Analyst Coverage Overlaps and Interfirm Information Spillovers

Tim Martens
Christoph J. Sextroh

COMPLIANCE WITH DATA POLICY FOR THE JOURNAL OF ACCOUNTING RESEARCH

May 2021

1. *A description of which author(s) handled the data and conducted the analyses.*

   All authors had access to all raw data throughout the entire project. Data management and analyses were performed by Tim Martens as described in more detail below (see #2 in particular).

2. *A detailed description of how the raw data were obtained or generated, including data sources, the specific date(s) on which data were downloaded or obtained, and the instrument used to generate the data (e.g., for surveys or experiments). We recommend that more than one author is able to vouch for the stated source of the raw data.*

   a) Patent and citation data
   The analysis is based on patent and citation data of U.S. firms. Tim Martens retrieved this data on January 12, 2018, from the Kogan, Papanikolaou, Seru, and Stoffman (2017) patent database. The database is available at: *https://iu.app.box.com/v/patents*

   b) Analyst data
   The analysis is based on IBES analyst data. Tim Martens retrieved the following data from the I/B/E/S database via WRDS:
   - Detail History Detail File with Actuals (March 15, 2018)
   - Recommendations Detail (July, 06, 2018)
   - IBES CRSP Link (Beta) (May, 16, 2018)
   - Summary History Summary Statistics (February 26, 2020)

   c) Firm data
   The analysis is based on CRSP and CRSP/Compustat Merged firm data. Tim Martens has retrieved the following data via WRDS:
   - CRSP/Compustat Merged Fundamentals Annual (March, 18, 2018)
   - CRSP/Compustat Merged Database Linking Table (February, 20, 2018)
   - CRSP/Compustat Merged Database Historial Segments (June, 13, 2018)
   - CRSP Stock Header Information (March, 07, 2018)
   - CRSP Daily Stock (February 27, 2020)

   d) Hoberg-Phillips similarity data
   The analysis includes the Hoberg-Phillips similarity measure. Tim Martens retrieved the data on July 1, 2018 from the 10-K Text-based Network Industry Classifications (TNIC) database at the SIC 2 level. The database is available at: *http://hobergphillips.tuck.dartmouth.edu/industryclass.htm*

e) SIC2 data
The analysis includes the labels of the individual SIC2 categories. Tim Martens retrieved this data on March 26, 2019 from the NY State University SIC2 database. The database is available at: *https://mckimmoncenter.ncsu.edu/2digitsiccodes/*

f) Brokerage closure/merger data
The analysis includes the brokerage closure/merger data available in Appendix A from Kelly, B., and A. Ljungqvist, 2012. Testing asymmetric-information asset pricing models. Review of Financial Studies 25, 1366–1413.

g) Common investor data
The analysis includes the Gilje et al. (2020) common ownership data. Tim Martens retrieved the data on March 30, 2020 via WRDS.

h) Common inventors data
The analysis includes the Patentsview inventor data. Tim Martens retrieved the data on March 6, 2019 at https://www.patentsview.org/

i) Common alliance data
The analysis includes SDC Platinum alliance data. Tim Martens has received the data from Hans Frankort (Cass Business School) on March 12, 2020 who retrieved the data at Cass Business School.

j) Examiner and applicant citation data
The analysis includes Patentsview examiner and applicant citation data. Tim Martens retrieved the data on March 6, 2019 at https://www.patentsview.org/

k) Institutional investor data
The analysis includes Thomson Reuters institutional 13F holdings stock ownership summary data. Tim Martens retrieved the data on February 24, 2020 via WRDS.

3. *If the data are obtained from an organization on a proprietary basis, the authors should privately provide the editors with contact information for a representative of the organization who can confirm data were obtained by the authors. The editors would not make this information publicly available. The authors should also provide information to the editors about the data sharing agreement with the organization (e.g., non-disclosure agreements, any restrictions imposed by the organization on the authors, such as restrictions to publish certain results).*

   All the data used in this study is either available from non-commercial databases or from commercial data providers.

4. *A complete description of the steps necessary to collect and process the data used in the final analyses reported in the paper. For experimental and survey papers, we require information about the instructions and instruments used to generate the data, subject eligibility and/or selection, as well as any exclusion criteria. The full set of instructions and instruments can be provided in the online appendix.*

We describe the data used in the main analysis in section 3 of the paper and define all variables in appendix A of the paper. The steps necessary to process the data are described in the computer programs referred to in part 5 of this Data Description Sheet.

5. *The computer programs or code used to convert the raw data into the final dataset used in the analysis plus a brief description that enables other researchers to use this program. The purpose of this requirement is to facilitate replication and to help other researchers understand in detail how the raw data were processed, the final sample was formed, variables were defined, outliers were treated, etc. This code or programming is in most circumstances not proprietary. However, we recognize that some parts of the code or data generation process may be proprietary, including from the authors' perspective. Therefore, instead of the code or program, researchers can provide a detailed step-by-step description of the code or the relevant parts of the code such that it enables other researchers to arrive at the same final dataset used in the analysis. In such cases, the authors should inform the editors upon initial submission, so that the editors can consider an exemption from the code sharing requirement. Whenever feasible, authors should also provide the identifiers (e.g., CIK, CUSIP) for their final sample. Authors should consult our FAQ Sheet on the JAR website for further details.*

   We use R to convert the raw data and perform all analyses. We use Stata to calculate the Oster test. Computer programs that enable other researchers to replicate the data preparation and arrive at the final dataset are provided in the file *MS JAR 2021 Code Package.pdf*. Identifiers (PERMNOs) and the final sample composition of annual firm pair observations are provided in the file *identifiers.csv*.

6. *An assurance that the data and programs will be maintained by at least one author (usually the corresponding author) for at least six years, consistent with National Science Foundation guidelines.*

   Both authors will maintain all data and programs for at least six years, consistent with National Science Foundation guidelines.