

# Replication Materials for “Winning by Default: Why is There So Little Competition in Government Procurement?”

Karam Kang and Robert A. Miller

This document provides descriptions for each file in the replication materials. The materials consist of three folders, entitled `data`, `estimation`, and `figures_and_tables`. The `data` folder includes all data files used for estimating the model and replicating the figures and tables in the paper. The `estimation` folder walks the reader through all files for the estimation procedure. Lastly, the `figures_and_tables` folder contains code to create figures and tables.

## Data

This section describes the `data` folder.

### Dataset list

All datasets in the `data` folder are provided in the Stata data file format (DTA), and all variables in each file are labeled with a short, informative description.

Data file	Source	Notes	Provided
<code>data/agency_employment.dta</code>	FedScope	Aggregated at the contracting agency level	Yes
<code>data/contractlist.dta</code>	FPDS	Derived from <code>data/contracts.csv</code>	Yes
<code>data/contracts.csv</code>	FPDS	Downloadable from <a href="http://usaspending.gov">usaspending.gov</a>	No
<code>data/contracts_attributes.dta</code>	FPDS	Derived from <code>data/contracts.csv</code>	Yes
<code>data/contracts_key.dta</code>	FPDS	Derived from <code>data/contracts.csv</code>	Yes
<code>data/cpiindex.dta</code>	BLS	Saved in the DTA format	Yes
<code>data/establishments_state.dta</code>	Census	Aggregated at the state level	Yes
<code>data/final_data.dta</code>	Multiple	Combines multiple data sources, serves as input for constructing <code>data/final_sample.dta</code> and creating Tables 1 and A2	Yes

Data file	Source	Notes	Provided
<code>data/final_sample.dta</code>	Multiple	Dataset used in the main analysis; input for creating all tables and figures except Tables 1 and A2	Yes
<code>data/house_assignments.dta</code>	Charles Stewart	Saved in the DTA format	Yes
<code>data/pscode_match.dta</code>	FPDS	Derived from <code>data/manuals/PSC Manual.pdf</code>	Yes
<code>data/senate_assignments.dta</code>	Charles Stewart	Saved in the DTA format	Yes
<code>data/solicitation_data.dta</code>	Authors	Collected from <code>fbo.gov</code>	Yes

## Code to generate datasets

- `data/contracts.do`: This DO-file cleans the procurement data from `usaspending.gov` and produces `data/contracts_key.dta` and `data/contracts_attributes.dta`.
- `data/final_data.do`: This DO-file creates key variables for the analysis and merges various datasets to construct `data/final_data.dta`. See Appendices A.1.1.–A.1.6 for the description of how each of the variables used in the estimation was constructed.
- `data/final_data_addendum.do`: This DO-file corrects an error in constructing “`ncont3_pas`” variable, which measures the experience of a contracting agency in dealing with similar contracts to a given contract. This error was found after the editorial process was completed. We find that the main results in the paper are robust to this error. The datasets provided in the replication materials include the variable with the error, but we provide this correction DO-file for those who are interested in using the variable for their own research.
- `data/final_sample.do`: This DO-file creates the data used in the main analysis, `data/final_sample.dta` and `estimation/main/data_main.csv`.
- `data/sensitivitydata_alt*.do`: These two DO-files create the datasets used in the sensitivity analysis, `estimation/sensitivity/data_alt*.csv`.
- `data/sensitivitydata_sub.do`: This DO-files creates the datasets used in the sensitivity analysis, `estimation/sensitivity/data_sub*.csv`.

## Further notes

- Datasets used for the main analysis: Two DTA-files, `data/final_data.dta` and `data/final_sample.dta`, are used for the main analysis in the study. The former dataset includes information on all definitive contracts or purchase orders of 2000–2018 with nominal size of at least \$150,000, in total of 628,652 contracts. The latter dataset is based on the former (see `data/final_sample.do`), containing the final sample of 6,981 contracts.
- Procurement contract datasets: There are four DTA-files derived from the FPDS procurement contract datasets, which we downloaded from `usaspending.gov` in July 2018. After saving the raw datasets from the government website as `data/contracts.csv` and running `data/contracts.do` will generate `data/contracts_key.dta` and `data/contracts_`

`attributes.dta`. Creating `data/contractlist.dta` is straightforward (simply a list); we constructed `data/pcode_match.dta` based on the information on the history of Product and Service Code (PSC) changes, as specified in the manual provided in the replication materials for reference (`data/manuals/PSC Manual.pdf`).

