

Getting Started I: *Introduction to Econometrics*

1. What is econometrics? Here are four descriptions. Which one did I author?

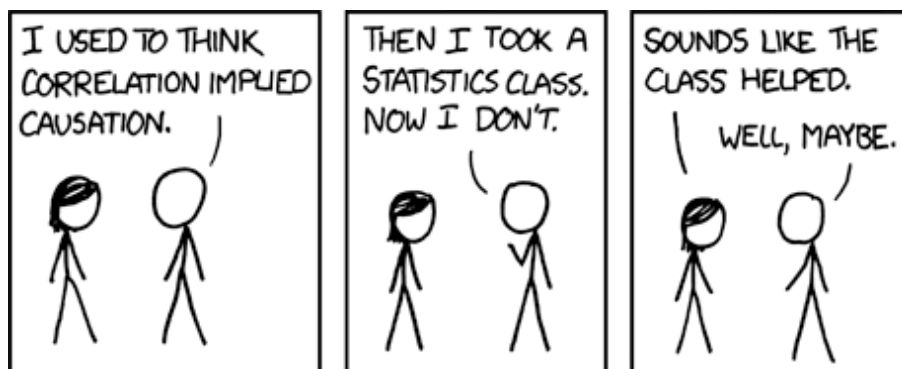
- a. *Econometrics is the application of statistical methods to economic data in order to give empirical content to economic relationships. More precisely, it is "the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference". Wikipedia*
- b. *Econometrics uses economic theory, mathematics, and statistical inference to quantify economic phenomena. In other words, it turns theoretical economic models into useful tools for economic policymaking. IMF*
- c. *Econometrics is the branch of economics concerned with the use of mathematics to describe, model, prove, and predict economic theory and systems. Wikibooks*
- d. *Econometrics is a data-driven analytic tool, which focuses on estimating relationships and forecasting impacts/outcomes. It is at the intersection of optimization, linear algebra, probability and statistics... and knowledge, understanding and common sense.*

2. Estimating relationships

- a. Correlation drives econometrics
- b. To what extent are changes in x (alone) associated with changes in y
 - i. *Ceteris Paribus* (everything else the same)
 1. Are you really controlling for all other factors? What did you miss?
 - ii. Where's the user's manual? *Oops, no manual!*
 1. What did you leave out of the model? ... and does it matter?

3. Causality v. correlation:

- a. It's so tempting to make causal inferences, but never forget, correlation is not causation.
- b. *Spurious correlation*. Some examples: <http://www.tylervigen.com/>
- c. *Examples*: How about: #firefighters (or fire trucks) and damage from fires? state average SAT and ACT scores? What do the data show? Don't just theorize; look at the data!



Introduction to Econometrics

4. Data:

- a. **Randomized controlled studies:** The *Gold Standard* ... if only
- b. **Observational studies:** Our fate. The real world is not so simple!
 - i. How representative is your data sample? How reliable are the data?
 - ii. What additional data do you wish you had?

5. The two biggest challenges in doing Econometrics

- a. **#1 Data Integrity:** Are the data reliable and representative? Or do you have to worry about erroneous data and/or **Sample Selection Bias**?

- b. **#2 Omitted Variable Impact/Bias (Endogeneity):**
What data (or variables or explanatory factors) did you leave out (omit) from your estimated model? How does that omission impact the estimated relationships?

#1: Data Integrity

#2: Endogeneity

#3: ... not close!

6. Our focus: *Least Squares Econometrics*

- i. For a given dataset, different candidate models will generate different predicted values, or estimates. Since we have data we can compare each model's *predicted*s with the *actual*s. We call the difference the *residual*: $residual = actual - predicted$.
- ii. Success comes with lots of small residuals (predicted close to actuals). Since residuals can be positive or negative, we measure success by summing the squares of the residuals, which gives us Sum Squared Residuals: $SSR = \text{sum of } (residual)^2$'s.
- iii. With Least Squares, we chose the model that best fits the data in the sense that SSRs are minimized... which is why we call it *least squares*.

7. What's the objective? Different objectives call for different models.

- a. Favorite coefficient models (the *variable in the spotlight*; you really only care about one relationship; *Kitchen Sink* models: don't skimp on those explanatory variables!; *more is more*)
- b. Behavioral/descriptive models (developing insights and better understanding of relationships; *parsimony preferred*)
- c. Forecasting models (predicting ahead, or *out-of-sample*; *less is more*)

8. *Econometrics in Action*

- a. Starting Point: What is the question? What is the issue? What are the relationships that you are hoping to better understand?
- b. You have to start somewhere!
 - i. You have a brain! Use it! Know your subject! That will inform your analysis.
 - ii. Theory, knowledge, understanding and plain common sense weigh in.
- c. Hone in on a specific hypothesis to be tested.

Introduction to Econometrics

d. Bring on the data

i. Data Integrity is the #1 concern in econometrics! Beware *Data GIGO*!¹

1. How reliable and/or representative are your data?
2. You can never spend enough time looking at your data ... Bad data? Biased sample? How to spot issues? (Use your brain!)
3. What do you do if you spot issues?

- a. Try to get better data/sample ... laziness is no excuse!
- b. And if you can't get better data/sample, you can sometimes fix things using *oh so sophisticated* methods...

- c. And even when you can't do that... maybe you can still say something about directions and magnitudes of effects/relationships?

- d. So cheer up... it's not hopeless!

ii. Building datasets is hard unrewarding work... and an incredibly valuable skill!

