Applied Econometrics 06: Linear regression with multiple regressors

Hosin Song

Spring 2016
Ewha Womans University

Where are we now?

- Motivation of this Class
- Review of Probability and Statistics
- Linear Regression with one regressor
- Linear Regression with multiple regressors

Why do we have to care about Multiple Regressors?

- A single regressor may not be realistic.

  More factors may affect the dependent variable Y.

- Probably you may care to know more detailed information to affect the test score.

- Omitted Variable Problem!

  It is one reason why we should care about multiple regressors.

Omitted Variable Problem: Income equation

- Regression of annual income (y) on dummy $D_{college}$

$$\text{where } D_{college} = \begin{cases} 1 & \text{if he or she has college degree,} \\ 0 & \text{otherwise.} \end{cases}$$

- Single regressor model: $E(y_i|X_i) = \beta_0 + \beta_1 D_{college\,i}$ (1)

- But, there are other factors which can determine the income such as ability, job experience and sex.

- To account for more factors, we need to consider multiple regressors:

  Say, $E(y_i|X_i) = \beta_0 + \beta_1 D_{college\,i} + \beta_2 \mathrm{IQ}_i + \beta_3 \mathrm{EXPER}_i + \beta_2 D_{sex\,i}$ (2)

- Roughly speaking, both $\hat{\beta}_1$ from (1) and (2) would be different provided that $corr(D_i, \text{other } X_i) \neq 0$.

Omitted Variable Problem: Test Score equation

- Regression of tests score(Y) on student-teacher ratio only may give you misleading results (say, incorrect $\hat{\beta}_1$)

  if you ignore some potentially important determinants of test scores.

- Omitted X than the test score might include the following factors:
  - school characteristics such teacher quality and computer usage
  - student characteristics such as family background

- The OLS estimator of β_1 would be different if you add some explanatory variables which are correlated with the student-teacher ratio.

Omitted Variable Bias

Omitted Variable Bias

If the regressor (say, student-teacher ratio) is correlated with a vaiable that has been omitted from the analysis and that determines, in part, the dependent variable(say, test scores), then the OLS estimator will have **omitted variable**.

- Truth: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + u_i$.
- Suppose, in practice, you regress Y on constant and X_1, and obtain $\hat{\beta}_1$. Then,

  Your OLS estimator $\hat{\beta}_1$ is not consistent if X_1 and X_2 are correlated.

Omitted Variable Bias

In general, suppose the true model is

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots \beta_K X_{iK} + u_i. \text{ (say, } K \geq 3.)$$

If you use only two regressors out of K regressors, and estimate the OLS estimators $\hat{\beta}_j, j = 1, 2,$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i.$$

then your OLS estimator $\hat{\beta}_1$ and $\hat{\beta}_2$ will not be consistent unless the unused X_{ik}'s, $k = 3, \dots, K$, are uncorrelated with the used regressors X_{i1} and X_{i2}.

Omitted Variable Bias: Cobb-Douglas Function

- Truth: $Q_i = C * L_i^{\beta_1} Q_i^{\beta_2} \epsilon_i$

  $\Rightarrow \ln Q_i = \beta_0 + \beta_1 \ln L_i + \beta_2 \ln K_i + u_i.$

  $\Rightarrow Q_i^* = \beta_0 + \beta_1 L_i^* + \beta_2 K_i^* + u_i.$

- If we drop $\ln K$ and use $\ln L$, then apply OLS to the following linear regression model

  $\ln Q_i = \beta_0 + \beta_1 \ln L_i + \text{error term}.$ That is,

  $\Rightarrow Q_i^* = \beta_0 + \beta_1 L_i^* + \text{error term}.$

- Then, the resulting OLS estimator $\hat{\beta}_1$ will not be consistent provided that K^* is correlated with L^*. (i.e., $\hat{\beta}_1$ has omitted variable bias.)

Omitted Variable Bias

Examples of Omitted Variable Bias in the test score example

Think of variables that (1) are correlated with student-teacher ratio and that (2) in part determines test scores.

- Percentage of English learners ? (1) Yes, (2) Yes $\Rightarrow$ Omitting may induce omitted variable bias.
- Time of day of the test? (1) No, (2) Yes $\Rightarrow$ Omitting does not induce omitted variable bias.
- Parking lot space per pupil? (1) Yes, (2) No $\Rightarrow$ Omitting does not induce omitted variable bias.

