

Ec140 - Regression as Conditional Expectation

Fernando Hoces la Guardia
07/15/2022

Regression Journey

- Regression as Matching on Groups. Ch2 of MM up to page 68 (not included).
- Regression as Line Fitting and Conditional Expectation. Ch2 of MM, Appendix. (Part II today)
- Multiple Regression and Omitted Variable Bias. Ch2 of MM pages 68-79.
- Regression Inference, Binary Variables and Logarithms. Ch2 of MM, Appendix + others.

Regression Journey

- Regression as Matching on Groups. Ch2 of MM up to page 68 (not included).
- **Regression as Line Fitting and Conditional Expectation. Ch2 of MM, Appendix. (Part II today)**
- Multiple Regression and Omitted Variable Bias. Ch2 of MM pages 68-79.
- Regression Inference, Binary Variables and Logarithms. Ch2 of MM, Appendix + others.

Regression as Conditional Expectation: Today's Goal

- Similar to line fitting, this topic will be pretty abstract.
- In a similar spirit, this material is useful but not essential.
- Put limited effort into understanding it (~today's class +1-4hrs looking at your notes and reading the appendix of Ch2) but if it doesn't sink in, don't get discouraged and move forward into the next topic.
- For Regression as Conditional Expectations our goals:
 1. Connect the concept of regression with conditional expectations.
 2. Show how regression is a simple difference in means within groups defined by regressors.
- Last mainly theoretical topic of the course!

Regression as Conditional Expectation: Outline

- Conditional Expectation and the Conditional Expectation Function (CEF).
 - Definitions
 - Example
 - Why should we care about it (regardless of regression)?
- Connection between CEF and regression.
- Use CEF to solidify our understanding of how regression as a simple difference in groups, for groups defined by values of regressors.
- Application: CEF and Regression With Binary Treatments

Regression as Conditional Expectation: Outline

- Conditional Expectation and the Conditional Expectation Function (CEF).
 - **Definitions**
 - Example
 - Why should we care about it (regardless of regression)?
- Connection between CEF and regression.
- Use CEF to solidify our understanding of how regression as a simple difference in groups, for groups defined by values of regressors.
- Application: CEF and Regression With Binary Treatments

Conditional Expectation and Conditional Expectation Function

- Earlier in the semester we defined the a Conditional Expectation as the population average of one variable, given some value taken by another variable.

$$\mathbb{E}(Y_i|X_i = x) = \sum_y y P(Y_i = y|X_i = x)$$

- This average does two things:
 1. it summarizes the distribution (the data) of Y_i , given x .
 2. it provides information how Y_i changes as we move x . It gave us our first way of thinking about associations (before RCTs and regressions!). As a thought exercise, let's think about the distribution of happiness (Y_i) and the result of some sports outcome (X_i) (or other leisurely activity).

