# The Oligarch Vanishes: Defensive Ownership, Property Rights, and Political Connections

## ReadMe

John S. Earle, Solomiya Shpak, Anton Shirikov, and Scott Gehlbach

August 5, 2023

The files included here are necessary to replicate all analyses, figures, and tables in the article and the online appendix. This ReadMe describes the scripts used to prepare and analyze the data on oligarch ownership and political connections and specifies the order in which the scripts should be executed.

**Note: after downloading the full set of files, create the following additional empty subfolders in the main directory of the data set: Network Data; Ownership Data; Regressions Data; Results. The files created during the analysis will be placed in these folders.**

There are four main sets of scripts that correspond to the four steps in the analysis:

(1) Scripts to extract ownership links from existing firm registries.
(2) Scripts to reconstruct ownership networks and generate the key variables of interest (measures of defensive ownership).
(3) Scripts to create samples with firm-level data for the analysis.
(4) Scripts to estimate regressions and generate tables for the paper.

The scripts are to be executed in that exact order. In each of the four sections below, we list the specific scripts and describe the purpose of each script. (The first, third, and fourth sets of scripts are written in Stata. The second is written in R.)

The short version is as follows. There are three scripts in the main directory that need to be executed in succession:

- ***owner_extraction_main_script.do*** (step 1)
- ***network_main_script.R*** (step 2)
- ***analysis.do*** (steps 3 and 4)

The script ***figures_tables.R*** produces additional figures and tables; it should be run after executing ***network_main_script.R*** (step 2).

The generated results (tables and figures) are placed in the folder *Results*.

The rest of the document describes these steps in greater detail.

To run the R scripts, the package *pacman* should be installed; all other necessary packages are installed/loaded automatically via *pacman* when the scripts are executed. To run the Stata scripts, packages *outreg2*, *ranktest*, and *ivreg2* need to be installed.

**Step 1. Extracting ownership links**

This set of scripts extracts the ownership records from two firm registries, JSCReg and SReg (see the corresponding SReg.dta and JSCReg.dta files). JSCReg is a registry of joint-stock companies, and SReg is a registry of all Ukrainian firms (see the paper for other details). Each entry in these registries is an ownership link, and it contains a firm's identifying code, its name, its owner's name and code, date of entry, and other information. Our main algorithm starts with the set of oligarch-owned firms and extracts their ownership records; then, for each owner identified in this way, we extract its owners, and we repeat these iterations until we reach the end owners (no more new owners can be added). We also perform the same steps for all the firms included in JSCReg.

Step 1 is performed by executing the Stata script ***owner_extraction_main_script.do***. This is the master script that runs all the other scripts to extract ownership links for three selected dates in 2002, 2004, and 2006. Running this file in Stata produces a set of files in the subfolder *Ownership Data*. Files called "all_entities_fin_*.dta" contain the sets of ownership links used later to construct ownership networks for the corresponding dates (in 2002, 2004, and 2006) based on JSCReg and SReg or on JSCReg only. Files called "Delo_UP_present_*.dta" contain information on whether each of the initial oligarch-owned firms can be found in JSCReg and SReg on corresponding dates.

The following scripts are run automatically when *owner_extraction_main_script.do* is executed (**that is, they are not to be executed manually and separately from the main script**):

- *owner_search_Delo_UP.do*: This script extracts ownership links for the initial set of oligarch-controlled firms ("Delo/UP firms;" see the paper for details), first extracting such links from JSCReg and then from SReg, repeating the process as many times as needed. This script, in turn, executes a series of other scripts, listed immediately below, to extract and clean the data and prepare the network analysis:
    - *extract_owners_JSCReg.do*: This script extracts all ownership links from JSCReg for a specified set of firm identifiers (represented in the variable "okpo") as of a specified date. The ownership links are saved, and additional information is passed to *extract_owners_SReg.do* to extract links from SReg.
    - *JSCReg_clean.do*: This script removes some owners found via *extract_owners_JSCReg.do* that we deem to be irrelevant (see the paper for details).
    - *extract_owners_SReg.do*: This script extracts all ownership links from SReg for a specified set of firm identifiers in a specified range of dates.
    - *SReg_clean_corporate.do*: This script removes some corporate owners found via *extract_owners_SReg.do* that we deem to be irrelevant (see the paper for details).
    - *SReg_clean_individual.do*: This script removes some individual owners found via *extract_owners_SReg.do* that we deem to be irrelevant (see the paper for details).
    - *network_preparation_Delo_UP.do*: This script cleans up the ownership links for network analysis: removes duplicate owners, creates unified owner identifiers, establishes unified identifiers for important individual owners (oligarchs), adds information on oligarchs and political connections.

- *owner_search_full_sample_JSCReg_only.do*: This script extracts all ownership links from JSCReg for a specified date (analogous to *extract_owners_JSCReg.do* but for all firms in JSCReg) and cleans up these ownership links for the subsequent network analysis.

**Step 2. Reconstructing ownership networks**

This set of scripts takes the data obtained on Step 1 and creates network graphs for oligarch-controlled firms, which are used to trace ownership back to oligarchs (or to establish the absence of links to oligarchs) or to foreign firms. Based on this network analysis, the scripts generate three measures of defensive ownership: no oligarch in chain; foreign in chain; offshore in chain. At this step, we also generate plots based on network graphs, provided in the paper.