Omitted Variable Bias (cont,)

Omitted Variable bias $\Rightarrow$ Violation of #1 LS Ass. $E(u_i|X_i) = 0$

Recall that the error term u_i in the linear regression model with a single regressor represents all factors, other than X_i, that are determinants of Y_i.

If one of these factors is correlated with X_i, this means that the error term is correlated with X_i. So $E(u_i|X_i) = 0$ does not hold.

A Formula for Omitted Variable Bias

The first OLS assumption does not hold when $\rho_{Xu} \neq 0$. Then, the OLS estimator $\hat{\beta}_1$ is not consistent any longer: $\hat{\beta}_1 \xrightarrow{p} \beta_1 + \rho_{Xu} \frac{\sigma_u}{\sigma_X} \neq \hat{\beta}_1$.

- To overcome omitted variable bias, we need multiple regression model which exploits more relevant regressors.

Inconsistency under Omitted Variable Problem : $\hat{\beta}_1 \xrightarrow{p} \beta_1 + \rho_{Xu}\frac{\sigma_u}{\sigma_X} \neq \hat{\beta}_1$.

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^n (X_i - \overline{X})^2} \\
&= \frac{\sum_{i=1}^n (X_i - \overline{X})\left(\beta_1 (X_i - \overline{X}) + (u_i - \overline{u})\right)}{\sum_{i=1}^n (X_i - \overline{X})^2} \\
&= \beta_1 + \frac{\sum_{i=1}^n (X_i - \overline{X})(u_i - \overline{u})}{\sum_{i=1}^n (X_i - \overline{X})^2} = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})(u_i - \overline{u})}{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2} \\
&= \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})(u_i - \overline{u})}{\sqrt{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2}\sqrt{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2}} \frac{\sqrt{\frac{1}{n}\sum_{i=1}^n (u_i - \overline{u})^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^n (u_i - \overline{u})^2}} \\
&= \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})(u_i - \overline{u})}{\sqrt{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2}\sqrt{\frac{1}{n}\sum_{i=1}^n (u_i - \overline{u})^2}} \frac{\sqrt{\frac{1}{n}\sum_{i=1}^n (u_i - \overline{u})^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2}} \\
&\xrightarrow{p} \beta_1 + \frac{\sigma_{Xu}}{\sigma_X \sigma_u}\frac{\sigma_u}{\sigma_X} = \beta_1 + \rho_{Xu}\frac{\sigma_u}{\sigma_X} \neq \beta_1
\end{aligned}
$$

Multiple Regression Model

Multiple Regression Model

$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + u_i, \quad i = 1, \ldots, n.$

- Y_i : the depenedent variable of the i^{th} observation.
- $X_{i1}, X_{i2}, \ldots, X_{ik}$: k regressors of the i^{th} observation.
- u_i is the error term the i^{th} observation.
- The population regression function:

  $E(Y|X_{i1} = x_1, X_{i2} = x_2, \ldots, X_{ik} = x_k) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k.$

Multiple Regression Model (cont.)

- The population regression function:

  $$E(Y|X_{i1} = x_1, X_{i2} = x_2, \ldots, X_{ik} = x_k) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k.$$

- β_1 is the (slope) coefficient on X_1, β_2 is the (slope) coefficient on X_2, and so on.
- Alternatively, $\beta_0, \beta_1, \ldots, \beta_k$ are called the parameters.
- The coefficient β_1 is the expected change in Y_i resulting from changing X_{1i} by one unit, holding $X_{2i}, \ldots, X_{ki}$ constant.

  The coefficients $\beta_2, \ldots, \beta_k$ on the other X's can be interpreted similarly.

- The intercept β_0 is the expected value of Y when all the X's equal to 0. The intercept can be thought of as the coefficient on a regressor X_{0i}, that equals 1 for all i.

Example: Multiple Regression with 2 regressors $k = 2$

$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + u_i$, $i = 1, \ldots, n$, where

- Population Regression function:

  $$E(Y|X_{i1} = x_1, X_{i2} = x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

- β_1 is the effect of a unit change in X_{1i} on Y holding X_2 constant (or controlling for X_2),

  that is,$\beta_1 = \frac{\Delta Y}{\Delta X_1}$ holding X_2 constant

  $\Rightarrow$ β_1 is the partial effect of X_1 on Y holding X_2 fixed.

Homoskedasticity and Heteroskedasticity

Homoskedasticity and Heteroskedasticity

The error term u_i in $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_1 X_{ik} + u_i$ is **homoskedastic** if $Var(u_i|X_{i1}, \ldots, X_{ik})$ is constant and thus does not depend on the values of $X_{i1}, \ldots, X_{ik}$.

