Recall from last week that an econometric model is one with three components: Theory, Math, & Stats.

To motivate this, consider a practical example: Keynes consumption function

Keynes theorized that higher income leads to higher consumption, but there is an level of income beyond which consumption starts to plateau. thus leading to higher savings.

This was his theory. In math we can translate these statements into the following properties:

$C = f(x)$ and $0 < \frac{dC}{dx} < 1$

$C :=$ Consumption , $x :=$ income

Moreover, to capture the logic that higher income eventually leads to rising savings, we say that the Average Propensity to Consume (APC) falls as income rises

$$\frac{d\,APC}{dx} = \frac{d\left(\frac{C}{x}\right)}{dx} < 0 \quad \text{(Quotient rule)}$$

$$= \frac{x \cdot C' - C \cdot x'}{x^2}$$

$$= \left(\frac{1}{x}\right)\left[C' - \frac{C}{x}\right]$$

$$= \left(\frac{1}{x}\right)[MPC - APC] < 0$$

$\Rightarrow$ MPC < APC

A simple specification for $C = f(x)$ is a linear one $C = \alpha + \beta x$, which satisfies Keynes' "laws" if
$\alpha > 0$ and
$0 < \beta < 1$

This equation is deterministic, but there are always random variations or disturbances in economic relationships. What can we do to transform this mathematical model into a statistical one?

We add an error term to ensure that the "allegedly random,

unexplained factor is truly unexplainable" (Greene, ch 2 . p 2)

$$C = \alpha + \beta X + \varepsilon$$

This is an econometric model.

The stochastic term, ε, is crucial. It converts a deterministic claim (which can be invalidated by a single contradictory example) into a probabilistic claim about expected outcomes. An implication of this is that now a large number of contradictory observations are needed to invalidate it. Thus, while less precise, an econometric model is more robust than a math

one.

Multiple Linear Regression Model (MLRM)

This often the starting point of any empirical research. It will be your workhorse for a good part of your career in economics or any data analysis role where you need to explain your findings to stakeholders.

The MLRM studies the relationship between a dependent variable and a set of independent variables.

$$y = f(x_1, x_2, \cdots, x_k) + \varepsilon$$

$f(\vec{x})$:= Population regression equation

$\vec{x}$:= Regressors or covariates

$\vec{\varepsilon}$:= Errors or Disturbances

y := Regressand

Earnings vs Education - A thought experiment

On average, we would expect that

educ ↑ → earnings ↑

$$earn_i = \beta_0 + \beta_1 educ_i + \varepsilon \quad (1)$$

Do you think this is accurate? What other factors may influenced an observed level of earnings and education? Would you expect a younger or older person to have more income? This logic

suggests Age is relevant

$$earn_i = \beta_0 + \beta_1 educ_i + \beta_2 Age_i + \varepsilon \quad (2)$$

Would you expect that, on average, educ and Age are positively or negatively correlated? Positive. Then, if we didn't "control" for Age would our model under- or over-estimate the marginal contribution of educ, β_2, on earnings? Overestimate.

Lastly, do you think earnings increase linearly with Age? That is, would you expect your earnings to continously increase at the same rate as you get older?
Probably not. Eventually you'll retire

and your income plateau. So, there is a diminishing marginal contribution of Age. How would you update the model to account for this?

$$earn_i = \beta_0 + \beta_1 educ_i + \beta_2 Age_i + \beta_3 Age^2 + \varepsilon \quad (3)$$

Given our theory, what sign would you expect for each parameter?

$\beta_0 \geq 0$, $\beta_1 > 0$, $\beta_2 > 0$, $\beta_3 < 0$

With this econometric model, we can now ask questions like "how do earnings of two individuals of same age, but different amount of education, compare?" We can run this thought experiment even if we don't have such data!

Here ε is the "unexplained component" of earnings that doesn't come from education or age (e.g., maybe intelligence, luck, etc) that could also impact earnings but does not impact any of our regressors.

Do you think this holds in our model?

Selection bias is probably present. Highly motivated individuals may seek more education. Also, they may have habits that, on average, leads them to higher earning opportunities. But how can we account for "motivation"?

To deal with this, we need a set of assumptions about how our data set

is produced by the DGP.

Gauss-Markov Assumptions

Linearity := $\vec{y} = X\vec{\beta} + \vec{\varepsilon}$

"linear relationship between x and y + additive disturbance"

Linearity is about the parameters and error term, not the relationship amongst the variables.

The following are all linear

$y = \alpha + \beta x + \varepsilon$

$y = \alpha + \beta \cos(x) + \varepsilon$

$y = \alpha + \beta \frac{1}{x} + \varepsilon$

* Same transform. can be applied to y

$$y = \alpha + \beta \ln(x) + \varepsilon$$

The following are not

$$y = \alpha + \beta^2 x + \varepsilon \qquad y = \alpha + \frac{1}{\beta} x + \varepsilon$$

$$y = \alpha + \beta x \cdot \varepsilon$$

Let me take a quick tangent to teach you something I wish I learnt in college.

