

# Appendix A Model Appendix

## A.1 Setup

The firm's optimization problem is to choose the number of inventions,  $n$  to maximize profits from inventions:

$$\max_n \left\{ \pi = v \left( An - \frac{bn^2}{2} \right) - \frac{dn^2}{2} \right\}$$

where,  $v \in (0, 1)$  is the share of value that the firm  $i$  can appropriate from its inventions. We assume that the firm faces an increasing cost of commercialization function,  $\frac{dn^2}{2}$ . The profit of the firm not only depends on the value-captured and cost of commercialization, but also the quality of inventions. The quality of the  $n^{\text{th}}$  invention is given by  $A - bn$ ,  $A > 0, b > 0$ , so that overall value is simply  $Q(n; A) = An - \frac{bn^2}{2}$ . The first order and second order conditions for an interior optimum are:

$$\text{FOC: } v(A - bn) - dn = 0$$

$$\text{SOC: } -(vb + d) < 0$$

The optimal number of inventions is given by  $n^* = \frac{A}{\frac{d}{v} + b}$ . Note that  $\frac{\partial n}{\partial A} = \frac{1}{\frac{d}{v} + b} > 0$  and  $\frac{\partial n}{\partial v} = \frac{Ad}{(bv + d)^2} > 0$ . Also note that a decrease in costs (commercialization) has the same effect on  $n^*$  as an increase in the appropriability parameter  $v$ . That is, we can capture an increase in appropriability as the decrease in the ratio  $\frac{d}{v}$ , and  $\frac{\partial n}{\partial \frac{d}{v}} = -\frac{A}{(\frac{d}{v} + b)^2} < 0$

## A.2 Comparative statics

### A.2.1 Average quality

At  $n^*$ , average quality (AQ) is:  $\frac{An^* - \frac{bn^{*2}}{2}}{n^*} = A \frac{bv + 2d}{2(bv + d)}$ . Average quality increases with  $A$  and decreases with  $v$ .

$$\frac{\partial AQ}{\partial A} = \frac{bv + 2d}{2(bv + d)} > 0$$
$$\frac{\partial AQ}{\partial v} = -\frac{Abd}{2(bv + d)^2} < 0$$

### A.2.2 Average value

At  $n^*$ , average value (AV) is:  $\frac{v(An^* - \frac{bn^{*2}}{2}) - \frac{dn^2}{2}}{n^*} = \frac{vA}{2}$ . Average value increases with both  $A$  and  $v$ .

$$\frac{\partial AV}{\partial A} = \frac{v}{2} > 0$$

$$\frac{\partial AV}{\partial v} = \frac{A}{2} > 0$$

A fall in  $d$  has qualitatively the same effect as an increase in  $v$ . An increase in  $v$  or a decrease in  $d$  increases average value on one hand but decrease average quality on the other. Henceforth we shall focus on only on changes in  $A$  or  $v$ .

### A.3 Markets for technology (MFT)

Consider first the case of out-licensing. All potential licensees have an appropriability parameter  $\bar{v}$ . Clearly, firms with  $v > \bar{v}$  will not license because there are no gains from trade. Suppose firm  $i$  can capture a fraction  $\gamma, 0 < \gamma \leq 1$  of the total value generated. Let  $\tilde{v} = v + \lambda\gamma(\bar{v} - v), v < \bar{v}$ , &  $\tilde{v} = v, v > \bar{v}$ . The firm's optimization problem is

$$\max_n \left\{ \pi = \tilde{v} \left( A - \frac{bn^2}{2} - \frac{dn^2}{2} \right) \right\}$$

The optimal number of inventions is:  $n^{MFT} = \frac{A\tilde{v}}{d + b\tilde{v}}$ . Note that  $n^*$  increases with  $\lambda$ , so MFT increases investment in invention for firms that are likely to license.

#### A.3.1 Average quality

Average quality (AQ) is:  $A \frac{b\tilde{v} + 2d}{2(b\tilde{v} + d)}$ ,  $v < \bar{v}$  and is  $A \frac{bv + 2d}{2(bv + d)}$ ,  $v > \bar{v}$

$$\frac{\partial AQ}{\partial A} = \frac{b\tilde{v} + 2d}{2(b\tilde{v} + d)} \quad v < \bar{v}$$

$$= \frac{bv + 2d}{2(bv + d)} \quad v > \bar{v}$$

$$\frac{\partial AQ}{\partial v} = -\frac{Abd(1 - \gamma\lambda)}{2(b\tilde{v} + d)^2}, \quad v < \bar{v}$$

$$= -\frac{Abd}{2(bv + d)^2} \quad v > \bar{v}$$

Firms with  $v < \bar{v}$  will license their inventions, and firms, with  $v \geq \bar{v}$ , will choose to commercialize their inventions rather than licensing them out. For low  $v$  firms,  $\frac{\partial AQ}{\partial A}|_{n^{MFT}} < \frac{\partial AQ}{\partial A}|_{n^*}$  from the previous section. Similarly,  $\frac{\partial AQ}{\partial v}|_{n^{MFT}} < \frac{\partial AQ}{\partial v}|_{n^*}$  for low  $v$  firms. For firms with  $v \geq \bar{v}$ , there will be no change in  $\frac{\partial AQ}{\partial A}$  and  $\frac{\partial AQ}{\partial v}$ . Empirically, we should expect the relationship between average quality and firm size to weaken in an active MFT.

#### A.3.2 Average value

Average value (AV) is:  $\frac{\tilde{v}A}{2}$ , where  $\tilde{v} = v + \lambda\gamma(\bar{v} - v), v < \bar{v}$ , &  $\tilde{v} = v, v > \bar{v}$ .

$$\begin{aligned}
\frac{\partial AV}{\partial A} &= \frac{\tilde{v}}{2} > 0, & v < \bar{v} \\
&= \frac{v}{2}, & v > \bar{v} \\
\frac{\partial AV}{\partial v} &= \frac{A(1 - \lambda\gamma)}{2}, & v < \bar{v} \\
&= \frac{A}{2}, & v > \bar{v}
\end{aligned}$$

For low  $v$  firms,  $\frac{\partial AV}{\partial A}|_{n^{MFT}} > \frac{\partial AV}{\partial A}|_{n^*}$  from the previous section and  $\frac{\partial AV}{\partial v}|_{n^{MFT}} < \frac{\partial AV}{\partial v}|_{n^*}$ . For high  $v$  firms, with  $v \geq \bar{v}$ , there will be no change in  $\frac{\partial AV}{\partial A}$  and  $\frac{\partial AV}{\partial v}$  in an MFT. Empirically, we should expect the relationship between average value and firm size to weaken in an active MFT.

## A.4 In-licensing

Suppose a firm can in-license inventions with an average quality of  $\mathfrak{q}$ . It follows that it will not develop internal inventions with quality lower than  $\mathfrak{q}$ . Define  $n(\mathfrak{q}) = \frac{A - \mathfrak{q}}{b}$ . The firm will never develop more than  $n(\mathfrak{q})$  projects internally. By definition, the quality of the marginal invention produced is  $\mathfrak{q}$ .

The firm will in-license if the value of the  $n(\mathfrak{q})^{th}$  internal project exceeds its marginal cost of development – if  $v\mathfrak{q} \geq dn(\mathfrak{q})$ . This is algebraically equivalent to  $\mathfrak{q} \geq \frac{dA}{bv + d} = \frac{A}{1 + \frac{bv}{d}}$ . That is, firms with lower inventive capability  $A$ , lower development cost  $d$ , and higher appropriability  $v$  are more likely to in-license. When a firm in-licenses, the optimal number of projects,  $n^*$ , is given by  $n^* = \frac{v\mathfrak{q}}{d}$ .

When a firm in-licenses, the average quality of its *internally generated* inventions is  $A - n(\mathfrak{q})\frac{b}{2} = \frac{1}{2}(A + \mathfrak{q})$ . As long as the firm does not in-license, its average quality falls with size if size increases appropriability but once the number of inventions reaches  $n(\mathfrak{q})$ , average quality will no longer fall with size. The average value of its internally generated inventions is  $\frac{v}{2}(A + \mathfrak{q})$ . If the size premium is due to superior appropriability, average value increases with size, and will increase linearly once size increases such that the number of inventions reaches  $n(\mathfrak{q})$ .

### A.4.1 Average quality

For  $n^* < n(\mathfrak{q})$ , average quality (AQ) is  $A\frac{bv + 2d}{2(bv + d)}$ , and for  $n^* > n(\mathfrak{q})$ , it is  $\frac{A + \mathfrak{q}}{2}$ .

$$\begin{aligned}
\frac{\partial AQ}{\partial A} &= \frac{1}{2} \frac{bv + 2d}{(bv + d)} & n^* < n(\mathbf{q}) \\
&= \frac{1}{2} & n^* > n(\mathbf{q}) \\
\frac{\partial AQ}{\partial v} &= -A \frac{2bd}{(2bv + 2d)^2} & n^* < n(\mathbf{q}) \\
&= 0 & n^* > n(\mathbf{q})
\end{aligned}$$

Once the firm starts in-licensing, average quality will not fall with  $v$ . For firms that do not in-license, the change in average quality w.r.t  $v$  is as before. Averaging across firms, the relationship between  $v$  and average quality is smaller in absolute terms (i.e, is less negative). On the other hand, once the firm starts in-licensing, the average quality of the internally produced inventions increases at a constant rate with  $A$ , which is smaller than for firms that do not in-license. On average, the relationship between average quality and  $A$  is weaker.

#### A.4.2 Average value

For  $n^* < n(\mathbf{q})$ , average value (AV) of internally generated inventions is  $\frac{Av}{2}$ , for  $n^* > n(\mathbf{q})$ , it is  $v \frac{A + \bar{q}}{2} - \frac{d}{2} \frac{A - \bar{q}}{b}$ .

$$\begin{aligned}
\frac{\partial AV}{\partial A} &= \frac{v}{2} & n^* < n(\mathbf{q}) \\
&= \frac{v}{2} - \frac{d}{2b} & n^* > n(\mathbf{q}) \\
\frac{\partial AV}{\partial v} &= \frac{A}{2} & n^* < n(\mathbf{q}) \\
&= \frac{A + \bar{q}}{2} & n^* > n(\mathbf{q})
\end{aligned}$$

If  $n^* < n(\mathbf{q})$ , the firm will not in-license and average value increases with  $A$ . Once the firm starts in-licensing, average value increases with  $A$  but at a slower rate because  $\frac{d}{2b} > 0$ . On average, the possibility of in-licensing weakens the relationship between average value and size, when the size premium is due to inventive capability.

If  $n^* > n(\mathbf{q})$ , the firm will not in-license and average value increases with  $v$ . Once the firm starts in-licensing, average value increases with  $v$  as well, but at a higher rate because  $\bar{q} > 0$ . On average, the possibility of in-licensing strengthens the relationship between average value and size, when the size premium is due to appropriability.

## Appendix B Data Appendix

### B.1 Sample construction

We compile our dataset of 2,786 firms and information on their patents by merging several datasets. We begin with firms in Duke Innovation & SCientific Enterprise Research Network (DISCERN) (Arora, Belenzon and Sheer, 2021). DISCERN includes Standard & Poor’s CRSP/Compustat firms with at least one patent assigned to it and with positive R&D expenditures in at least one year between the period 1980 and 2015. DISCERN extends and improves the historical NBER patent dataset to dynamically match Compustat companies with patents, accounting for name changes and ownership changes. We restrict the firms in DISCERN to manufacturing firms, in the Standard Industrial Classification (SIC) code range of 2000 to 3999. We merge this dataset with Compustat to obtain firm accounting information such as sales, assets, and employees as well as industry codes (i.e., SIC, NAICS).

### B.2 Patent value and quality

For each patent, we add the measure of private value of invention from Kogan et al. (2017). The authors estimate the value of a patent using an event study that captures abnormal market returns in response to a patent issuance within a three-day window. While the original paper provides data until 2010, we supplement it with recent data up to 2015 from Stoffman, Woepffel and Yavuz (Forthcoming). We add forward citations for each patent as a measure of invention quality from PATSTAT, maintained by the European Patent Office. Forward citations are normalized by the average forward citations in a three-digit Cooperative Patent Classification (CPC) and grant year. We compile information on grant year and CPC from patentsview.org maintained by the USPTO.

To derive the average value of a firm’s inventions for a given year, we divide the total value of the firm’s patents applied for in that year by the total number of patents applied for in the same year. Similarly, to derive the average quality of a firm’s inventions for a given year, we divide the total number of citations received by the firm’s patents applied for in that year by the total number of patents that the firm applied for in the same year.

### B.3 Markets for technology and commercialization capabilities

The ability to sell or license patents enable firms to rely less on their own existing customer base to commercialize their inventions. We use patent assignment data from the USPTO (<https://www.uspto.gov/ip-policy/economic-research/research-datasets/patent-assignment-dataset>) to construct a measure of propensity to trade inventions. Using the USPTO’s patent assignment dataset, we count the number of patents traded within each CPC class for a given year. We then derive the firm-year level measure by computing, for each firm and year, the average number of patents traded across the CPC classes weighted by the share of each firm’s patents in those CPC classes. We use an indicator variable that takes the value of 1 for the top quartile of this measure and 0, otherwise to use for our analysis. In addition to

patent reassignment, we follow [Arora, Cohen and Walsh \(2016\)](#) and identify industries with a high potential for licensing.

We use a measure of use of capabilities to commercialize new inventions from the division of innovative labor (DOIL) survey from [Cohen, Lee and Walsh \(2021\)](#) to test whether it moderates the association of firm size with patent value and quality. Question 9.2 of the survey asks respondents if they used any new types of equipment or hired new employees to commercialize their new or significantly improved goods and services. We take the average of the responses to this yes/no question in a six digit NAICS and supplement it to our firm sample. We construct the variable of interest, ‘Commercialization Assets’ as an indicator variable that takes the value 1 if the firms are above the median of the distribution of the industry average of this measure and 0 otherwise.

## B.4 Diversification

### B.4.1 Herfindahl–Hirschman Index

We construct a measure of a firm’s diversification using the number of different three-digit CPCs that a firm patents in. Keeping firm sales fixed, the more diversified the firm’s patent portfolio is, the more limited will the firm be in each area of operation in terms of scale, and hence, we would expect the relationship of firm sales with patent value and quality from value-capture to be moderated. This measure is constructed by multiplying the number of CPC classes that a firm patents by  $(1-\text{HHI}_{CPC3})$ , where  $\text{HHI}_{CPC3}$  is the Herfindahl–Hirschman Index across the CPC classes that the firm patents in. by multiplying the number of CPC classes that a firm patents by  $(1-\text{HHI}_{CPC3})$ , where  $\text{HHI}_{CPC3}$  is the Herfindahl–Hirschman Index across the CPC classes that the firm patents in. We construct the variable of interest, “Diversification” as an indicator variable that takes the value 1 if the firms are above the 30<sup>th</sup> percentile of the measure of diversification this measure and 0 otherwise.

### B.4.2 Stirling Index

We construct this measure using equation 9 in [Stirling \(2007\)](#) reproduced here:

$$D = \sum_{ij(i \neq j)} d_{ij} \cdot p_i \cdot p_j$$

where the measure of diversification  $D$  is the sum of over all CPC class pairs of a firm over the product of the distance between each pair of CPC class,  $d_{ij}$ , and the share of firm’s patents in CPC classes  $i$  and  $j$ . The distance between each CPC class pair,  $d_{ij}$ , is calculated as a range between 0 and 2, where 2 is the distance between two completely distinct CPC classes, 1 is the distance between CPC classes whose CPC section is the same and 0 is for firms that patent only in one CPC class. The final firm-year measure is log-transformed.

## B.5 Instrumental variable

We use tariffs charged by US on imports from all other countries from the World Integrated Solution’s (WITS) UN TRAINS dataset to instrument for firm sales. We use the simple average of ad valorem tariff rates from all national tariff lines, within a four digit SIC code for the years 1989 to 2015 from this dataset. To account for variations in tariffs across different industries, we demean each tariff rate by the average tariff in an SIC across the sample period. This demeaned measure is weighted by each firm’s sales share in an SIC using primary SIC codes from Standard & Poor’s historical Compustat segment files. We clean the Compustat segments file by removing all geographical segments and keeping firm-year segments where the source date is equal to the data date.

To illustrate, Abbott laboratories, a manufacturing firm in SIC 3845 (Electromedical and Electrotherapeutic Apparatus), has sales shares in three different primary industries based on historical Compustat segments file. These are SIC 2834 (Pharmaceutical Preparations) with a 54% share, SIC 2835 (In Vitro and In Vivo Diagnostic Substances) with a 32% share, and SIC 3842 (Orthopedic, Prosthetic, and Surgical Appliances and Supplies) with 14% share of the total US 21,848 million dollars sales in the year 2013. Simple average of tariffs (deviations from mean tariff across all sample years) for these three SICs in 2013 are -.0475 for SIC 2834, -0.29 for SIC 2835, and -0.97 for SIC 3842. Thus, the instrument for sales (US 20,247 million dollars) in 2014 for Abbott Laboratories is  $(0.54 \cdot -0.86 + 0.32 \cdot -0.29 + 0.14 \cdot -0.97) / 3 = -0.231$ .

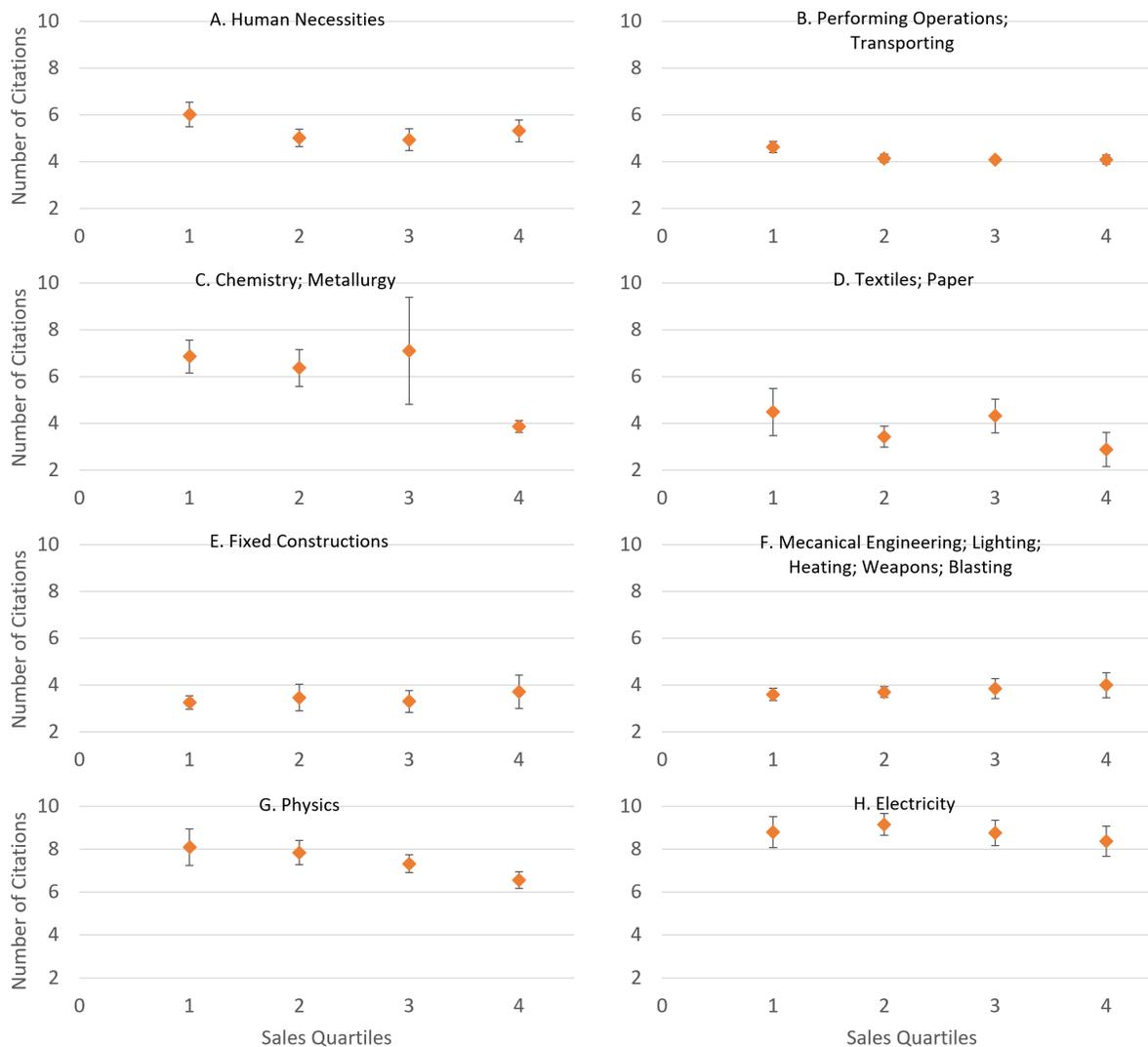
Column 2 of Appendix Table C.4 reports the first-stage results from the 2SLS analysis - i.e., the relationship between import tariffs and sales (firm size). Prior studies have shown that the relationship between sales and import tariffs is likely to be quite nuanced. While higher import tariffs can protect domestic firms from foreign competition, they can also raise production costs for domestic firms due to higher costs of intermediate inputs (Fan et al., 2021; Flaaen and Pierce, 2019; Handley, Kamal and Monarch, 2020). Furthermore, the studies point out that retaliatory tariffs can also impact the domestic firms’ competitiveness in foreign markets (Flaaen and Pierce, 2019). In a related stream of research examining the effect of the “China shock” following its accession to the WTO, studies have shown that the impact of increased import competition can have both “horizontal” and “vertical” components, having different effects on final goods and intermediate goods markets (Aghion et al., 2021; Autor, Dorn, Hanson, Pisano and Shu, 2020). They also point out that the effect may vary across high- and low-productivity firms (Aghion et al., 2021; Autor, Dorn, Hanson, Pisano and Shu, 2020). In our analysis, the first stage results show a negative relationship between import tariffs and firm sales, with an F-statistic of around 27.<sup>38</sup>

---

<sup>38</sup>The low p-value of the Wu-Hausman test suggests the presence of endogenous regressors in the OLS model.

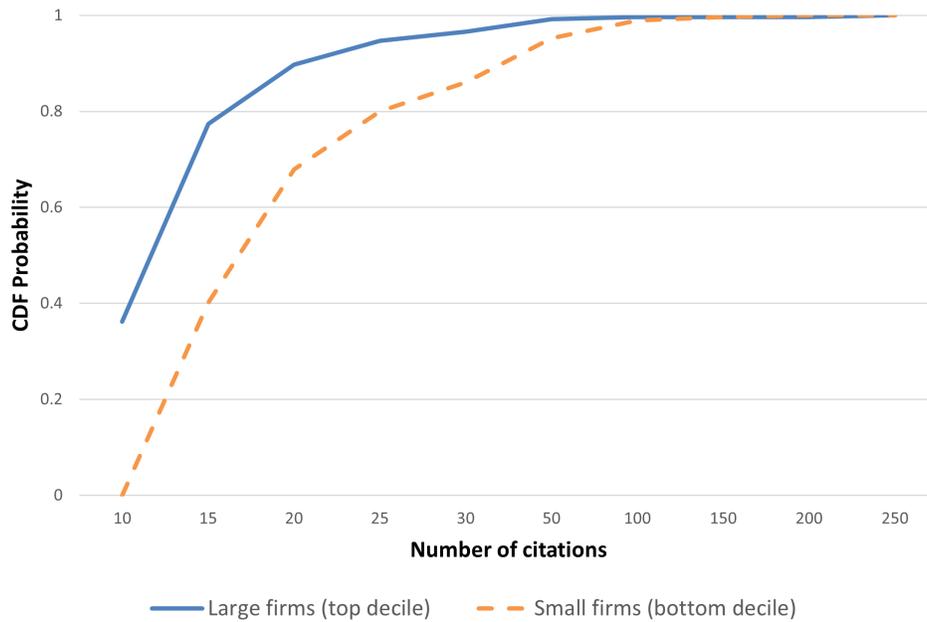
## Appendix C Additional Tables and Figures

Figure C.1: Firm Size and Inventive Capability - Technology Areas



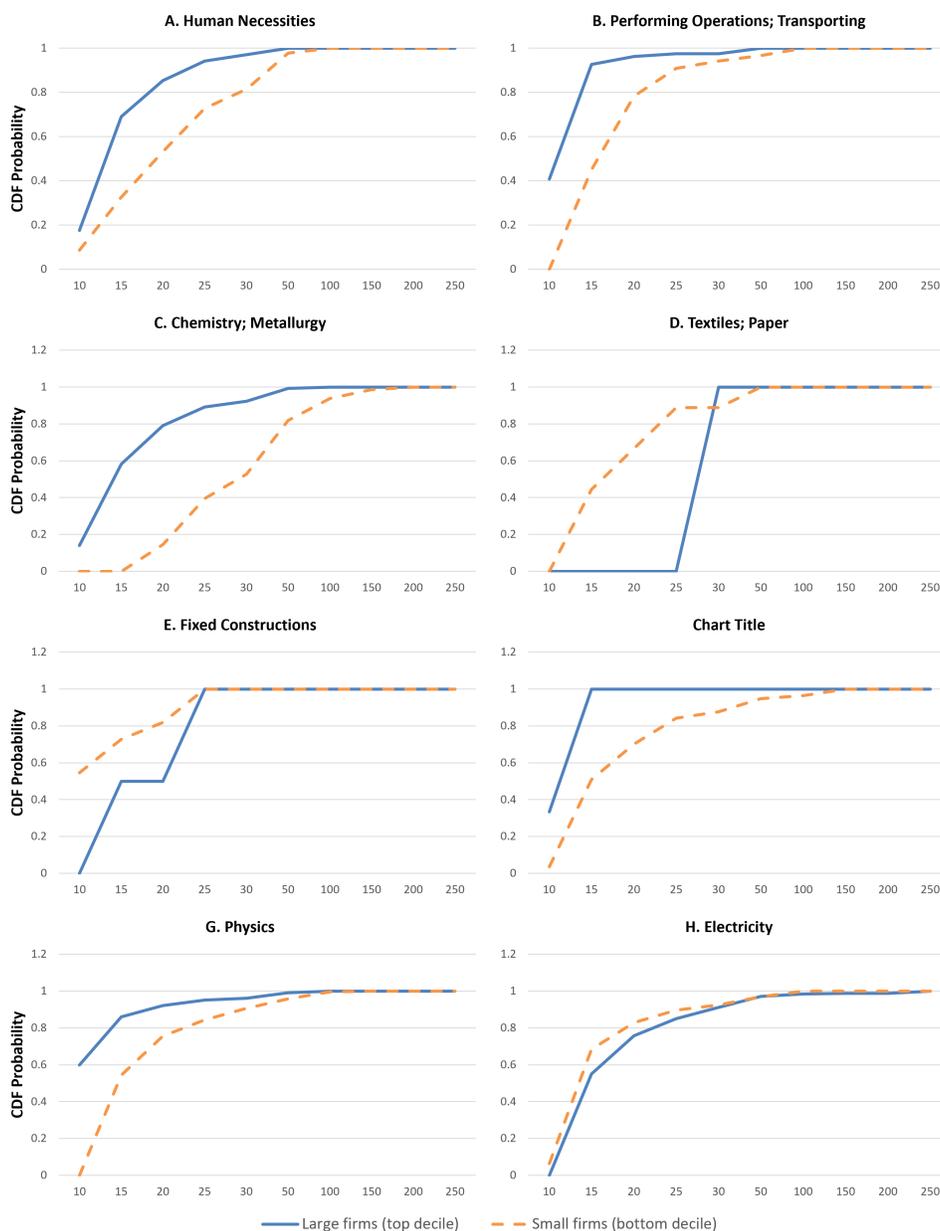
*Notes:* The figure presents the average quality of the highest quality patents across all CPC classes by firm size (i.e., sales quartiles) while controlling for the number of patents each firm produces in each CPC class. The X-axis is sales quartiles proxying firm size, and the Y-axis is the number of forward citations that highest quality quality patents receive in a given firm size bin, CPC class, and year.