Otherwise, the error term is **heteroskedastic**.

OLS Principle and OLS Estimator

OLS Principle and OLS Estimator

The OLS estimators $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ are the values of $b_0, b_1, \ldots, b_k$ that minimize the sum of squared predicted mistakes

$$\Omega_n = \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{i1} - \ldots - b_k X_{ik})^2.$$

- F.O.C.: $\frac{\partial \Omega_n}{\partial b_i} = 0$ for $i = 0, 1, \ldots, k$.

  Such $b_i's$ are OLS estimators, and they are denoted by $\hat{\beta}_i$.

- Predicted (or fitted) Value $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik}$, for all i

- Resudual $\hat{u}_i = Y_i - \hat{Y}_i$ for all i

Measures of Fit in Multiple Regression

Measures of Fit in Multiple Regression

- The R^2: $R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS}$ where

  Explained Sum of Squares is $ESS = \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2$, and

  the Total Sum of Squares is $TSS = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$

- Including more regressors makes R^2 nondecreasing.(larger or at least the same as before.)

- For the reason above, the "Adjusted R^2"(or $\overline{R}^2$) is used.

$$\overline{R}^2 = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS} = 1 - \frac{\frac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}_i^2}{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y}_i)^2}.$$

"Adjusted R^2" (or $\overline{R}^2$): Correction for R^2

$$\overline{R}^2 = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS} = 1 - \frac{\frac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}_i^2}{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y}_i)^2}.$$

- Note: R^2 does not decrease when a new variable is added since SSR does not increase.

  Adding a variable improves the fit of the model in terms of R^2, even though it has nothing to do with the quality of analysis.

- One way to correct such problem is to deflate or reduce the R^2 by some factor which is related to the number of regressors.

  The adjusted R^2 (or $\overline{R}^2$) is such example.

"Adjusted R^2" (or $\overline{R}^2$): Correction for R^2 (cont.)

$$\overline{R}^2 = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS} = 1 - \frac{\frac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}_i^2}{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y}_i)^2}.$$

1. $\frac{n-1}{n-k-1} > 1$, so $R^2 > \overline{R}^2$
2. Adding a regressor has two opposite effects on $\overline{R}^2$:

   Whether the $\overline{R}^2$ increases or decreases depends on which of these two effects is stronger.

   1. SSR falls $\Rightarrow \overline{R}^2 \uparrow$,
   2. $\frac{n-1}{n-k-1} \uparrow \Rightarrow \overline{R}^2 \downarrow$
3. $\overline{R}^2$ can be negative. It happens when $\frac{SSR}{TSS} > \frac{n-k-1}{n-1}$.

The Least Squares Assumptions in Multiple Regression

LS Assumptions in Multiple Regression

Assumption 1: $E(u_i|X_{i1}, X_{i2}, ..., X_{ik}) = 0$

Assumption 2: $(Y_i, X_{i1}, X_{2i}, ..., X_{ik}), i = 1, \ldots, n$, are i.i.d.

Assumption 3: $0 < E(X_{i1}^4) < \infty, ..., 0 < E(X_{ik}^4) < \infty$, and $0 < E(Y_i^4) < \infty$. This means that large outliers are unlikely.

Assumption 4: No Perfect Multicollinearity.

Multicollinearity

Perfect Multicollinearity

Perfect multicollinearity exists if one of regressors is a perfect linear function of the other regressors.

Assumption 4 requires that there is no such relationship among regressors so that LS estimation works.

- $Y_1 = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$.

  Suppose that $X_{1i} = 2X_{2i}$, then it is not possible to compute OLS estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ uniquely. Why?

  If so, the regression model $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$
  $\Leftrightarrow Y_1 = \beta_0 + (2\beta_1 + \beta_2) X_{2i} + u_i = \beta_0 + \gamma X_{2i} + u_i$.

- Thus, β_1 and β_2 are not unique, in the sense that any pair (β_1, β_2) gives the same Y_i provided that $2\beta_1 + \beta_2 = \gamma$.

Multicollinearity (cont.)

- **Dummy variable Trap**

$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + u_i$ where Y is income, X_{i1} is years of schooling and $X_{i2} = D$ is a dummy variable.

Note that $D = 1$ if female, and 0 otherwise.

Including both D **and** $1 - D$ in the regression equation brings about the perfect multicollinearity problem.