A linear model is often thought as a first approximation of some underlying non-linear relationship. At first, it might seem very restrictive. But there is actually

quite a bit we can do to help capture curvatures while preserving this linearity assumption.

Translog model

a) $y = g(x_1, \cdots, x_K)$

b) $\ln(y) = \ln(g(\cdot)) = f(\cdot)$

note $x_K = e^{\ln(x_K)}$, hence $f(\cdot)$ is a function of the logs of x.

c) $\ln(y) = f(\ln(x_1), \cdots, \ln(x_K))$

We can expand this in a 2nd order Taylor series around the intercept point $x = [1, 1, \cdots, 1]^T$

$$\ln(y) = f(\vec{0}) + \sum_{i=1}^{K} \frac{\partial f(\cdot)}{\partial \ln(x_i) \mid \ln(\vec{x})=0}$$

$$+ \frac{1}{2} \sum_{i=1}^{K} \sum_{\ell=1}^{K} \frac{\partial^2 f(\cdot)}{\partial \ln(x_i) \ln(x_\ell) \mid \ln(\vec{x})=0}$$

$$\cdot \ln(x_i) \ln(x_\ell) + \varepsilon$$

Since $f(\cdot)$ and its derivatives evaluated at a fixed value are constant, we can interpret them as coefficients

$$\ln(y) = \beta_0 + \sum_{i=1}^{K} \beta_i \ln(x_i)$$

$$+ \frac{1}{2} \sum_{i=1}^{K} \sum_{\ell=1}^{K} \gamma_{i\ell} \ln(x_i) \ln(x_\ell) + \varepsilon$$

which is linear in the parameters and the error term. Also, note that the log-linear model (more on this in future lessons) is a special case $\gamma_{i,l} = 0$.

By the end of the semester, you will know how to estimate and interpret this type of model.

Full Rank: $X \in \mathbb{R}^{n \times k}$ and $Rank(X) = k$

> "No exact linear relationships among the regressors"

If $Rank(X) = k$ then the columns of X are linearly independent, and there are at least k

observations

This is also called the "identification problem". It is necessary for estimation of the parameters

If there are two linearly dependent regressors, then one will have redundant variation and we will not be able to learn anything from the data. In fact, software programs will drop them.

Exogeneity: $E[\varepsilon | X] = \vec{0}$

> "Disturbances must have a conditional expected value of 0 at every observation"

The idea is that no regressors should convey any meaningful information for predicting ε_i. In other words, the errors are truly random draws from same population.

$$E[\varepsilon | X] = 0 \Rightarrow E[\varepsilon] = 0 \wedge Cov(\varepsilon, X) = 0$$

$$\wedge E[y | X] = X\beta$$

Assumptions 1-3 make the linear regression model.

$$y \sim X = E[y | X]$$

So, if $E[\varepsilon | X] \neq 0$ then $E[y | X] \neq X\beta$ Hence exogeneity is

needed to ensure that $X\beta$ is the conditional mean.

I won't go into detail, but make a note that as long as you have an intercept term, α, in your regression then this assumption is not very restrictive (Greene, Ch 2, p 14)

Spherical Disturbances (aka homoskedasticity and no-autocorrelation)

$$Var(\varepsilon_i | X) = \sigma^2, \forall i: 1 \ldots n$$

+

$$Cov(\varepsilon_i, \varepsilon_j | X) = 0, \forall i \neq j$$

"Deviations of observations from their expected value are uncorrelated."

The combination of these two assumptions can be expressed in matrix form

$$E[\varepsilon\varepsilon^T | X] = \begin{bmatrix} \sigma^2 & & 0 \\ & \sigma^2 & \\ 0 & & \sigma^2 \end{bmatrix} = \sigma^2 I$$

It is called spherical because if $\Sigma = \sigma^2 I$ in the multivariate normal distribution then the equation $f(x) = c$ is the formula for a "ball" in n-dimensional space centered at μ and with radius σ

Exogenous DGP

"The process that generates data X is unrelated to the one generating ε"

