ECON 626: Applied Microeconomics

Lecture 2:

Regression Basics and Heteroskedasticity

Professors: Pamela Jakiela and Owen Ozier

One piece of probability/statistics

Variance of mean

Recall: For any **independent** random variables X and Y:

$$Var(X+Y) = Var(X) + Var(Y)$$

Recall: For any constants a and b and random variable X:

$$Var(aX+b) = a^2 Var(X)$$

Thus, for independent random variables X_i each with variance σ^2:

$$Var\left(\sum_{i=1}^{i=N} X_i\right) = N\sigma^2$$

And so

$$Var\left(\frac{1}{N}\sum_{i=1}^{i=N} X_i\right) = \frac{1}{N}\sigma^2$$

Linear Algebra (quick review)

Multiplication

$$k \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} k \cdot a_1 \\ k \cdot a_2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = [1 \cdot 3 + 2 \cdot 4] = [11] = 11$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Multiplication, continued (handout version)

$$\begin{bmatrix} D_{11} & 0 \\ 0 & D_{22} \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} D_{11}a & D_{11}b \\ D_{22}c & D_{22}d \end{bmatrix}$$

Suppose we want $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ to be the **inverse** of $\begin{bmatrix} D_{11} & 0 \\ 0 & D_{22} \end{bmatrix}$.

That is,

$$\begin{bmatrix} D_{11}a & D_{11}b \\ D_{22}c & D_{22}d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Four (simple) equations, four unknowns.

$$\begin{bmatrix} D_{11} & 0 \\ 0 & D_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{D_{11}} & 0 \\ 0 & \frac{1}{D_{22}} \end{bmatrix}$$

So inverting a diagonal matrix is easy. A general formula?

Inverses (2x2 case)

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{\underbrace{ad - bc}_{\text{determinant}}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Transpose

$$\text{Let } A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}. \text{ Let } B = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \text{ AB} = ?$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 15 \end{bmatrix}$$

BA = "conformability error" but B'A' =

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 6 & 15 \end{bmatrix}$$

Thus, B'A' =(AB)'. (Note: A' is sometimes written A^T.)

Regression

Recall the basic regression (estimation) formula

$$\hat{\beta} = (X'X)^{-1}X'y$$

But what is $(X'X)^{-1}$? What does $(X'X)^{-1}X'$ do to y?

Matrices' easy interpretation in "treatment" context

Suppose that we are interested in the relationship between outcome Y_i and a treatment indicator D_i. Regress the outcome on... the treatment indicator and a constant.

$$X_i = [D_i \;\; 1]$$

Suppose that half of N observations have $D_i = 1$ and half have $D_i = 0$.

$$X = \begin{bmatrix} D_1 & 1 \\ D_2 & 1 \\ ... & ... \\ D_{\frac{N}{2}} & 1 \\ D_{\frac{N}{2}+1} & 1 \\ D_{\frac{N}{2}+2} & 1 \\ ... & ... \\ D_N & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ ... & ... \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \\ ... & ... \\ 1 & 1 \end{bmatrix} \; ; \; Y = \begin{bmatrix} Y_1 \\ Y_2 \\ ... \\ Y_{\frac{N}{2}} \\ Y_{\frac{N}{2}+1} \\ Y_{\frac{N}{2}+2} \\ ... \\ Y_N \end{bmatrix}$$

Matrices' easy interpretation in "treatment" context

Recall that we're after

$$\hat{\beta} = (X'X)^{-1}X'y$$

Set up $X'X$:

$$\begin{bmatrix} 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 & 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ \dots & \dots \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \\ \dots & \dots \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{N}{2} & \frac{N}{2} \\ \frac{N}{2} & N \end{bmatrix}$$

Matrices' easy interpretation in "treatment" context

Recall,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

So:

$$\begin{bmatrix} \frac{N}{2} & \frac{N}{2} \\ \frac{N}{2} & N \end{bmatrix}^{-1} = \frac{1}{\frac{N^2}{2} - \frac{N^2}{4}} \begin{bmatrix} N & -\frac{N}{2} \\ -\frac{N}{2} & \frac{N}{2} \end{bmatrix} = \frac{2}{N} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$$

