FIN 620
Emp. Methods in Finance

Lecture 2 – Linear Regression II

Professor Todd Gormley

Today’s Agenda

- Quick review
- Finish discussion of linear regression
  - Hypothesis testing
    - Standard errors
    - Robustness, etc.
  - Miscellaneous issues
    - Multicollinearity
    - Interactions
- Presentations of "Classics #1"

Background readings

- Angrist and Pischke
  - *Sections 3.1-3.2, 3.4.1*
- Wooldridge
  - *Sections 4.1 & 4.2*
- Greene
  - *Chapter 3 and Sections 4.1-4.4, 5.7-5.9, 6.1-6.2*

Announcements

- Exercise #1 (which is optional) covers the material from today and last class

Quick Review *[Part 1]*

- When does the CEF, $E(y|x)$, we approx. with OLS give causal inferences?
- How do we test for whether this is true?

Quick Review *[Part 2]*

- What is interpretation of coefficients in a log-log regression?
- What happens if rescale log variables?

Quick Review *[Part 3]*

- How should I interpret coefficient on x_1 in a multivariate regression? And what two steps could I use to get this?
  - **Answer** = …
  - Can get same estimates in two steps by first partialing out some variables and regressing residuals on residuals in second step

Linear Regression – *Outline*

- The CEF and causality (very brief)
- Linear OLS model
- Multivariate estimation
- Hypothesis testing
  - Heteroskedastic versus Homoskedastic errors
  - Hypothesis tests
  - Economic versus statistical significance
- Miscellaneous issues

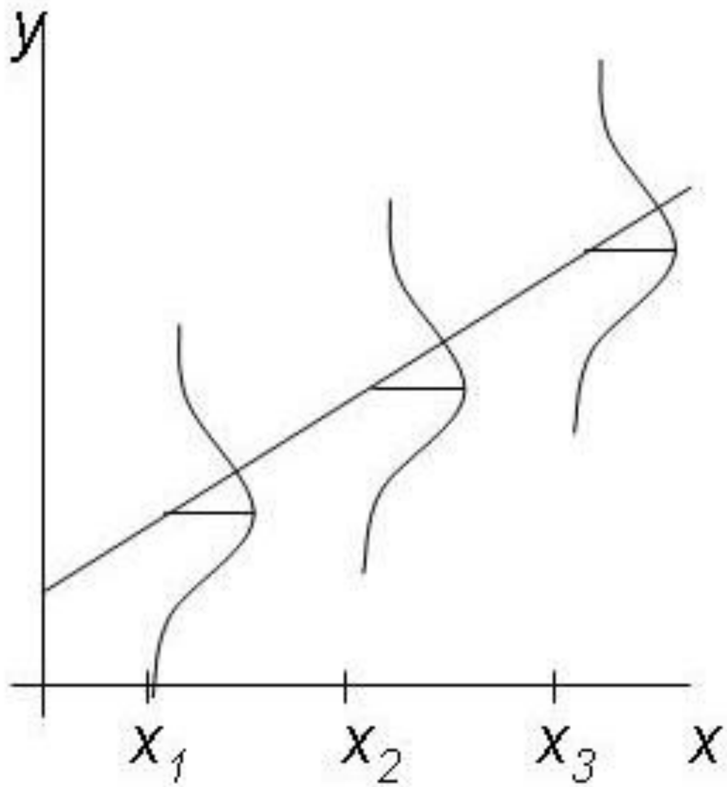
Hypothesis testing

- Before getting to hypothesis testing, which allows us to say something like "our estimate is statistically significant," it is helpful to first look at OLS variance
  - Understanding it and the assumptions made to get it can help us get the right standard errors for our later hypothesis tests