CEF: Definition

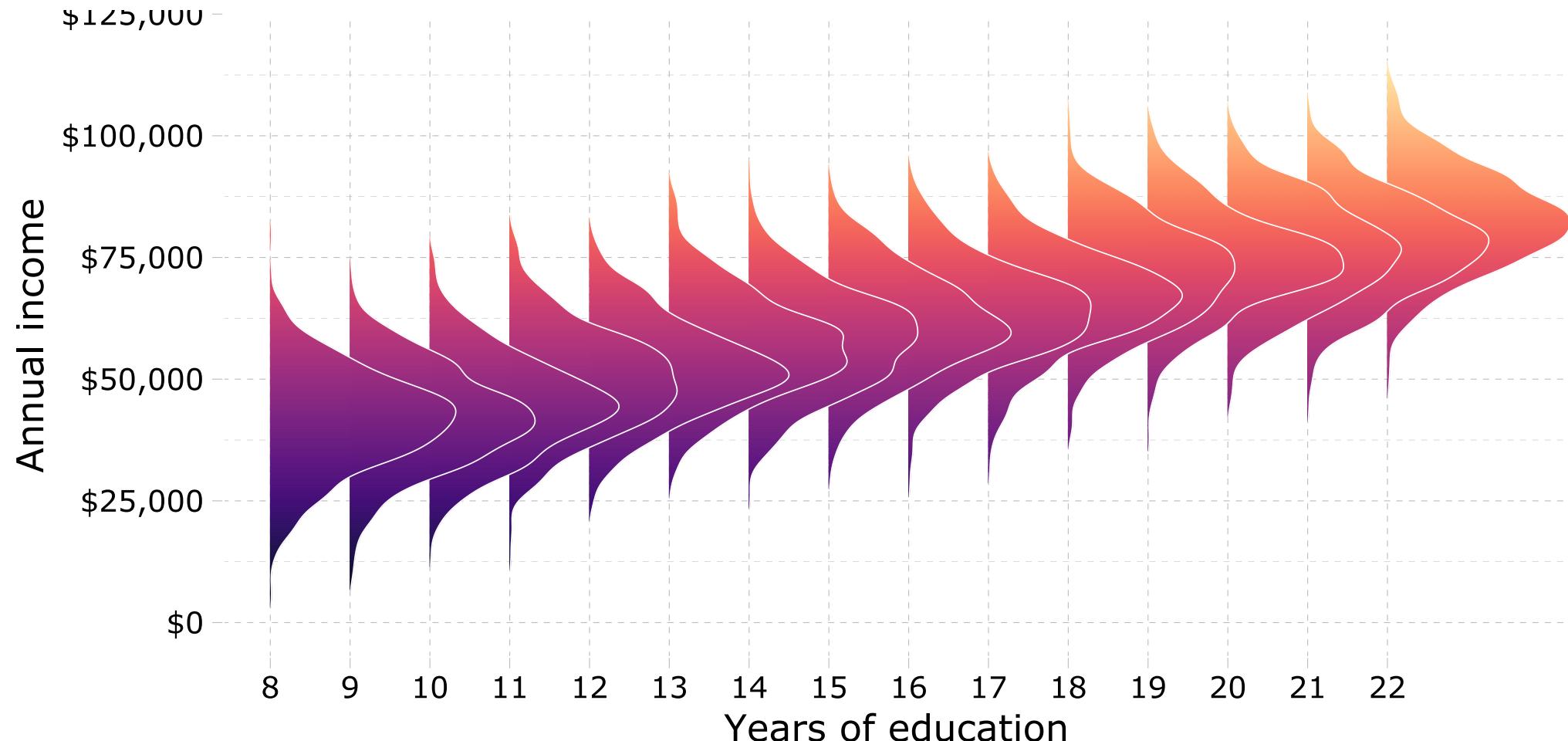
- When we look at a CE, we naturally tend to focus on Y_i . But if our goal is to describe Y_i , the relevant section to vary in the CE its not Y_i , its X_i .
- For this reason when we want to use the concept CE to describe the data (Y_i), with one number (the average), across a range of values of x , we define the **Conditional Expectation Function (CEF)**, as the collection of all such CEs (averages).

Examples:

- $E[\text{Income}_i \mid \text{Education}_i]$
- $E[\text{Birth weight}_i \mid \text{Air quality}_i]$
- $E[\text{Wage}_i \mid \text{Gender}_i]$

