Online Appendix for "M&A and Innovation: A New Classification of Patents"

Zhaoqi Cheng, Ginger Zhe Jin, Mario Leccese, Dokyun Lee, Liad Wagman

A Details on merging PatentsView, M&A data, and data of public firms

We merge the 2010-2019 US patent data with 2010-2021 M&A data from Standard & Poor's (S&P) Global Market Intelligence. We then summarize the connections among acquirers and targets based on the zones defined in Section 2 of the main text of the paper. This exercise showcases the potential utility of combining patent and business ownership data.

We merge the patent data from US PTO's PatentsView platform with 451 Research—a tech M&A database maintained and operated by S&P Global Market Intelligence—(henceforth, S&P), and the Center for Research in Security Prices (CRSP), which records historical descriptions and market data on companies listed in the AMEX, NYSE American, NYSE Arca, and NASDAQ exchanges at the security-level.

In the S&P database, each observation is an M&A transaction associated with a change in majority ownership. In total, it covers 46,216 M&A transactions involving 21,039 acquirers recorded between 2010 and 2021. All targets are firms belonging to the Information, Communication and Energy

Technology (ICET) space but acquirers can operate in any sector. Jin, Leccese and Wagman (2022*b*) show that while the Worldwide Mergers, Acquisitions, and Alliances Database offered by Refinitiv's SDC covers every sector of the economy, it is less comprehensive than the S&P database for majority control deals involving ICET targets.

S&P classifies the acquiring and acquired companies into a hierarchical technology taxonomy that has 4 levels, with level-1 being the broadest tech category (resembling an industry, such as "Application Software" and "Internet Content and Commerce," in some cases similar to 4-digit NAICS codes such as 5112 and 5191), and level-4 being the narrowest (resembling a market niche, such as "Benefit and Payroll Management" and "Video-On-Demand Servers").[1] To our best knowledge, the S&P taxonomy is based on business descriptions rather than patent activities. Since this taxonomy is only available for firms involved in ICET M&As between 2010 and 2021, we do not utilize this taxonomy when we merge the S&P data with PatentsView.

The S&P database additionally provides each tracked firm's headquarter location, founding date and number of employees, whether a firm is publicly traded, a business description, and the consummation date for each acquisition.[2]

One challenge in the data merge is that a firm may have multiple subsidiaries and some of them file patents as different assignees in the PatentsView data.

[1]All level-1 "parent" categories in the S&P technology taxonomy have level-2 "children" categories, but not all level-2 categories have further children levels. In total, there are about two dozen tech categories and two hundred verticals, yielding an average of approximately nine verticals per tech category. Each firm in the S&P database is assigned a primary category, representing the firm's core business, which includes a level-1, a level-2, and, if available, level-3 and level-4 classifications. Firms may also be assigned one or more secondary categories (organized analogously in the taxonomy).

[2]See Jin, Leccese and Wagman (2022*b*) for additional details on the S&P database.

To address this, we first associate each assignee with its legal entity (e.g., NewsCorp is part of Dow Jones, AWS is part of Amazon) using parent company and/or company alias available in S&P. Because different assignees may belong to the same parent, an entity can be bigger than an assignee. For entity name matching, we do not employ a fuzzy matching strategy as it generates many false positive results. To avoid interpreting the acquisition of a subsidiary of a larger entity as that of the entity itself, we attach the same identifier to any parent and its subsidiaries appearing as an acquirer in S&P, but treat each S&P target as a separate entity. Moreover, to keep track of changes in patent portfolios because of M&As, we code any patent that a target filed after the year of the M&A deal as part of its acquirer's patent portfolio after the deal. Another challenge is to reconcile different names of the legal entity across the S&P and PatentsView data. To address it, we keep only the essential part of an entity's name, by trimming non-alphanumeric characters and suffixes from the full name. For example, "Dow Jones & Co. Inc." would become simply "Dow Jones."

