Readme

This folder contains replication files for “Patents, data exclusivity and the development of new drugs” published in The Review of Economics and Statistics

Authors:

Fabian Gaessler, MPI Munich, fabian.gaessler@ip.mpg.de

Stefan Wagner, ESMT Berlin, stefan.wagner@esmt.org.

Date:

7 September 2020

Summary:

This archive contains all do files to replicate the figures and tables in the paper as well as the web appendix.

Parts of the used data have been obtained under licenses that prevent publication. This applies to

1. Cortellis database from Clarivate Analytics (see below)
2. PATSTAT (Spring 2017 version) from the European Patent office (see below)

We gladly provide all information of the source of the data so that researchers know how to obtain access to this proprietary data.

Moreover, these data will be deposited at the Max Planck Institute for Innovation and Competition (MPI). Local data access is open to external researchers. For more information, see: http://www.ip.mpg.de/en/research/innovation-and-entrepreneurship-research/data-access.html

Version and software:

The code was tested on Stata/SE 15.1 for Windows (64-bit).

The time to run all do-files is about 8 hours on an Intel Core i7-6600U CPU @ 2.60GHz with 20 GB RAM.

Necessary Stata packages (latest version as of August 2020) that must be installed by the user:

binscatter, estout, esttab, ivreg2, ppmlhdfe, reghdfe

Instructions and input data files:

All (license-free) data and all do-files are included in the main directory and its subfolders.

Executing the master do-file “run_all.do” calls additional do-files in lower folders that reproduce all tables and figure published in the paper and its web appendix without further intervention.

Here we provide a summary of the do-files called within “run_all.do”. Further and more detailed comments are in included in that do-files directly.

*** Data preparation (main structure)

1) Preparation of Cortellis data:

- Drug characteristics: "$path\Cortellis_2018\01_0_gen_drug_details.do"
    - draws on 15 files obtained from Cortellis ("$path\cortellis_2018\cortellis drug_details\raw\Drugs_Results (0).xls" to "$path\cortellis_2018\cortellis drug_details\raw\Drugs_Results (14).xls") exported from Cortellis website (in April 2018) (not included but accessible at MPI)
- Drug history: "$path\Cortellis_2018\02_0_gen_drug_history.do"
    - draws on 71 dataset obtained from Cortellis ("$path\cortellis_2018\cortellis drug_history\raw\Drugs_Results (0).xls" to "$path\cortellis_2018\cortellis drug_history\raw\Drugs_Results (70).xls”) exported from Cortellis website (in April 2018) (not included but accessible at MPI)
- Drug indications: "$path\Cortellis_2018\03_0_gen_indications.do"
    - draws on indications data ("$path\cortellis_2018\cortellis indications\Hermosilla \CMS32_DESC_SHORT_DX.txt", "$path\cortellis_2018\cortellis indications\Hermosilla\indications_categorized.txt", "$path\cortellis_2018\ cortellis indications\Hermosilla\CortellisIndicationsICD9Crosswalk _ProfessionallyCoded_Nov2014.txt") (not included but accessible at MPI or available from their creator Manuel Hermosilla, mh@jhu.edu) and icd9 data ("$path\cortellis_2018\cortellis indications\Krieger\indications _icd9_coded_extended.dta") (not included but accessible at MPI or available from their creator Joshua Krieger, jkrieger@hbs.edu)
- Company characteristics: "$path\Cortellis_2018\04_0_gen_companies.do"
    - draws on 33 datasets obtained from Cortellis ("$path\cortellis_2018\cortellis companies\raw\Companies_Results (0).xls" to "$path\cortellis_2018\cortellis companies\raw\Companies_Results (32).xls", exported from Cortellis website (in April 2018) (not included but accessible at MPI) and a complete geocoding of company addresses contained in Cortellis ("$path\cortellis_2018\cortellis companies\geocoded\companies_addresses.csv") (included)
- Patent information: "$path\Cortellis_2018\05_0_gen_patent_drugs.do"
    - draws on 36 datasets obtained from Cortellis ("$path\cortellis_2018\cortellis patent_drugs\raw\Patents_Results (0).xls" to "$path\cortellis_2018\cortellis patent_drugs\raw\Patents_Results (35).xls”), exported from Cortellis website (in April 2018) (not included but accessible at MPI)
- Trials information: "$path\Cortellis_2018\06_0_gen_clinical_trials.do"
    - draws on 35 datasets obtained from Cortellis ("$path\cortellis_2018\cortellis patent_drugs\raw\Clinical_Trials_Results (0).xls" to "$path\cortellis_2018\cortellis patent_drugs\raw\Clinical_Trials_Results (34).xls”), exported from Cortellis website (in April 2018) (not included but accessible at MPI)

2) Preparation of PATSTAT data

- Country information: "$path\patstat\00_0_gen_countries.do"
    - draws on country-level data from PATSTAT ("$path\patstat\raw\tls801_part01.txt") (not included but accessible at MPI)
- Identifying drug patents: "$path\patstat\01_0_gen_cortellis_patents_link.do"
    - draws on a manually created dataset that matches Cortellis patent numbers with PATSTAT ("$path\patstat\raw\cortellis_appln_id.csv") (included) and patent variables extracted from PATSTAT ("$path\patstat\raw\PATSTAT_basic_042017.dta") (not included but accessible at MPI)

*** Analysis

3) Creation of final dataset and output as reported in the paper and web appendix

- Drug development and oppositions: "$path\paper\01_0_cr_patents_drug_history.do"
    - draws on a pre-processed dataset on EPO post-grant opposition outcomes and examiner participation based on PATSTAT and EPO Register data ("$path\patstat\oppositions\analysis_opposition_final.dta") and a pre-processed dataset on EPO post-grant opponents based on EPO Register data ("$path\patstat\oppositions\opponents_n.dta") (not included but accessible at MPI)
- Descriptives: "$path\paper\02_0_analysis_descriptives.do"
    - generates descriptive statistics as reported in paper and web appendix
- Regressions: "$path\paper\03_0_analysis_regressions.do"
    - generates regression results reported in the paper and the web appendix. One robustness check draws on a pre-processed dataset on claim reduction due to opposition from PATSTAT ("$path\patstat\oppositions\B2_classified_applnid.dta") (not included but accessible at MPI)

#Data sources

If not stated otherwise, all data have been obtained from the following sources:

Cortellis – information on pharmaceutical development histories and information on the associated patents (https://clarivate.com/cortellis/)

PATSTAT – patent data (https://www.epo.org/searching-for-patents/business/patstat.html)