If so, $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + u_i$ where $X_{2i} = D_i$ and $X_{3i} = 1 - D_i$. Then, we know that $X_{2i} + X_{3i} = D_i + 1 - D_i = 1$.

Imperfect Multicollinearity

Imperfect Multicollinearity

Imperfect multicollinearity means that two or more of the regressors are highly correlated with another regressor.

Since it is not perfect multicollinearity, there is no problem in computing the OLS estimators.

But, the standard error of OLS estimators is larger under imperfect multicollinearity than under no imperfect multicollinearity.

- $Y_1 = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \Rightarrow \sigma^2_{\hat{\beta}_1} = \frac{1}{n}(\frac{1}{1-\rho^2_{X_1,X_2}})\frac{\sigma^2_u}{\sigma^2_{X_1}}$ under homoskedasticity.

  Imperfect multicollineaity(or near perfect multicollinearity) give you less precise OLS estimator.

Imperfect Multicollinearity (cont.)

- Example 1:

  1. Y: testscore
  2. X: student-teacher ratio, percent of English learners in the school, number of first generation immigrants in the district.

     Note: high correlation between the percent of English learners and the number of first generation immigrants

- Example 2:

  1. Y: income,
  2. X: years of schooling, job experience, age, and sex

     Note: high correlation between the job experience and the age

Multiple Linear Regression with Vetors and Matrices

Regression equation

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_K x_{iK} + u_i = x_i'\beta + u_i, \text{ for } i = 1, \ldots, n \quad (1)$$

Note that x_i and θ are column vectors. That is, $x_i = [x_{i1}\ x_{i2}\ \ldots x_{iK}]'$ where "′" indicates a transposition.
If you stack y_i and x_i and u_i over observation i, then you have a matrix form as follows.

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u} \quad (2)$$

Note

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1K} \\ x_{21} & x_{22} & \ldots & x_{2K} \\ \ldots & & & \\ x_{n1} & \ldots & & x_{nK} \end{bmatrix} = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \text{ and } \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}.$$

Note that $x_i \in \mathbb{R}^K$, $y_i \in \mathbb{R}$ and parameter vector $\beta \in \mathbb{R}^K$.

OLS Estimators of Multiple Linear Regression with Vetors and Matrices

- $\mathbf{y} = \mathbf{X}\beta + \mathbf{u} \Rightarrow \mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\beta + \mathbf{X}'\mathbf{u} \Rightarrow \frac{\mathbf{1}}{\mathbf{n}}\mathbf{X}'\mathbf{y} = \frac{\mathbf{1}}{\mathbf{n}}\mathbf{X}'\mathbf{X}\beta + \frac{\mathbf{1}}{\mathbf{n}}\mathbf{X}'\mathbf{u}$
- Taking Expectation given $\mathbf{X}$, then $\mathbf{E}(\frac{\mathbf{1}}{\mathbf{n}}\mathbf{X}'\mathbf{y}) = \mathbf{E}(\frac{\mathbf{1}}{\mathbf{n}}\mathbf{X}'\mathbf{X})\beta + \mathbf{E}(\frac{\mathbf{1}}{\mathbf{n}}\mathbf{X}'\mathbf{u})$.
- Note:
    1. $E\left[\frac{1}{n}\mathbf{X}'\mathbf{X}\right] = E\left[\frac{1}{n}\sum_i^n x_i x_i'\right] = E(x_1 x_1')$ by identical random variable.
    2. $E\left[\frac{1}{n}\mathbf{X}'\mathbf{X}\right] = E\left[\frac{1}{n}\sum_i^n x_i y_i\right] = E(x_1 y_1)$ by identical random variable.
    3. $E\left[\frac{1}{n}\mathbf{X}'\mathbf{u}\right] = E\left[\frac{1}{n}\sum_i^n x_i u_i\right] = E(x_1 u_1) = 0$ by identical random variable and OLS assumption $E(u_1|x_1) = 0$.

- No multicollinearity assumption means that the matrix $\frac{1}{n}\mathbf{X}'\mathbf{X}$(or $E[x_1 x_1']$) is invertible. Hence,
- $\beta = (E(\frac{\mathbf{1}}{\mathbf{n}}\mathbf{X}'\mathbf{X}))^{-1}E(\frac{1}{n}\mathbf{X}'\mathbf{y}) = E(x_1 x_1')^{-1}E(x_1 y_1)$.
- $\hat{\beta} = \widehat{E(x_1 x_1')}^{-1}\widehat{E(x_1 y_1)} = (\frac{1}{n}\sum_i^n x_i x_i')^{-1}\frac{1}{n}\sum_i^n x_i y_i = (\frac{\mathbf{1}}{\mathbf{n}}\mathbf{X}'\mathbf{X})^{-1}(\frac{\mathbf{1}}{\mathbf{n}}\mathbf{X}'\mathbf{y})$ $= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ by sample analogy.

