1. Complete the FHK replication exercise (separate distribution). Do you think it is useful to be able to reproduce results in this way?

The outputs of replication with different code are shown in the Appendix of this document.

Yes, but a more useful way is write the code totally by ourselves, instead of repeating the authors, even if the results are robust to deleting replicates of ‘gvkey’, no CL-2, and winsorization and so on. In that way, we can avoid making the same mistakes that the authors may do.

1. How easy do you think it would be to reproduce the results of FHK if you did not have the code I gave to you?

It is difficult. At least, it will take much longer time for me to figure out the code.

1. Look at the datasheet and code supplements provided for Volume 58, Issue 1 of the *Journal of Accounting Research* [here](https://research.chicagobooth.edu/arc/journal-of-accounting-research/online-supplements/volume-58). How would you rate each of these in terms of (i) ability to reproduce results from data provided and (ii) ability to produce the data using the code and documentation provided?

**Paper 1: Mihir N. Mehta, Suraj Srinivasan, and Wanli Zhao**

(1) The data provided only include ‘year’, ‘gvkey\_acq’, ‘gvkey\_tar’. I cannot reproduce results from these data, because the dependent variables, independent variables, and controls are not included.

(2) The document “MSZ datasheet” describes that the authors manually review the fuzzy matches, hand-collect merger antitrust review approval announcement date by searching multiple sources, determine the reason for each committee member departure case by by reviewing some datasets, and use several keywords to match BoardEx. These steps are difficult to produce the data because the multiple sources may change and there may be some subjective judgments in the manual collection of data. Who will spend the same effort as the authors do in hand-collecting data, just to reproduce the results?

**Paper 2: Pietro Bonetti, Miguel Duro, and Gaizka Ormazabal**

(1) The data provided only include ‘CUSIP’. I cannot reproduce results from the data.

(2) The document “MDO datasheet” describes that the authors obtain ownership data from Amadeus-Bureau van Dijk. I think BvD provides the latest ownership, but does not provide history data. I am not sure whether this influences their data collection. BTW, is there any dataset that provides history ownership of small private firms (not large private firms that issue debts and then disclose information)?

The code describes how the authors clean the data, but do not describe how they run empirical analyses.

**Paper 3: Lisa DeSimone, Rebecca Lester and Kevin Markle**

(1) The authors do not provide any data, so I cannot reproduce results.

(2) It is good that the authors provide the website of the data source, but the data source of the key dependent variable ‘tax evasion’ is ambiguous. The authors compile the dataset from the University of Michigan, the OECD, KPMG, the Tax Foundation, and Trading Economics, but it is unclear about the specific data source in these organizations, so it may be difficult to reproduce the data.

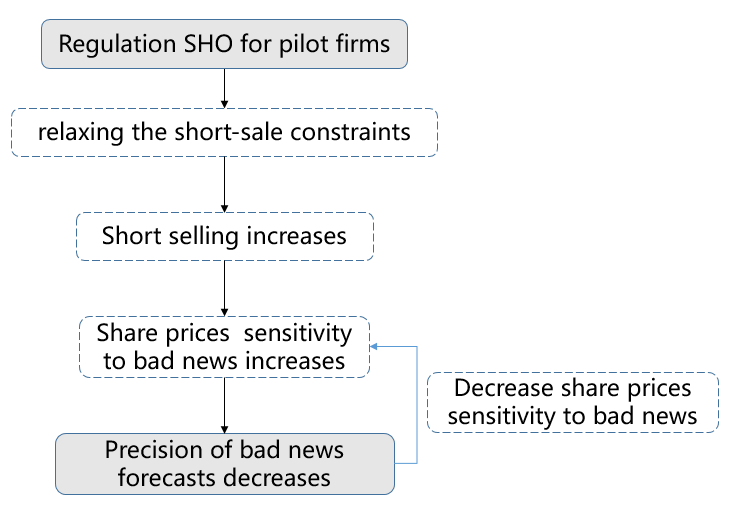
The authors also paid for the data on Swiss Impex site. Who will pay just to reproduce the results?