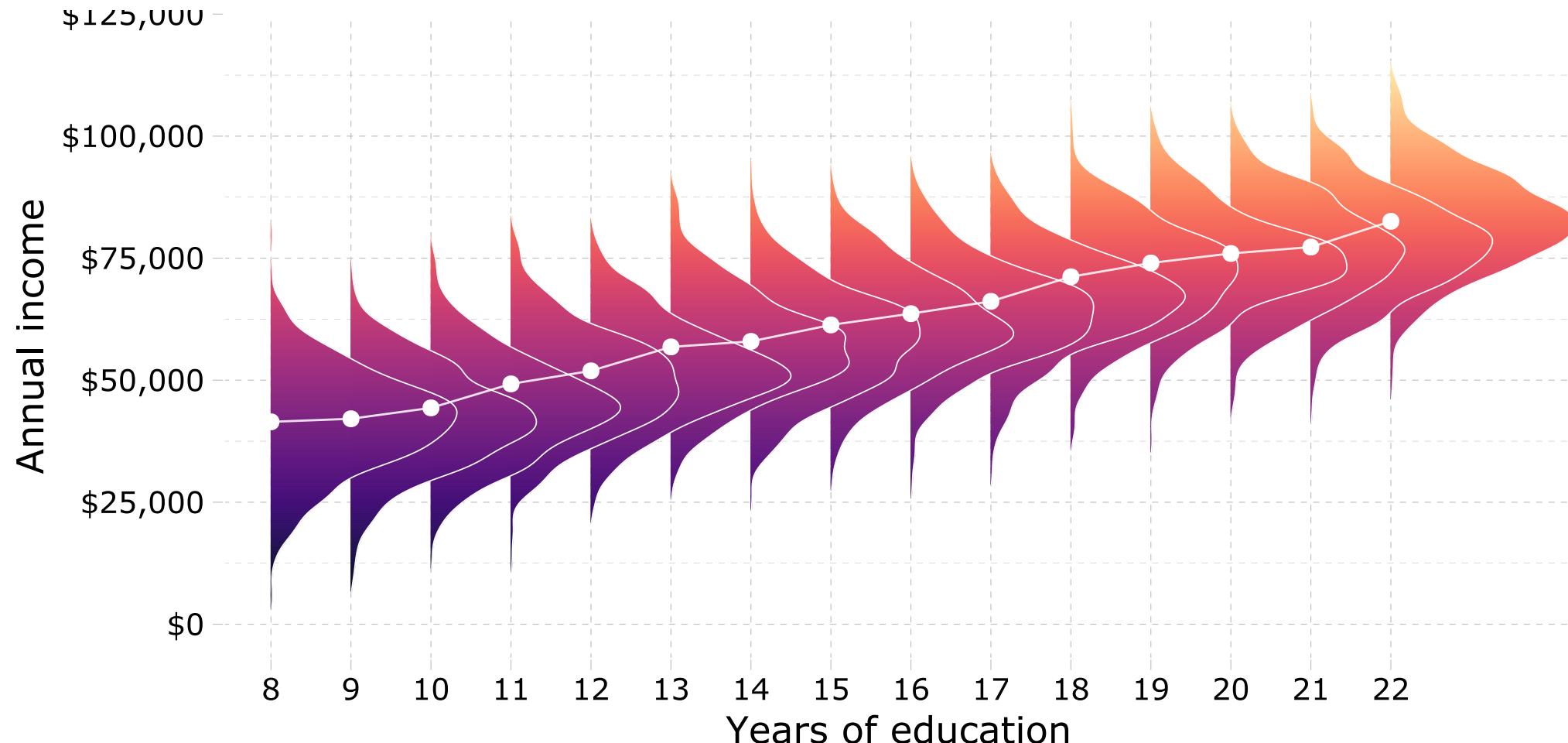
Graphically (From another of [Ed Rubin's Classes](#))...

