_1 - b_2$. Make the substitution: $(\mathbf{y} - \mathbf{x}_3) = b_1(\mathbf{x}_1 - \mathbf{x}_3) + b_2(\mathbf{x}_2 - \mathbf{x}_3) + \mathbf{e}$ <u>Problem</u>: $\min_b \sum_{i=1}^n ((y_i - x_{i3}) - b_1(x_{i1} - x_{i3}) - b_2(x_{i2} - x_{i3}))^2$ (3) Equality. If $b_3 = b_2$, then $\mathbf{y} = b_1 \mathbf{x}_1 + b_2 \mathbf{x}_2 + b_2 \mathbf{x}_3 + \mathbf{e}$ $= b_1 \mathbf{x}_1 + b_2(\mathbf{x}_2 + \mathbf{x}_3) + \mathbf{e}$ <u>Problem</u>: $\min_b \sum_{i=1}^n (y_i - b_1 x_{i1} - b_2(x_{i2} + x_{i3}))^2$

Restricted Least Squares

• Theoretical results provide insights and the foundation of several tests.

- Programming problem with J restrictions (**R** is Jxk and **q** is kx1): Minimize wrt **b** $S = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ s.t. **R** $\boldsymbol{\beta} = \mathbf{q}$
- Quadratic programming problem

 ⇒ Minimize a quadratic criterion s.t. a set of linear restrictions.
 Concave programming problem, all binding constraints. No need for Kuhn-Tucker
 Solve using a Lagrangean formulation.

 The Lagrangean approach (the 2 is for convenience with is λ Jx1)

$$\operatorname{Min}_{\mathbf{b},\boldsymbol{\lambda}} \quad L^* = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + 2 \ \boldsymbol{\lambda}' (\mathbf{R} \ \boldsymbol{\beta} - \mathbf{q}) \\ = (\mathbf{y}'\mathbf{y} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}) + 2 \ \boldsymbol{\lambda}'(\mathbf{R} \ \boldsymbol{\beta} - \mathbf{q})$$



Linear Restrictions

• Q: How do linear restrictions affect the properties of the least squares estimator? Model (DGP): $y = X\beta + \varepsilon$ Theory (information): $R\beta - q = 0$ Restricted LS estimator: $b^* = b - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(Rb - q)$ 1. Unbiased? $E[b^*|X] = \beta - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}E[(Rb - q) | X] = \beta$ 2. Efficiency? $Var[b^*|X] = \sigma^2(X'X)^{-1} - \sigma^2(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}$ $Var[b^*|X] = Var[b|X] - a nonnegative definite matrix < Var[b|X]$ 3. b^* may be more "precise." Precision = Mean squared error = variance + squared bias.

Linear Restrictions 1. $\mathbf{b}^* = \mathbf{b} - \mathbf{Cm}$, $\mathbf{m} = \text{the "discrepancy vector" } \mathbf{Rb} - \mathbf{q}$. Note: If $\mathbf{m} = \mathbf{0} \implies \mathbf{b}^* = \mathbf{b}$. (Q: What does $\mathbf{m} = \mathbf{0}$ mean?) 2. $\lambda = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{q}) = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{m}$ When does $\lambda = \mathbf{0}$? What does this mean? 3. Combining results: $\mathbf{b}^* = \mathbf{b} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\lambda$ 4. Recall: $\mathbf{e'e} = (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb}) \le \mathbf{e}^{*'}\mathbf{e}^* = (\mathbf{y} - \mathbf{Xb}^*)'(\mathbf{y} - \mathbf{Xb}^*)$ \Rightarrow Restrictions cannot increase $\mathbf{R}^2 \implies \mathbf{R}^2 \ge \mathbf{R}^{2*}$

Linear Restrictions – Interpretation

• Two cases

- Case 1: Theory is correct: $\mathbf{R}\boldsymbol{\beta} \mathbf{q} = \mathbf{0}$ (restrictions hold). \mathbf{b}^* is unbiased & $\operatorname{Var}[\mathbf{b}^* | \mathbf{X}] \leq \operatorname{Var}[\mathbf{b} | \mathbf{X}]$
- Case 2: Theory is incorrect: $\mathbf{R}\boldsymbol{\beta} \mathbf{q} \neq \mathbf{0}$ (restrictions do not hold). \mathbf{b}^* is biased & $\operatorname{Var}[\mathbf{b}^* | \mathbf{X}] \leq \operatorname{Var}[\mathbf{b} | \mathbf{X}].$

• Interpretation

- The theory gives us information.

Bad information produces bias (away from "the truth.") Any information, good or bad, makes us more certain of our answer. In this context, *any* information reduces variance.



What about ignoring information (theory)?
 Not using the correct information does not produce bias.
 Not using information foregoes the variance reduction.

Testing in Economics



"The three golden rules of econometrics are test, test and test." David Hendry (1944, England)



"The only relevant test of the validity of a hypothesis is comparison of prediction with experience." Milton Friedman (1912-2006, USA)

24