- Agency employment dataset (`data/agency_employment.dta`): We downloaded the federal government employment data, the FedScope Employment Cube, in August 2018, and Appendix A.1.7 provides a detailed description on how to construct (1) a variable that measures the level of contracting officers' government experience in an agency, and (2) another variable that counts the number of contracting officers hired by the agency. These variables are straightforward to construct, and the only tricky part is to match the government agency names between this dataset and the procurement dataset. We created an agency ID variable, called `agencyid`, to match these two datasets.
- Consumer price index (CPI) dataset (`data/cpiindex.dta`): The CPI index data was sourced from US Department of Labor Bureau of Labor Statistics. The link is <http://www.usinflationcalculator.com/inflation/consumer-price-index-and-annual-percent-changes-from-1913-to-2008/>. We re-weighted the CPI indices by setting the index of December 2010 as 100.
- Number of establishments dataset (`data/establishments_state.dta`): We downloaded the original dataset from County Business Patterns (<https://www.census.gov/programs-surveys/cbp/data/datasets.html>) in October 2018. We aggregated the number of establishments at the state and the year level. See Appendix A.1.8 for more information.
- Congressional assignment datasets (`data/house_assignments.dta` and `data/senate_assignments.dta`): We downloaded the congressional committee data collected by Charles Stewart III and Jonathan Woon. The link is [http://web.mit.edu/17.251/www/data\\_page.html](http://web.mit.edu/17.251/www/data_page.html), and you may look at the section called "Congressional Committees, Modern Standing Committees, 103rd–115th Congresses (updated to November 17, 2017)." The DTA-files are basically identical to the Excel files provided in the website, except the file format. The collectors of these datasets state that "[t]he data are made available for academic use, but may not be used commercially."
- Online advertisement dataset (`data/solicitation_data.dta`): We built this dataset on the public notices posted on the federal business opportunity website ([www.fbo.gov](http://www.fbo.gov)) regarding the procurement contracts in our sample. Appendix A.1.9 provides further details.
- Manuals: Two PDF files, `data/manuals/FPDSNGDataDictionary.pdf` and `data/manuals/USASpendingDataDictionary.pdf`, provide the definition of each variable in the FPDS dataset. The information on the PSC is presented in `data/manuals/PSCManual.pdf`.

## Programs

This section describes the `estimation` and the `figures_and_tables` folders.

## Computational requirements

The code provided here is written for Matlab and Stata. Portions of the code (for optimization) use Knitro/MATLAB (see the reference at [https://www.artelys.com/docs/knitro/3\\_referenceManual/knitromatlabReference.html](https://www.artelys.com/docs/knitro/3_referenceManual/knitromatlabReference.html)). The code was last run on a quad-core Intel-based desktop with MacOS

version 10.15.7. Computation took about three hours for estimation, about two weeks for generating all bootstrap results, and about two days for obtaining all sensitivity results.

## Description of programs

- Programs in `estimation/main` will generate the results of the main estimation. The file `estimation.m` will import the final sample data, `data_main.csv`, run the estimation, and generate the estimation result file, `estimation.mat`. The functions used in the estimation are provided in the folder (`fun_*.m` and `halton.m`).
- Programs in `estimation/bootstrap` will generate all bootstrap results. The file `bootstrap_simdata.m` will create bootstrap samples. Sequentially running `bootstrap_steps12.m` and `bootstrap_steps345.m` will produce the estimation results of each bootstrap sample. The file `results.m` will generate `results.mat`, which summarizes the bootstrap results. The functions used in the estimation are provided in the folder (`fun_*.m` and `halton.m`).
- Programs in `estimation/sensitivity` will generate the results of fifteen different sensitivity analyses. The main dataset in the folder, entitled as `data_main.csv`, is used for nine of these sensitivity exercises. The remaining six sensitivity analyses are based on `data_alt*.csv` and `data_sub*.csv`. For each sensitivity analysis, one can run the corresponding file `sens_v*.m`, generating the respective results summarized in `sens_v*.mat`. The functions used in the sensitivity analyses are provided in the folder (`fun_*.m` and `halton.m`).
- Programs in `figures_and_tables` will generate all figures and tables in the study. Running `figure2.m` and `figureA2.m` will generate Figures 2 and A2, respectively. Running `tableA11.m` will generate `tableA11.csv`, which serves as input for `tables.do`. Running the latter DO-file will create all statistics for Tables 1–4, A1–A8, and A11. It will also create Figure A3. Running `tables.m` will generate statistics for Tables 4–6, A9–A10, and A12–A13.

## Instructions to replicators

- Run `estimation/main/estimation.m` to estimate the parameters of the main model specification. It will generate `estimation/main/estimation.mat`.
- Create an empty folder `estimation/bootstrap/estimates`, and save `estimation/main/estimation.mat` as `estimation/bootstrap/estimates/est0.mat`. Also, save the same file as `estimation/sensitivity/sens_v0.mat`.
- Create an empty folder `estimation/bootstrap/simdata`, and run `estimation/bootstrap/bootstrap_simdata.m`. It will generate bootstrap samples in `estimation/bootstrap/simdata`.
- Run `estimation/bootstrap/bootstrap_steps12.m` and `estimation/bootstrap/bootstrap_steps345.m` sequentially. It will generate estimation results for each bootstrap sample in `estimation/bootstrap/estimates`.
- Run `estimation/bootstrap/results.m`, which will generate `estimation/bootstrap/results.mat`.
- Run `estimation/sensitivity/sens_v*.m`, which will generate `estimation/sensitivity/sens_v*.mat`.
- Run `estimation/sensitivity/sensitivity_results.m`, which will generate `estimation/sensitivity/sensitivity_results.mat`.
- Run `figures_and_tables/tableA11.m`, which will generate `figures_and_tables/tableA11.csv`.
- Run `figures_and_tables/tables.do`, `figures_and_tables/tables.m`, `figures_and_tables/figure2.m`, and `figures_and_tables/figureA2.m` to generate figures and tables.