Asymptotics of OLS estimator $\hat\beta = (\hat\beta_1, ..., \hat\beta_K)'$

- $\hat\beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$.
- $\sqrt{n}(\hat\beta - \beta) = \sqrt{n}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} = (\frac{\mathbf{1}}{\mathbf{n}}\mathbf{X}'\mathbf{X})^{-1}\frac{1}{\sqrt{n}}\mathbf{X}'\mathbf{u}$
  $= (\frac{1}{n}\sum_i^n x_i x_i')^{-1}\frac{1}{\sqrt{n}}\sum_i^n x_i u_i$.

  Note that
  1. $\frac{1}{n}\sum_i^n x_i x_i' \xrightarrow{p} E(x_1 x_1')$ by LLN.
  2. $\frac{1}{\sqrt{n}}\sum_i^n x_i u_i \xrightarrow{d} N(0, E(u_1^2 x_1 x_1'))$ by CLT.

- Therefore, $\sqrt{n}(\hat\beta - \beta) \xrightarrow{d} N\left(0, V_{\sqrt{n}(\hat\beta-\beta)}\right)$

  where $V_{\sqrt{n}(\hat\beta-\beta)} = E(x_1 x_1')^{-1}E(u_1^2 x_1 x_1')E(x_1 x_1')^{-1}$.

- Sample Counterparts:
  $\hat V_{\sqrt{n}(\hat\beta-\beta)} = (\frac{1}{n}\sum_i^n x_i x_i')^{-1}\frac{1}{n}\sum_{i=1}^n \hat u_i^2 x_i x_i'(\frac{1}{n}\sum_i^n x_i x_i')^{-1}$
  $= (\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}\frac{1}{n}\sum_{i=1}^n \hat u_i^2 x_i x_i'(\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}$. Hence,
  $\hat V_{(\hat\beta-\beta)} = \frac{1}{n}(\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}\frac{1}{n}\sum_{i=1}^n \hat u_i^2 x_i x_i'(\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}$, $\hat u_i = y_i - x_i'\hat\beta$.

- If the error terms are homoskedastic (i.e., $V(u_i|x_i) = \sigma_u^2$),

$$V_{\sqrt{n}(\hat{\beta}-\beta)} = E(u_1^2)E(x_1x_1')^{-1} = \sigma_u^2 E(x_1x_1')^{-1}.$$

$$V_{(\hat{\beta}-\beta)} = \frac{1}{n}E(u_1^2)E(x_1x_1')^{-1} = \frac{1}{n}\sigma_u^2 E(x_1x_1')^{-1}.$$

- Sample Counterparts:

$$\hat{V}_{\sqrt{n}(\hat{\beta}-\beta)} = \hat{\sigma}_u^2(\frac{1}{n}\sum_i^n x_ix_i')^{-1} = (\frac{1}{n}\sum_{i=1}^n \hat{u}_i^2)(\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}.$$

$$\hat{V}_{(\hat{\beta}-\beta)} = \frac{1}{n}\hat{\sigma}_u^2(\frac{1}{n}\sum_i^n x_ix_i')^{-1} = \frac{1}{n}(\frac{1}{n}\sum_{i=1}^n \hat{u}_i^2)(\frac{1}{n}\mathbf{X}'\mathbf{X})^{-1}.$$

- $SE(\beta_k)$ is $\sqrt{\frac{\hat{V}^{kk}_{\sqrt{n}(\hat{\beta}-\beta)}}{n}}$ where $\hat{V}^{kk}_{\sqrt{n}(\hat{\beta}-\beta)}$ is k_{th} diagonal element of $\hat{V}_{\sqrt{n}(\hat{\beta}-\beta)}$.
- Alternatively, $SE(\beta_k)$ can be expressed as $\sqrt{\hat{V}^{kk}_{(\hat{\beta}-\beta)}}$ where $\hat{V}^{kk}_{(\hat{\beta}-\beta)}$ is k_{th} diagonal element of $\hat{V}_{(\hat{\beta}-\beta)}$.

Test in Multiple Regression

Test of Single Restriction

$H_0 : \beta_k = c$ vs $H_1 : \beta_k \neq c$. Apply the t-test (or p-value).

