

Why macroeconomics sucks

Eric Blair

6 October 2008

[Or more specifically, why time series analyses are not to be trusted.]

I've often mentioned here that time series analyses need to be eyed with great suspicion. Here, I give a detailed explanation why.

How it's supposed to work:

- Write down a model of how variables influence each other.
- Then, gather data to test the model, using a limited number of facts we know about the world (primarily central limit theorems).
- To test, estimate parameters in the model; the facts about the real world give us some idea of the likelihood that the parameter we care about is not zero.

The worst macro models and time series screw up every single step in the above chain—sometimes many times over.

Write down a model There are more data points than any of us could possibly count. However, there are only so many causal stories that we humans have been able to write down. Given a time series to explain, like GDP, and a hundred completely random variables, like sales of Barbies, beats-per-minute of Billboard's #1 song, Pantone numbers for the colors on Ikea's catalog, you are guaranteed with near-certainty that at least one of those variables has a strongly significant correlation with the variable to be explained. Here, apophenia kicks in, and after you've seen that BPM and GDP are correlated, you'll have no problem inventing a model for it.

In short, the model has to come first, and has to be a serious attempt at explaining the world. But you knew that.

[There's also the problem that it is very difficult to write down a model for which there isn't another model with the causation the other way 'round. Timing won't necessarily save us: Christmas card sales cause Christmas, as the saying goes. But that's another deeper problem.]

No, really, write down a model Now, one problem with real data is that everything moves at once. Thus, as soon as you say 'A causes B', somebody will brusquely interject that no, C causes B, and A is just a bystander in it all. Therefore, macroeconomic papers often control for 'the usual variables'.

Reapplying the principle from the last section, don't write down a regression without a model to back it up—and having a model to back up half of it is as bad as having no model at all.

You no doubt felt that the section about how having a model is important is self-evident, and most serious macro papers will start off with a good model. But the statistics with which the model is estimated will almost always include spare variables which aren't in the model. With micro papers, it's a mixed bag; with macro papers, this is the norm.

Without a model to say anything about the extra variables, you've got a lot of leeway to screw around. If C, D, and E don't reshape B the way you want, try C squared, log of D, and D times E ('D interacted with E', as the lingo goes. Used in this manner, this term is meaningless. Does D times E appear in your model?). We throw out the parameters estimated for these extra control variables, and some take that to mean that we don't have to bother with a model for them. That alibi is the downfall of serious time series analysis, and covers a great deal of the empirical macro literature.

Using facts about the world I used to think that stats is just an arbitrary list of customs that we made up so we have common means of arbitrating our disputes. But no, there are solid foundations to it.

Flip a coin a hundred times, write down the number of heads. Repeat the hundred-flip procedure a few hundred times. You now have a list of numbers between zero and a hundred, and you can plot their frequency. Most of the numbers you wrote down will be near fifty, and things will taper off as you get to the ends—a bell curve.

By which I do not just mean a curve which is fat in the middle and wide at the ends. I mean: $p(x) = \exp(-(x/50 - 1)^2)/\sqrt{\pi}$. This is the sort of precision we need to be able to say that one thing differs from another with 95.28% certainty. In fact, it's the sort of thing we need to say anything at all, because we only have a few coin flips, but we're trying to say something about what would happen with an indefinite number of future coin flips. Having a mathematical theorem about the probabilities in the limit means we don't have to take guesses.

For many other processes, there are other very specific things we can say about the resulting probability distribution. Statistics relies desperately on these few tricks we have at our disposal, generally known as the Central Limit Theorems (CLT). What do you do if the process you've modeled doesn't fit the assumptions of a CLT? Then you can't say how confident you are that the value for A that you drew differs from zero, because you don't know the distribution of A down to a square root of pi. If you draw another set of values for A, maybe it'll be distributed like the few that you've already drawn, but maybe it'll be totally different.

For example, for many types of game, if you have two players repeating the game a thousand times, the distribution of actions that player one took will have nothing at all to do with the distribution of actions that player two took, because the distributions are not independent—what player two does is directly related to what player one does. Experimental game theorists know this, and set the unit of one observation to be a whole run of the game. If you want enough data, don't have a hundred people play the game together and call it a hundred data points, but run the entire experiment with all

new people a hundred times.

Back to time series. The model claims that the variable of interest is related to the various other things we wrote down, plus or minus an error term each period. As with the game playing above, the CLT will only apply when those error terms are independent and identically distributed. Independent: the error last period had nothing to do with the error this period; identically distributed: the distribution we're drawing the errors from doesn't change with time.

For a time series, these assumptions are untenable. It is very difficult to invent a story for what those error terms mean that reasonably fits the independence and identical distribution assumption. [Yes, I know you allowed for different variances via the var-covar matrix, but why am I supposed to believe you that the mean is constant or follows a linear trend that you can soak up with the coefficient on date?]

What that means is that we can't apply the central limit theorems unless we make hefty assumptions about the world outside the model: everything in the world can be reduced to one variable, whose mean is constant, or at best, whose mean moves at a constant, linear step every period. That error is probably the mean of a hundred variables, many of which are moving up with time and many of which are moving down with time—no CLT on earth is going to tell you anything about what that series of errors will look like (even though there are CLTs to say something about the mean of unordered and independent draws from a hundred outside variables).

[OK, one last attempt at explaining it: for any variable you gather over a hundred periods, I can find you a hundred other unrelated variables, somewhere, that sum up in some reasonable linear combination to exactly the data series you gathered. In many lab or micro settings, this isn't the case, but it holds for the consistently-arranged variables of all macro time series studies. So if you include the hundred variables which replicate your pet variable, your pet variable will lose significance (especially since collinearity means your data matrix is now singular and $(X'X)^{-1}$ blows up). So you exclude the hundred variables—but now your error term is based on a process which is heavily influenced by your variable, and IID goes up in flames again. Darned if ya do, darned if ya don't, and there's no mathematical formula to tell you how to select variables for inclusion or exclusion, so it's a crap shoot, which means the parameter estimates you get are a crap shoot. The sane thing to do would be to use the model, but the model doesn't mention but a few variables, and the other typical time-seriesesque variables just get included by fiat or custom.]

Sure, it happens often enough that the error term really is well behaved, and everything that isn't in the model has a neutral effect. It is often true here in the real world that A does cause B. But most statistical analyses in macroeconomic and time series studies do little or nothing to help ferret that out, because they need to make a dozen arbitrary outside-the-model assumptions for the statistical test of the model to work out. Without backing up the assumptions in mathematical results, they are as up for question as the model itself. Yeah, sometimes it's OK to fudge the assumptions (which are never 100% true to begin with), but it's nice to at least pretend to take them seriously.

Policy implications Be Bayesian. When the paper says 'Therefore, the coefficient on A is significant with 99.99986% certainty', read that to mean 'there's one more piece of evidence that there might be something special about A.' With a hundred of 'em, we can maybe start believing that A really is special. But any one time series, no matter

the R^2 or the number of stars by the coefficients, can only provide a limited amount of evidence, because unless it is done right, it flaunts too many of the fundamental assumptions underlying statistics.