

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

All authors handled the data and conducted the analyses.

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.

We obtained our raw data from a variety of sources, outlined below. All data are publicly available.

- *Patent application data, including the date of the application and application number, downloaded from the USPTO's PatentsView database on 11/13/2023 (pg_published_application.tsv).*
- *Crosswalk between patent applications and granted patents downloaded on 7/31/2023 (pg_granted_pgpubs_crosswalk.tsv).*
- *A list of applicants and inventors listed on each application downloaded on 11/14/2023 (pg_inventor_not_disambiguated.tsv and pg_applicant_not_disambiguated.tsv).*
- *Applicant's locations downloaded on 11/14/2023 (pg_location_not_disambiguated.tsv)*
- *A link between firm names, GVKEYs, and PERMNOs (gvkey permno link.dta) was downloaded from the Compustat/CRSP databases in WRDS on 8/8/2023.*
- *Application_data.dta containing information on the applications' class, subclass, and examiner name downloaded from <https://www.uspto.gov/ip-policy/economic-research/research-datasets/patent-examination-research-dataset-public-pair> on 9/13/2023.*
- *Citations for granted patents were downloaded from the USPTO's PatentsView database on 3/2/2024 (g_us_patent_citation.tsv).*
- *A link between granted patents and firms was downloaded on 8/11/2023 from Kogan, L., Papanikolaou, D., Seru, A. and Stoffman, N.'s data repository, which can be found at (<https://github.com/KPSS2017/Technological-Innovation-Resource-Allocation-and-Growth-Extended-Data>). We also use the file to measure the market value of patents.*
- *A crosswalk between zip code and congressional district downloaded on 1/3/2021 (<https://github.com/OpenSourceActivismTech/us-zipcodes-congress>).*
- *Congressional members, their districts, their committee assignments, and their tenure were obtained from Charles Stewart III's website (http://web.mit.edu/17.251/www/data_page.html) on 1/3/2021.*
- *Crosswalk between zip code and city downloaded from <https://www.huduser.gov/apps/public/uspsscrowswalk/login?> on 8/3/2023.*
- *The text of firm's 10-Ks were downloaded using WRDS SEC filing search (<https://wrds-www.wharton.upenn.edu/text-search/wrds-sec-filing-search/>) on 2/22/2024.*
- *Supplemental firm-state operation distribution data from Garcia and Norli (2012) was downloaded on 5/3/2021 (<https://leeds-faculty.colorado.edu/garcia/page3.html>).*
- *Link between gvkey and cik (gvkey-cik) downloaded on 5/5/2021 from WRDS.*
- *The claim and description text for granted patents and applications was downloaded from the USPTO's PatentsView database on 8/7/2023.*

- *Patent litigation data downloaded from USPTO's Patent Litigation Docket Reports Data on 4/2/2024 (<https://bulkdata.uspto.gov/data/patent/litigation/2020/patents.dta.zip>).*
- *Transactions.dta containing communications between the applicants and the examiners was obtained from <https://www.uspto.gov/ip-policy/economic-research/research-datasets/patent-examination-research-dataset-public-pair> on 9/13/2023.*
- *Office Action data (office_actions.dta) downloaded 9/13/2023 from <https://www.uspto.gov/ip-policy/economic-research/research-datasets/office-action-research-dataset-patents>*
- *Firm characteristics were downloaded from the Compustat/CRSP databases in WRDS on 2/11/2021.*
- *Patent approval date downloaded from USPTO (https://s3.amazonaws.com/data.patentsview.org/download/g_patent.tsv.zip) on 4/25/2024.*

All authors are able to vouch for the sources of the raw data.

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). In particular, the authors should indicate if an organization or data provider imposes restrictions on the publication of the results, has not given the authors full control of the relevant data, requires that the results must be reviewed or approved prior to public release of the paper or publication.

We do not obtain any proprietary data. The databases described above are accessed and licensed through the institutions where we worked. These include New York University, University of Michigan, and Bocconi University.

To be provided in the paper or the online appendix:

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 provide a detailed description of the data collection and processing in Sections 3 and 4 of the manuscript.

To be provided *upon acceptance* of the paper and prior to publication:

5. After downloading or obtaining the raw data, all manipulations of the data should be done via computer programs. The code for these manipulations should be included in the code submitted upon acceptance (see below). No manipulations of raw data can take place manually or outside the computer code provided. If compliance with this requirement is not feasible, the authors need to

explain and disclose any manipulations of the raw data (e.g., manually created variables or file conversions). When feasible, we also encourage the authors to share the code that downloads the data.

All manipulations of the data are done via computer programs and are contained in the “CMZ_code.do” file, which also provides the code to download the data.

6. The computer programs (i.e., code) used to (1) convert the raw data into the final dataset used in the analysis, (2) to execute the statistical or econometric analysis, and (3) to generate the tables or to produce the output used in constructing tables of the manuscript. A brief description that enables other researchers to understand and run the code should be provided. 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, and which commands were used in the analysis, 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 disclosing the proprietary portion 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 results that the authors obtained and presented in their manuscript. In such cases, the authors should inform the editors upon initial submission, so that the editors can consider an exemption allowing the step-by-step description. Whenever feasible, authors are required to provide the identifiers (e.g., CIK, CUSIP) for their final sample. Authors should consult our FAQ Sheet on the JAR website for further details.

The code file, “CMZ_code.do,” includes all the steps for obtaining the raw data, sampling procedures, statistical analysis, and output generation. All the data in our study come from either data subscriptions from our schools or publicly-available sources.

Note that Stata packages change over time which can cause slight differences in results. For example, the REGHDFE package is regularly updated, and occasionally these updates can have material impacts such as: (1) changes to the syntax of a REGHDFE call; (2) changes to the calculations of clustered standard errors. Slight differences are to be expected, depending on vintage.

We also provide the final sample identifier file, “CMZ_identifiers.dta,” which contains unique company identifiers (CUSIP) used in the main analysis.

7. A comprehensive log file that shows the execution of the entire code. This log file should cover all the steps that convert the raw data into a final dataset and the execution of all statistical and econometric analyses presented in the tables of the manuscript. The portion of the log file that shows proprietary code or data may be masked. In this case, the reader should be referred to the step-by-step description provided as per the requirements in Item 6.

“CMZ_log_file.log” shows the execution of the entire code.

8. 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.

We as a team will maintain the data and programs according to the guideline.