The conditional distributions of Y_i for $X_i = x$



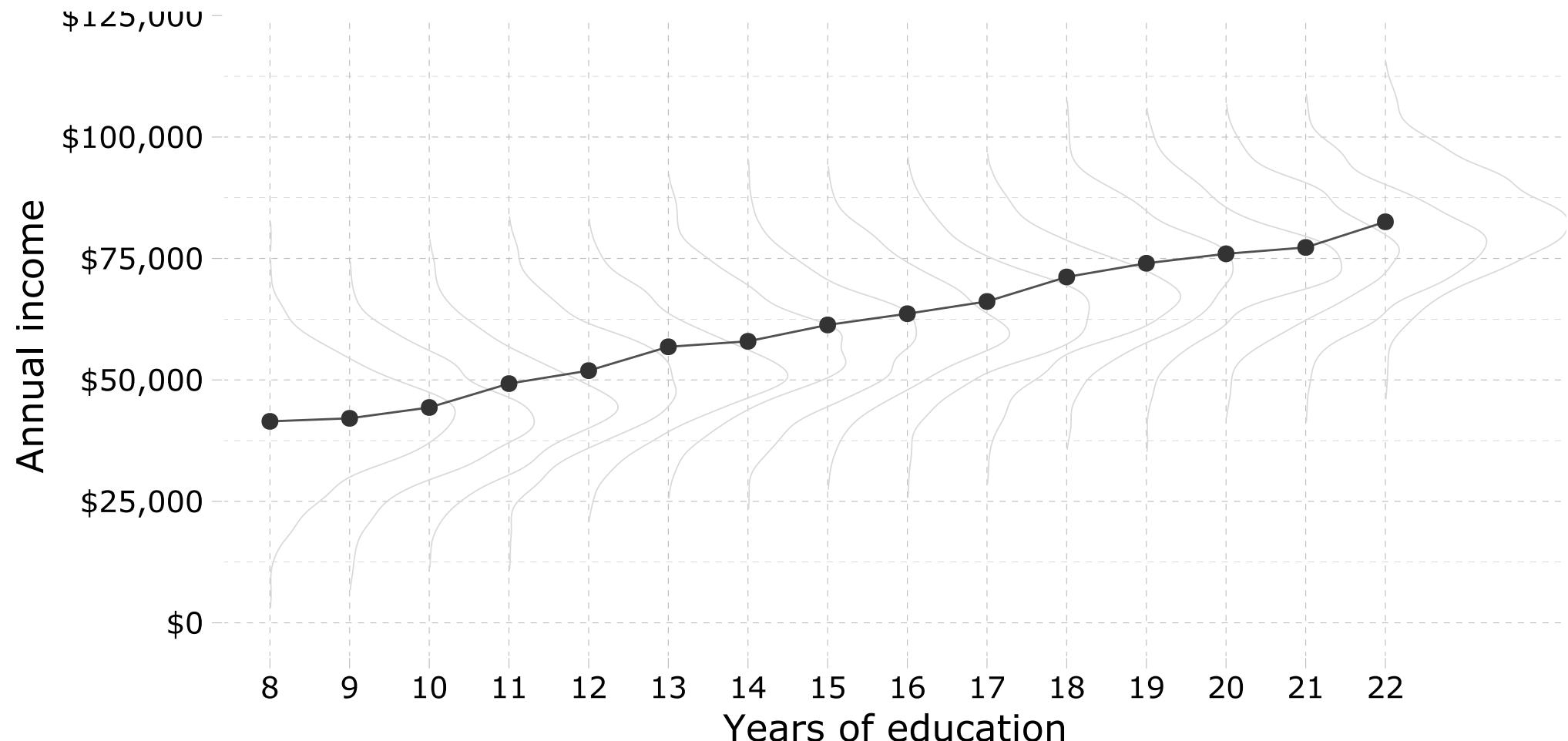
The conditional distributions of Y_i for $X_i = x$

The CEF, $E[Y_i | X_i]$, connects these conditional distributions' means.



The conditional distributions of Y_i for $X_i = x$

Focusing in on the CEF, $E[Y_i | X_i]$...



The CEF is a Function of One or More Xs

- **The CEF is a function of the variables and values we are conditioning on.** For the expression:

$$CEF(X_i = x) = \mathbb{E}(Y_i | X_{1i}) = \mathbb{E}(Y_i | \underbrace{X_{1i} = x}_{\text{this is what varies}})$$

- In a somewhat confusing convention (to me!), you will usually see that the CEF is presented as the expression in the middle above, in terms of the variable (X_{1i}), and not the possible values (x).
- The next key step is to realize that, the same way we can condition over one variable (X_{1i}), we can condition over many other variables too:

$$\mathbb{E}(Y_i | X_{1i} = x_1, X_{2i} = x_2, \dots, X_{Ki} = x_K,) = \mathbb{E}(Y_i | X_{1i}, X_{2i}, \dots, X_{Ki})$$

CEF With Multiple Xs (in English) 1/2

$$\mathbb{E}(Y_i | X_{1i} = x_1, X_{2i} = x_2, \dots, X_{Ki} = x_K) = \mathbb{E}(Y_i | X_{1i}, X_{2i}, \dots, X_{Ki})$$