Our name match was done on the whole universe of the S&P data (2010-2021) and PatentsView data (1976 to present). Out of the 46,216 M&A deals in the S&P data, we were able to match 37.42% of the acquirers, 16.42% of the target, and 8.96% of both the acquirer and the target in the same deal. Because the same acquirer or target may appear in multiple deals, our match rate on unique acquirers and targets is lower (22.05% and 15.15%, respectively). The low match rate is not surprising, as not all firms have patent filings even if they operate in the broadly defined technology space.

In addition, we merge PatentsView with CRSP. This merge was done indirectly by merging our PatentsView data with the patent data that Kogan et al. (2017) constructed for publicly-traded firms.[3] This allows us to identify public firms that do not appear in the S&P data.

For our analysis, we further restrict the sample up to 2019 because very few patents filed in 2020-2022 appear in the PatentsView data due to regular time of patent examination. Since we are interested in comparing the patent stock of acquirers and targets, we drop all entities that appear to have a zero patent stock in all years between 2010 and 2019, where patent stock at year t is defined as the total number of granted patents that the entity had filed between $t-20$ and $t-1$. This entails dropping 3.24% of the entities.[4]. Note that this percentage is different from the 5.10% in the main text of the paper because there we focus on a smaller sample of entities, which do not include public firms that do not appear in the S&P data. We also drop few cases where the acquirer and the target carry the same entity name.

Table A.1 summarizes the dataset resulting from merging PatentsView with S&P and CRSP. To describe the merged data, this table categorizes entity firms into four groups. Acquirers–Never–Target, Target–Never–Acquirer and Acquirers–and–Targets are entities that appear in both PatentsView and S&P, with the difference that entities in the first (second) group appear in S&P only as acquirers (targets), while entities in the third group appear as both acquirers and targets. For example, a company that has made an acquisition after 2010, but has also sold a child company after 2010, would

[3] In particular, we use the online appendix of Kogan et al. (2017) at `http://mitsloan.mit.edu/shared/ods/documents?PublicationDocumentID=5894` and their Github data at `GitHubdata:https://github.com/KPSS2017/Technological-Innovation-Resource-Allocation-and-Growth-Extended-Data`.

[4] An entity that has zero patent stock throughout 2010-2019 may appear in the universe of the PatentsView data if it had filed a patent before 1990.

belong to the third group. All observations in these three groups have their acquirer and target names matched between S&P and PatentsView. The last group (Public-Non-S&P) is composed by publicly-traded entities which are somewhat active in the patent space but did not engage in any majority control acquisitions after 2010 as recorded by S&P. As a result, these entities appear in both PatentsView and CRSP but not in S&P.

Note that the above grouping is time-invariant: if an entity (or any of its subsidiaries) appears as an acquirer at one time but as a target at another time, it is tagged as Acquirers–and–Targets throughout the data when we generate summary statistics for Table A.1. This explains why many target entities belong to the Acquirers–and–Targets group rather than the Targets–Never–Acquirer group. In fact, manual checks show that many well-known acquirers such as Yahoo and Google fall in the Acquirers–and–Targets group because Google acquired Motorola Mobility in 2011 but sold it in pieces to two different acquirers in 2012, while Yahoo made a number of acquisitions before it was acquired by Verizon in 2017.

Table A.1 reports three panels. The top panel refers to entities that had a positive patent stock between 2010 and 2019. Since we define the patent stock at year t as the portfolio of patents filed between $t-20$ and $t-1$, the top panel includes entities that have developed any patent between 1990 and 2019. The middle and bottom panels are relative to entities that filed any patent between 2010 and 2019. By definition, this is a subset of entities in the first panel.

Table A.1 suggests a positive correlation between tech M&A and patent development activity across types of entities. In fact, most entities that