Test of Multiple Restrictions: F- test

Suppose there are $q(< K)$ restrictions.

$$H_0 : \beta_1 = c_1, \beta_2 = c_2, ..., \beta_q = c_q$$

versus

$$H_1 : \text{ at least one } \beta_i \neq c_i \text{ for } i = 1, ..., q.$$

Apply F-test since the test statistic here F-statistic follows $F_{q,\infty} = \chi^2_q / q$. Use the critical value in F-stat table depending on your significance level.

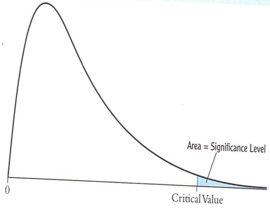
χ^2 table

TABLE 3 Critical Values for the χ^2 Distribution

Degrees of Freedom	Significance Level 10%	5%	1%
1	2.71	3.84	6.63
2	4.61	5.99	9.21
3	6.25	7.81	11.34
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.64	12.59	16.81
7	12.02	14.07	18.48
8	13.36	15.51	20.09
9	14.68	16.92	21.67
10	15.99	18.31	23.21
11	17.28	19.68	24.72
12	18.55	21.03	26.22
13	19.81	22.36	27.69
14	21.06	23.68	29.14
15	22.31	25.00	30.58
16	23.54	26.30	32.00
17	24.77	27.59	33.41
18	25.99	28.87	34.81
19	27.20	30.14	36.19
20	28.41	31.41	37.57
21	29.62	32.67	38.93
22	30.81	33.92	40.29
23	32.01	35.17	41.64
24	33.20	36.41	42.98
25	34.38	37.65	44.31
26	35.56	38.89	45.64
27	36.74	40.11	46.96
28	37.92	41.34	48.28
29	39.09	42.56	49.59

$F_{m,\infty}$ table

TABLE 4 Critical Values for the $F_{m,\infty}$ Distribution



Degrees of Freedom	Significance Level 10%	5%	1%
1	2.71	3.84	6.63
2	2.30	3.00	4.61
3	2.08	2.60	3.78
4	1.94	2.37	3.32
5	1.85	2.21	3.02
6	1.77	2.10	2.80
7	1.72	2.01	2.64
8	1.67	1.94	2.51
9	1.63	1.88	2.41
10	1.60	1.83	2.32
11	1.57	1.79	2.25
12	1.55	1.75	2.18
13	1.52	1.72	2.13
14	1.50	1.69	2.08
15	1.49	1.67	2.04
16	1.47	1.64	2.00
17	1.46	1.62	1.97
18	1.44	1.60	1.93
19	1.43	1.59	1.90
20	1.42	1.57	1.88
21	1.41	1.56	1.85
22	1.40	1.54	1.83
23	1.39	1.53	1.81
24	1.38	1.52	1.79
25	1.38	1.51	1.77
26	1.37	1.50	1.76
27	1.36	[illegible]	[illegible]

STATA Practice

- Suppose the true model is
  $testscr = \beta_0 + \beta_1 str + \beta_2 pct\_EL + error\ \ term$
  reg testscr str el_pct , vce(r)
- Single reressor model $testscr = \beta_0 + \beta_1 str + error\ \ term$
  reg testscr str, vce(r) [Compare both $\hat{\beta}_1$!]
- See adjusted R^2
  reg testscr str el_pct , vce(r)
  ereturn list [Look for e(r2_a) Or]
  display e(r2_a)
- If you want to see the variance of covariance of $\hat{\beta}$
  vce

STATA Practice (cont.)

- F-test $H_0 : \beta_1 = c_1, \beta_2 = c_2$ vs $H_1 : \beta_i \neq c_1$ or $\beta_2 \neq c_2$
- **test (name of x1==** c_1**) (name of x2==** c_2**)**
- Estimate the equation first: reg testscr str el_pct, vce(r)
- For the test of $H_0 : \beta_1 = 2$ vs. $H_1 : \beta_1 \neq 2$, do t-test .

  test str==2 or test str
- For the test of $H_0 : \beta_1 = 2, \beta_2 = 3$ vs. $H_1 : \beta_1 \neq 2$ or $\beta_2 \neq 3$, do F-test.

  test (str==2) (el_pct==3)
- Question: $H_0 : \beta_1 = \beta_2$ vs $H_0 : \beta_1 \neq \beta_2$. t-test or F-test? How to test in STATA?