1. You can easily spend 90%+ of your time building your datasets... don't skimp at this stage... your data drives your results!
2. Why didn't you bring more data to the party? Laziness is no excuse!
3. You can always do more and do better.
4. When do you put your pencil down?

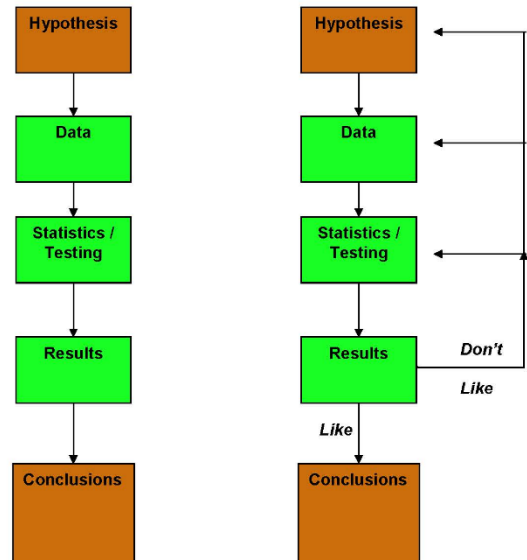
iii. LHS (Dependent) and RHS (Independent) Variables

1. The *dependent variable* is the variable/factor to be *explained/predicted* by the model (LHS variable; typically called y)
2. The *independent/explanatory/control variables/covariates* provide the data that you'll be using to generate *predictions* (RHS variables; typically called the x 's)

iv. SLR v. MLR models

1. Simple Linear Regression (SLR) Models: You only have one explanatory (RHS) variable. I know these models are easier to work with... but are you the laziest person ever? Do you really think that there's just one likely explanatory factor? These are great for introducing the topic of econometrics... but after that...

The Scientific Method: Fantasy v. Reality



¹ *GIGO*: Garbage In; Garbage Out.

Introduction to Econometrics

2. Multiple Linear Regression (MLR) Models: Bring on the explanatory variables! ... the more the merrier! The econometrics is a bit more complicated... *but maybe life is complicated!*
 - v. Endogeneity (Omitted Variable Impact/Bias) is the #2 concern in doing econometrics! ... You can't worry enough about this! Will your results stand up when z is added to the model?
 1. *Damn those confounding variables!* (correlated with x's and y's... and left out of (omitted from) the model. Example: coffee consumption and health ... and tobacco.
 2. Focus on what's in your model, but focus more on what you left out... gazillions of explanatory factors.. maybe one was important? How do you know without taking a look? Laziness is no excuse!
- 
- e. *Running regs*: Estimate the model
 - i. But not just one model: robust (sensitivity) analysis
 - ii. Keep asking:
 1. Do the results make sense? Do I believe them? Or will everyone laugh at me when I brag about them? What will the critics say? Don't forget the smell test!
 2. Are the data reliable and representative? (#1 Data Integrity)
 3. Did I leave any important explanatory factors out of the Model? (#2 Endogeneity)
 4. What have I learned? What mysteries remain?
 - iii. Running regs is easy to do once you have data... and hard to do well!
 1. You can always do more and do better.
 2. When do you put your pencil down?
 - iv. Econometrics as storytelling... every model answers some question... but is it the right question? ... or one you care about?
 - v. In the end you'll have a preferred model... but the compelling analysis doesn't rely on a single model
 1. Be careful though... How many regressions did you run? (see *skepticism* below)
 - f. *Goodness of Fit* metrics: How well does your model fit the data? How much of the variation in the dependent LHS variable (y) is explained by the model?
 - g. Estimated effects/coefficients: signs and significance
 - i. **First:** Look first at the signs of the estimated effects. Are they positive or negative? What is the direction of the estimated relationship? Does the sign of the estimate make sense? And if not, then what are you going to do about it?

Introduction to Econometrics

1. Signs reflect (*partial*) correlations in the data... correlations that control for everything else in the model.
- ii. **Second:** Then look second at the specific estimated coefficients... Are you impressed? Are they statistically significant? Can you reject the Null Hypothesis of no/zero/nada/zilch effect... @ some credible significance level?
- h. Economic significance (meaningfulness) v. statistical significance
 - i. It might be a statistically significant effect... But does it really matter? Is the estimated impact meaningful? Should anyone care? When you brag about your results, will everyone just laugh at you?
 - ii. Just because the effect is precisely estimated doesn't mean that it's large in magnitude... or worthy of attention.
 - iii. Testing for statistical significance is relatively easy... determining whether an estimated effect is meaningful is not so easy, and very much judgmental and *ad hoc* (common sense; eyeball test; elasticities; *beta* regressions)
- i. Results/conclusions/hypotheses (not) rejected
 - i. How comfortable are you generalizing your conclusions? That was, after all, the driving force behind your analysis.
 - ii. And what about that smell test? ... will everyone just laugh at you?

9. Be a complete skeptic!

- a. Bring a healthy skepticism to all published research... and be pleasantly surprised!

10. Art v science

- a. [Data Matters](#) ... and Art matters too!



11. The biggest challenge: *What do I do next?*

- a. Running regs is easy to do once you have some data to work with... and hard to do well!
- b. Don't forget: ... *knowledge, understanding and common sense*. Know your subject; use your brain!
- c. Don't be motivationally challenged. You can always do more! But when do you put your pencil down?

