Econ 210 - Multivariate Linear Regression

Sidharth Sah[1]

September 8, 2023

[1]Thanks to Azeem Shaikh and Max Tabord-Meehan for useful material

Introduction

- We will now expand our discussion of linear regression to allow for multiple regressors - i.e. multiple X variables
- This will open up new possibilities including "controlling" for confounding variables
- In order to do so, we will need to use matrix notation

Linear Algebra Review

- An $n \times m$ *matrix* is a rectangular array of numbers with n rows and m columns. If A is a matrix, then component $a_{i,j}$ of A is the number in the ith row and jth column of A
- An $n \times m$ *matrix* is *square* if $n = m$
- A square matrix is *symmetric* if $a_{i,j} = a_{j,i} \ \forall \ i, j$
- A square matrix is diagonal if $a_{i,j} = 0$ if $i \neq j$
- The diagonal matrix with all 1's along the diagonal and n rows and columns is called the *identity matrix*, denoted I_n

Linear Algebra Review

- An $n \times 1$ matrix is called an n-dimensional *column vector* and a $1 \times m$ matrix is called an m-dimensional *row vector*
- If A is an $n \times m$ matrix, its *transpose*, A', is equivalent to an $m \times n$ matrix C, where $c_{i,j} = a_{j,i}$
- If A and B are matrices of the same size, then for $C = A + B$, $c_{i,j} = a_{i,j} + b_{i,j}$

Linear Algebra Review

- For $n \times k$ matrix A and $k \times m$ matrix B, $C = AB$ is an $n \times m$ matrix with $c_{ij} = \sum_{l=1}^{k} a_{i,l} b_{l,j}$
- Note that matrix multiplication is not commutative - i.e. it need **not** be the case that $AB = BA$
- Note that matrix multiplication and addition are both continuous functions
- For n-dimensional column vectors $a = (a_1, ..., a_n)'$, $b = (b_1, ..., b_n)'$, $a'b = \sum_{i=1}^{n} a_i b_i$ is called the *inner product*

Invertible Matrices

- Say we want to solve

$$Ax = b$$

for x, where A is $n \times n$ and both x and b are $n \times 1$. If A is Invertible, meaning there exists matrix A^{-1} such that $AA^{-1} = A^{-1}A = I_n$, then there is a solution:

$$A^{-1}Ax = A^{-1}b \Rightarrow x = A^{-1}b$$

Linear Independence

- We know that A is invertible when the columns of A are Linearly Independent
- Suppose that $a_1, \ldots, a_k$ are n-dimensional vectors. The vectors are Linearly Dependent if there exist a set of scalars $c_1, \ldots, c_k$, that are not all 0, such that:

$$\sum_{i=1}^{k} c_i a_i = 0$$

If no such set of scalars exists, then $a_1, \ldots, a_k$ are linearly independent

Random Matrices

- A Random Matrix is a matrix whose elements are random variables
- For random matrix X, $E[X]$ is the matrix of the expectations
- For random matrices X, Y and non-random matrices A, B:

  (i) $E[AX + B] = AE[X] + B$

  (ii) $E[X + Y] = E[X] + E[Y]$

  so long as the relevant operations are defined

Variances of Random Vectors

- For random, n-dimensional column vector X,

$$Var(X) = E[(X - E[X])(X - E[X])']$$

- Note that this is an $n \times n$ matrix, where the element in the ith row and jth column is $Cov(X_i, X_j)$
- For non-random matrix A, non-random column vector b, and random column vector x,

$$Var(Ax + b) = AVar(X)A'$$

so long as the relevant operations are defined

Law of Large Numbers

- The Law of Large Numbers generalizes to random matrices. Let $X_1, ..., X_n$ be iid $\sim X$ where X is a random matrix. Suppose the second moment of each element of X exists. Then,
$$\overline{X}_n \xrightarrow{p} E[X]$$
- This follows immediately from the univariate LLN - we take the sample mean of each component of the X_i's, and we know each of those converge

Continuous Mapping Theorem

- This will again be used in conjunction with the <u>Continuous Mapping Theorem</u>: Suppose $X_n \xrightarrow{p} x$ and $Y_n \xrightarrow{p} y$. For any g that is continuous and defined at (x, y),
$$g(X_n, Y_n) \xrightarrow{p} g(x, y)$$
- This holds for random matrices x, y and again holds for any finite number of arguments for g
- Note that matrix multiplication, matrix addition, and taking inverses of invertible matrices are all continuous functions

Central Limit Theorem

- The Central Limit Theorem also generalizes. Let $X_1, ..., X_n$ be iid $\sim X$ where X is random column vector. Suppose the second moment of each element of X exists. Then,

$$\sqrt{n}(\overline{X}_n - E[X]) \xrightarrow{d} N(0, Var(X))$$

- Where in this case $N(0, Var(X))$ is the multivariate normal distribution

Multivariate Normal Distribution

- Say that $X \sim N(m, V)$ and we have non-random A and b. Then, $Ax + b$ is also normal, if those operations are defined
- $Ax + b$ has the distribution:

$$Ax + b \sim N(Am + b, AVA')$$

Slutsky's Lemma

- We'll again frequently use the CLT in conjunction with Slutsky's Lemma: Suppose $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} y$, where y is non-random. Then,

  (i) $X_n Y_n \xrightarrow{d} Xy$

  (ii) $X_n + Y_n \xrightarrow{d} X + y$

  (iii) $X_n Y_n^{-1} \xrightarrow{d} Xy^{-1}$

  whenever the relevant operations are defined

MLR Setting

- Let $(Y, X_1, ..., X_k, U)$ be a rv and $\beta_0, ..., \beta_k$ be parameters such that:

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k + U$$

- Defining $X = (1, X_1, ..., X_k)'$ (will sometimes call $X_0 = 1$) and $\beta = (\beta_0, \beta_1, ..., \beta_k)'$ allows us to re-express this as:

$$Y = X'\beta + U$$

Interpretations of MLR

- We'll continue to interpret this regression equation using the same three interpretations we've seen:
  (i) Linear conditional expectation
  (ii) Best linear predictor/best linear approximation to conditional expectation
  (iii) Causal model

Linear Conditional Expectation Interpretation

- Suppose that $E[Y|X] = X'\beta$. Then, we'll **define** $U = Y - E[Y|X]$, so by construction:

$$Y = X'\beta + U$$

- Then,

$$E[U|X] = E[UX] = E[U] = 0$$

- As X is a vector, the middle statement implies $E[UX_j] = 0 \ \forall \ j$
- Again, a *descriptive* interpretation

Best Linear Predictor/Best Linear Approximation to Conditional Expectation Interpretation

- Define β as satisfying

$$\min_b E[(Y - X'b)^2]$$

- As before, this β will be equivalent to solving

$$\min_b E[(E[Y|X] - X'b)^2]$$

- **Define** $U = Y - X'\beta$. Then, by construction:

$$Y = X'\beta + U$$

- Again a *descriptive* interpretation

U Under the Best Linear Predictor/Best Linear Approximation to Conditional Expectation Interp

- Taking first-order conditions of the first minimization problem with respect to each of the k components of b will suggest:

$$\frac{d}{db_j}E[(Y - X'b)^2] = E[-2X_j(Y - X'b)]$$
$$= E[X_jU] = 0 \; \forall \; 0 \leq j \leq k$$

- Thus, can say $E[XU] = E[U] = 0$

Causal Model Interpretation

- Suppose $Y = g(X, U)$ where g is a causal model, X are the observed determinants, and U are the unobserved determinants
- The causal effect of X_j on Y holding all else equal is $\frac{dY}{dX_j}$
- If we suppose that

$$g(X, U) = X'\beta + U$$

then $\frac{dY}{dX_j} = \beta_j$

U Under the Causal Model Interpretation

- Can still always normalize β_0 such that $E[U] = 0$
- However, any statement regarding $E[U|X]$ or $E[X_j U]$ is a *substantive* assumption about the data

Recovering Causal β_1 Using Controls

- Suppose we have a causal model:

$$Y = \beta_0 + \beta_1 D + U$$

  where $E[DU] \neq 0$
- Then, the simple linear regression $\hat{\beta}_1$ is not consistent for the causal parameter β_1
- Now suppose we have some additional variable X such that,

$$E[U|D, X] = E[U|X]$$

  aka U is mean independent of D *conditional* on X. Now we *can* get at causal β_1 (if $E[U|X]$ is linear)

Recovering Causal β_1 Using Controls

- Further assume that $E[U|X]$ is linear, so we can represent it as $E[U|X] = \gamma_0 + \gamma_2 X$ (where γ_0, γ_2 are descriptive parameters)
- Then, define

$$
\begin{aligned}
V &= U - E[U|D, X] \\
&= U - E[U|X] \\
&= U - \gamma_0 - \gamma_2 X
\end{aligned}
$$