Easy way:

$$\left(\frac{N}{2} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \right)^{-1} = \frac{2}{N} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}^{-1} = \frac{2}{N} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$$

Matrices' easy interpretation in "treatment" context

Recall we wanted to find: $\hat{\beta} = (X'X)^{-1}X'y$. What about $X'y$?

$$\begin{bmatrix} 0 & 0 & ... & 0 & 1 & 1 & ... & 1 \\ 1 & 1 & ... & 1 & 1 & 1 & ... & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ ... \\ Y_{\frac{N}{2}} \\ Y_{\frac{N}{2}+1} \\ Y_{\frac{N}{2}+2} \\ ... \\ Y_N \end{bmatrix} = \begin{bmatrix} \sum_{i=\frac{N}{2}+1}^{N} Y_i \\ \\ \sum_{i=1}^{N} Y_i \end{bmatrix}$$

$$= \begin{bmatrix} \sum_T Y_i \\ \\ \sum_T Y_i + \sum_C Y_i \end{bmatrix}$$

Matrices' easy interpretation in "treatment" context

We can now compute:

$$\hat{\beta} = (X'X)^{-1}X'y$$

$$\frac{2}{N}\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} \sum_T Y_i \\ \\ \sum_T Y_i + \sum_C Y_i \end{bmatrix} = \frac{2}{N}\begin{bmatrix} 2\sum_T Y_i - \sum_T Y_i - \sum_C Y_i \\ \\ -\sum_T Y_i + \sum_T Y_i + \sum_C Y_i \end{bmatrix}$$

$$= \frac{2}{N}\begin{bmatrix} \sum_T Y_i - \sum_C Y_i \\ \\ \sum_C Y_i \end{bmatrix} = \begin{bmatrix} \bar{Y}_T - \bar{Y}_C \\ \\ \bar{Y}_C \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \hat{\beta}$$

We just ran a regression. Now, on to the standard error!
What will the dimensions of the variance-covariance matrix be?

Homoskedastic error

Recall the basic regression (estimation) formula

What is the variance of $\hat{\beta} = (X'X)^{-1}X'y$?
First, re-write $\hat{\beta}$:

$$y = X\beta + e$$

$$\hat{\beta} = (X'X)^{-1}X'y \quad = (X'X)^{-1}X'(X\beta + e)$$
$$= \beta + (X'X)^{-1}X'e$$

Structure of the error term, homoskedasticity

Ways of writing second term, $(X'X)^{-1}X'e$:

$$(X'X)^{-1}X'u \qquad (CT\ 4.11)\ with\ E[u|X] = 0\ (assumption\ ii\ p.73)$$

$$\left[\sum X_i X_i'\right]^{-1} \sum X_i e_i \qquad (AP\ p.45)\ with\ E[X_i e_i] = 0\ (mechanically)$$

Before proceeding to estimate variance, independent observations (CT p.73 assumption ii) (assumptions and implication):

$$E[uu'|X] = \Omega = Diag[\sigma_i^2]$$

Under homoskedasticity, $\sigma_i^2 \equiv \sigma^2\ \forall i$. Thus,

$$\hat{\Omega} = \hat{\sigma}^2 I$$

A reasonable estimator, $\hat{\sigma}^2$, for σ^2: $\frac{1}{N-K} \sum_N u_i^2 = \frac{1}{N-K} \left(\sum_T \hat{u}^2 + \sum_C \hat{u}^2 \right)$.
(In this example, $K = 2$.)

