Douglas Dion
June 22, 2013

SOME REMARKS ON EITM

I. The Old Nixon

Let me confess this at the outset -- I am an EITM newb. Despite lots of effort on Granato's part, I am still not sure I understand what the "three steps" mean. I barely have a handle on stationarity, can dimly remember seeing a reference to the Erlang distribution, and know with certainty that I could never come up with Achen's model of rational myopia. So when it comes to practice, I am way behind. I can't even claim to be among the people who founded this movement, who saw need for EITM years ago.

I wasn't part of the original set of individuals who came together to create EITM. True, I wasn't invited to be, but had I been asked, I probalby would have declined. And I can lay out for you the argument I would have made:

1. Fundamental basis of all evaluation is truth and beauty.
2. While the two never conflict, it is easier to be wrong about truth than about beauty.
3. Theory is beautiful and statistical work is ugly.
4. By Bayes' Rule, then, statistical work is likely to be untrue.

Given this, I thought it was especially important to avoid contaminating theory with statistics. I even went so far as to submit a statistical model of filibuster duration that had no covariates besides time, because (I felt) the important feature of the model was the distributional form, and that anything else that was included was some *ad hoc*.

But I was so much older then – I'm younger than that now.

II. The New Nixon

From those sad beginnings, I have somehow ended up spending part of the last two summers at the University of Houston participating in the Hobby Center's EITM Institute, the first semester simply to offer suggestions regarding graduate student presentations, the second to not only offer comments but also (with Professor Gail Buttorff of the University of Kansas) to teach one of the daily sessions.

The question then is what has changed. In a nutshell,

1. They finally asked me.
2. They were willing to pay.

I am of course joking. I hate going anywhere, so being asked, while necessary, is hardly sufficient to get me out Iowa City. As to the second point, I regularly tell heads of my department that I am not in this for the money, and they repeatedly demonstrate through my annual salary increases that they believe me.

Fact is, wouldn't be here unless thought there was something intellectually important in what we are doing.

And the point to me is this: there aren't formal models and statistical techniques. Nor are there even formal models and statistical models. *There are just mathematical models*. That's all there is.

To me, real goal of EITM is getting people to recognize that obvious fact.

The best illustration of this point that I know of is in a terrific article published in 1975 by Morris Fiorina.[1] Fiorina (among other things) discusses the fundamental Stokes' critique of the spatial model.[2] I quote it at length:

> Years ago Donald Stokes (1963) published an excellent critique of the Downsian spatial model. He pointed out that it was a serious oversimplifica- tion to assume that an issue space remained constant over time, to assume that voters perceive the space identically, and to assume that voters agree upon the actual candidate locations and vice versa. All very true. But how many of those occasionally engaged in multidimensional scaling (and how many more of those who read such research) are fully aware that these techniques typically make corresponding assumptions about the manner in which the data are generated?4 The point is simple. Even those whose hands are buried deepest in the data make very big and questionable assumptions in their research. Many who throw stones at formal models reside in glass houses themselves

Of course, the causal direction moves both ways: if every statistical model is a formal theory in disguise (as seems consistent with Fiorina), then every formal theory is a statistical model as well.

III. The Difficulty of Seeing the Obvious

This raises an important question: if this fact is so obvious, as I claim, then why did it take so long to recognize it, and why do we have to have things like special summer institutes to get people to think this way.

The reason is that we not only live in, but are trained in a bifurcated world.

On the one hand you have theories. Theories involve deriving logical impliciations from a set of assumptions. It is a world which exists in the mind of the analyst, and often has only a weak connection to reality.

On the other hand, you have data. Data is real, in the sense that you can trace where it came from. Unfortunately, data comes into the world with a variety of possible ailments (heterogeneity, ill-behaved dynamics) which need to be diagnosed and then cured.

Seeing the world in any other way is difficult, even for those of us who are strongly committed to the EITM ideal. This is hardly surprising: the first generation never lose their accents. And these conflicts will continue as long as we fail to appreciate the error that drives these views in the first place. Which is this: we have this crazy idea that statistical techniques are about the real world, about real things, and that formal theories are about some world that doesn't exist. If I had to identify a dividing line (and this is often difficult), I would say it wasn't a utility function or an error term, but measurement. Formal theories don't worry about how one measures something.

---

[1] Fiorina, Morris). 1975. "Formal Models in Political Science." *American Journal of Political Science* 19(1): 133-159.

[2] Stokes, Donald E. 1963. "Spatial Models of Party Competition." *American Political Science Review* 57(2): 368-377. Despite being an analysis of the spatial model, there is no reference to the work of Duncan Black.

This view – that statistics is about the real world -- causes no end of confusion for beginning graduate students. How, for example, could a regression coefficient have a standard error? How could it be uncertain? And the errors in a regression model: how can they be normally distributed (for example)? They aren't even probabilistic. We know exactly what the error is, after all: it is the different between the observed *y* and the predicted *y*.

Only reason why have standard errors is because behind all statistical technology is a formal (or better, *mathematical*) model of how the world works. I for one do not consider it an especially intuitive model. Nor do I think it is even particularly realistic. Assuming things like errors being independent and identically distributed are much more offensive to my view of reality than the assumptions of the spatial model will ever be. I hear politicians talk about left and right all the time, but I have never heard one yet mention *iid* or muse on whether the errors are orthogonal to the regressors.

Without the mathematical models of the world, statistical results mean nothing. The mathematical model is the only reason we care about things like mean and variance in the first place.

But the mathematical models that undergirds statistics is hidden from most students, and there is a vast industry of software companies whose job it is to make sure that you don't have to think about it. Very easy in political science to run Markov Chain Monte Carlos and not to know the function for the variance of a sum of two random variables.

My view, which has been confirmed by presentations so far, has to do with vocabulary. Take a simple example: lag operators. Thought I understood what a lag operator was – I figured it was like an expectations operator, a kind of a function (which of course it is). So when I saw a presentation where L's were crossed out, I went crazy (needlessly, as it turns out). This happens a number of times. My general reaction to a theoretical statistical presentation is always that something is missing. I am sure that statistical types feel the same way about formal models. We each have our own "shorthand."

In my mind, the best thing EITM could do for political science is to establish a consistent system of notation and conventions of presentation that would make it possible to easily move between statistical and formal analysis. And the best way to do that, I think, is by thinking of them both as mathematical models.