Editorial: Realistic Academic Standards and the Value of Replications

Avanidhar Subrahmanyam

UCLA Anderson School of Management, USA; asubrahm@anderson.ucla.edu

> He that is without sin among you, let him first cast a stone at her
>
> The Bible, King James Version, John 8:7

When we sign up for a PhD in finance, many of us are fairly inexperienced at what research is all about. It is only when we actually get our hands dirty that we realize that compelling research is not just about finding a topic but also about telling a compelling narrative that can potentially influence how others think about our topic. Subjectivity is essential in developing this narrative, because the story woven inevitably depends on design choices as well as reporting choices. During my PhD program a senior professor in my department gave me the advice that "you can be selective in what you report (in doing empirical work)."

That left me wondering what he meant. Evidently, one does not have to report every empirical result. But, how should scholars decide what to report, given that they and only they see all of the output? Reporting mixed results runs the risk of destroying the "compelling" nature of the narrative. Reporting every result runs the risk of having to write up a convoluted mess that no-one will want to drudge through. So what should one do? One past AFA president gave me the following "rule of thumb:" "If a result survives 2/3 of your specifications, that's good enough and you don't need to worry about reporting the remaining 1/3."

Luckily I didn't have to get into these issues early in my career, because I was mostly doing theory. As I embarked on empirical work, it became painfully obvious to me that pure objectivity is a mirage. Subjectivity is an inevitable part of doing research, as is making choices on what to report. We all have to rely on our internal compasses to ensure that we remain honest and report "largely" reliable results.

I would like to divide up dilemmas arising from reporting choices in scholarship into three parts:

(i) Outright misrepresentation, which involves reporting objectively untrue results.

(ii) Choosing to report a result that survives a handful of specifications out of several runs.

(iii) Suppressing those results that are unfavorable to the thesis being advanced, where the thesis is largely supported by the data.

In my view only (i) and (ii) represent clear unethical behavior and (iii) is often unavoidable. While I think the baseline occurrences of (i) or (ii) are low to begin with, our efforts should focus on reducing them further. Before I provide suggestions on how to accomplish this, let me mention some cautionary notes that emphasize the need for incentives, rather than "cheap talk" on the subject.

First, human biases apply to our community, too. One common bias is the confirmation bias, where the human mind looks for evidence to confirm a world-view already held. This suggests that an "efficient market type" may seek out evidence that confirms this view, whereas a behavioral type may seek out the reverse. This bias alone can probably account for a variety of robustness issues in research, because many non-robust results may be due to researchers having selectively looked for confirming evidence, possibly unconsciously. This is not unethical behavior. It is important to distinguish confirmation bias from (ii) and (iii) but often it is hard to tell if researchers had nefarious motives or were simply interpreting evidence through the lens of their world view.

Second, consider the incentives we have set for our younger colleagues. They have six years to publish five or six papers in the top journals and if they manage to do this and make an impact, they may earn many hundred thousands of dollars a year for the rest of their working lives. It is not surprising that desperate young researchers can lose their compass momentarily, given that the top journals are asking them to produce "compelling" work. That is, it is one thing for us successful academics to ride the high horse and preach appropriate ways of doing research. It is quite another thing to put ourselves in the shoes of our young colleagues. Many now famous empiricists may even remember that they also made a few reporting calls that were on the margins of (iii). Moreover, who are we to preach after tenure? Did we never favor our own students when we had power to do so? As editors, did our own career concerns never influence us to accept papers of a famous scholar? Did we never unfavorably tilt a referee report on a worthy paper critical of our work? Did we ever selectively use evidence in expert witnessing work as part of our testimony?

Based on the above arguments, I would argue that we should not preach via cheap talk. If we do, we are hypocrites. I also argue our profession should focus less on retroactively (and possibly unfairly) scapegoating individuals or past papers (unless we find cases of (i) and egregious cases of (ii)); and more on setting proper incentives for the future.

In terms of setting future incentives, here are some observations that may help. First, as referees, should we insist on significant results? It is really OK if not every result goes the way of the paper's central thesis. Second, should we require innumerable robustness checks? Why not worry more about the importance of the topic, rather than whether the results make a "splash" by all lining up one way?

Third, as editors, why should we insist on putting authors through the wringer by making them do more and more runs? As members of tenure committees, should we place a little emphasis on whether the author has exercised intellectual honesty (e.g., via reporting results objectively and honestly), rather than on splash and impact?

In sum, we should improve our governance for the future in productive ways and not retroactively in unproductive ways. We should work towards fine tuning an *ex ante* incentive system that minimizes the incidences of (i) and (ii).

I do believe that authors of all published papers in our leading journals should make available the data and programs for replicating the main results. What will this requirement accomplish? First, this requirement will eliminate (i). Second, most of us are mature enough to know the difference between (ii) and (iii), and making the data and programs available will allow us to ascertain which results are robust. Third, the *ex ante* incentives for (ii) will be reduced. Finally, honest errors will be easily discovered. When applied uniformly to everyone, a code and data sharing requirement, ensures that the more famous and productive scholars will not escape the need to provide reliable results, no matter how many papers they write.

I hope that the CFR going forward will not only point out shortcomings of published work, but also play a useful role in delineating more robust results. However, it can only do so if the data and programs are becoming universally available for public consumption.