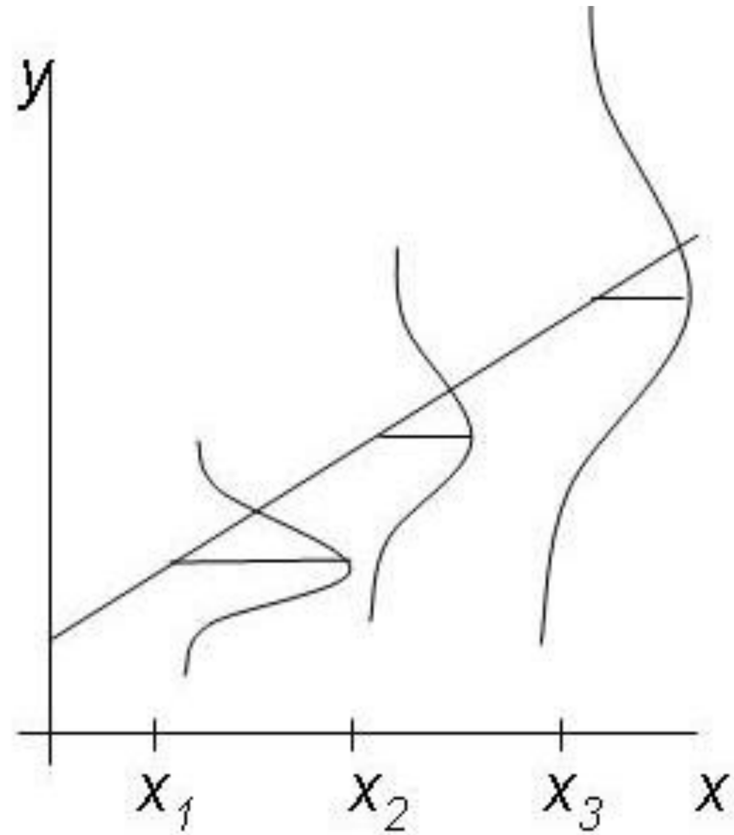
Variance of OLS Estimators

- Homoskedasticity implies $\mathrm{Var}(u|x) = \sigma^2$
  - I.e., Variance of disturbances, *u*, does <u>not</u> depend on level of observed *x*
- Heteroskedasticity implies $\mathrm{Var}(u|x) = \mathrm{f}(x)$
  - I.e., Variance of disturbances, *u*, does depend on level of *x* in some way

Variance visually…



Homoskedasticity



Heteroskedasticity

Which assumption is more realistic?

- In investment regression, which is more realistic, homoskedasticity or heteroskedasticity?

$$\textit{Investment} = \alpha + \beta Q + u$$

  - **Answer:** Heteroskedasticity seems like a much safer assumption to make; not hard to produce stories on why homoskedasticity is violated

Heteroskedasticity (HEK) and bias

- Does heteroskedasticity cause bias?
  - **Answer** = No! $E(u|x)=0$ (which is what we need for unbiased estimates) is something entirely different. Hetereskedasticity just affects SEs!
  - Heteroskedasticity just means that the OLS estimate may no longer be the most efficient (i.e., precise) linear estimator
- **So, why do we care about HEK?**

Default is homoskedastic (HOK) SEs

- Default standard errors reported by programs like Stata assume HOK
  - If standard errors are heteroskedastic, statistical inferences made from these standard errors might be incorrect…

How do we correct for this?

Robust standard errors (SEs)

- Use “robust” option to get standard errors (for hypothesis testing) that are robust to heteroskedasticity
    - Typically increases SE, but usually won’t make that big of a deal in practice
    - If standard errors go down, could have problem; use the larger standard errors!
    - We will talk about clustering later…

Using WLS to deal with HEK

- Weighted least squares (WLS) is sometimes used when worried about heteroskedasticity
  - WLS basically weights the observation of x using an estimate of the variance at that value of x
  - Done correctly, can improve *precision* of estimates

WLS continued… a recommendation

- Recommendation of Angrist-Pischke *[See Section 3.4.1]*: **don't bother with WLS**
  - OLS is consistent, so why bother?
    Can just use robust standard errors
  - Finite sample properties can be bad [and it may not actually be more efficient]
  - Harder to interpret than just using OLS [which is <u>still</u> best linear approx. of CEF]

Linear Regression – *Outline*

- The CEF and causality (very brief)
- Linear OLS model
- Multivariate estimation
- Hypothesis testing
  - Heteroskedastic versus Homoskedastic errors
  - Hypothesis tests
  - Economic versus statistical significance
- Miscellaneous issues

Hypothesis tests

- This type of phrases are common: "The estimate, $\hat{\beta}$, is statistically significant"
  - What does this mean?
  - **Answer** = "Statistical significance" is generally meant to imply an estimate is statistically different than zero

But where does this come from?

Hypothesis tests *[Part 2]*

- When thinking about significance, it is helpful to remember a few things…
  - Estimates of β_1, β_2, etc. are functions of random variables; thus, they are random variables with variances and covariances with each other
  - These variances & covariances can be estimated [See textbooks for various derivations]
  - Standard error is just the square root of an estimate's estimated variance

Hypothesis tests *[Part 3]*

- Reported *t*-stat is just telling us how many standard deviations our sample estimate, $\hat{\beta}$, is from zero
  - I.e., it is testing the null hypothesis: $\beta = 0$
  - *p*-value is just the likelihood that we would get an estimate $\hat{\beta}$ standard deviations away from zero by luck if the true $\beta = 0$

!

Hypothesis tests *[Part 4]*

- See textbooks for more details on how to do other hypothesis tests; E.g.
  - $\beta_1 = \beta_2$
  - $\beta_1 = \beta_2 = \beta_3 = 0$
  - **Given these are generally easily done in programs like Stata, I don't want to spend time going over the math**

!

