Readme File

These files contain the data and codes used to generate the principal findings in Fang, Huang and Karpoff (JF, 2016), on the Regulation SHO pilot program's effect on discretionary accruals of the treated pilot firms relative to nonpilot firms.

The key dataset containing the list of pilot vs. nonpilot firms has not been changed since 2011. We have shared this dataset upon request beginning in 2014, and in total have shared it with more than 120 researchers. We are posting the data and codes now to make it easier for researchers to evaluate the claim in a new paper by Black et al. (2019) that our principal findings cannot be replicated. Contrary to this claim, we believe that the findings that pilot firms have lower discretionary accruals during the treatment period are robust and easy to replicate. Replicability is important in empirical research. We applaud Black et al. for their replication efforts and welcome additional replication efforts from interested parties. In full disclosure, the Black et al. team has had our pilot data and some of our codes posted below since October 2015. Independently, Massa, Zhang, and Zhang (RFS 2015, Table 4) also find that pilot firms have lower discretionary accruals during the pilot period.

The codes and data available here allow you to replicate the basic findings of Fang et al. (2016), and to probe the robustness of these findings by considering alternative accrual measures and/or model specifications. Before using these data, however, please make sure to obtain permission from the Russell Company to access the Russell 2004 3000 index. By accessing these data, you also must agree that you have legal access to WRDS.

For ease of navigation, we divided the data and codes into two parts:

The first part – the folder titled "Codes and Data for Discretionary Accruals" – includes:

1. A SAS dataset titled "pilot" containing the list of 2004 Russell 3000 index, which we manually matched to CRSP. These data have not been changed since 2011.

2. A SAS dataset titled "pmda" containing four versions of the performance matched discretionary accrual measure. This dataset was created using Compustat in September 2012 and has not been changed since then.

3. A SAS dataset titled "ff_sic_match" containing Fama-French 49 industry classification codes. This data was downloaded from Kenneth French's data library.

4 & 5. A SAS program titled "SAS codes for replication_Sep 2019" and a STATA do file titled "Stata codes for replication_Sep 2019." These are streamlined versions of the codes used in our original research (to make it easier to follow and to incorporate both 2012 and current versions of the Compustat data). Please save all files to your C drive (or whichever local drive that you would like to use and change the directory in both files accordingly), and remote login to WRDS. You then should be able to access the three datasets listed above in SAS and reproduce Table 3 in our paper using the four versions of performance matched discretionary accrual measures created in 2012, as well as eight additional

measures of performance-matched discretionary accruals using the current Compustat. The results hold across the board and, in fact, are stronger using the current Compustat data. Please follow the notes in the STATA codes for definitions of the 12 different measures of discretionary accruals and instructions here on how to install cluster2 (https://www.kellogg.northwestern.edu/faculty/petersen/htm/papers/se/se_programming.htm) if you have not already done so.

6. A SAS program titled "Codes_Sep 2012," which is the original set of codes written to create the four versions of performance matched discretionary accrual measures in 2012. If you happen to have an excerpt of the 2012 Compustat data, you are welcome to run them as well.

The second part – the folder titled "Codes and Data for Total Accruals" – includes:

1. A SAS dataset titled "pilot" containing the list of 2004 Russell 3000 index. This is the same as item 1 above.

2 & 3. Two SAS programs titled "Total Accruals" and "PSMatching," respectively. We wrote these programs in response to a question posed by Hemang Desai in 2014 and then again in 2015 about the replicability of our principal findings using total accruals instead of discretionary accruals. The results are robust to using total accruals. We thank Hemang for this suggestion, and sent these data, codes, and results (i.e., all of the files in this folder) to him and Jeff Yu in October 2015.

4 & 5. Two excel files containing the results using total accruals (unmatched) and total accruals (matching using the level of total accruals and the change in total accruals from 2002 to 2003). The results are again robust.

References:

Black, Bernard, Hemang Desai, Kate Litvak, Woongsun Yoo, and Jeff Yu, 2019. The Reg SHO reanalysis project: reconsidering of Fang, Huang and Karpoff (2016) on Reg SHO and earnings management. *Working Paper*.

Fang, Vivian W., Allen H. Huang, and Jonathan M. Karpoff, 2016. Short selling and earnings management: A controlled experiment. *Journal of Finance* 71: 1251-1294.

Massa, Massimo, Bohui Zhang, and Hong Zhang, 2015. The invisible hand of short selling: Does short selling discipline earnings management? *Review of Financial Studies* 28: 1701-1736.