

{11/04/20.Johlson}

## Elephants in the Room: Accounting Seminars

To avoid professional mishaps, it helps to be aware of not-to-be touched topics. Members of a community generally recognize the possibility that raising certain topics, while already familiar to everyone, can cast people of importance in a poor light, or worse yet, make the community itself look bad. In colloquial parlance: such topics are the elephants in the room.

This document exemplifies accounting research elephants. The general idea is to avoid issues that touch on known weaknesses of accepted research practice. If expressed sharply enough, a question raised about a paper's methodology can remind seminar participant that almost all of our research output does not achieve much at all; put simply, the methodological foundations are too fragile. This causes uncomfortable reactions ("All of us know research-publications is a game, and we do not need to be reminded."). Any questions that implicitly put some general underlying research methodology issues into focus accordingly annoy people because nothing gets resolved, and it tends to accentuate a negative view on what our research accomplishes. Some topics can even challenge the status of powerful individuals. As a sage once quipped: "Though partaking in games can be relaxing and enjoyable, it is well to note that real stakes change everything."

### Five Elephants.

#### **1. Referring to the absence of a Fama - MacBeth analysis.**

To make a seminar aware that the paper did not apply or report on the Fama - MacBeth method is a no-no. Years ago researchers learned that this method is all too likely to disappoint: it tends to produce clear-cut null-acceptances. Thus FM is a tough hurdle to pass unless the underlying story is solid. So why bring it up when we already know that most stories tested are shaky, at best? (To be sure, the contemporary "preferred" method of analysis pools all data and use FEs; it undoubtedly helps when the goal is null-rejection!)

Examples of offensive questions.

*“I notice that you did not report on results using FM. Why is that since you easily could have done so?”*

*“If you estimate annual regressions, what percentage of the 28 years of data would you get the correct sign on the main variable? If you did not check, what do you think would be the case; of course, feel free to guess.”*

## **2. Asking whether a RHS variable contributes to explaining the dependent variable.**

To request a speaker to address this matter is offensive for the simple reason that, in general, most people in the room would know that the answer is “at best, the variable’s contribution is marginal”. An answer along these lines would not make the presenter feel or look good since it would contradict the paper’s broader message. In turn, it raises the subject of the prevalence of farfetched hypotheses and how to deal with its consequences without overstepping ethical bounds. People do not want to be reminded about this painful subject. So why bring it up, even indirectly? Thus, the contemporary convention of not reporting on findings addressing the effective lack of a variable’s relevance was settled more than a decade ago; if you cannot get out of the box, do not walk into it. If pushed by a seminar questioner, the presenter may simply provide a bland response like “The variable in question is statistically significant, and my understanding is that such an outcome, per accepted practice, allows me to conclude that the variable is relevant”.

Examples of offensive questions:

*“By how much would your R-square decrease after having deleted variable X on the RHS, in Table 2? What would happen is not clear from the table, but I presume you checked?”*

*“Per your table of descriptive statistics, your variable X and the dependent variable correlate to the tune of 6%, which I do not believe anyone would claim supports X’s relevance. Why should I believe that X can be of any*

*greater relevance in your main regression because you add 15 controlling variables?”*

### **3. It takes more than Stars to settle the matter.**

To ask a presenter to deal with the statistical issue related to large N and a t-statistic that seems relatively small makes everyone uncomfortable: after all, looking at a few stars should suffice to declare victory. It IS what social scientists with an empirical bent have been doing for decades. Of course, at the same time everybody knows that *something* seems seriously remiss when N runs into the tens of thousands and the t-stat approximates, say, 4. Statisticians, plus our intuition, tell us that under such circumstances classical statistics can mislead. But why argue about long-established conventions in a regular seminar which do not focus on methodological issues? Makes no sense: we have all learned to live with “stars are stars, no need to raise the role of N”.

