Data Description Sheet

Item 1. A description of which author(s) handled the data and conducted the analyses.

All authors handled the data used in this paper and Guoman She conducted the analyses.

Item 2. A detailed description of how the raw data were obtained or generated, including data sources, the date(s) on which data were downloaded or obtained, and the instrument used to generate the data (e.g., for surveys or experiments). We recommend that more than one author is able to vouch for the stated source of the raw data.

We provide detailed information about each database used in the most recent version of our manuscript below:

Conference Calls: We obtain our sample of quarterly earnings conference call transcripts over the 2002 to 2012 period from Thomson Reuters. We most recently collected data from this source on October 2017. The transcripts are in XML format. We parse the date of the call, firm identifiers and the types of conference calls from XML-friendly headers. We then link the earnings conference call transcripts with Compustat based on the firm identifiers. The transcripts also provide tags to identify different segments of the call (e.g., presentation and discussion) and individuals. We parse the text of presentation by managers using a Perl script and calculate the proportion of product-market-related related words in each call.

Supply Contracts: We obtain supplier contracts from SEC Edgar. We first obtain the URL address of annual, quarterly and current reports filed by nonfinancial firms incorporated in the U.S. from WRDS, then download all material business contracts filed as Exhibit 10 through the 10-K, 10-Q and 10-K. We use a Perl script to parse the title of each contract and retain only supply contracts of which the filer is the seller or reseller. We collected data from this source between June 2016 and January 2017. Appendix A3 of the submitted manuscript provides detailed information about the process.

Import and export data: We obtain the U.S. import and export data at the SIC industry level from http://faculty.som.yale.edu/peterschott/sub_international.htm. We most recently collected data from this source on January 2017. We also gather the information about the imports of final goods from the World Input-Output Database, available at http://www.wiod.org/database/int_suts13, and the crosswalk file from International SIC to U.S. SIC using the concordance table provided by Professor Jon Haveman (https://www.macalester.edu/research/economics/page/haveman/Trade.Resources/tradeconcordances.html).

Cartel Enforcement Database: We gather data on actual convicted cartel activity from Connor (2014) and link it to Compustat using company name. We bought an updated version of this dataset from Professor Connor on January 2017.

Financial Information: We gather financial information from Compustat database, stock price information from CRSP database. We also gather institutional ownership data from Thomson Reuters, Census HHI and Census industry sales from the U.S. Bureau (https://www.census.gov) and pair-wise product similarity score from Hoberg-Phillips data library (http://hobergphillips.usc.edu/). We most recently collected data from this source on March 2019.

Producer Price Index: We gather the Producer Price Index (PPI) data from the Bureau of Labor Statistics (https://www.bls.gov/ppi/data.htm), and convert the NAICS code to SIC using the crosswalk files obtained from http://www.ddorn.net/data.htm. We most recently collected data from this source on June 2017.

SEC Server Request Records: We obtain the server request records from the EDGAR Log File Data Set available on the SEC's Web servers. We determine the type and filing time of firms' filings using the EDGAR Master File. We most recently collected data from this source on January 2017.

Rule of Law: We obtain the indicator of the rule of law from The Worldwide Governance Indicators (WGI) project (*http://info.worldbank.org/governance/wgi/index.aspx#home*). We most recently collected data from this source on June 2017.

Competitors and Clients: We collect the news announcement related to new clients from CapitalIQ Key Development Data on WRDS. We collected data from this source on November 2016. We identify a firm's competitors using the Factset Revere relationships database on WRDS. We use the relationships identified by Factset Revere from April 2003 to December 2012. We most recently collected data from this source on December 2017.

Country-level Characteristics: We follow Dong et al. (2019) and collect country-level characteristics, including gross domestic production (GDP), unemployment rates, exports of goods and services, and domestic credit to private sector from the World Bank. The Chinn-Ito index is collected from http://web.pdx.edu/~ito/Chinn-Ito_website.htm. We most recently collected data from this source on July 2017.

Item 3. If the data are obtained from an organization on a proprietary basis, the authors should privately provide the editors with contact information for a representative of the organization who can confirm data were obtained by the authors. The editors would not make this information publicly available. The authors should also provide information to the editors about the data sharing agreement with the organization (e.g., non-disclosure agreements, any restrictions imposed by the organization on the authors, such as restrictions to publish certain results).

All data used in this paper is available through standard commercial databases except the cartel data that we purchased directly from Professor John Connor at Purdue University. Professor Connor can be reached at jconnor@purdue.edu and we are willing to share our

purchase agreement (funded using a RGC Grant while all three authors were at the HKUST).

4. A complete description of the steps necessary to collect and process the data used in the final analyses reported in the paper. For experimental and survey papers, we require information about the instructions and instruments used to generate the data, subject eligibility and/or selection, as well as any exclusion criteria. The full set of instructions and instruments can be provided in the online appendix.

Raw data were obtained from the sources listed in Item #2. The steps necessary to collect and process the data used in the final analyses reported in the paper are described in section 4 of the paper.

5. The computer programs or code used to convert the raw data into the final dataset used in the analysis plus a brief description that enables other researchers to use this program. The purpose of this requirement is to facilitate replication and to help other researchers understand in detail how the raw data were processed, the final sample was formed, variables were defined, outliers were treated, etc. The code or programming is in most circumstances not proprietary. However, we recognize that some parts of the code or data generation process may be proprietary, including from the authors' perspective. Therefore, instead of the code or program, researchers can provide a detailed step-by-step description of the code or the relevant parts of the code such that it enables other researchers to arrive at the same final dataset used in the analysis. In such cases, the authors should inform the editors upon initial submission, so that the editors can consider an exemption from the code sharing requirement. Whenever feasible, authors should also provide the identifiers (e.g., CIK, CUSIP) for their final sample. Authors should consult our FAQ Sheet on the JAR website for further details.

All data manipulation and construction of the datasets for analyses were done using a combination of the following three languages: Perl, SAS and STATA. The final analyses were conducted using STATA. We describe the processes used to create the main dataset in Section 4.2 of the paper. For further details, see the programs enclosed (1process conference calls.pdf; 2main.do; 3additional variable construction.sas). The GVKEY identifying firms included in our final sample are reported in "Final sample (GVKEY).csv". The CID identifying conference call transcripts included in our final sample are reported in "Earnings call (CID) used in tests.csv".

6. An assurance that the data and programs will be maintained by at least one author (usually the corresponding author) for at least six years, consistent with National Science Foundation guidelines.

We agree to maintain all data and programs for at least six years suggested by the National Science Foundation.

References

Connor, J. “The Private International Cartels (PIC) Data Set: Guide and Summary Statistics, 1990-2013.” (2004). Working Paper.
Dong, A.; M. Massa; and A. Zaldokas. “The Effects of Global Leniency Programs on Margins and Mergers.” RAND Journal of Economics 50 (2019): 883-915.