**Paper 4: Frank Zhou and Yuqing Zhou**

(1) The data provided only include datadate, rdq, permno. I cannot reproduce results from these data.

(2) All data are obtained from WRDS. The authors group different steps into different files, and describe the data cleaning clearly. I think we can reproduce data using the code and documentation provided.

1. Draw a causal diagram linking the Reg SHO pilot treatment to the outcome variables studied in Li and Zhang (2015).



1. Given your causal diagram, is the research design used in Li and Zhang (2015) subject to the criticism of BHLYY with regard to conditioning on post-treatment outcome variables? Can you suggest a way to modify to Li and Zhang (2015) to address this concern?

Yes. BHLYY criticize that “*Absent evidence for a causal channel running from relaxing short sale rules, through higher short interest and lower share price for pilot firms, any observed differences between pilot and control firms are unlikely to be due to the experiment and are instead likely to be false positives*”.

In Li and Zhang (2015), we cannot observe *higher short interest and lower share price for pilot firms*. The change of observed share prices sensitivity to bad news can also not validate the channel that mangers fear lower prices.

A structural model may help to make the causal relationship clearer, but it cannot solve BHLYY’s criticism. If in the real world, *higher short interest and lower share price for pilot firms* indeed do not happen, a structural model is also an imagined ‘story’.

However, why managers must observe the outcome and then fear? They can just make decisions based on their expectation, though this expectation may be wrong. Of course, if so, their decision is not in an equilibrium. After they find that their expectation is wrong, they may reverse their over-reacted decisions.

Besides, the conclusion of the first order effect “short selling” seems inconsistent. Li and Zhang (2015) argue that “*As expected,* ***short selling activities*** *increased significantly for the pilot stocks relative to other stocks during the suspension period (e.g., Diether, Lee, and Werner [2009], Grullon, Michenaud, and Weston [2012], Angelis, Grullon, and Michenaud [2013])*”. However, BHLYY criticize that Diether et al (2009) lacks a plausible causal channel. Heath et al. (2019) find that Regulation SHO did not significantly alter the level of short selling, measured by short interest divided by shares outstanding (Page 28) after adjusting critical values. I think Li and Zhang (2015) do not explain clearly the measure of short selling activities and what kind of short selling activities increased.

1. Do you believe that the concern of Simmons et al. (2011) that “undisclosed flexibility in data collection and analysis allows presenting anything as significant” applies to accounting research? What do you think about their proposed solutions?

Yes. I think these solutions are useful but these are just common issues that authors and reviewers know. The scientific degree of the research depends on whether the disclosure or judgment is convincing enough, and whether authors and reviewers think about their research carefully. Otherwise, the disclosure will become generic and boilerplate.

I was wondering why there are media monitoring research integrity in science (https://forbetterscience.com/about/ ), but not in accounting. I think the reason depends on who cares about whether the results are robust or not.

**Requirement for authors:**

R1: decide the rule for terminating data collection ex ante. In empirical studies, authors usually say that they collect data and then conduct tests. It seems that the end of the sample period is usually three or four years before the publication date. Or, the sample period starts in the year when dataset is available. I think it is difficult to decide the rule ex ante, but the authors should disclose how they judge the start and end date of the sample period.

R2 “20 observations” and R3 “all variables collected” do not apply for empirical studies, in which there are more than 20 observations and the variables are just in the dataset. Researchers just make decisions on which variables should be used.

R4 “Failed manipulations”, R5 “eliminated observations” and “covariates” are useful. Researchers should report whether their results are robust to different research designs and choices.

**Guidelines for reviewers**:

G1, G2, and G3 ask reviewers to accept imperfect results and to ensure the results robustness. G4 “conducting an exact replication” does not apply for empirical studies because the code are just there. Unlike participants in experiments, raw data in empirical studies are less likely to change.

Appendix:

