This is the README file for the paper

Collusion through Joint R&D: An Empirical Assessment

REStat Volume 96, Issue 2 [May 2014] MS #14767

The analysis presented in the paper is done at two levels of aggregation: 1) the firm-level and the 2) firm-segment-level. For both of them we generated a corresponding estimation sample starting from the Compustat North America Industrials database and the Compustat Segment database, respectively. In both cases we matched this data with the NCRA-RJV database provided by Nicholas Vonortas, which holds information on RJVs and its participants under the National Cooperative Research Act (1985–1999), and the NBER U.S. Patent Citations Data File.

Since Compustat is a licensed database, we are not allowed to share the data. This data, however, can be easily accessed by acquiring a subscription to Compustat. We used the version of the Compustat North America Database from September 2005, while we accessed the COMPUSTAT Segment database on December 2011.

We obtained the NCRA-RJV database from Nicolas Vonortas under a non-disclosure clause. We therefore cannot provide the original data, which can be directly requested to Prof. Vonortas. However, we provide the key RJV network variables constructed based on the original data and used for the regression.

In this replication file, we provide three directories

1. The directory “\REStat\_MS14767\_Vol96(2)\Data preparation Compustat” contains all do files we used to generate the final estimation sample at the firm-level as well as the corresponding log files. We do not provide the data to run these do files as they are mostly proprietary. we provide a master do file 0\_raw\_to\_final.do which runs the 12 do files necessary to generate the final estimation sample. This do files also contains short description of the content of each of the 12 do files. These steps are the following:

1. Preparing the COMPUSTAT North America database
2. Cleaning the data so to only focus on American Firms
3. Generating the key variables base on Compustat
4. Create RJV panel from the original RJV-level data
5. Generating the RJV panel only based on American firms
6. Creating the RJV Participation variables
7. Creating the variables measuring links among firms
8. Creating of link variables measuring links among firms based on Market Shares
9. Managing patent data
10. Creating the relevant patent variables
11. Generating patent stock variables
12. Creating of final sample

The output of these steps is the Stata data file "RESTATestimation\_sample.dta".

2. The directory “\REStat\_MS14767\_Vol96(2)\Data preparation Compustat Segment” contains all do files we used to generate the final estimation sample at the firm-segment-level as well as the corresponding log files. We do not provide the data to run these do files as they are mostly proprietary. We provide a master do file 0\_raw\_to\_final.do which runs the 12 do files necessary to generate the final estimation sample for the segment database. This do files also contains short description of the content of each of the 12 do files. These steps are the following:

1. Preparing the COMPUSTAT Segment database
2. Cleaning the data so to only focus on American Firms
3. Generating the key variables base on Compustat
4. Create RJV panel from the original RJV-level data
5. Generating the RJV panel only based on American firms
6. Creating the RJV Participation variables
7. Creating the variables measuring links among firms
8. Creating of link variables measuring links among firms based on Market Shares
9. Managing patent data
10. Creating the relevant patent variables
11. Generating patent stock variables
12. Creating of final sample

The output of these steps is the Stata data file "RESTAT\_estimation\_sample\_segment.dta".

1. The directory “\REStat\_MS14767\_Vol96(2)\Estimation” contains the two estimation samples (.dta files) together with the corresponding do files to replicate all the tables and figures presented in the paper. The log files as well as tables are also reported.