Linear Regression – *Outline*

- The CEF and causality (very brief)
- Linear OLS model
- Multivariate estimation
- Hypothesis testing
  - Heteroskedastic versus Homoskedastic errors
  - Hypothesis tests
  - Economic versus statistical significance
- Miscellaneous issues

Statistical vs. Economic Significance

- **These are not the same!**
  - Coefficient might be statistically significant, but economically small
    - You can get this in large samples, or when you have a lot of variation in x (or outliers)
  - Coefficient might be economically large, but statistically insignificant
    - Might just be small sample size or too little variation in x to get precise estimate

Economic Significance

- **You should always check economic significance of coefficients**
  - E.g., how large is the implied change in y for a standard deviation change in x?
  - And importantly, is that plausible? If not, you might have a specification problem

Linear Regression – *Outline*

- The CEF and causality (very brief)
- Linear OLS model
- Multivariate estimation
- Hypothesis testing
- Miscellaneous issues
  - Irrelevant regressors & multicollinearity
  - Binary models and interactions
  - Reporting regressions

Irrelevant regressors

- What happens if include a regressor that should <u>not</u> be in the model?
  - We estimate $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$
  - However, real model is $y = \beta_0 + \beta_1 x_1 + u$
  - **Answer:** We still get a consistent estimate of all the β, where $\beta_2 = 0$, but our standard errors might go up (making it harder to find statistically significant effects)… see next few slides

Variance and of OLS estimators

- Greater variance in your estimates, $\hat{\beta}_j$, increases your standard errors, making it harder to find statistically significant estimates

- **So, useful to know what increases** $Var\left(\hat{\beta}_j\right)$

Variance formula

- Sampling variance of OLS slope is…

$$Var\left(\hat{\beta}_j\right) = \frac{\sigma^2}{\sum_{i=1}^{N}\left(x_{ij} - \bar{x}_j\right)^2 \left(1 - R_j^2\right)}$$

for $j = 1,\ldots, k$, where R_j^2 is the R^2 from regressing x_j on all other independent variables including the intercept and σ^2 is the variance of the regression error, u

Variance formula – *Interpretation*

- How will more variation in x affect SE? *Why?*
- How will higher σ^2 affect SE? *Why?*
- How will higher R_j^2 affect SE? *Why?*

$$Var\left(\hat{\beta}_j\right) = \frac{\sigma^2}{\sum_{i=1}^{N}\left(x_{ij} - \bar{x}_j\right)^2\left(1 - R_j^2\right)}$$

Variance formula – Variation in x_j

- More variation in x_j is good; smaller SE!
  - Intuitive; more variation in x_j helps us identify its effect on y!
  - This is why we always want larger samples; it will give us more variation in x_j

Variance formula – Effect of σ^2

- More error variance means bigger SE
  - Intuitive; a lot of the variation in y is explained by things you didn't model
  - Can add variables that affect y (even if not necessary for identification) to improve fit!

Variance formula – Effect of R_j^2

- **However**, more variables can also be bad if they are highly collinear
  - Gets harder to disentangle effect of the variables that are highly collinear
  - This is why we don't want to add variables that are "irrelevant" (i.e., they don't affect y)

Should we include variables that do explain y and are highly correlated with our x of interest?

Multicollinearity *[Part 1]*

- Highly collinear variables can inflate SEs
  - But it does <u>not</u> cause a bias or inconsistency!
  - Problem is just one of a having too small of a sample; with a larger sample, one could get more variation in the independent variables and get more precise estimates

Multicollinearity *[Part 2]*

- Consider the following model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

where x_2 and x_3 are highly correlated

  - $Var\left(\hat{\beta}_2\right)$ and $Var\left(\hat{\beta}_3\right)$ may be large, but correlation between x_2 and x_3 has no direct effect on $Var\left(\hat{\beta}_1\right)$
  - If x_1 is uncorrelated with x_2 and x_3, the $R_1^2 = 0$ and $Var\left(\hat{\beta}_1\right)$ unaffected

Multicollinearity – **Key Takeaways**

- It doesn't cause bias
- Don't include controls that are highly correlated with independent variable of interest if they aren't needed for identification *[i.e.,* $E(u|x) = 0$ *without them]*
  - But obviously, if $E(u|x) \neq 0$ without these controls, you need them!
  - A larger sample will help increase precision

Linear Regression – *Outline*

- The CEF and causality (very brief)
- Linear OLS model
- Multivariate estimation
- Hypothesis testing
- Miscellaneous issues
  - Irrelevant regressors & multicollinearity
  - Binary models and interactions
  - Reporting regressions

Models with interactions

- Sometimes, it is helpful for identification, to add interactions between x's
  - Ex. – theory suggests firms with a high value of x_1 should be more affected by some change in x_2
  - E.g., see Rajan and Zingales (1998)
- The model will look something like…

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u$$

Interactions – *Interpretation [Part 1]*

- According to this model, what is the effect of increasing x_1 on y, holding all else equal?