Recovering Causal β_1 Using Controls

- Plug this into the original causal model:

$$Y = \beta_0 + \beta_1 D + V + \gamma_0 + \gamma_2 X$$
$$Y = \underbrace{\beta_0 + \gamma_0}_{=\tilde{\beta}_0} + \underbrace{\beta_1}_{=\tilde{\beta}_1} D + \underbrace{\gamma_2}_{=\tilde{\beta}_2} X + V$$
$$Y = \tilde{\beta}_0 + \tilde{\beta}_1 D + \tilde{\beta}_2 X + V$$

- Now have a multivariate regression where $E[V|D,X] = 0$. (As will show later), this means we can estimate $\tilde{\beta}_1$ consistently, where $\tilde{\beta}_1$ is equivalent to the original, causal β_1
- Note that $\tilde{\beta}_2$ is **not** causal - it is equivalent to γ_2, which we only ever defined descriptively

Interaction Effects

- Interaction Effects allow us to capture how certain variables may "interact" with one another in affecting another variable
- Assume we have a causal model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U$$

where

$$\begin{aligned} Y &= \text{house price} \\ X_1 &= \text{square footage} \\ X_2 &= \begin{cases} 1 \text{ if in a city} \\ 0 \text{ if otherwise} \end{cases} \end{aligned}$$

Interaction Effects

- This model implies that the effect of house size on price doesn't depend on the house's location - it's always just β_1. We could think this is quite odd - if space if more at a premium in cities, shouldn't there be a greater effect of house size on price in cities?
- NB: We're using the causal model interp for expositional clarity - we can imagine going through a similar exercise for descriptive parameters

Interaction Effects

- We can allow for flexibility by running two regressions - one for houses that are in cities (X_1^{City}, Y^{City}) and one for houses that are not (X_1^N, Y^N):

$$Y^{City} = \beta_0^{City} + \beta_1^{City} X_1^{City} + U$$
$$Y^N = \beta_0^N + \beta_1^N X_1^N + U$$

- Now we have two parameters on size where $\beta_1^{City} \neq \beta_1^N$ would indicate that the effect of size on price varies by location

Interaction Effects

- A more concise and equivalent solution would be to run one regression that includes the interaction X_1X_2:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + U$$

- Now we'll have that the effect of size on price is:

$$\beta_1 + \beta_3 X_2 = \begin{cases} \beta_1 + \beta_3 = \beta_1^{City} \text{ if } X = 1 \\ \beta_1 = \beta_1^{N} \text{ if } X = 0 \end{cases}$$

- This idea generalizes to continuous X_2 - would have a continuum of effects of X_1, $\beta_1 + \beta_3 X_2$, varying across all possible values of X_2

Requirements to Calculate β

- We now turn to discussion of how to calculate β from moments of X and Y, where X and Y satisfy:

  $$Y = X'\beta + U$$

  For this, we will maintain the following assumptions:

  (a) $E[XU] = 0$
  (b) $E[XX'] < \infty$
  (c) No perfect colinearity in X

- We are agnostic about how we arrive at (a)
- (b) is analogous to the prior assumption of $Var(X) < \infty$. (c) is the analogue to the assumption of $0 < Var(X)$

Perfect Colinearity

- Perfect Colinearity occurs if all of the components of X are linearly related, ie there exists a $c \neq 0$ such that:

$$c'X = 0$$

or, equivalently, there exists $0 \leq j \leq k$ such that:

$$X_j = c_0 X_0 + ... + c_{j-1} X_{j-1} + c_{j+1} X_{j+1} + ... + c_k X_k$$

Perfect Colinearity

- If there is **no** perfect colinearity, $E[XX']$ is invertible, so we can solve for β (given the other assumps). To prove this, suppose X is not perfectly colinear but $E[XX']$ is not invertible. Then, by definition of invertibility, there exists $c \neq 0$ such that

$$\begin{aligned} E[XX']c &= 0 \\ c'E[XX']c &= 0 \\ E[(c'X)(X'c)] &= 0 \\ E[(c'X)^2] &= 0 \\ c'X &= 0 \text{ (always)} \end{aligned}$$

- The last line is a contradiction, so it then follows that not being perfectly colinear implies $E[XX']$ is invertible

Perfect Colinearity

- Avoiding colinearity is almost never a “serious” problem - just have to be careful about combinations of certain kinds of variables
- Say we’re interested in immigration, and want to include in our regression:

$$X_1 = \begin{cases} 1 \text{ if born in US} \\ 0 \text{ if not} \end{cases} \qquad X_2 = \begin{cases} 1 \text{ if born outside US} \\ 0 \text{ if not} \end{cases}$$

- Including both (and $X_0 = 1$) induces colinearity, because:

$$X_1 = 1 - X_2$$

- Only keep one of them

Calculating β

- Under our new maintained assumptions, we can now calculate β from moments of X and Y. Specifically, we plug the definition of U into $E[XU] = 0$:

$$
\begin{aligned}
E[X(Y - X'\beta)] &= 0 \\
E[XY] - E[XX']\beta &= 0 \\
E[XY] &= E[XX']\beta \\
\Rightarrow \beta &= E[XX']^{-1}E[XY]
\end{aligned}
$$

- Notice that we made use of the no perfect colinearity assumption in the last line

Requirements to Estimate β

- We now discuss how to estimate β from finite samples. For this, we'll continue to maintain the following assumptions:
  (a) $E[XU] = 0$
  (b) $E[XX'] < \infty$
  (c) No perfect colinearity in X
- We also add that $(X_1, Y_1), ..., (X_n, Y_n)$ are iid $\sim (X, Y)$ where X and Y satisfy: $Y = X'\beta + U$

Estimating β

- From the above, a natural estimator for β is given by:

$$\hat{\beta}_n = (\frac{1}{n}\sum_{i=1}^{n} X_i X_i')^{-1}\frac{1}{n}\sum_{i=1}^{n} X_i Y_i$$

- This is the Ordinary Least Squares estimator of β

Ordinary Least Squares

- The name ordinary least squares (again) indicates that $\hat{\beta}_n$ solves:

$$\min_b \frac{1}{n} \sum_{i=1}^{n} (Y_i - X_i' b)^2$$

- Thus, the OLS estimator will also satisfy the following first-order conditions of the minimization problem:

$$\frac{1}{n} \sum_{i=1}^{n} X_i (Y_i - X_i' b) = 0$$

Residuals

- We again call:

$$\hat{Y}_i = X_i'\hat{\beta}_n$$

  the fitted or predicted values. The amounts these are off by are called the Residuals:

$$\hat{U}_i = Y_i - \hat{Y}_i = Y_i - X_i'\hat{\beta}_n$$

- By implication of the previous FOC's,

$$\sum_{i=1}^{n} X_i \hat{U}_i = 0$$

Residuals

- We again call:

$$\hat{Y}_i = X_i'\hat{\beta}_n$$

the fitted or predicted values. The amounts these are off by are called the Residuals:

$$\hat{U}_i = Y_i - \hat{Y}_i = Y_i - X_i'\hat{\beta}_n$$

- By implication of the previous FOC's,

$$\sum_{i=1}^{n} X_i \hat{U}_i = 0$$

R^2

- Can still use R^2 to measure the quality of fit of a regression:
$$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS}$$
where ESS, TSS, and SSR are defined identically to before
- It will still be the case that $0 \leq R^2 \leq 1$, with $R^2 = 0$ and $R^2 = 1$ having the same implications as before
- It is still the case that R^2 is a descriptive measure, and doesn't have anything to do with validating causal models

R^2

- However, there is now a new concern - R^2 will mechanically increase with additional regressors
- Remember that we said that $\hat{\beta}_n$ satisfies:

$$\min_b \frac{1}{n} \sum_{i=1}^{n} (Y_i - X_i' b)^2$$

- By the definitions of residuals and SSR, this means that:

$$SSR = \sum_{i=1}^{n} \hat{U}_i^2 = \min_b \sum_{i=1}^{n} (Y_i - X_i' b)^2$$

R^2

- Now imagine running a regression with k regressors and calling the residuals from that regression $\hat{U}_i$. Run another regression with the same k regressors plus one additional regressors, and calling the new residuals $\hat{U}_i^*$. Then,

$$\min_b \sum_{i=1}^{n}(Y_i - X_i'b)^2 \geq \min_{b,b_{k+1}} \sum_{i=1}^{n}(Y_i - X_i'b - X_{i,k+1}b_{k+1})^2$$

$$\sum_{i=1}^{n} \hat{U}_i^2 \geq \sum_{i=1}^{n} \hat{U}_i^{*2}$$

$$SSR \geq SSR^*$$

$$R^2 \leq R^{2*}$$

Adjusted R^2

- For this reason, some people prefer Adjusted R^2, $\overline{R^2}$, which "penalizes" additional regressors:

$$\overline{R^2} = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS}$$

- $\overline{R^2}$ may increase *or* decrease with additional regressors
- Caveat: $\overline{R^2} < 1$, but $\overline{R^2}$ can be negative (unlike R^2)

Unbiasedness of OLS

- Along with the normal maintained assumptions, assume that $E[U|X] = 0$. Then, the OLS estimator is unbiased:

$$E[\hat{\beta}_n] = \beta$$

Unbiasedness of OLS

- Start by taking a transformation of $\hat{\beta}_n$ so that β is in the expression:

$$\begin{aligned}\hat{\beta}_n &= (\frac{1}{n}\sum_{i=1}^{n} X_i X_i')^{-1}\frac{1}{n}\sum_{i=1}^{n} X_i Y_i \\ &= (\frac{1}{n}\sum_{i=1}^{n} X_i X_i')^{-1}\frac{1}{n}\sum_{i=1}^{n} X_i(X_i'\beta + U_i) \\ &= \beta + (\frac{1}{n}\sum_{i=1}^{n} X_i X_i')^{-1}\frac{1}{n}\sum_{i=1}^{n} X_i U_i\end{aligned}$$

- Next we'll show that $E[\hat{\beta}_n|X_1, ..., X_n] = \beta$

Unbiasedness of OLS

$$\begin{aligned}
E[\hat{\beta}_n|X_1,...,X_n] &= \beta + E[(\frac{1}{n}\sum_{i=1}^{n} X_i X_i')^{-1}\frac{1}{n}\sum_{i=1}^{n} X_i U_i|X_1,...,X_n] \\
&= \beta + (\frac{1}{n}\sum_{i=1}^{n} X_i X_i')^{-1}\frac{1}{n}\sum_{i=1}^{n} X_i E[U_i|X_i] \\
&\qquad ((Y_i, X_i) \perp (Y_j, X_j),\ i \neq j) \\
&= \beta + (\frac{1}{n}\sum_{i=1}^{n} X_i X_i')^{-1}\frac{1}{n}\sum_{i=1}^{n} X_i E[U|X] \\
&\qquad ((Y_i, X_i) \sim (Y, X)) \\
&= \beta
\end{aligned}$$

Unbiasedness of OLS

- Finally, we can close the bow with the Law of Iterated Expectations:

$$
\begin{aligned}
E[\hat{\beta}_n] &= E[E[\hat{\beta}_n | X_1, ..., X_n]] \\
&= E[\beta] \\
&= \beta
\end{aligned}
$$

Consistency of OLS

- Along with the normal maintained assumptions, assume that $E[Y^4], E[X_j^4] < \infty$. (Can drop the assumption of $E[U|X] = 0$. Assuming $E[XU] = 0$ is good enough for consistency). Then, the OLS estimator is consistent:

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

- To get convergence in probability, we'll show each "piece" converges with the WLLN and then "stitch them back together" with CMT

Consistency of OLS

- Because we have an iid sample and the relevant moment conditions are satisfied, by WLLN:

$$\frac{1}{n}\sum_{i=1}^{n} X_i X_i' \xrightarrow{p} E[XX'] \text{ and } \frac{1}{n}\sum_{i=1}^{n} X_i Y_i' \xrightarrow{p} E[XY]$$

- Because $E[XX']^{-1}$ is invertible (b/c of no perfect colinearity), $E[XX']^{-1}E[XY]$ is a continuous function, so, by CMT:

$$\hat{\beta}_n = (\frac{1}{n}\sum_{i=1}^{n} X_i X_i')^{-1}\frac{1}{n}\sum_{i=1}^{n} X_i Y_i \xrightarrow{p} E[XX']^{-1}E[XY] = \beta$$

Omitted Variable Bias

- Suppose we had a causal model where:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U$$

  and $E[U] = 0$, $E[X_1 U] = 0$, and $E[X_2 U] = 0$ - we can estimate the parameters consistently
- Say we instead estimated:

$$Y = \beta_0^* + \beta_1^* X_1 + U^*$$

- What would β_1^* converge to?

