This is the replication read me file for the paper *How Wide is the Firm Border?*. For any questions regarding the replication of this paper, please send an e-mail to any of the following addresses: atalayecon@gmail.com, hortacsu@uchicago.edu, mli@compasslexecon.com, or Chad.Syverson@chicagobooth.edu.

Enghin Atalay, Ali Hortaçsu, Mary Jialin Li, and Chad Syverson

August 14, 2019.

Access to the Data

The three primary datasets used for this study are the Census of Manufacturers, the Commodity Flow Survey, and the Longitudinal Business Database. Access to these data is administered by the Census Bureau, via the Center for Economic Studies (CES) program. For approved projects only, the data can be accessed via the Census Research Data Centers. See https://www.census.gov/ces/rdcresearch/howtoapply.html for information on how to apply for access to the Census data.

Programs

As discussed in "data_compile.do", portions of the code rely on datasets (mainly sicTable01.csv) which were produced in our earlier paper, *Vertical Integration and Input Flows.* As a result, it may be necessary to consult the replication files for our earlier paper, which can be found at https://www.aeaweb.org/articles?id=10.1257/aer.104.4.1120 .

1) data_compile.do : This file constructs the intermediate datasets (primarily the dataset of sending-establishment-by-destination-ZIP code pairs, and the dataset of merged and incidentally merged establishments.)

2) data_analyze.do : This file produces all of the tables and figures contained in Sections 2-4, and Appendix C.

3) gmmiv.do : This file performs the simulations discussed in Appendix D.

4) gmm_poiend.ado . This file is called by data_analyze.do and gmmiv.do . It programs our GMM estimator (discussed in Online Tables 16-17.)