The variance-covariance matrix

For the variance, we write this quadratic form of the estimation error:

$$\left((X'X)^{-1}X'u\right) \ \left((X'X)^{-1}X'u\right)'$$
$$(X'X)^{-1}X'u\ u'X(X'X)^{-1}$$
$$(X'X)^{-1}X'\hat{\Omega}X(X'X)^{-1} \qquad (CT\ 4.21)$$

Under homoskedasticity, our estimate, $\hat{\Omega} = \hat{\sigma}^2 I$.

$$(X'X)^{-1}X'\hat{\sigma}^2 IX(X'X)^{-1}$$
$$(X'X)^{-1}X'IX(X'X)^{-1}\hat{\sigma}^2$$
$$(X'X)^{-1}X'X(X'X)^{-1}\hat{\sigma}^2$$
$$(X'X)^{-1}\hat{\sigma}^2$$
$$\frac{2}{N}\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}\hat{\sigma}^2 = \left(\frac{2}{N(N-K)}\right)\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}\left(\sum_T \hat{u}^2 + \sum_C \hat{u}^2\right)$$

We have an estimator for the basic variance-covariance matrix under homoskedasticity. What about heteroskedasticity?

Heteroskedasticity

Structure of the error term, revisited

Ways of writing second term, $(X'X)^{-1}X'e$:

$$(X'X)^{-1}X'u \qquad (CT\ 4.11)\ with\ E[u|X] = 0\ (assumption\ ii\ p.73)$$

$$\left[\sum X_i X_i'\right]^{-1} \sum X_i e_i \qquad (AP\ p.45)\ with\ E[X_i e_i] = 0\ (mechanically)$$

Before proceeding to estimate variance, independent observations (CT p.73 assumption ii) (assumptions and implication):

$$E[uu'|X] = \Omega = Diag[\sigma_i^2]$$

So a reasonable estimator:

$$\hat{\Omega} = Diag[\hat{u}_i^2]\ (CT\ notation)\ = Diag[\hat{e}_i^2]\ (AP\ notation)$$

Why heteroskedasticity?

For the variance, we write this quadratic form of the estimation error:

$$\begin{aligned}
& \left((X'X)^{-1}X'u\right)\ \left((X'X)^{-1}X'u\right)' \\
& (X'X)^{-1}X'u\ u'X(X'X)^{-1} \\
& {\color{red}(X'X)^{-1}}X'\hat{\Omega}X{\color{red}(X'X)^{-1}} \qquad (CT\ 4.21) \\
= \quad & \left(\sum x_i x_i'\right)^{-1} \sum \hat{u}_i{}^2 x_i x_i' \left(\sum x_i x_i'\right)^{-1} \qquad (CT\ 4.21)
\end{aligned}$$

Note that AP 3.1.7 is written in expectations, in a formulation that leads to the variance of $\sqrt{N} \cdot \hat{\beta}$, just as CT 4.17 does:

$$(\textit{notation swap}) \quad E[X_i X_i']^{-1} E[X_i X_i' e_i^2] E[X_i X_i']^{-1} \qquad (AP\ 3.1.7)$$

Why heteroskedasticity?

$$X'\hat{\Omega}X$$

$$= \begin{bmatrix} 0 & 0 & ... & 0 & 1 & 1 & ... & 1 \\ 1 & 1 & ... & 1 & 1 & 1 & ... & 1 \end{bmatrix} Diag[\hat{u}_i^2] \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ ... & ... \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \\ ... & ... \\ 1 & 1 \end{bmatrix}$$

My big fat greek ... diagonal matrix

$$\hat{\Omega}X = \begin{bmatrix} \hat{u}_1^2 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & \hat{u}_2^2 & \dots & 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \hat{u}_{\frac{N}{2}}^2 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \hat{u}_{\frac{N}{2}+1}^2 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \hat{u}_{\frac{N}{2}+2}^2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & \hat{u}_N^2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ \dots & \dots \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \\ \dots & \dots \\ 1 & 1 \end{bmatrix}$$

Why heteroskedasticity?