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u$$

  - **Answer:**

$$\Delta y = \left(\beta_1 + \beta_3 x_2\right)\Delta x_1$$

$$\frac{dy}{dx_1} = \beta_1 + \beta_3 x_2$$

Interactions – *Interpretation [Part 2]*

- If $\beta_3 < 0$, how does a higher x_2 affect the partial effect of x_1 on y?

$$\frac{dy}{dx_1} = \beta_1 + \beta_3 x_2$$

  - **Answer:** The increase in y for a given change in x_1 will be smaller in levels (<u>not</u> necessarily in absolute magnitude) for firms with a higher x_2

Interactions – *Interpretation [Part 3]*

- Suppose, $\beta_1 > 0$ and $\beta_3 < 0$ … what is the sign of the effect of an increase in x_1 for the <u>average</u> firm in the population?

$$\frac{dy}{dx_1} = \beta_1 + \beta_3 x_2$$

  - **Answer:** It is the sign of $\frac{dy}{dx_1}\Big|^{x_2=\bar{x}_2} = \beta_1 + \beta_3 \bar{x}_2$

A **very** common mistake! *[Part 1]*

- Researcher claims that "since $\beta_1 > 0$ and $\beta_3 < 0$, an increase in x_1 increases y for the average firm, but the increase is less for firms with a high x_2"

$$\frac{dy}{dx_1}\Big|^{x_2=\bar{x}_2} = \beta_1 + \beta_3\bar{x}_2$$

  - Wrong!!! The average effect of an increase in x_1 might be negative if $\bar{x}_2$ is very large!
  - β_1 only captures partial effect when $x_2 = 0$, which might not even make sense if x_2 is never 0!

A **very** common mistake! *[Part 2]*

- To improve interpretation of β_1, you can reparameterize the model by demeaning each variable in the model, and estimate

$$\tilde{y} = \delta_0 + \delta_1 \tilde{x}_1 + \delta_2 \tilde{x}_2 + \delta_3 \tilde{x}_1 \tilde{x}_2 + u$$

where
$$\tilde{y} = y - \mu_y$$
$$\tilde{x}_1 = x_1 - \mu_{x_1}$$
$$\tilde{x}_2 = x_2 - \mu_{x_2}$$

A **very** common mistake! *[Part 3]*

- You can then show… $\Delta y = \left(\delta_1 + \delta_3 \tilde{x}_2\right)\Delta x_1$

  and thus, $$\frac{dy}{dx_1}\Big|^{x_2=\mu_2} = \delta_1 + \delta_3\left(x_2 - \mu_2\right)$$

  $$\frac{dy}{dx_1}\Big|^{x_2=\mu_2} = \delta_1$$

- Now, the coefficient on the demeaned x_1 can be interpreted as effect of x_1 for avg. firm!

The main takeaway – Summary

- **If you want to coefficients on <u>non-interacted</u> variables to reflect the effect of that variable for the "average" firm, demean all your variables before running the specification**

- Why is there so much confusion about this? Probably because of indicator variables…

Indicator (binary) variables

- We will now talk about indicator variables
  - Interpretation of the indicator variables
  - Interpretation when you interact them
  - When demeaning is helpful
  - When using an indicator rather than a continuous variable might make sense

Motivation

- Indicator variables, also known as binary variables, are quite popular these days
  - Ex. #1 – Sex of CEO (male, female)
  - Ex. #2 – Employment status (employed, unemployed)
  - Also see in many diff-in-diff specifications
    - Ex. #1 – Size of firm (above vs. below median)
    - Ex. #2 – Pay of CEO (above vs. below median)

How they work

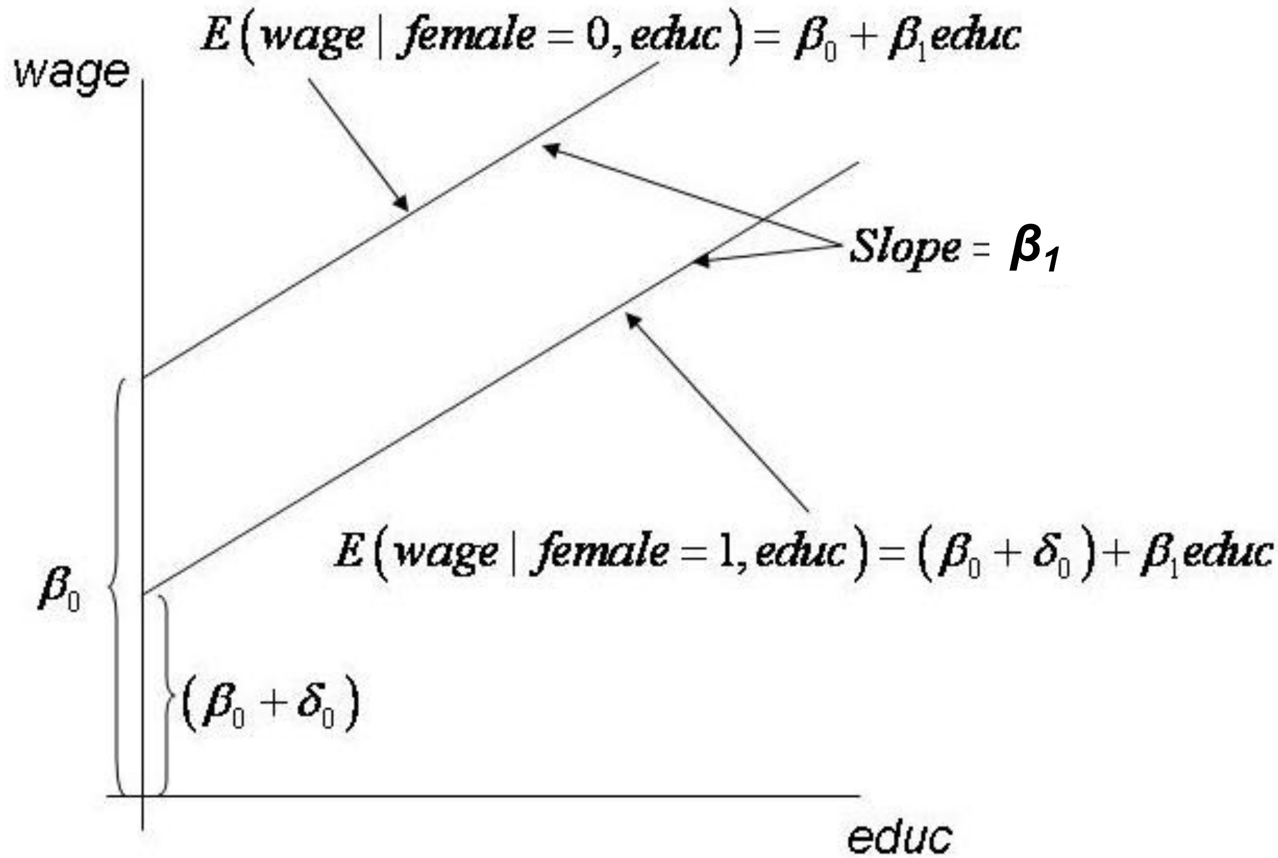
- Code the information using dummy variable
  - Ex. #1: $Male_i = \begin{cases} 1 & \text{if person } i \text{ is male} \\ 0 & \text{otherwise} \end{cases}$
  - Ex. #2: $Large_i = \begin{cases} 1 & \text{if Ln(assets) of firm } i > \text{median} \\ 0 & \text{otherwise} \end{cases}$
- Choice of 0 or 1 is relevant only for interpretation

Single dummy variable model

- Consider $wage = \beta_0 + \delta_0 female + \beta_1 educ + u$
- δ_0 measures difference in wage between male and female given same level of education
    - $E(wage|female = 0, educ) = \beta_0 + \beta_1 educ$
    - $E(wage|female = 1, educ) = \beta_0 + \delta_0 + \beta_1 educ$
    - Thus, $E(wage|f = 1, educ) - E(wage|f = 0, educ) = \boldsymbol{\delta_0}$
- Intercept for males $= \beta_0$, females $= \beta_0 + \delta_0$

Single dummy just shifts intercept!

- When $\delta_0 < 0$, we have visually…



Single dummy example – Wages

- Suppose we estimate the following wage model

$$Wage = -1.57 - 1.8female + 0.57educ + 0.03exp + 0.14tenure$$

  - Male intercept is -1.57; it is meaningless, **why?**
  - How should we interpret the 1.8 coefficient?
    - **Answer:** Females earn $1.80/hour less then men with <u>same</u> education, experience, and tenure

Log dependent variable & indicators

- Nothing new; coefficient on indicator has % interpretation. Consider following example…

$$\ln(price) = -1.35 + 0.17\ln(lotsize) + 0.71\ln(sqrft) \\ +0.03bdrms + 0.054colonial$$

  - Again, negative intercept meaningless; all other variables are never all equal to zero
  - **Interpretation** = colonial style home costs about 5.4% more than "otherwise similar" homes

Multiple indicator variables

- Suppose you want to know how much lower wages are for married and single females
  - Now have 4 possible outcomes
    - Single & male
    - Married & male
    - Single & female
    - Married & female
  - To estimate, create indicators for three of the variables and add them to the regression

But, which to exclude?

