David Youngberg

Acct/Busa/Econ 222—Bethany College

**Lecture 10: Extensions in Econometrics**

1. Predicting the future
	1. So far we’ve figured out how to estimate a relationship between several independent variables and one dependent variable.
	2. We also discussed how to do a time series: where the data spans across time periods rather than places (e.g. countries and states).
	3. One thing we can do is use this information to predict the future, either by:
		1. Estimating future values of independent variables based on the year. For example, estimating the cost of Social Security by estimating how many elderly individuals there will be.
		2. Putting “year” into the regression directly as an independent variable.
		3. The former method is more accurate, but more complicated. The latter method is often sufficient for a general estimate of, say, sales or population.
		4. Note that the former method is just the latter method several times over.
	4. To estimate the data, we just treat “year” like any other variable. If the coefficient for year is 0.2 and we want to know what the world will be like in 2025, we estimate by multiplying 0.2 and 2025. After considering the other variables (if any) and adding the intercept we have an estimate of the future.
2. Forecasting warnings
	1. Forecasting is powerful, but not all-powerful. Like regressions, it comes with some assumptions, namely that the pattern of data won’t change.
	2. The problem to keep in mind with forecasting is naively forecasting. Consider this comic:



* + 1. Will every word we say be “sustainable” in 2109? Of course not, but the naïve forecast would suggest otherwise.
	1. Another thing to keep in mind is *seasonal variation*—patterns of change in a time series within a year—and *cyclical variation*—patents of change in a time series longer than a year.
		1. A cyclical variation might be a pattern of boom and bust.
		2. A seasonal variation might consumer sales and the holiday season.
		3. These variations are important because if your sample size is too small, you won’t get an accurate picture: you might be predicting based on Christmas season data, for example.
	2. Seasonality can lead to another problem: autocorrelation.
		1. Regressions assume that there is no pattern in the residuals of the data.
		2. But if you’re estimating a line, regular variations will create regular residuals. A pattern.
		3. This is called *autocorrelation*—successive residuals are correlated. It is also called serial correlation.
		4. Running a regression with autocorrelation will give you biased results. There is a test of autocorrelation (called Durbin-Watson statistic) and corrections for it, but we won’t get into that here. Just something to be aware of in the future.
	3. There’s a lot more to forecasting which is why we have a whole course about it—ECON 350: Business Cycles and Forecasting—but this should give you a good idea of the challenges that come with this part of econometrics.
1. The difficulty of experiment
	1. We have econometrics because economists (and others) can’t run controlled experiments.
		1. It is impractical—and ethically dubious—to run experiments in economics, especially for questions concerning policy.
		2. To run an experiment on, say, if immigration helps an economy, we would need to organize two self-contained societies with identical populations (both in size and composition) and then forcibly immigrate a sizable number of people into one of the societies. Depending on the nature of the immigration question, we might have to also force an equal number of non-immigrants into the other society.
	2. So we must rely on real-world data and do the best we can. One of the themes I hope you picked up in writing your paper is that, as analysts, we are prisoners of the data.
	3. One way to use the data is regression analysis. There are lots of different ways to use the analysis, and lots of different complications which regressions address the complications of reality. We will now explore some of the more exotic ways.
2. Discontinuity
	1. Again, imagine the ideal experiment: two identical groups with one undergoing some kind of treatment. Then we see how the treatment is different from the control to measure the impact of the treatment.
	2. While we can’t directly do that, sometimes a single population has a line and only one side of the line is treated.
		1. Because treatment is absolute (it affects everyone on one side equally, no matter how close they are to the edge) and it’s unlikely that someone barely on one side is that different from someone just barely on the other side, we have, effectively, a controlled experiment.
		2. Example: How much does graduating with honors matter for future earnings potential? To graduate with honors, you need at least 3.35 GPA. It’s probably true that a student with a 3.35 GPA isn’t that different than one with a 3.34 GPA. But only one of them gets the treatment. By comparing the earning potentials of many such small distinctions, we can estimate an effect.
		3. Example: In 2008, the Romania government provided vouchers to buy computers to every family below the poverty line. Economists Ofer Malamud and Cristian Pop-Eleches examine the effects. Since people just below the poverty line aren’t that different than people just above it, the authors could determine that while the program gave poor people computers…



Test scores fell…

 

Because everyone just played computer games.



Oops.

1. Diff-in-diff
	1. Sometimes there’s a policy change or some other sudden event and you want to know if it did what it was suppose to do and how effective it was. For example, the college builds a new learning facility and wonders how that impacted student GPA.
	2. It’s tempting to use discontinuity (check right before when the policy took effect and right after) but this gets tricky since other things could be happening in the meantime. GPA might have changed between the year before and the year after construction for many reasons.
	3. You might then note that the policy change only affecting one area, such as a state or in this case a college. By comparing the college with the policy change and a similar college without it, you can measure the effect. But even similar colleges could have different average GPAs for many different reasons.
	4. Difference-in-difference (diff-in-diff) combats this problem by doing both.

|  |  |  |
| --- | --- | --- |
|  | **Augustana College** | **Bethany College** |
|  | *Before* | *After* | *Before* | *After* |
|  | 2.7 | 2.8 | 2.9 | 3.2 |
| **Difference in Time** | +0.1 | +0.3 |
| **Diff-in-Diff** | +0.2 |

* + 1. In this hypothetical, we can see that GPAs are on the rise, by compared differences across time, we can hold these trends constant.
	1. Another diff-in-diff example
		1. <http://www.theigc.org/article/cycling-school-increasing-high-school-enrollment-girls-bihar>
1. Indexing numbers
	1. Suppose it’s 1980 and you’re interested in investing in real estate. If you examine nominal real estate prices, you’d find the average price doubled in the past ten years. You might get excited about your investment, but you’d be wrong. Why?
		1. Inflation! Adjusted for inflation, home prices between 1970-1979 were steady. In 1980, they were declining.
	2. It’s hard to compare time periods not just because of inflation but also other macroeconomic trends. Some numbers are also tricky to compare due to their large size. So we use an index number.
		1. An *index number* expresses the relative change in price, quantity, or value compared to a base period.
		2. Note that all inflation-adjusted values are index to a particular year, as in “in 2000 dollars.”
	3. Indexing takes the base period and converts it to 100 by dividing by the value by the based period and multiplying by 100. Other values are also divided by the based period and multiplied by 100.
		1. Sometimes it’s multiplied by 10 or 1,000, or 1. As long as the other periods have the same treatment as the based period.
	4. You can also compare patterns of different time periods with indexing. Here’s a graph of how consumption progressed in different recessions, with consumption at the start of each recession indexed to 1.[[1]](#footnote-1)



* + 1. For the record, this graph is two-and-a-half years old. Things are looking better now.
	1. Since indexes are based on an easy to read number, getting percent growths are easy to get. From 100 to 130 is 30% growth.

**Lab Section**

1. Chapter 15
2. Chapter 16
3. Homework
	1. Chapter 15 1-2
	2. Chapter 16 1-2
1. <http://economistsview.typepad.com/economistsview/2009/11/will-consumption-growth-return-to-its-prerecession-level.html> [↑](#footnote-ref-1)