As you read a paper: the usual questions

Finance 8823, July 2022, Murray Z. Frank

Every research paper is distinct. But there is also a great deal of common thinking that underlies most empirical corporate finance papers. There are standard issues that you need to watch for. While phrased in various ways, these questions routinely show up in seminars and in referee reports. You should get used to thinking about these for each paper. Several of these match ideas in the classic little book by Huff and Geis (1954).

A hallmark of high quality modern empirical work is that it starts with a tight question that is worth answering. Then we get extremely careful and precise interpretations of the estimates. A great deal of effort is devoted to avoiding what Feynman (1974) famously called Cargo Cult Science – research that has roughly the right form, but not the right substance.

Your empirical model may be simple, or it may be complex; but either way think carefully about what the parameters mean. How might the data actually be saying something other than the obvious interpretation? If you like the obvious interpretation, how can you rule out the alternatives? Can you find evidence beyond your initial idea, to help substantiate your interpretation? Can you find evidence that your ideas generalize to other time periods or to other countries or other datasets?

1. What is the question?
    a. Why should anyone care about the answer to this question?
    b. How does the research question relate to existing theoretical and empirical literature? What is new?
2. What data are used in the paper?
    a. How was the original data gathered/created? Is this a source of bias?
    b. How did the authors obtain access to the data?
    c. How exactly are the key variables defined? What are the units of measurement?
    d. Key variables have names and definitions. How does the empirical definition relate to the economic concept we would usually associate with that name? How does this affect the interpretations?
3. What is the overall empirical strategy?
    a. Suppose that you had access to an ideal data set for this paper's question, what would the data look like? What empirical approach would you use?
    b. How does the data set in the paper compare to your imaginary ideal data set? What limitations does that create? How does the paper's empirical strategy cope with those limitations?

    c. How is the model in the paper identified? Is there an important observational equivalence to worry about?
    d. What creates exogenous variation? Are the exogeneity claims reasonable?
    e. What features are key to determining the reported parameter values? Is the data or the model more important to the inferences?
4. What specific econometric methods are in use?
    a. Do these methods seem like a good approach to answer the question?
    b. What are the key estimating equations? How are those equations justified?
    c. How is the error term justified?
    d. What might be important but unobservable?
    e. What instruments are in use? Do they seem plausible? Can we tell?
    f. How exactly does the estimated model connect to the theory?
5. Internal validity. What are the main results in the paper?
    a. What is the "punchline" for the paper?
    b. Think about units of measurement. What exactly is the economic meaning or interpretation of the key parameters that are being estimated?
    c. Are the reported parameter magnitudes plausible? How can we tell?
    d. What are the broader economic implications of the evidence?
    e. What did you learn from reading this paper? How does it fit in the literature?
6. External validity. Are the results generalizable?
    a. Outside the original dataset and time period, what do the results imply for finance? How credible is the idea that the results apply elsewhere?
    b. Does the paper offer direct evidence on an external data set, or on a hold out sample?
    c. Does the paper suggest good reasons why we might or might not expect the result to apply beyond the original data?
7. Normative implication. Does the paper say that anyone should act differently? This is often about government policy.
    a. Are there any policy implications claimed in the paper?
    b. Can you see policy implications that the authors did not claim?
8. What questions does this paper leave unanswered?
    a. Can you see a reasonable way to try to answer those questions?
    b. Are those unanswered question interesting enough to be worth your effort to try to answer them?

You need to read the assigned papers before class. It is also a good idea to reread them soon after class to see if you pick up on things that you did not notice before. Talking about the papers with other PhD students is a very good idea.

A cautionary note: Our literature cares a great deal about **identification**. But what does that actually mean? This is the source of a surprising amount of confusion.

1. By definition, a model is identified if distinct values of the parameters correspond to distinct distributions of the data under the researcher's maintained assumptions. There is no other model that is observationally equivalent under those assumptions. This is a property of a model structure. It is not a property of an estimator. See Andrews, Gentzkow, and Shapiro (2020) section 4, or Lewbel (2019) for extensive explanations.
2. But you will also hear the term used informally: "Here, the phrase "how a parameter is identified" refers instead to a more intuitive notion that can be roughly phrased as follows: "What are the key features of the data, or the key sources of (assumed) exogenous variation in the data, or the key a priori theoretical or statistical assumptions imposed in the estimation, that drive the quantitative values of the parameter estimates, and strongly influence the substantive conclusions drawn from the estimation exercise?" Keane (2010)
3. When the term identification used, make sure you know which meaning is intended. They do not mean the same thing at all. Both are worth knowing about.

Andrews, I., Gentzkow, M. and Shapiro, J.M., 2020. Transparency in structural research. Journal of Business & Economic Statistics, 38, 4, 711-722.

Feynman, R. P. 1974, Cargo cult science. Engineering and Science, 37(7), 10-13.

Huff, D. and I. Geis, 1954, How to lie with statistics, reissued 1993 Norton.

Keane, M.P., 2010. Structural vs. atheoretic approaches to econometrics. Journal of Econometrics, 156, 1, 3-20.

Lewbel, A., 2019. The identification zoo: Meanings of identification in econometrics. Journal of Economic Literature, 57, 4, 835-903.