Table A.1: Summary statistics for merged data

	Acquirers Never–Target	Targets Never–Acquirer	Acquirers–and–Targets	Public Non–S&P	Total
Firm-level regardless of patent filings in 2010-2019					
# of unique entities	1,979	3,840	1,157	4,388	11,364
Share of public firms	0.39	0.12	0.44	1	0.54
Year founded	1987.12	1998.77	1990.31	1994.98	1994.17
Patent portfolio size in 2010	157.57	22.87	489.47	141.31	142.87
# of zones with any patent in 2010	3.01	2.59	3.00	3.62	3.15
Firm-level conditional on filing patents in 2010-2019					
# of unique entities	1,527	2,032	892	1,986	6,437
# of assignees per entity	1.08	1.00	1.27	1.26	1.12
Total # of patents filed (2010-2019)	269,099	25,504	346,443	234,242	875,288
Average # of patents filed (2010-2019)	176.23	12.55	388.39	117.95	135.98
# of zones with any patents files (2010-2019)	6.28	3.44	5.79	7.14	5.81
Average # of patent files per zone per year	12.52	2.47	27.77	8.24	12.04
Share of public firms	0.40	0.11	0.47	1	0.50
Year founded	1986.96	2000.37	1989.54	1994.95	1993.79
Patent portfolio size in 2010	203.68	39.26	633.13	276.57	243.92
# of zones with any patent in 2010	3.37	3.14	3.34	4.04	3.58
Average # of backward patent citations	20.15	14.88	9.73	24.86	16.89
Average # of backward scientific citations	7.10	8.32	5.60	14.36	8.39
Average # of forward citations	2.48	3.61	2.14	3.56	2.64
Average patent originality by CPC symbols	0.50	0.56	0.51	0.54	0.52
Zone-level conditional on filing patents in 2010-2019					
Median # of patents filed per zone per entity	37.00	8.20	17.81	45.32	23.06
Avg # of patents filed per zone per entity	181.83	30.72	277.27	149.27	152.71
Std. Dev. # of patents filed per zone per entity	578.14	92.14	1117.58	401.09	625.39
# of unique entities that filed patents per zone	56.05	39.93	30.95	81.08	52.00
Concentration of patent filing by entities	4705.21	6106.52	4874.97	4440.92	5031.91

Notes: # of patents per zone is conditional on having any patents in the zone. Acquirers and targets are identified by data that appear in both PatentsView and S&P; Public–Non–S&Ps are identified by data that appear in both PatentsView and CRSP but not in S&P.

developed patents between 1990 and 2009 and made a majority control acquisition of a technology target after 2010 were also active in the patent space between 2010 and 2019, while the same cannot be said about public companies that did not engage in S&P deals after 2010. In addition, by comparing Table 1 in the main text and Table A.1 in this appendix, one can see that roughly 29.92% of all the 2,142,570 patents filed between 2010 and 2019 were filed by entities active in technology acquisition after 2010 (Acquirers Never–Target, Targets Never–Acquirer, Acquirers–and–Targets), another 10.93% were filed by public companies that do not appear in any S&P–recorded M&A deals, and the remaining 59.15% were filed by entities not included in Table A.1, i.e. private or non-profit entities that never appeared in the S&P data. Regardless of whether we condition our sample on entities filings a patent between 2010 and 2019, we find that Target–never–Acquirers are the entities with the smallest and most concentrated patent portfolio as of 2010 (based on the number of zones with any patent in 2010), while Acquirers–and–Targets display the largest patent portfolio as of 2010 and Public–Non–S&Ps are the most expansive in zone coverage. Moreover, Target–never–Acquirers are on average the youngest and the least likely to be public.

The middle and bottom panels shed some lights on in–house R&D activity between 2010 and 2019 across groups of entities. In particular, Acquirers–and–Targets are the entities with the highest number of filings, with 346,443 patents in total, corresponding to an average of more than 388 patents per entity. However, public entities which did not engage in S&P-recorded M&A after 2010 reached more zones with their patents, and hence had the lowest concentration of patent filings. Consistently, for each zone, the number of entities filing patents and the median number of filings per zone were the largest for Public-Non-S&P entities. However, we find that the average number of patents filed per zone was the highest for the Acquirers–and–Targets group, and there is large heterogeneity in the dispersion of patent activity across entities within this group.

Furthermore, the average number of backward, backward scientific, and forward patent citations, together with patent originality by CPC symbols, characterize filings' quality across group of entities. The statistics reported in the middle panel of Table A.1 shows that Targets–Never-Acquirers tend to be more original and have a higher forward citations than Acquirers–Never–Targets, Acquirers–and–Targets, and Public-Non-S&P, but this last group is the one with the highest number of backward scientific and backward patent citations.

