

A1. Overview: Market for Innovation Sample Construction

We employ several publicly available data sources to construct the various patent sales-based variables used in our analyses (e.g., *No. Patent Sales*, *No. Patent Sales (private-and-private)*, *No. Patent Sales (private-and-public)*, etc.). This online appendix describes each data source and the procedures used to construct our variables. The patent data sources we use include the USPTO Patent Assignment Dataset, the PatentsView dataset, and patent data matched to CRSP identifiers from Kogan et al. (2017). We then combine these datasets to identify the types of organizations engaging in a patent transaction (i.e., public or private firms) as well as any transactions taking place as part of an M&A. Finally, we aggregate the number of patent sales to the technology-class year level as our main dependent variable of interest.

A2. USPTO Patent Assignment Dataset

a. Dataset Description

The United States Patent and Trademark Office (USPTO) maintains a dataset to record the assignment of both granted patents and patent applications in order to have a complete history of patent ownership transfers. Transacting parties notify the USPTO of their patent ownership transfers, and these parties have a strong incentive to do so given that only assignments recorded with the USPTO can be used as evidence of a purchase in courts (Serrano (2010)).

The USPTO maintains the USPTO Patent Assignment Dataset in both .dta and .csv formats that contain transactions recorded with the patent office from 1970 and updates the data every year or two. The data are currently hosted [here](#). We employ the 2019 version of the dataset in our paper. The USPTO Patent Assignment Dataset is comprised of six separate datasets that are linked with the unique identifier known as the Reel Frame ID Number (*rf_id*), identified at the transaction level. The “assignment”, “assignor”, “assignee”, and “documentid” datasets are all primary records maintained by the USPTO, while the “assignment_conveyance” and “documented_admin” datasets are constructed from the primary records by economists and statisticians, including the USPTO’s Chief Economist, Alan Marco (see Marco, Myers, Graham, D’Agostino and Apple (2015)). Figure OA1 presents a schema for the six databases associated with the USPTO Patent Assignment Dataset.

The “assignment” dataset includes information at the transaction level, including the correspondent that notified the USPTO of the transaction, the description of the type of assignment and the date the transaction was recorded by the patent office. The “assignor” and “assignee” datasets include the name of the assignee (the patent’s new owner) and the assignor (the patent’s old owner), the address of the assignee and the execution date between the parties. Each patent transaction (unique *rf_id*) can have multiple assignees and/or assignors. The “documentid” dataset includes the identifiers for the patent documents included in the transaction, such as the application number, pre-grant publication number, and granted patent number. Frequently, patents are transferred in a group, as opposed to only individual patents.

Economists and statisticians at the USPTO have constructed the “assignment_conveyance” dataset to identify specifically employer assignments. Patents can only be granted to individuals, though inventor employees often assign patent rights to their employer. Marco, Myers, Graham, D’Agostino and Apple (2015) describe the procedures used to identify employer assignments in Section 4.2.2 of their paper. They find that during their sample period from 1970 – 2014, approximately 82% of transactions and 52% of patents relate to employer assignments. We use this constructed field to remove employer assignments in our sample, as subsequently described. Finally, the “documented_admin” dataset was also constructed by Marco et al. (2015) to better match patent “documented” numbers in the raw data to other administrative datasets.

b. Procedures

To test our hypotheses, we seek to identify patent transfers that result from sales between firms and not internal transfers recorded with the USPTO due to employer assignments or administrative name changes. The following describes how we identify transfers between firms (i.e., sales) in the market for innovation following Marco et al. (2015).

1. We begin with the “assignor” dataset that includes the execution date for the transfer (exec_dt). We then remove duplicate execution dates at the transaction level (i.e., duplicates of exec_dt within an rf_id), retaining the latest execution date, so that there is a unique execution date for every transaction. We remove duplicates consistent with the caution in Marco et al. (2015) Section 5.1 regarding duplicate recording.
2. To build our primary sample of patent transfers, we start with the “documentid” dataset that is defined at the patent document-transaction level. We then merge in the execution date from step 1 and the conveyance type from the “assignment_conveyance” dataset using rf_id as the merge key.
3. In order to identify patent assignments that most accurately reflect the market for innovation, we first remove employer assignments. The identification of employer assignments is as described by Marco et al. (2015) Section 4.2.2. Table OA1 presents the frequency of the remaining non-employer assignment conveyance types.
4. We then follow Serrano (2010) and remove assignments due to administrative events, including corrections and name changes. Specifically, we retain conveyance types “assignment”, “govern” and “merger” and drop conveyance types “correct”, “namechg”, “missing”, “other”, “security” and “release”.

Marco et al. (2015) caution users of the USPTO Patent Assignment Dataset that there is a recording lag between when a sale takes place and when the USPTO is notified of the sale (see Section 5.2 of their paper). They find that the average recording lag is 182 days and that the shortest recording lags are for the “assignment” conveyance type, which make up 83% of our sample.¹ To be conservative, we end our sample in 2017 such that year t+1 includes 2018 patent sales. Because we employ the 2019 version of the USPTO Patent Assignment Dataset, we have sufficient hindsight to ensure our sample of sales is reasonably complete.

¹ Marco et al. (2015) find that “merger” conveyance types have the longest recording lags and that those lags appear to increase over time. Our results are robust to excluding M&A sales (see Table 12 Panel A) and thus we conclude that longer recording lags for M&A transactions in later periods do not significantly bias our inferences.