- In the same way $E[Y|X = 13]$ can represent the answer to the question “what is the average earning for individuals with 13 years of education?”, we can also ask “what is the average earnings for individuals with 13 years of education, and residing in California?” or $E[Y|X_1 = 13, X_2 = 6]$ where 6 represents a numerical code for the state of California.
- For this case, the CEF would be written as $E[Y|X_1, X_2]$

CEF With Multiple Xs (in English) 2/2

- As the number of X 's grow, it becomes unwieldy notation, hence from now on we use X to refer to "one or more variables", leaving us with a general expression for the CEF:

$$\mathbb{E}(Y_i|X)$$

- As quick trick to never forget that the CEF depends on X (and not on Y), every time you see the expression $\mathbb{E}(Y_i|X)$ replace it in your mind with some function $g(X)$.

Regression as Conditional Expectation: Outline

- Conditional Expectation and the Conditional Expectation Function (CEF).
 - Definitions
 - **Example**
 - Why should we care about it (regardless of regression)?
- Connection between CEF and regression.
- Use CEF to solidify our understanding of how regression as a simple difference in groups, for groups defined by values of regressors.
- Application: CEF and Regression With Binary Treatments

CEF: Example

- Let's bring this abstract concept down to earth for the case of private/public colleges.
- Before drawing the connection to regression, we can still think of the conditional expectation of our variable of interest (Y_i = earnings), and how it changes for all our other (explanatory) variables:

$$\mathbb{E}(Y_i|X) = \mathbb{E}(Y_i|P_i, GROUP_i, SAT_i, \ln PI_i)$$

- Where $GROUP_i$ represents the collection of 150 binary variables.
- Notice until now, we have not said anything about the functional form of the CEF, to make it more explicit, let's bring the $g()$ notation:

$$\mathbb{E}(Y_i|P_i, GROUP_i, SAT_i, \ln PI_i) = g(P_i, GROUP_i, SAT_i, \ln PI_i)$$

Regression as Conditional Expectation: Outline

- Conditional Expectation and the Conditional Expectation Function (CEF).
 - Definitions
 - Example
 - **Why should we care about it (regardless of regression)?**
- Connection between CEF and regression.
- Use CEF to solidify our understanding of how regression as a simple difference in groups, for groups defined by values of regressors.
- Application: CEF and Regression With Binary Treatments

But Why Should we Care About the CEF?

- The CEF is good summary of relationship between Y and X because:
 - (Informally) Averages are good ways of summarizing random variables (and $Y|X$ is a RV)
 - (Formally) The CEF is the best predictor of $Y|X$ in the sense that it solves the Minimum Mean Squared Error (MMSE) problem ($\arg \min_{m(X)} \{E(Y - m(X))^2\}$). In English: given X , our best guess about Y , defined as the one that will render the smallest mistake, will be its CEF.
 - This last point sounds similar to regression, but up to this point we have not run any regressions!

Regression as Conditional Expectation: Outline

- Conditional Expectation and the Conditional Expectation Function (CEF).
 - Definitions
 - Example
 - Why should we care about it (regardless of regression)?
- **Connection between CEF and regression.**
- Use CEF to solidify our understanding of how regression as a simple difference in groups, for groups defined by values of regressors.
- Application: CEF and Regression With Binary Treatments

Connection Between CEF and Regression 1/4

- Today we have seen that the CEF is good for summarizing relationships in the data
- Before today we have also seen that Regression provides simple and powerful insights about relationships in the data
- Let's connect the two.
- Here we focus on the regression estimates for the population (remember we can think of regression as minimizing a population or a sample problem). These parameters correspond to:

$$\{\alpha, \beta\} = \arg \min_{a,b} \left\{ \mathbb{E}[(Y_i - a - bX_i)^2] \right\}$$

Connection Between CEF and Regression 2/4

- Now we are allowing to have more than one regressor, so let's redefine β as the collection of all parameters that solve the minimization (what before was $\alpha, \beta, \delta, \gamma, \dots$), and \mathbf{X} as the collection of all regressors. This allows us to write the above expression in a more compact way (representing more regressors!):

$$\beta = \arg \min_b \left\{ \mathbb{E}[(Y_i - \mathbf{X}_i b)^2] \right\}$$

(as we increase the number of regressors the minimization requires linear algebra in addition to calculus. This is the solution of OLS with multiple variables:

$\beta = E[\mathbf{X}_i \mathbf{X}'_i]^{-1} E[\mathbf{X}_i \mathbf{Y}_i]$ where each element represents matrices of different dimensions. This expression is the matrix version of $\beta = \frac{Cov(\mathbf{X}_i, \mathbf{Y}_i)}{Var(\mathbf{X}_i)}$ and of $\alpha = \mathbb{E}[\mathbf{Y}_i] - \mathbf{b} \mathbb{E}[\mathbf{X}_i]$ combined!)

Connection Between CEF and Regression 3/4

- Basic claim connecting CEF and Regression: if you are interested in the CEF (as summary of relationships), we should be interested in the coefficients that solve the regression (minimization) problem.
- Let's review two properties to support this claim:
- **Property 1:** If the CEF is linear, a regression estimation (minimization) will obtain the CEF parameters ("If the CEF is linear, regression will find it"). But linearity of CEF is a strong assumption, CEF is linear when underlying data is normal, this is not very common. Using MM words: "If the CEF is linear, regression will find it".

Connection Between CEF and Regression 4/4

- **Property 2:** If the CEF is not linear, regression estimates will find the best linear approximation to the CEF in the sense that the regression coefficients will be the solution to a minimization problem with respect to $\mathbb{E}[Y_i|X]$, which is different from Y_i ($\arg \min_b (E(\mathbb{E}[Y_i|X] - X'b)^2)$). Using MM words: "If the CEF is not linear, regression finds a good approximation to it".
- (if you are interested in the proofs for these check [this slides](#), it all starts from the LIE, or Adam's Law!)
- Let's go back to our example to see this properties in practice.

Connection Between CEF and Regression 5/4

- We define the CEF for the private/public example as follows:

$$\mathbb{E}(Y_i|P_i, GROUP_i, SAT_i, \ln PI_i) = g(P_i, GROUP_i, SAT_i, \ln PI_i)$$

- Now, to explore property 1, we will **assume** that the CEF of $\ln Y_i$ is linear in X's:

$$\mathbb{E}(\ln Y_i|X) = \theta_1 + \theta_2 P_i + \sum_{j=1}^{150} \theta_{3,j} GROUP_{ji} + \theta_4 SAT + \theta_5 \ln PI_i$$

- What *property 1* tells us is that in this case, the OLS regression for β (define as the collection of all parameters) will produce as result the *same* parameters of the CEF above (i.e, $\beta = \theta$).

Connection Between CEF and Regression 6/6

- However, linearity can be a strong assumption for the CEF. Here is where property 2 strengthen the connection between CEF and regression.
- Going back the CEF being any kind of function $g()$, not necessarily linear. Property 2, says that the estimate from the linear regression will provide a good approximation to the this CEF.
- Final comment regarding the connection between regression and CEF: they are connected, but they are not the same. Whenever you see an expression for $\mathbb{E}(\ln Y_i | X)$ that looks very similar to $\ln Y_i$ remember that the former is an average of the later.

Regression as Conditional Expectation: Outline

- Conditional Expectation and the Conditional Expectation Function (CEF).
 - Definitions
 - Example
 - Why should we care about it (regardless of regression)?
- Connection between CEF and regression.
- **Use CEF to solidify our understanding of how regression is a simple difference in groups, for groups defined by values of regressors.**
- Application: CEF and Regression With Binary Treatments

Back to Regression as Matching

- Now we can look the regression output from Regression as Matching and interpret coefficient of interest as an estimate of β , an approximation to the difference between to conditional expectations:

$$\beta = \mathbb{E}(Y_i|P_i = 1, X_{-P}) - \mathbb{E}(Y_i|P_i = 0, X_{-P})$$

- Where we have replace $GROUP_i, SAT_i, lnPI_i$ with X_{-P} to simplify notation.