- We must exclude one of the four because they are perfectly collinear with the intercept, but does it matter which?
  - **Answer:** No, not really. It just effects the interpretation. Estimates of included indicators will be *relative* to excluded indicator
  - For example, if we exclude "single & male," we are estimating partial change in wage relative to that of single males

But, which to exclude? *[Part 2]*

- **Note:** if you don't exclude one, then statistical programs like Stata will just drop one for you automatically. For interpretation, you need to figure out which one was dropped!

Multiple indicators – *Example*

- Consider the following estimation results…

$$\ln(wage) = 0.3 + 0.21 marriedMale - .20 marriedFemale$$
$$-0.11 singleFemale + 0.08 education$$

  - I omitted single male; thus, intercept is for single males
  - And can interpret other coefficients as…
    - Married men earn ≈ 21% more than single males, all else equal
    - Married women earn ≈ 20% less than single males, all else equal

Interactions with Indicators

- We could also do prior regression instead using interactions between indicators
  - I.e., construct just two indicators, 'female' and 'married' and estimate the following

$$\ln(wage) = \beta_0 + \beta_1 female + \beta_2 married + \beta_3 \left(female \times married \right) + \beta_4 education$$

  - **How will our estimates and interpretation differ from earlier estimates?**

Interactions with Indicators *[Part 2]*

- Before we had,

$$\ln(wage) = 0.3 + 0.21 marriedMale - .20 marriedFemale$$
$$-0.11 singleFemale + 0.08 education$$

- Now, we will have,

$$\ln(wage) = 0.3 - 0.11 female + 0.21 married$$
$$-0.30\left(female \times married\right) + 0.08 education$$

> - **Question:** Before, married females had wages that were 0.20 lower; how much lower are wages of married females now?

Interactions with Indicators *[Part 3]*

- **Answer:** It will be the same!

$$\ln(wage) = 0.3 - 0.11 female + 0.21 married \\ -0.30(female \times married) + \ldots$$

  - Difference for married female = –0.11+0.21–0.30 = -0.20; the same as before

- **Bottom line** = you can do the indicators either way; inference is unaffected

Indicator Interactions – *Example*

- Krueger (1993) found…

$$\ln(wage) = \hat{\beta}_0 + 0.18compwork + 0.07comphome$$
$$+0.02\left(compwork \times comphome\right) + \ldots$$

  - Excluded category = people with no computer
  - How do we interpret these estimates?
    - How much higher are wages if have computer at work? **≈18%**
    - If have computer at home? **≈7%**
    - If have computers at both work and home? **≈18+7+2=27%**

Indicator Interactions – *Example [part 2]*

- Remember, these are just approximate percent changes… To get true change, need to convert
    - E.g., % change in wages for having computers at both home and work is given by **100*[exp(0.18+0.07+0.02) – 1] = 31%**

Interacting Indicators w/ Continuous

- Adding dummies alone will only shift intercepts for different groups
- However, if we <u>interact</u> these dummies with <u>continuous</u> variables, we can get different slopes for different groups as well
  - **See next slide for an example of this**