Step 2 is performed by executing the R script ***network_main_script.R***. **Note: the best way to do this is to open the script in RStudio from the folder without any active instances of R or RStudio running in the background. That is, any such instances need to be closed prior to opening the script (otherwise, the code may be executed in the wrong directory).**

This script runs all the other scripts that construct the ownership network for each of the dates of interest (in 2002, 2004, and 2006) and generate the defensive ownership measures, both for the main set of oligarch-controlled firms and for all firms in JSCReg; it also produces Figure 1. The script creates a series of data files saved in the folder *Network Data*:

- "DUP_hiding_*.dta" contain measures of defensive ownership for the main set of oligarch-controlled firms (for corresponding dates in 2002, 2004, and 2006)
- "DUP_foreign_owners_*.dta" contain auxiliary information for these firms (lists of foreign entities found in ownership chains of oligarch-controlled firms)
- "JSCReg_hiding_*.dta" contain measures of defensive ownership data for all the firms in JSCReg (for corresponding dates in 2002, 2004, and 2006)
- "JSCReg_foreign_owners_*.dta" contain auxiliary information for these firms (lists of foreign entities found in firms' ownership chains).

The following scripts are run automatically when ***network_main_script.R*** is executed (**they are not to be executed manually or separately from the main script**):

- *create_network_Delo_UP.R*: This script loads a specified list of ownership links (from files "all_entities_fin_*.dta") and creates a network graph based on these links for the main list of oligarch-controlled firms. The script also performs some additional operations such as cleaning owner names and adding labels for selected network nodes.
- *create_network_full_JSCReg.R*: Same as *create_network_Delo_UP.R*, but for the full set of firms in JSCReg.
- *defensive_ownership_measures_Delo_UP.R*: This script generates measures of defensive ownership by tracing ownership links through the network graph.

- *defensive_ownership_measures_Delo_UP_alt_oligarch.R*: Same as the script *defensive_ownership_measures_Delo_UP.R*, but with an alternative specification of oligarchs within oligarch groups (Table A5 in the paper; see the paper for details).
- *defensive_ownership_measures_Delo_UP_JSCReg_only.R*: Same as the script *defensive_ownership_measures_Delo_UP.R*, but using only JSCReg data.
- *defensive_ownership_measures_full_JSCReg.R*: Same as the script *defensive_ownership_measures_Delo_UP.R*, but for the full set of firms in JSCReg. **Note: executing this script could take several hours.**
- *figure_1.R*: The script to generate example plots of oligarch ownership networks, shown in Figure 1 in the paper (saved in the subfolder "Plots").

In addition, the script ***figures_tables.R*** needs to be executed. This script runs the following three scripts (they can also be run separately if needed):

- *figures_2_3.R*: The script to generate maps in the paper (Figures 2 and 3).
- *figure_4*.R: The script to produce plots of regression residuals (Figure 4).
- *tables_5_A1_A2.R*: The script to generate summary statistics on defensive ownership for Tables 5, A1, and A2 in the paper.

**3. Constructing samples with firm data for the analysis**

All scripts for creating the samples (step 3) and running the estimation (step 4) are included in the master script ***analysis.do***. **They are not to be run separately**; that is, to perform both step 3 and step 4, only ***analysis.do*** needs to be executed.

The following Stata do-files are used to construct the samples for the analysis (using network data generated in Step 2 above and stored in folder *Network data*, as well as economic data stored in the folder *Prepared data*):

- *baseline0406.do:* The script to generate the baseline sample. This data is further used to produce results in Tables 2, 3, 4, 5 (Columns 1 and 2), Table 6 (baseline), Table 7 (Columns 1 and 3), Table A3, Table A4, Table A6, A7, and A8.
- *baseline0204.do* and *JSCReg0204.do:* The scripts to generate the baseline and JSCReg samples used in placebo regressions in Table 8.
- *basline0406altown.do:* The script to generate the sample with an alternative definition of who is an oligarch. This sample is used in Table A5.
- *JSCReg0406.do:* The script to generate the JSCReg samples used in Table 5 (Columns 3 and 4), Table 6 (JSCReg), Table 7 (Columns 2 and 4),
- *allJSCReg0406.do:* The script to generate the sample of all JSCReg firms. This sample is used in Tables A9 and A10.

All generated data files are saved in the folder *Regressions data.*

## 4. Estimation results

Before performing Step 4, please install the following commands in Stata:
*Outreg2*
*Ivreg2*
*Ranktest*

There is a series of scripts to generate Tables 1 through 8 and Tables A2 through A10: Table1.do, Table2_3.do, etc. As with step 3, these scripts are included in the script ***analysis.do*** and **are not to be run separately**. The generated Excel tables are placed in *Results*, with names corresponding to table numbers in the paper. Note that there are Excel files with the names ending with "r" such as Table 4r, TableA4r, TableA5r, TableA6r, and TableA8r. These tables report reduced form coefficients from the reduced form equations in Tables 4, A4, A5, A6, and A8, respectively. *Table7_orange_base.do* creates an additional version of Table 7.