$$\hat{\Omega}X = \begin{bmatrix} 0 & \hat{u}_1^2 \\ 0 & \hat{u}_2^2 \\ ... & ... \\ 0 & \hat{u}_{\frac{N}{2}}^2 \\ \hat{u}_{\frac{N}{2}+1}^2 & \hat{u}_{\frac{N}{2}+1}^2 \\ \hat{u}_{\frac{N}{2}+2}^2 & \hat{u}_{\frac{N}{2}+2}^2 \\ ... & ... \\ \hat{u}_N^2 & \hat{u}_N^2 \end{bmatrix}$$

Why heteroskedasticity?

$$X'\hat{\Omega}X = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 & 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} 0 & \hat{u}_1^2 \\ 0 & \hat{u}_2^2 \\ \dots & \dots \\ 0 & \hat{u}_{\frac{N}{2}}^2 \\ \hat{u}_{\frac{N}{2}+1}^2 & \hat{u}_{\frac{N}{2}+1}^2 \\ \hat{u}_{\frac{N}{2}+2}^2 & \hat{u}_{\frac{N}{2}+2}^2 \\ \dots & \dots \\ \hat{u}_N^2 & \hat{u}_N^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=\frac{N}{2}+1}^{N} \hat{u}_i^2 & \sum_{i=\frac{N}{2}+1}^{N} \hat{u}_i^2 \\ \sum_{i=\frac{N}{2}+1}^{N} \hat{u}_i^2 & \sum_{i=1}^{N} \hat{u}_i^2 \end{bmatrix} = \begin{bmatrix} \sum_T \hat{u}^2 & \sum_T \hat{u}^2 \\ \sum_T \hat{u}^2 & \sum_T \hat{u}^2 + \sum_C \hat{u}^2 \end{bmatrix}$$

Why heteroskedasticity?

$$(X'X)^{-1}X'\hat{\Omega}X(X'X)^{-1} \quad (CT\ 4.21)$$

$$\left(\frac{2}{N}\right)\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} \sum_T \hat{u}^2 & \sum_T \hat{u}^2 \\ \sum_T \hat{u}^2 & \sum_T \hat{u}^2 + \sum_C \hat{u}^2 \end{bmatrix}\left(\frac{2}{N}\right)\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$$

$$= \left(\frac{4}{N^2}\right)\begin{bmatrix} \sum_T \hat{u}^2 & \sum_T \hat{u}^2 - \sum_C \hat{u}^2 \\ 0 & \sum_C \hat{u}^2 \end{bmatrix}\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$$

$$= \left(\frac{4}{N^2}\right)\begin{bmatrix} \sum_T \hat{u}^2 + \sum_C \hat{u}^2 & -\sum_C \hat{u}^2 \\ -\sum_C \hat{u}^2 & \sum_C \hat{u}^2 \end{bmatrix}$$

CT p.75: DOF correction w/ empirical (not theoretical) basis, N/(N-K)

All formulas together, before plugging in K.

Estimated coefficients:

$$\hat{\beta} = \begin{bmatrix} \bar{Y}_T - \bar{Y}_C \\ \\ \bar{Y}_C \end{bmatrix}$$

Estimated VCV matrix under homoskedasticity:

$$\left(\frac{2}{N(N-K)}\right) \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \left(\sum_T \hat{u}^2 + \sum_C \hat{u}^2\right)$$

Estimated VCV matrix under heteroskedasticity:

$$\left(\frac{4}{N(N-K)}\right) \begin{bmatrix} \sum_T \hat{u}^2 + \sum_C \hat{u}^2 & -\sum_C \hat{u}^2 \\ \\ -\sum_C \hat{u}^2 & \sum_C \hat{u}^2 \end{bmatrix}$$

All formulas together.