Figure C.2: CDF of top 1% cited patents for large and small firms



*Notes:* The figure reports the Cumulative Distribution Function of patent quality for large firms (top 10% by firms size) and for small firms (bottom 10% by firm size). Firm size is average firm sales across the sample period. Size quartiles are constructed with equal number of patents rather than an equal number of firms. For example, the solid line represents 1,902 patents of 24 large firms, while the dashed line represents 2,104 patents of 462 small firms. Patent quality is the number of forward citations received by patents produced by each firm for each year. Patent citation is normalized by CPC class-year mean.

Figure C.3: CDF of top 1% cited patents for large and small firms by technology area



*Notes:* The figure reports the Cumulative Distribution Function (CDF) of patent quality for large firms (top 10% by average firms sales) and for small firms (bottom 10% by average firm sales). Firm size is average firm sales across the sample period. Size quartiles are constructed with equal number of patents rather than an equal number of firms. Patent quality is the number of forward citations received by patent produced by each firm for each year. Patent citation is normalized by CPC class-year mean.

Table C.1: Correlation Matrix for firm level Variables

	Avg. quality	Avg. pvt value	% multi-juris patents	MFT	MFC	Diversification	Centered US tariffs	Firm sales	Emp-loyees	Market cap.	Pat. stock	Pub. stock	# of patents
Average Quality	1												
Average Private Value	0.01	1											
Share of multi-juris. patents	0.03	0.12	1										
MFT	-0.02	0.01	0.05	1									
MFC	-0.02	-0.06	-0.06	0.09	1								
Diversification	-0.03	0.14	0.21	-0.03	0	1							
US Tariffs (simple average)	0.03	-0.06	-0.13	-0.22	0.02	-0.01	1						
Firm Sales	-0.01	0.35	0.09	0.01	-0.07	0.47	-0.03	1					
Employees	-0.03	0.25	0.12	-0.04	-0.01	0.64	-0.02	0.62	1				
Market Capitalization	0.02	0.52	0.09	0.07	-0.02	0.34	-0.06	0.72	0.36	1			
Patent Stock	-0.02	0.15	0.11	0.15	0.04	0.59	-0.04	0.47	0.47	0.47	1		
Publication Stock	0.02	0.39	0.09	0.03	-0.02	0.38	-0.04	0.4	0.35	0.56	0.53	1	
# of Patents	-0.01	0.12	0.12	0.12	0.04	0.62	-0.03	0.42	0.43	0.49	0.81	0.42	1

Notes: This table shows the Pearson pairwise correlation of key firm level variables.

Table C.2: Summary Statistics by Industries with potential for licensing

	Pharmaceuticals		Instruments		Transport		Medical Equipment		Other Industries	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Average Quality	1.42	2.45	1.43	1.89	1.30	1.03	1.66	2.04	1.24	1.21
Average Private Value (\$ mn)	21.77	53.61	7.89	15.65	13.38	21.73	8.14	17.95	12.63	36.57
Share of multi-juris. patents	0.90	0.30	0.80	0.40	0.85	0.35	0.84	0.37	0.80	0.40
Firm Sales (\$ mn)	1,443.70	5,973.83	1,128.89	3,480.36	10,733.39	28,108.30	545.15	1,194.84	3,344.66	14,451.30
Reassignments	341.10	225.30	415.44	588.08	157.86	305.40	355.46	279.26	357.73	774.90
Commercialization Assets	0.08	0.09	0.18	0.08	0.08	0.11	0.26	0.09	0.19	0.17
Diversification	1.36	2.60	2.51	4.84	8.20	11.71	0.97	1.80	3.66	6.20
Market Capitalization (\$ mn)	5,663.18	23,025.81	1,737.58	4,988.13	5,143.84	11,949.38	1,456.35	4,311.08	4,313.47	20,763.50
Patent Stock	90.71	287.33	95.48	278.09	289.55	665.21	46.84	100.06	179.30	662.11
Publication Stock	255.95	907.02	56.80	158.14	145.91	362.58	21.90	96.46	71.11	332.26
US Tariffs (simple average)	-0.17	0.94	0.11	1.56	0.06	0.87	-0.02	2.19	0.12	1.00
# of Patents	18.32	47.39	19.19	47.29	59.47	146.31	10.79	19.69	39.33	142.65
No. of firms	524	524	318	318	124	124	176	176	1,641	1,641
Sample Period	1980-2015	1980-2015	1980-2015	1980-2015	1980-2015	1980-2015	1980-2015	1980-2015	1980-2015	1980-2015
No. of Obs	4,202	4,202	2,970	2,970	1,583	1,583	1,561	1,561	16,599	16,599