B Analysis of the overlap between acquirer and target

Since a thorough examination of a technology-driven M&A deal requires antitrust agencies to examine the correlation of innovative activities between the acquirer and the target, we take a first look at the extent to which the patent activities of the merging parties overlap at the time of acquisition. Unlike Table A.1, this exercise requires an entity's *time-variant* status as of the deal time. By definition, it is also conditional on acquirers and targets that have a non-zero patent portfolio at the consummation of the M&A deal, which further restricts the sample.

For this analysis, we focus on the sample of 2,955 S&P merger deals during 2010-2019 satisfying the following conditions: (i) the acquirer has a positive patent stock some time between 2010 and 2019; (ii) the target has a positive patent stock some time between 2010 and the year of its first acquisition after 2010; (iii) both acquirer and target names are matched between the S&P and PatentsView data. Condition (ii) is related to the fact that we reclassify any patent that a target filed after the year of the M&A deal as part of its acquirer's patent portfolio after the deal. This implies that we cannot include in this sample the subsequent M&A deals involving targets

Table B.1: Summary statistics as of the time of acquisition

	Data conditional on filings in 2010-2019		Unconditional sample	
# of M&A deals	1,589		2,955	
	Acquirers	Targets	Acquirers	Targets
# of unique entities	778	1,575	1,283	2,934
% of Acquirers–and–Targets	42.93	26.60	43.80	20.69
Share of public firms	0.62	0.21	0.57	0.18
Year founded	1981.12	1998.62	1982.69	1997.97
Patent stock per zone	32.97	1.49	28.42	0.82
Average number of zones with patents	3.77	2.97	3.55	2.65
Concentration of patent portfolio	4188.33	5656.57	4348.54	6120.38

that were acquired in a previous year. However, we are able to include the (few) cases in which the same target participated into multiple deals in the year of its first acquisition between 2010 and 2019.

Table B.1 presents summary statistics of acquirers and targets at the time of acquisition. Our analysis of acquirers and targets include 2,955 M&A deals recorded in S&P, with both acquirer and target names matched with PatentsView. However, not all of these parties filed patents during 2010-2019. When we condition on patent filing, the sample reduces to 1,589 deals. Either conditional or unconditional, Table B.1 shows that, at the time of acquisition, targets are younger, are less likely to be public, span a lower number of zones, have a smaller patent stock per zone, and thus hold a more concentrated patent portfolio in terms of zone coverage.

To characterize the zone overlap between acquirer and target, we represent each merging party as a 172×1 vector of patents across zones at the time of the M&A deal, and construct three measures of pairwise similarity between acquirer and target: (i) Cosine similarity, which, for any pair of vectors X and Y, can be computed as $\frac{X \cdot Y}{||X||||Y||}$. (ii) Jaccard similarity, which we define as the ratio of the number of zones in which both acquirer and target have 'enough' patents, divided by the number of different zones in which acquirer or target has 'enough' patents.[5] In other words, if we define A (T) as the set of zones in which the acquirer (target) has enough patents, then one can compute the Jaccard similarity measure as $\frac{|A \cap T|}{|A \cup T|}$. (iii) Overlap coefficient, defined as the ratio of the number of zones in which both acquirer and target have enough patents, divided by the number of different zones in which the smaller between acquirer and target has enough patents, i.e. $\frac{|A \cap T|}{\min(|A|,|T|)}$.

Figure B.1 shows the distribution of the similarity measures across M&A deals. The distributions of cosine and Jaccard similarities show that most deals involve acquirers and targets active in different TBZs—although both measures point to a non–trivial fraction of deals between entities with a similarity score close to one. The picture is somewhat different if one looks at the distribution of the overlap coefficient. In fact, the largest share of deals involves acquirers and targets with a score close to one. This suggests that often the target is only active in a subset of the TBZs in which the acquirer is active, or vice versa. To see this, consider a case in which a target is active only in TBZ #1, while the acquirer is active with the same intensity in TBZs #1, #2, #3, #4 and #5. In this case, the Jaccard similarity would equal 0.2, while the overlap coefficient would be 1. If instead the target was also active in TBZ #6, the Jaccard similarity would drop to 0.17 and the overlap

[5]To avoid counting zones which are not particularly relevant for the innovative activity of an entity, we define enough patents in a zone as the case in which the vector element relative to that zone is at least 0.1.