Continuous Interactions – *Example*

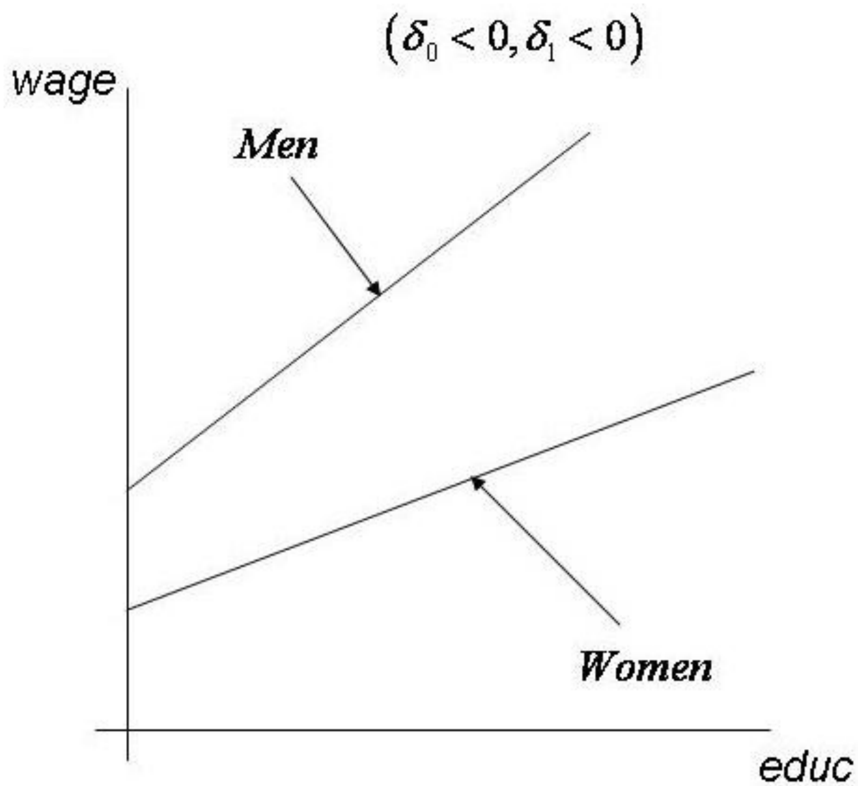
- Consider the following

$$\ln(wage) = \beta_0 + \delta_0 female + \beta_1 educ + \delta_1 \left(female \times educ \right) + u$$

  - What is intercept for males? β_0
  - What is slope for males? β_1
  - What is intercept for females? $\beta_0 + \delta_0$
  - What is slope for females? $\beta_1 + \delta_1$

Visual #1 of *Example*

$$\ln(wage) = \beta_0 + \delta_0 female + \beta_1 educ + \delta_1 (female \times educ) + u$$

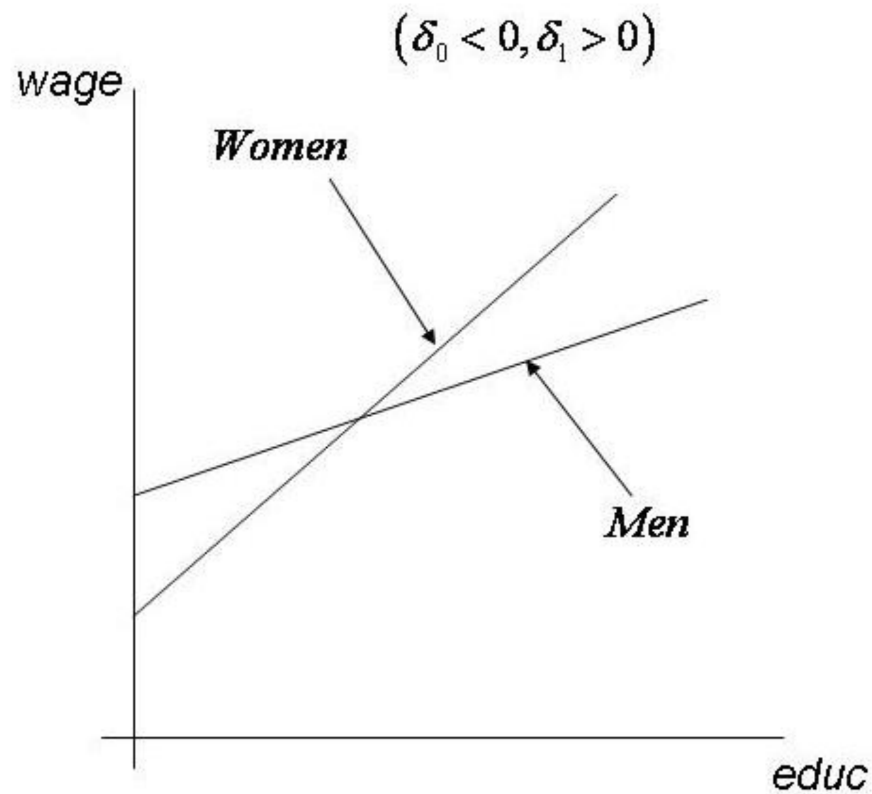


In this example…

- Females earn lower wages at all levels of education
- Avg. increase per unit of education is also lower

Visual #2 of *Example*

$$\ln(wage) = \beta_0 + \delta_0 female + \beta_1 educ + \delta_1 (female \times educ) + u$$



In <u>this</u> example…

- Wage is lower for females but only for lower levels of education because their slope is larger

Is it fair to conclude that women eventually earn higher wages with enough education?

Cautionary Note on Different Slopes!

- Crossing point (where women earn higher wages) might occur outside the data *(i.e., at education levels that don't exist)*
  - Need to solve for crossing point before making this claim about the data

$$Women: \ln(wage) = \beta_0 + \delta_0 + (\beta_1 + \delta_1) educ + u$$
$$Men: \ln(wage) = \beta_0 + \beta_1 educ + u$$

  - **They equal when *educ* = δ_0/δ_1**

Cautionary Note on Interpretation!