*Notes:* This table presents summary statistics by industries with high potential for licensing.

Table C.3: Distribution of forward citations of top 1% patents of small and large firms by technology area

	(1)	(2)	(3)	(4)	(5)	(6)
	Num. Patents in top 1%	Num. patents in top 1% for small firms	Num. patents in top 1% for large firms	Mean for small firms	Mean for large firms	K-S Test: H0: Small=Large; HA: CDF of citations for Small firms lies below CDF for Large firms
All	9261	926	885	21.9	14.0	Reject
Human Necessities	922	92	68	21.3	14.5	Reject
Performing Operations; Transporting	1200	120	81	19.1	11.7	Reject
Chemistry; Metallurgy	1413	144	129	39.0	16.6	Reject
Textiles; Paper	80	9	1	20.5	25.5	Do not Reject
Fixed Constructions	108	11	8	11.8	18.9	Do not Reject
Mechanical Engineering; Lighting; Heating; Weapons; Blasting	568	57	3	22.4	9.9	Reject
Physics	2334	237	229	19.2	11.4	Reject
Electricity	2631	263	247	16.3	20.3	Do not Reject

*Notes:* This table shows the number of patents, mean and test results of Kolmogorov-Smirnov tests of top 1% cited patents of small (bottom sales decile) and large (top sales decile) firms. The null hypothesis Kolmogorov-Smirnov tests is that of equality of CDFs of citations for small and large firms. This null is either rejected or not rejected in favor of the alternate hypothesis that the CDF of quality for small firms lies below the CDF of quality for large firms, specified in the last column.

Table C.4: Additional test of the model - quality

Dependent variable	log(1+Avg. cites)	log(1+Sales) <sub>t-1</sub>	log(1+Avg. cites)
	(1)	(2)	(3)
VARIABLES	OLS	1st Stage	2nd Stage
log(1+Sales) <sub>t-1</sub>	-0.026 (0.008)		-0.239 (0.102)
log(1+Patent stock) <sub>t-1</sub>	-0.019 (0.009)	0.155 (0.020)	0.014 (0.019)
log(1+Publication stock) <sub>t-1</sub>	-0.010 (0.010)	0.155 (0.025)	0.023 (0.019)
log(1+Market capitalization) <sub>t-1</sub>	0.002 (0.006)	0.345 (0.017)	0.076 (0.035)
(Centered) tariffs <sub>t-2</sub>		-0.073 (0.013)	
Cragg-Donald F-stat		102.40	
Kleibergen-Paap F-stat		29.80	
Firm dummies	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes
SE Clustered	Firm	Firm	Firm
Mean of DV	1.29	1.29	1.29
Observations	14,242	14,242	14,242
R-squared	0.49		-0.11

*Notes:* This table presents 2SLS results on the relationship between invention quality and firm size. *Centered US tariffs* is the average tariffs charged by US on its imports from the rest of the world. This measure is demeaned within each four digit SIC and weighted by the share of each firm's sales across SIC's (based on Compustat's segment data). Clustered standard errors are reported in parentheses.

Table C.5: The Relationship between Firm Size and Invention Value - Technology Area Level

Dependent variable	ln(Average patent value)				
	(1)	(2)	(3)	(4)	(5)
VARIABLES	Baseline	Firm FEs	Tech Area FEs	Stock	Market Cap
$\log(1+\text{Sales})_{t-1}$	0.406 (0.012)	0.169 (0.024)	0.170 (0.024)	0.247 (0.024)	0.056 (0.021)
$\log(1+\text{Patent stock})_{t-1}$				-0.172 (0.023)	-0.160 (0.022)
$\log(1+\text{Publication stock})_{t-1}$				0.020 (0.023)	0.013 (0.021)
$\log(1+\text{Market capitalization})_{t-1}$					0.292 (0.015)
Firm dummies	No	Yes	Yes	Yes	Yes
Tech area dummies	No	No	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes
SE Clustered	Firm	Firm	Firm	Firm	Firm
Mean of DV	18.93	19.00	19.00	19.45	19.44
Observations	62,662	62,380	62,380	59,810	59,681
R-squared	0.52	0.85	0.85	0.85	0.86

*Notes:* This table presents the relationship between average private value of inventions and firm size. The sample consists of 2,786 public firms (from Compustat) and covers years 1980 through 2015. Average private value is from [Kogan et al. \(2017\)](#) and is based on abnormal stock market returns in response to patent issuance. Firm sales is lagged by one year. Publication stock and Patent stock are from [Arora, Belenzon and Sheer \(2021\)](#). Clustered standard errors are reported in parentheses.

