

INSTRUCTIONS TO REPLICATE THE RESULTS IN: ‘Detecting Bidder Groups in Collusive Auctions’

The folder *Dataset_Do_and_M_Files* contains this ReadMe.pdf that you are reading as well as 2 other folders: *Stata* and *Matlab*. Their content is as follows:

- The folder *Matlab* contains the Matlab programs performing the bid and participation tests. The folder contains two subfolders, one for each test.
 - The folder *PartiticipationTest*: contains the following files:
 - * *MasterPT.m*: This is the main routine to execute the participation test. It loads the dataset, sets various parameters in relation to the dataset and calls the subroutine executing the test.
 - * *ValidationData.xls* and *SuspectCartels.csv*: Are two datasets. *ValidationData.xls* is a dataset of auctions with the associated firm-specific and auction-specific identifiers. The dataset provided for the replication is the Validation data. *SuspectCartels.csv* is a dataset of suspect coordinating groups formed by the firms participating in *ValidationData.xls*. In the dataset provided for replication, the groups are the 8 cartels of the Turin case.
 - * *ParticipationTest.m*: This is the main subroutine that *MasterPT.m* calls to execute the participation test.
 - * *frequency_categories.m*, *frequency_miles_capital.m*, *info_auction_bins.m1.m* and *participation_xmatlab.mat*: These are three ancillary subroutines and one data files that are required to perform the participation test. These file serve to create the comparison groups for this test by selecting firms on the basis of their legal qualifications to bid, capitalization and distance.
 - To replicate the results of the participation test for the Validation data:
 1. Unzip the folder *PartiticipationTest* and verify it contains all the files listed above.
 2. Set the Matlab path to this folder.
 3. Execute the file *MasterPT.m*. This will replicate the results of the participation test for the Validation data (i.e., it will reproduce Figure 2 in the data). Running the code multiple times will produce slightly different results each times, this is due to the number of resampling being set at 100 to speed up the execution of the whole code. Set the parameter *bootstrap_winner* to reduce the randomness. The code allows to set a few other parameters, see instructions at the beginning of *MasterPT.m*.
 - The folder *BidTest*: contains the following files:
 - * *MasterBT.m* and *MasterBT_withbins.m*: These are the main routines to execute the bid test without conditioning on covariates (*MasterBT.m*) or conditioning on them (*MasterBT_withbins.m*). It loads the dataset, sets various parameters in relation to the dataset and calls the subroutine executing the test.

- * *ValidationData.xls*, *SuspectCartels.csv* and *cartel_groups*: the first two are the datasets described above. The third file is a dataset reporting the subset of cartel members used for the test.
 - * *dataset_4auctions_new.mat* and *dataset_8auctions_new.mat*: matlab datasets with tuples of auctions to be analyzed;
 - * *frequency_categories.m*, *frequency_miles_capital.m*, *info_auction*, *participation_xmatlab*: subfiles used for the results with bins.
- To replicate the results of the bid test for the Validation data:
 1. Unzip the folder *BidTest* and verify it contains all the files listed above.
 2. Set the Matlab path to this folder.
 3. open *MasterBT.m* if you want to replicate the results without bins (Table A1 and the odd numbered columns of Table 5) or *MasterBT_withbins.m* if you want to replicate the results with bins (Table A2 and the even numbered columns of Table 5).
 4. set variable *cartel_interest_superflex* equal to the cartel number you want to analyze (an integer between 1 and 8).
 5. Run the file. Running the code multiple times will produce slightly different results each times, this is due to the number of resampling being set at 100 to speed up the execution of the whole code. Set the parameter *bootstrap_winner*, *n_couples_auctions_used_max* and *n_dozen_firms_used* to reduce the randomness. The code allows to set a few other parameters, see instructions at the beginning of the master Matlab file.
 - The folder *Stata* contains the Stata programs performing all the analyses in the paper complementing the bid and participation test. This folder contains the following subfolders:
 - The folder *Table1_SumStats_MainData* contains the Main data as well as the code producing the summary statistics reported in Table 1 of the paper.
 - The folder *Table2_ProbitOLS_MainData* contains a version of the Main data expanded to account for all the firms that were potential participants in each auction (see note to Table 2 in the manuscript) and shortened by excluding all auctions for which not all bids were observed. The folder contains also the do file performing the regressions shown in Table 2.
 - The folder *Table3and4_SumStats_ValidData* contains the Validation data as well as the code producing the summary statistics reported in Table 3 and 4 of the paper.
 - The folder *Table6and7_ProbitClustering_ValidData* contains a dataset reporting the type of linkages existing between each pair of firms appearing in the Validation data: *network_ValidationData.dta*. This dataset is used as input in the file *Table6.do* to perform the probit regressions shown in the paper (Table 6). This same dataset is used as input in the file *Clustering.do* to illustrate the functioning of our clustering approach for the construction of candidate groups for the tests.

The clustering algorithm creates the groups analyzed in Table 7 (when applied to the Validation data) as well as those analyzed in Table 8 (when applied to the Main data).

- The folder *Figure4and5* contains a dataset that combines the Main and Validation data and allow to replicate (through the file *Figure4and5.do*) both Figure 4 and Figure 5 in the paper.
- To replicate the analysis performed by the Stata codes:
 1. Enter into the desired subfolder of the *Stata* folder.
 2. Change the path in the initial lines of the do file;
 3. Run the *do* file.
- Finally, to replicate the datasets used for the analysis, various publicly available data shall be combined together. The data can be obtained from the following sources:

1. ABAs and FPAs in the Main data were manually extracted from the pdf files collected by the company Telemat Spa that collects them from the individual auctioneers and resells them to construction firms and other interested parties:

<http://www.telemat.it/>.

2. The ABAs in the Validation data were manually collected from the documentation made available by the legal office of the Municipality of Turin with regard to the case *Turin Court of Justice, 1st criminal Section, sentence N. 2549/06 R.G., 04/28/2008*.

<http://www.comune.torino.it/en/>,

3. The data on firm characteristics comes from the Italian Registry of Firms:

http://www.infocamere.it/eng/about_us.htm#.

4. All measures of distance between firms and PAs are at their zip code level and were obtained through the freely available API of:

<http://classic.mapquest.com>.

5. The source for the data on public administrations is Italy's National Statistical Institute:

http://demo.istat.it/index_e.html.

Please let us know at fdc@bu.edu if you have any comment.