- Interpretation of non-interacted terms when using continuous variables is tricky
- E.g., consider the following estimates

$$\ln(wage) = 0.39 - 0.23\,female + 0.08educ - .01(female \times educ)$$

  - Return to *educ* is 8% for men, 7% for women
  - But, at the average education level, how much less do women earn? **[–0.23 – 0.01 × avg. *educ*]%**

Cautionary Note *[Part 2]*

- Again, interpretation of non-interacted variables does not equal <u>average</u> effect unless you demean the continuous variables
  - In prior example estimate the following:

$$\ln(wage) = \beta_0 + \delta_0 female + \beta_1 (educ - \mu_{educ}) + \delta_1 female \times (educ - \mu_{educ})$$

  - Now, δ_0 tells us how much lower the wage is of women at the average education level

Cautionary Note *[Part 3]*

- Recall! As we discussed in prior lecture, the slopes won't change because of the shift
  - Only the intercepts, β_0 and $\beta_0 + \delta_0$, and their standard errors will change

- **Bottom line =** if you want to interpret non-interacted indicators as the effect of indicators at the average of the continuous variables, you need to demean all continuous variables

Ordinal Variables

- Consider credit ratings: $CR \in (AAA, AA, ..., C, D)$
- If want to explain interest rate, *IR*, with ratings, we could convert *CR* to numeric scale, e.g., AAA = 1, AA = 2, … and estimate

$$IR_i = \beta_0 + \beta_1 CR_i + u_i$$

  - However, what are we implicitly assuming, and how might it be a problematic assumption?

Ordinal Variables continued…

- **Answer:** We assumed a constant linear relation between interest rates and CR
  - I.e., Moving from AAA to AA produces same change as moving from BBB to BB
  - Could take log interest rate, but is a constant proportional much better? Not really…
- A better route might be to convert the ordinal variable to indicator variables

Convert ordinal to indicator variables

- E.g., let $CR_{AAA} = 1$ if CR = AAA, 0 otherwise; $CR_{AA} = 1$ if CR = AA, 0 otherwise, etc.
- Then, run this regression

$$IR_i = \beta_0 + \beta_1 CR_{AAA} + \beta_2 CR_{AA} + \ldots + \beta_{m-1} CR_C + u_i$$

  - Remember to exclude one (e.g., "D")
- This allows IR change from each rating category *[relative to the excluded indicator]* to be of different magnitude!

Linear Regression – *Outline*

- The CEF and causality (very brief)
- Linear OLS model
- Multivariate estimation
- Hypothesis testing
- Miscellaneous issues
  - Irrelevant regressors & multicollinearity
  - Binary models and interactions
  - Reporting regressions

Reporting regressions

- Table of OLS outputs should generally show the following…
  - Dependent variable [clearly labeled]
  - Independent variables
  - Est. coefficients, their corresponding standard errors (or *t*-stat), and stars indicating level of stat. significance
  - R^2
  - # of observations in each regression

Reporting regressions *[Part 2]*

- In body of paper…
  - Focus only on variable(s) of interest
    - Tell us their sign, magnitude, statistical & economic significance, interpretation, etc.
  - Don’t waste time on other coefficients unless they are “strange” (e.g., wrong sign, huge magnitude, etc.)

Reporting regressions *[Part 3]*

- And last, but not least, don't report regressions in tables that you aren't going to discuss and/or mention in the paper's body
  - If it's not important enough to mention in the paper, it's not important enough to be in a table

Summary of Today *[Part 1]*

- Irrelevant regressors and multi-collinearity do not cause bias
  - However, they can inflate standard errors
  - So, avoid adding unnecessary controls
- Heteroskedastic variance does not cause bias
  - Just means the default standard errors for hypothesis testing are incorrect
  - Use 'robust' standard errors (if larger)

Summary of Today *[Part 2]*

- Interactions and binary variables can help us get a causal CEF
  - However, if you want to interpret non-interacted indicators it is helpful to demean continuous var.
- When writing up regression results
  - Make sure you put key items in your tables
  - Make sure to talk about both economic and statistical significance of estimates

In First Half of Next Class

- Discuss causality and potential biases
  - Omitted variable bias
  - Measurement error bias
  - Simultaneity bias
- Relevant readings – *see syllabus*

Assign papers for next week…

- Fazzari, et al (BPEA 1988)
  - Finance constraints & investment
- Morck, et al (BPEA 1990)
  - Stock market & investment
- Opler, et al (JFE 1999)
  - Corporate cash holdings

These classic papers in finance that use rather simple estimations and 'identification' was not a foremost concern

Do your best to think about their potential weaknesses…

Break Time

- Let’s take our 10-minute break
- We’ll do presentations when we get back