Estimated coefficients:

$$\hat{\beta} = \begin{bmatrix} \bar{Y}_T - \bar{Y}_C \\ \\ \bar{Y}_C \end{bmatrix}$$

Estimated VCV matrix under homoskedasticity:

$$\left(\frac{2}{N(N-2)}\right) \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \left(\sum_T \hat{u}^2 + \sum_C \hat{u}^2\right)$$

Estimated VCV matrix under heteroskedasticity:

$$\left(\frac{4}{N(N-2)}\right) \begin{bmatrix} \sum_T \hat{u}^2 + \sum_C \hat{u}^2 & -\sum_C \hat{u}^2 \\ \\ -\sum_C \hat{u}^2 & \sum_C \hat{u}^2 \end{bmatrix}$$

General case, treating fraction p.

$$(X'X) = N \begin{bmatrix} p & p \\ p & 1 \end{bmatrix}$$

$$(X'X)^{-1} = \frac{1}{p(1-p)N} \begin{bmatrix} 1 & -p \\ -p & p \end{bmatrix}$$

All formulas together, general case.

Estimated coefficients:

$$\hat{\beta} = \begin{bmatrix} \bar{Y}_T - \bar{Y}_C \\ \\ \bar{Y}_C \end{bmatrix}$$

Estimated VCV matrix under homoskedasticity:

$$\left(\frac{1}{p(1-p)N(N-2)}\right) \begin{bmatrix} 1 & -p \\ -p & p \end{bmatrix} \left(\sum_T \hat{u}^2 + \sum_C \hat{u}^2\right)$$

Estimated VCV matrix under heteroskedasticity:

$$\left(\frac{1}{p^2(1-p)^2 N(N-2)}\right) \begin{bmatrix} (1-p)^2 \sum_T \hat{u}^2 + p^2 \sum_C \hat{u}^2 & -p^2 \sum_C \hat{u}^2 \\ \\ -p^2 \sum_C \hat{u}^2 & p^2 \sum_C \hat{u}^2 \end{bmatrix}$$

Focus on the coefficient on treatment. (step 1)

Estimated coefficient: $\hat{\beta}_1 = \bar{Y}_T - \bar{Y}_C$

Estimated variance of $\hat{\beta}_1$ under homoskedasticity:

$$\frac{\sum_T \hat{u}^2 + \sum_C \hat{u}^2}{p(1-p)N(N-2)} = \frac{\sum_T \hat{u}^2 + \sum_C \hat{u}^2}{(N-2)} \left(\frac{1}{p(1-p)N} \right)$$

Estimated variance of $\hat{\beta}_1$ under heteroskedasticity:

$$\frac{(1-p)^2 \sum_T \hat{u}^2 + p^2 \sum_C \hat{u}^2}{p^2(1-p)^2 N(N-2)} =$$

$$\frac{\sum_T \hat{u}^2}{p^2 N(N-2)} + \frac{\sum_C \hat{u}^2}{(1-p)^2 N(N-2)}$$

Focus on the coefficient on treatment. (step 2)

Estimated coefficient: $\hat{\beta}_1 = \bar{Y}_T - \bar{Y}_C$

Estimated variance of $\hat{\beta}_1$ under homoskedasticity:

$$\frac{\sum_T \hat{u}^2 + \sum_C \hat{u}^2}{p(1-p)N(N-2)} = \underbrace{\frac{\sum_T \hat{u}^2 + \sum_C \hat{u}^2}{(N-2)}}_{\textit{variance}} \left(\underbrace{\frac{1}{pN}}_{\textit{averages}} \underbrace{+}_{\textit{difference}} \textit{of} \underbrace{\frac{1}{(1-p)N}}_{\textit{averages}} \right)$$

Estimated variance of $\hat{\beta}_1$ under heteroskedasticity:

$$\frac{(1-p)^2 \sum_T \hat{u}^2 + p^2 \sum_C \hat{u}^2}{p^2(1-p)^2 N(N-2)} =$$

$$\underbrace{\frac{\sum_T \hat{u}^2}{p(N-2)}}_{\textit{variance}} \underbrace{\left(\frac{1}{pN}\right)}_{\textit{averages}} \underbrace{+}_{\textit{difference}} \textit{of} \underbrace{\frac{\sum_C \hat{u}^2}{(1-p)(N-2)}}_{\textit{variance}} \underbrace{\left(\frac{1}{(1-p)N}\right)}_{\textit{averages}}$$