A3. PatentsView Dataset

The PatentsView database creates research-ready datasets with detailed documentation from the bulk USPTO patent data files. Support for PatentsView includes the Office of the Chief Economist at the USPTO. The related datasets provide information on granted patents since 1976 and for published patent applications since 2001 and is currently hosted [here](#). Importantly, this dataset includes the names of all organizations and individuals on the patent and identifies the entity type (e.g., US corporation, foreign corporation, federal government, US individual, etc.). We exploit the PatentsView data link between organization name and entity type to allow us to identify whether a party listed in the USPTO Patent Assignment Dataset is a company (as opposed to an individual or government entity, for example). We link the entity type per PatentsView to the entity identified in the USPTO Patent Assignment Dataset using the name matching procedures described in Section A5.

A4. Kogan et al. Dataset

As described in Kogan et al. (2017), the researchers create a dataset linking patent numbers to public firms, using CRSP permno as the firm identifier. Noah Stoffman includes a link to the data files on his website and the latest datasets are available [here](#). We employ the version of the data released in 2020 that includes patents granted through 2019. We use the Kogan et al. (2017) dataset to identify whether a corporation is a public firm as described in Section A5.

A5. Identifying Organization Types and Merging with Patent Assignment Data

To identify the types of entities transacting in the USPTO Patent Assignment Dataset, we create a dataset of unique assignor (assignee) names and use name matching procedures to match the assignor (assignee) name in the USPTO Patent Assignment Dataset to the entity names in the PatentsView data where entity type is defined. We further leverage the patent-level data in the Kogan et al. (2017) dataset to determine whether an entity identified as a company in the PatentsView data is a public company.

a. Procedures to Identify Organization Type

1. We create a dataset of unique assignor (assignee) names as listed in the USPTO Patent Assignment Dataset. We include only assignor (assignee) names that relate to patent transfers likely representing economic sales as described in Section A2-b.
2. We also create a dataset of unique entity names from PatentsView and identify if an entity is a private company, public company, or other.
 - i. To do so, we use the “rawassignee” table in PatentsView and retain unique organization names and entity types (variable names organization and type, respectively). PatentsView categorizes an entity as one of nine types; we retain organization names if PatentsView designates them as either a US Company or

Corporation (type=2) or a Foreign Company or Corporation (type=3). If the same organization name appears multiple times with different types, we retain the modal organization type such that each unique organization name has one type. Note that this step allows us to identify whether an entity is a company or not but does not identify whether a given company is public or private.

- ii. To identify whether the unique entity names in PatentsView are public firms, we first merge in all patent numbers associated with the PatentsView organization name. Next, we merge in the Kogan et al. (2017) dataset updated through 2019 on patent number. If a patent number is linked to a permno in the Kogan et al. (2017) dataset, we code the related entity as a public corporation and assume the remaining companies identified in step A5-2-i. are private corporations.
- iii. We identify NPEs using Stanford's NPE dataset found at <https://npe.law.stanford.edu>. We use the asserter categories obtained from the "cases" dataset where the asserter category takes on the value of one for "acquired patents". Miller 2018 in the Stanford Technology Law review states that "Category 1 includes any NPE primarily in the business of asserting patents it has acquired from other entities. We include in this category large patent aggregators such as Acacia and Intellectual Ventures."

b. Name Standardization Procedures

We then proceed to standardize the organization names in both the USPTO Patent Assignment Dataset and the PatentsView datasets in preparation for name matching. We use the applicable name standardization routines employed by Hall, Jaffe and Trajtenberg (2001) in creating the NBER patent dataset. The Stata .do files are currently hosted [here](#). The name standardization procedures include removing punctuation, standardizing common company keywords (e.g., all instances of either "laboratory" or "laboratories" replaced with "lab"), and removing designations of corporate form (e.g., "corp" and "inc").

The PatentsView dataset produced in step A5-a2 already retains only entities designated as a company or corporation. However, the USPTO Patent Assignment Dataset assignor and assignee files in step A5-a1 do not include a field for organization type. Thus, we also take steps to identify entities that are most likely companies as opposed to individuals prior to name matching. To do so, we use the Stata routine from the NBER project to flag assignor (assignee) names if designations of corporate form are present (prior to removing them as described above). We then retain only likely companies and corporations in the standardized assignment dataset before name matching.

c. Name Matching Procedures

We match the standardized assignment dataset to the standardized PatentsView dataset using the Stata routine "matchit". We use tokens as the string-matching method (i.e., match words as opposed to two-character bigrams). We apply the simple weighting transformation that weights tokens based on the frequency with which they occur in the data, thereby giving more weight to unique terms and less weight to commonly occurring words.

The Stata matching routine can identify more than one potential match for each unique assignor (assignee) name. Where multiple matches exist, we retain the match with the highest similarity score. Then for all remaining matches, we retain only those matches with a similarity score of 0.99 or greater.

A6. Number of Patent Sales

With an established link between organization names in the USPTO Patent Assignment Dataset and identification of whether the entity is a private firm, a public firm, or other entity, we aggregate patent sales to the technology class-year level. If there are multiple organization types represented among the assignors (assignees), we retain the modal organization type for the sale.