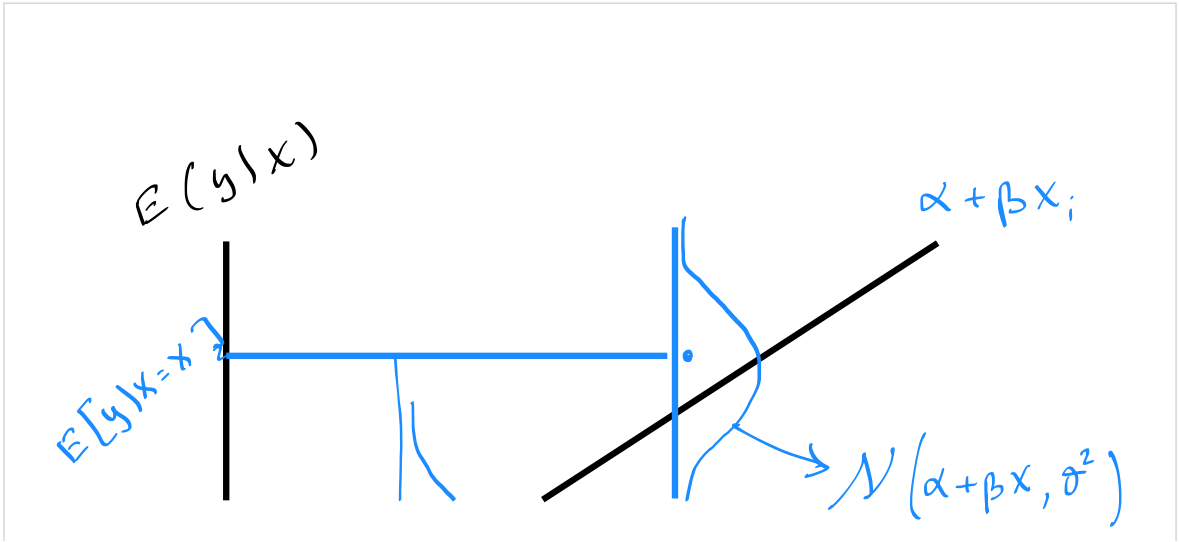
In an experimental setting (e.g., agriculture) X is deterministic (non-stochastic). So that the experimenter <u>chooses</u> X and observes y. Ex, Effect of fertilizer concentration (x) on some yield (y)

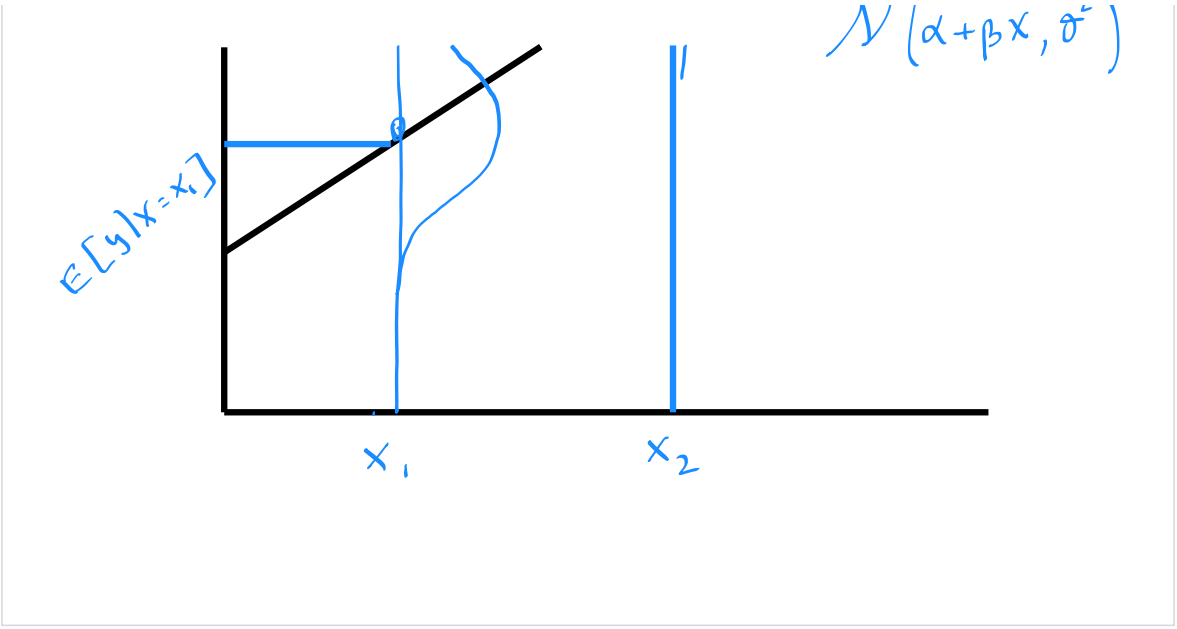
Since we don't have this luxury in Econ, we must deal with the inherent randomness of our variables. Yet, there

may be non-random variables in our models. (e.g., dummies, time trends, etc.).

The crucial detail of this assumption is that X may be a mix of random and non-random variables, but the source of these regressors is completely unrelated to the source of errors.

Graphically, we can visualize a MLRM with the necessary assumptions with the following visualization:





Key Takeaways

There are four key takeaways I want you to remember

1) The linearity condition is in the parameters β_i and the error term ε_i, not on the regressors x_i.

2) Regressions are static. You fit them to a sample of data. Hence the line fitted is only applicable for your data. This is just another way to express the conditional mean $E[y|x]$ Change the sample of data and the regression line will probably change. So, never interpret regression results as ones that must hold over time.

3) Certain transformations can be applied to linearize the data. Econometrics requires you understand what these transformations imply for the interpretation of your model. This means that you might need to sacrifice predictive power for

interpretability. In contrast with DS where you would only care about prediction power.

This is an Econometrics class, so you are expected to interpret and defend whatever transformations you choose to apply.

4) A Regression is not causal. Your narrative and theory foundation is what can make it causal.

At the end of the day, "causality" is proven only with a compelling narrative and proper validation of the assumptions we covered today. This can be a problem because

people are biased to believe simple stories. Recent work in CS / ML is experimenting with "automatic causal discovery" to reduce the need for human narrative. We'll see where that takes us...