Omitted Variable Bias

- From our SLR notes, we know that:

$$\beta_1^* \xrightarrow{p} \frac{Cov(X_1, Y)}{Var(X_1)}$$

- We can see that:

$$\begin{aligned} Cov(X_1, Y) &= Cov(X_1, \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U) \\ &= \beta_1 Var(X_1) + \beta_2 Cov(X_1, X_2) \\ \Rightarrow \beta_1^* &\xrightarrow{p} \beta_1 + \beta_2 \frac{Cov(X_1, X_2)}{Var(X_1)} \end{aligned}$$

- The second term is called Omitted Variable Bias, and its sign depends on the sign of $\beta_2 Cov(X_1, X_2)$

Limiting Distribution of OLS

- Along with the normal maintained assumptions, assume that $E[Y^4], E[X_j^4] < \infty, \forall\, j$. Then:

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} N(0, \Sigma)$$

where

$$\Sigma = E[XX']^{-1} Var(XU) E[XX']^{-1}$$

- To get convergence in distribution, we'll show each "piece" converges using CLT and WLLN and the "stitch them back together" with Slutsky

Limiting Distribution of OLS

- During proof of unbiasedness, we showed that:

$$\hat{\beta}_n = \beta + (\frac{1}{n}\sum_{i=1}^{n} X_i X_i')^{-1}\frac{1}{n}\sum_{i=1}^{n} X_i U_i$$

- This implies that:

$$\sqrt{n}(\hat{\beta}_n - \beta) = \underbrace{(\frac{1}{n}\sum_{i=1}^{n} X_i X_i')^{-1}}_{=A}\underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i U_i}_{=B}$$

Limiting Distribution of OLS

- $E[XU] = 0$ by assumption, so, using the CLT:

$$B = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i U_i = \sqrt{n}(\frac{1}{n} \sum_{i=1}^{n} X_i U_i - E[XU])$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i U_i \xrightarrow{d} N(0, Var(XU)) \qquad \text{(CLT)}$$

- We showed during proof of consistency that, by WLLN,

$$\frac{1}{n} \sum_{i=1}^{n} X_i X_i' \xrightarrow{p} E[XX']$$

Limiting Distribution of OLS

- Use Slutsky to recombine:

$$\sqrt{n}(\hat{\beta}_n - \beta) = \underbrace{(\frac{1}{n}\sum_{i=1}^{n} X_i X_i')^{-1}}_{=A} \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i U_i}_{=B}$$

$$\xrightarrow{d} E[XX']^{-1} N(0, Var(XU))$$

$$\xrightarrow{d} N(0, \underbrace{E[XX']^{-1} Var(XU) E[XX']^{-1}}_{=\Sigma})$$

Inference on OLS

- We want to test hypotheses/do inference about β. We'll focus on testing hypotheses regarding individual components of β (ex. $H_0 : \beta_j = 0$)
- In order to make the limiting distribution useful, we need to be able to estimate Σ. Under the same assumptions that gave us the limiting distribution, can show that:

$$\hat{\Sigma}_n = (\frac{1}{n}\sum_{i=1}^{n} X_i X_i')^{-1} \frac{1}{n}\sum_{i=1}^{n} X_i X_i' \hat{U}_i^2 (\frac{1}{n}\sum_{i=1}^{n} X_i X_i')^{-1}$$

is a consistent estimator of Σ

Inference on OLS

- Define e_j as a $(k+1) \times 1$ vector with a 1 in the jth position and 0's everywhere else. Thus, $e_j'\beta = \beta_j$

$$
\begin{aligned}
e_j'(\sqrt{n}(\hat{\beta}_n - \beta)) &= \sqrt{n}(\hat{\beta}_j - \beta_j) \\
&\xrightarrow{d} N(0, e_j'\Sigma e_j)
\end{aligned}
$$

  where $e_j'\Sigma e_j$ is a scalar, equal to the element in the jth row and jth column of Σ

Inference on OLS

- By Slutsky, because $\hat{\Sigma}_n$ is consistent:

$$\frac{\sqrt{n}}{\sqrt{e_j'\hat{\Sigma}_n e_j}}(\hat{\beta}_j - \beta_j) \xrightarrow{d} N(0,1)$$

- Call $SE(\hat{\beta}_j) = \sqrt{\frac{e_j'\hat{\Sigma}_n e_j}{n}}$ the standard error of $\hat{\beta}_j$ and we can do all the normal inference, with two-sided test statistic for $H_0 : \beta_j = \beta_{j,0}$:

$$T_n = |\frac{\hat{\beta}_j - \beta_{j,0}}{SE(\hat{\beta}_j)}|$$

and confidence interval at significance level α:

$$[\hat{\beta}_j \pm \Phi^{-1}(1 - \frac{\alpha}{2})SE(\hat{\beta}_j)]$$