TABLE 2.2
Private school effects: Barron's matches

	No selection controls			Selection controls		
	(1)	(2)	(3)	(4)	(5)	(6)
Private school	.135 (.055)	.095 (.052)	.086 (.034)	.007 (.038)	.003 (.039)	.013 (.025)
Own SAT score $\div 100$.048 (.009)	.016 (.007)		.033 (.007)	.001 (.007)
Log parental income			.219 (.022)			.190 (.023)
Female				-.403		-.395

Selectivity-group dummies	No	No	No	Yes	Yes	Yes

Notes: This table reports estimates of the effect of attending a private college or university on earnings. Each column reports coefficients from a regression of log earnings on a dummy for attending a private institution and controls. The results in columns (4)–(6) are from models that include applicant selectivity-group dummies. The sample size is 5,583. Standard errors are reported in parentheses.

Back to Regression as Matching

$$\beta = \mathbb{E}(Y_i | P_i = 1, X_{-P}) - \mathbb{E}(Y_i | P_i = 0, X_{-P})$$

- Notice that the right hand side **is** a difference between groups, where each group is has the same characteristics (given by X_{-P})
- Given properties 1 and 2, this equality will hold if we assume a linear CEF, or it will be a good approximation even if the CEF is not linear.

	TABLE 2.2 Private school effects: Barron's matches					
	No selection controls			Selection controls		
	(1)	(2)	(3)	(4)	(5)	(6)
Private school	.135 (.055)	.095 (.052)	.086 (.034)	.007 (.038)	.003 (.039)	.013 (.025)
Own SAT score $\div 100$.048 (.009)	.016 (.007)		.033 (.007)	.001 (.007)
Log parental income				.219 (.022)		.190 (.023)
Female					-.403	-.395

Selectivity-group dummies	No	No	No	Yes	Yes	Yes
<i>Notes:</i> This table reports estimates of the effect of attending a private college or university on earnings. Each column reports coefficients from a regression of log earnings on a dummy for attending a private institution and controls. The results in columns (4)–(6) are from models that include applicant selectivity-group dummies. The sample size is 5,583. Standard errors are reported in parentheses.						

- Conditional Expectation and the Conditional Expectation Function (CEF).
 - Definitions
 - Example
 - Why should we care about it (regardless of regression)?
- Connection between CEF and regression.
- Use CEF to solidify our understanding of how regression is a simple difference in groups, for groups defined by values of regressors.
- **Application: CEF and Regression With Binary Treatments**

CEF and Regression With Binnay Treatments as Single Regressor 1/3

- The CEF for Y_i conditioning only on a binary variable Z_i can take only two values:

$$\mathbb{E}(Y_i|Z_i) = \begin{cases} \mathbb{E}(Y_i|Z_i = 0) \\ \mathbb{E}(Y_i|Z_i = 1) \end{cases}$$

- Given this, we can write the CEF as a function of these two CEs:

$$\begin{aligned}\mathbb{E}(Y_i|Z_i) &= (1 - Z_i) \mathbb{E}(Y_i|Z_i = 0) + Z_i \mathbb{E}(Y_i|Z_i = 1) \\ &= \underbrace{\mathbb{E}(Y_i|Z_i = 0)}_{\theta_1} + Z_i \underbrace{(\mathbb{E}(Y_i|Z_i = 1) - \mathbb{E}(Y_i|Z_i = 0))}_{\theta_2} \\ &= \theta_1 + \theta_2 Z_i\end{aligned}$$

CEF and Regression With Binnay Treatments as Single Regressor 2/3

- Hence, we just prove that the CEF for this particular case (one binary conditioning variable) is linear.
- Now we can invoke property 2 of CEF and regression, and we have that a regression that estimates the following equation:

$$Y_i = \alpha + \beta Z_i + e_i$$

- Will estimate coefficients that corresponds to the CEF ($\alpha = \theta_1, \beta = \theta_2$).

CEF and Regression With Binary Treatments as Single Regressor 3/3

- Moreover, the estimate for β in a regression, will correspond to the difference in CEs:

$$\beta = \mathbb{E}(Y_i|Z_i = 1) - \mathbb{E}(Y_i|Z_i = 0)$$

- This is why we can write simple difference in groups in regression form!

Acknowledgments

- Ed Rubin's Graduate Econometrics
- MM