Examples of offensive questions:

*“To support your claim that X relates to Y you refer to the key regression showing two stars. But since you have 65,000 plus observation, does that not fall far short from minimum requirements? Given the sample size, don’t you think you would need at least a t-stat of 5 to back up your claim? “*

*“You have 55,000 plus observations, and thus it does not surprise that quite a few of your controlling variables are statistically significant. And in many cases much more significant than your variable of main interest. Yet a subset of these more powerful variables end up with what seems to be the wrong signs. Do you think something has gone wrong more broadly? “*

### **4. Referring to the possibility of using a hold-out sample.**

Setting aside the case when researchers hypothesizes about making money in financial markets, the accepted convention is that there is no need for a “fresh” holdout sample. The importance of a holdout sample comes into focus if a researcher worries about stating conclusions which are likely to be less than robust. But since worries about publishing dubious conclusions are quite rare, there are no apparent compelling imperatives to use hold-out

samples. Doing so will make the paper longer and it will look more pretentious. And worse, the findings produced by the hold-out sample can introduce ambiguities as to the correct overall conclusions. Not good at all; to worry or complain about current practice is amateurish. (Researchers in the know stick to the convention of posing only the kind of robustness tests that always work out.)

Examples of offensive questions:

*“I notice that the data you analyze end in 2016. Why not check whether you get the same results using the 3-4 more years of recent data? It can serve as an exceptionally “clean” and relevant hold-out sample; to me, it seems like a real opportunity that should be exploited.”*

*“Given that you have a relatively large N, could you not have split the sample into, say, three mutually exclusive subsets and then move on to check whether the three cases pretty much tell the same story?”*

##### **5. Issues related to “screen-picking” and “data-snooping”.**

To raise these matters is truly unacceptable. If attempted, the presenter might well perceive the question as an assault on her/his person, and in this regard many seminar participants will be sympathetic. Most people in the room have had their own experiences, often frustrating, massaging the data and staring at regressions until the procedures yield acceptable findings essential to bring the project to a successful fruition. And nowadays people can actually put to use software that generates millions of regressions, and where one gets to pick the one most pleasing (Yes, it pretty much ensures “economic significance” of a variable when needed). Nonetheless, active researchers do not like to discuss these kind of experiences; the topic is too sordid and, accordingly, one learns to live with it as a more or less painful private matter. (Reminder: a world-wise person does not roam around in a gambling establishment asking the players whether they lost some money recently.)

Offending questions, examples:

*“Your regression relies on some very strange RHS variables that seem to be important, yet they do not show up as part of the descriptive statistics. Can you elaborate on this matter? And why do you have such an abundance of interactive effects as independent variables?”*

*“Your regression findings show that quite a few of the RHS variables have the wrong signs, yet you do not confront this matter. For example, leverage has a negative sign though the dependent variable relates to cost of equity. Did you try various regressions to potentially produce the correct signs but failed? Or is the matter irrelevant because you only care about the sign and t-stat of the main variable on the RHS?”*

**A remark: Can one raise hardnosed questions about basic methodological matters without being offensive?**

The answer is affirmative as long as one treads carefully. One needs to keep in mind: (i) start out being positive and recognize some apparent merits of the presenter’s work, then (ii) proceed cautiously to raise the question without hinting that the question would not have been raised unless there are potentially serious weaknesses prowling in the background. In other words, only upon reflection will the presenter, as well as most seminar participants, recognize that the question turns on methodological deficiencies which might well indirectly submit that the paper’s conclusions are dubious.

Example:

*“Your implementation of the standard fixed effects model is impressive, and you make it clear that a serious analysis required a fair amount of work to ensure comprehensive controls that go beyond the prior literature. I think most of us fully appreciate these constructive aspects of your research. That said, you could perhaps do more to convince readers that the results are as compelling as you claim. My suggestion is that you implement the following two stage procedure. First, run your pooled FE regression without the main variable on the RHS. Second, for each year correlate the regression residual*

*with the main variable. Then evaluate your implied hypothesis that the great majority of correlation across the years is positive. It ought to be quite informative, do you not agree?*

**The Future:** Some aspects of accepted research practice tend to lead to a host of elephants which become too apparent and agonizing. A change in accepted research practice will be looked for. High on the list is the need to provide relief by moving away from the most awkward aspect of the current modus operandi: (i) telling a more or less eccentric story leading up to the statistical hypotheses, and (ii) where the data support the story by a one-sided (statistically significant) null rejections combined with robustness tests that never introduce doubts. Assuming that a change in accepted practice happens, one can expect (and perhaps even hope for) an “announcement per rumor” to the effect that “from now on it is OK to posit a two sided hypotheses”.