What's wrong with how we teach (and then practice) econometrics? What can we do about it?

Arnab Bhattacharjee (Heriot-Watt University) and Mark E. Schaffer (Heriot-Watt University)

Thursday 2 September, 11:30-13:00 BST

We consider three problems with the teaching of u/g econometrics and suggest how to address them.

Problem 1: Teaching statistical significance and "null hypothesis significance testing" (NHST). The statistics profession has faced this head-on in a big way, and the economics discipline has only recently started to take this on. For example, the American Statistical Association released a "Statement on Statistical Significance and P-Values" in 2016, with six principles including "Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold." Amrhein et al. (2019), in a *Nature* paper cosigned by over 800 researchers (including one of us), suggest that researchers "retire statistical significance". NHST is particularly harmful in economics, since we are typically interested in magnitudes rather than y/n questions. Finding that $\widehat{\epsilon_d} = 0.8$ and then testing $H_0{:}\,\epsilon_d = XXX$ is uninformative even the value XXX is chosen sensibly; we need to know whether the estimate of the demand elasticity is precise or noisy.

Solution: Follow the statisticians and "embrace uncertainty". We should teach interval estimation (confidence intervals) and the concept of "coverage" as the key learning outcomes. Reporting that the 95% CI is [0.75, 0.85] tells the student (and the researcher) almost everything they need to know (this is a fairly precise estimate for a commodity with inelastic demand). So would reporting that the 95% CI is [0.2, 1.4] (this a noisy estimate with little useful information). The frequentist concept of "coverage" is easier to convey than how to interpret p-values: in repeated samples, 95% of the time the estimated interval will contain the true value, in much the same way that 95% of the throws in a game of ring-toss will land around the stake – the difference is that in econometrics, you never find out if a particular throw of the ring (a particular estimation) was successful.

Problem 2: Teaching causality. At the u/g level, the main failure in econometrics teaching is to distinguish sufficiently clearly between predictive inference and causal inference. When we teach OLS, we typically start with the assumptions required for causal inference. The historical roots for this are the traditional focus on estimating structural parameters in economic models, dating back to the Cowles Commission and earlier. There are several problems here, among them the difficulties students (and researchers) have in interpreting control variables, the difficult transition from cross-section to the time-series setting, where forecasting (prediction) is central, and how to incorporate machine learning into our econometric syllabuses.

Solution: teach prediction first, and then causal inference. The interpretation of OLS in a predictive setting is much easier to teach and understand, as are the requirements for OLS to be an optimal predictor. This also facilitates subsequent teaching of OLS as a tool for causal inference. Examples such as “hospital treatment predicts health status” vs “hospital treatment has a causal effect on health status” are easier to convey once students are comfortable with OLS as a predictive tool. Teaching machine learning methods as part of predictive inference is very natural. So too is introducing time series data and forecasting.

Problem 3: Disciplinary diversity in Big Data econometrics across the three relevant disciplines – computer science, economics and statistics – is not taught well. Students are often left wondering how econometrics is different from statistics. Likewise, students would like to know what to expect that is different in a Big Data econometrics course vs a computer science course on machine learning, for example.

Solution: The central issue is that we do not often think through the lens of disciplinary diversity. We propose explaining the different disciplinary approaches using examples. For example: what would happen to output if an economy is hit by a positive 10% demand shock, a negative 10% supply shock and a 5% monetary policy shock? Computer Science would be very useful in discovering patterns in the past data, and can provide excellent predictions. But these data would likely not conform to the precise scenario under consideration. In any case, shocks are not observed, so one needs a clear definition in the context of a model. Statistics can help by interpreting shocks as error processes, and would aim to find out the “correct” reduced form data generating process, given a well-specified model. However, one still needs a structural model to make sense of the economic shocks. Economics is interested not only in the reduced form, but also in the causal structural model and counterfactual policy and shock scenarios. Clear articulation of this disciplinary diversity can go a long way.