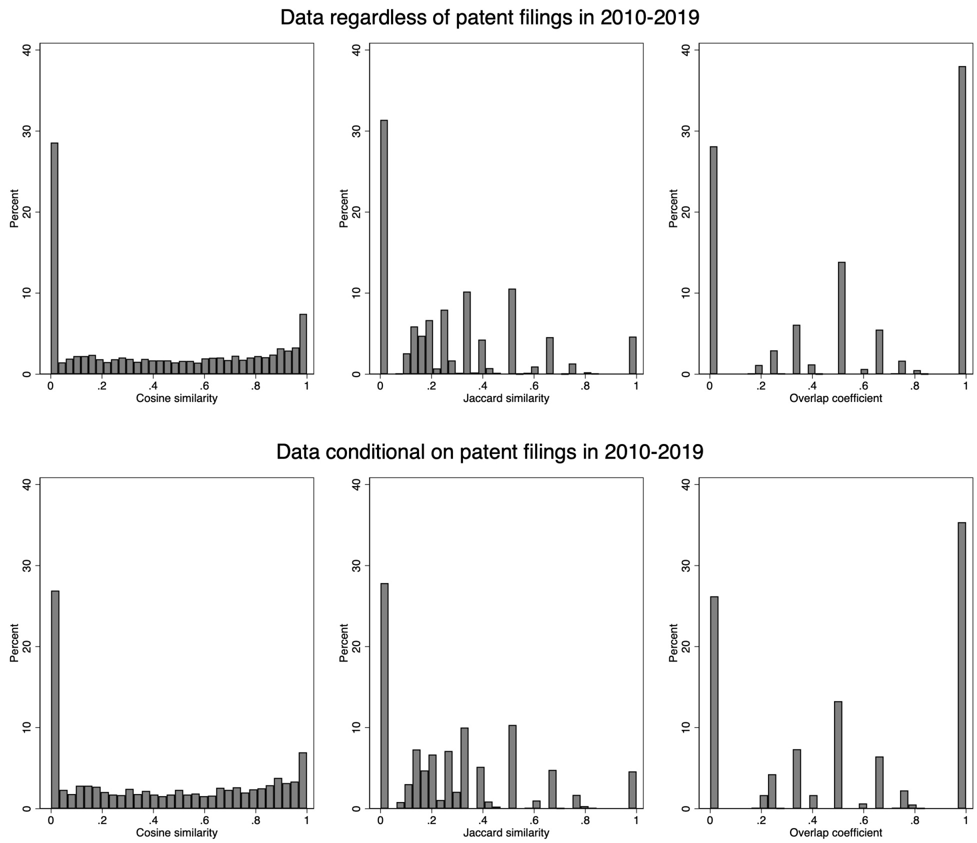

Figure B.1: Overlap in patent activity between acquirer and target at acquisition

coefficient would drop to 0.5.

Overall, Figure B.1 suggests that most deals involve either acquirers and targets active in different TBZs or targets that are active in a subset of TBZs in which the acquirer is active. These results are consistent with acquisitions being a way for firms to either strengthen their position in their core business areas, or to expand into new, potentially unrelated areas (Jin, Leccese and Wagman, 2022*a*). The non-trivial density of overlap coefficients between 0.2 and 0.8 also suggests that the overlap between the acquirer and the target is sometimes incomplete, as the target (acquirer) may operate in zones in which the other has little presence as of the time of the acquisition.

References

Jin, Ginger Zhe, Mario Leccese, and Liad Wagman. 2022*a*. “How do top acquirers compare in technology mergers? New evidence from an S&P taxonomy.” *International Journal of Industrial Organization*, 102891.

Jin, Ginger Zhe, Mario Leccese, and Liad Wagman. 2022*b*. “M&A and technological expansion.” *Available at SSRN 4009215.*

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. 2017. “Technological innovation, resource allocation, and growth.” *Quarterly Journal of Economics*, 132(2): 665–712.