Table C.6: The Relationship between Firm Size and Invention Quality - Technology Area Level

Dependent variable	log(1+Average forward citation)				
	(1)	(2)	(3)	(4)	(5)
VARIABLES	Baseline	Firm FEs	Tech Area FEs	Stock	Market Cap
log(1+Sales) <sub>t-1</sub>	-0.005 (0.002)	-0.034 (0.006)	-0.035 (0.006)	-0.021 (0.006)	-0.022 (0.006)
log(1+Patent stock) <sub>t-1</sub>				-0.021 (0.006)	-0.021 (0.006)
log(1+Publication stock) <sub>t-1</sub>				-0.007 (0.007)	-0.007 (0.007)
log(1+Market capitalization) <sub>t-1</sub>					0.001 (0.005)
Firm dummies	No	Yes	Yes	Yes	Yes
Tech area dummies	No	No	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes
SE Clustered	Firm	Firm	Firm	Firm	Firm
Mean of DV	1.32	1.32	1.32	1.31	1.31
Observations	62,662	62,380	62,380	59,810	59,681
R-squared	0.01	0.25	0.27	0.26	0.26

*Notes:* This table presents the relationship between invention quality (average forward citations) and firm size (firm sales) at the firm-year level. The sample consists of 2,786 public firms (from Compustat) and covers years 1980 through 2015. Forward citations are weighted by the average value in their cohort (CPC class and grant year). Firm sales is lagged by one year. Publication stock and Patent stock are obtained from [Arora, Belenzon and Sheer \(2021\)](#). Clustered standard errors are reported in parentheses.

Table C.7: Alternative measures of firm size

Dependent variable	ln(Avg. value)		log(1+Avg. cites)	
VARIABLES	(1)	(2)	(3)	(4)
ln(Employees) <sub>t-1</sub>	0.008 (0.021)		-0.019 (0.009)	
Large firm dummy <sub>t-1</sub>		0.127 (0.040)		-0.015 (0.017)
log(1+Patent stock) <sub>t-1</sub>	-0.130 (0.018)	-0.135 (0.018)	-0.016 (0.007)	-0.019 (0.006)
log(1+Publication stock) <sub>t-1</sub>	0.004 (0.019)	0.002 (0.019)	-0.016 (0.008)	-0.018 (0.007)
log(1+Market capitalization) <sub>t-1</sub>	0.302 (0.013)	0.298 (0.013)	0.003 (0.005)	-0.002 (0.005)
Firm dummies	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes
SE Clustered	Firm	Firm	Firm	Firm
Mean of DV	14.11	14.11	1.30	1.30
Observations	24,340	24,340	24,338	24,340
R-squared	0.89	0.89	0.44	0.44

*Notes:* This table presents robustness checks using an alternative measure of firm size (i.e., employee count). Large firm dummy takes a value of 1 if a firm is in the top half of the employee count distribution. Clustered standard errors are reported in parentheses.

Table C.8: Alternative measure of diversification - Stirling Index

Dependent variable	ln(Avg. value)	log(1+Avg. cites)
VARIABLES	(1)	(2)
log(1+Sales) <sub>t-1</sub>	0.057 (0.018)	-0.03 (0.007)
log(1+Diversification)	-0.033 (0.009)	0.018 (0.004)
log(1+Patent stock) <sub>t-1</sub>	-0.131 (0.017)	-0.019 (0.006)
log(1+Publication stock) <sub>t-1</sub>	-0.003 (0.018)	-0.012 (0.007)
log(1+Market Capitalization) <sub>t-1</sub>	0.29 (0.012)	0.003 (0.005)
Firm Dummies	Yes	Yes
Year Dummies	Yes	Yes
SE Clustered	Firm	Firm
Mean of DV	14.1	1.3
Observations	24,529	24,529
R-squared	0.89	0.44

*Notes:* This table presents the relationship between firm diversification (measured by Stirling index based on [Stirling \(2007\)](#)) and both invention value and quality. This index is constructed by taking the sum over the product of share of patents in each CPC class pair and the distance between each CPC pair of a firm. The distance between each CPC class pair is calculated as a range between 0 and 2, where 2 is the distance between two completely distinct CPC classes, 1 is the distance between CPC classes whose CPC section is the same and 0 is for firms that patent only in one CPC class. The final measure is log-transformed. Clustered standard errors are reported in parentheses.

Table C.9: Core technology area

Dependent variable	ln(Avg. value)	log(1+Avg. cites)
	(1)	(2)
VARIABLES		
log(1+Sales) <sub>t-1</sub>	0.054 (0.018)	-0.028 (0.008)
log(1+Patent stock) <sub>t-1</sub>	-0.138 (0.018)	-0.011 (0.007)
log(1+Publication stock) <sub>t-1</sub>	-0.003 (0.019)	-0.016 (0.008)
log(1+Market capitalization) <sub>t-1</sub>	0.285 (0.012)	0.005 (0.005)
Firm dummies	Yes	Yes
Year dummies	Yes	Yes
SE Clustered	Firm	Firm
Mean of DV	13.96	1.31
Observations	24,529	24,529
R-squared	0.88	0.40

*Notes:* This table presents the firm size relationship with average value and quality of inventions in the core technology area of each firm. Core technology area is defined as one in which a firm patents most frequently. Clustered standard errors are reported in parentheses.