1 Data description

1.1 Data Sources

- *Patents.* PatentsView USPTO, PATSTAT patent classification into NAICS by **?**, crosswalk between the two sectors built by Gianluca Tarasconi (2019).[1]
- *Concentration and Sales.* US Census extended data from **?**.[2]
- *Price Indices.* To deflate sales, I use NAICS-specific price indices from the BLS.
- *Market Regulations.* Mercatus RegData 4.0.[3]
- *Productivity.* I use output per worker from the Economic Census.

1.2 Constructed Data

- *Knowledge Markets.* I build knowledge markets using observed flows of inventors across projects classified through text analysis into different NAICS 4-digit codes by **?**. I maximize the modularity of the resulting network using the Louvain method to identify communities of NAICS that are connected by inventor flows over the period 1976-2015. I obtain about 10 non-singleton markets.
- *Effective Inventors.* As discussed above, I compute inventor productivity as the inventor fixed effect α_i from the fully-saturated regression:

  $$\#\text{Patents}_{cfit} = \alpha_i + \alpha_{cft} + \varepsilon_{cfit} \tag{1}$$

  at the level of CPC (Cooperative Patent Classification) class, c, assignee/company, f, inventor i and year t. I use the broadest CPC class to maximize the number of fixed effects that I can identify. The term α_{cft} is a CPC class by firm by time fixed effect.
- *Extended Regulation Series.* Mercatus RegData provides a count of restrictions imposed on a number of NAICS 4d-digit product markets, obtained by matching a set of keywords in NAICS descriptions to regulatory texts, and then taking the best match for each document. However, the available data does not include a set of codes for unspecified reasons.
  Therefore, I process the description of NAICS 4d codes and compute the cosine-similarity between all pairs of sectors. I build an estimate of sector-relevant restrictions for missing sectors by taking an average weighted by cosine similarity of sectors included in RegData. In particular, I include in the average the five most similar NAICS codes if similarity is larger than .2, and I

[1] See https://patentsview.org/forum/7/topic/143, https://rawpatentdata.blogspot.com/2019/07/patstat-patentsview-concordance-update.html

[2] Available at https://sites.google.com/site/drjankeil/data.

[3] https://www.quantgov.org/bulk-download.

attribute the regulations of the most similar sector otherwise. I chose this threshold by inspecting the similarity associated to various NAICS pairs (XXX provide examples in footnote XXX).

- *Share of self-citations and excess citation measures.* For each patent classified by **?**, i.e. with non-missing NAICS classification, I count the set of cited patents that belong to the citing patent's assignee. In the case of cited patents with multiple assignees, I consider half a count if the assignee is among them. The share of self-citation is given by this count divided by total citations. I construct five measures to correct self-citations for the assignee's importance in the relevant technology class of cited patents. For each citation made, excess self-citations are defined as $1 - Pr$ (self-citation). The various measures differ on how the probability of self-citation is computed. For the first three measures, I compute this probability as the assignee's share of total patents in the NAICS code attributed to the citing patent. In employ in turn the share of NAICS patents in the year, the previous five years, and the cumulative share from the beginning of the sample. The other two measures are based on the CPC classification at the group and subgroup levels (the lowest levels of detail in the classification). For this measure, the probability of self-citation is constructed for each citation by taking the share of patents by the assignee in the CPC (sub)group and year corresponding to the cited patents.[4]
  Finally, I aggregate all measures across assignees in the same NAICS 4-digit code using the number of patents in the relevant NAICS code by each assignee in each year.

- *Inventor productivity.* As a measure of inventor productivity, I use the average growth in output per worker divided by inventor's fixed effect. I choose this measure since it is the productivity measure available for most sectors.

- *Forward citations and patent generality.* See **?** and **?**.

1.2.1 Aggregation at Five-Year Census Frequency

Data from the Economic Census are available at five-year frequency for the years 1997-2017, which requires aggregating the other data at the same frequency. Since I am interested in the effect of concentration on the allocation of inventors, I average all variables related to inventors and productivity using the five-year window *starting* in the census year (e.g., 1997-2001 for 1997). In the IV regression I use product restrictions as an instrument for concentration, which motivates me to average restrictions in the five-year window *ending* in the census year (e.g., 1993-1997 for 1997). Since **?**'s matching only covers the period up to 2016, I run all specifications in long-differences over the time frame 1997-2012.

2 Results

I present four main findings that apply to the period 1997-2012:

[4] This procedure is close in spirit to the approach followed in Akcigit and Kerr's (2018) Appendix C.

1. Effective inventors have become more concentrated in specific technology classes and economic sectors;

2. Sectors with increased concentration have attracted a growing share of relevant inventor types. The IV analysis suggest that the rise in inventor shares is the result, and not the cause, of increased concentration;

3. Growth in the share of relevant inventors is negatively correlated with inventor productivity, as measured by average growth in output per workers divided by effective inventors employed;

4. Growth in the share of relevant inventors is positively correlated with the share of self-citations and excess self-citations, as well as concentration of inventors at the top within sectors.

The first finding emerges from the computation of Gini coefficients of effective inventors across technologies and sectors. Findings 2-4 come from long-difference regressions over the period 1997-2012. Regressions are weighted by sector sales in 2012 for findings 2-3, which rely on Census sector-level measures, with robust standard errors. I present both unweighted and unweighted results for finding 4, since these variables do not represent aggregates according to sales, and rely on patent data only.

For findings 2-4, I run long-difference regressions at the product market level for the period 1997-2012:

$$\Delta \text{Outcome}_p = \beta \Delta \text{Indep.Var.}_p + \gamma' \Delta \text{Controls}_p + \varepsilon_p, \tag{2}$$

where Δ denotes the long-difference operator. Throughout the analysis, I use the change in log-real sector sales or the change in log- real sector sales per company as controls, in order to capture effects on the outcome stemming from increases in the real size of sectors or firms. I also run the within-knowledge-market specification:

$$\Delta \text{Outcome}_p = f_k \mathbf{1}\left\{p \in k\right\} + \beta \Delta \text{Indep.Var.}_p + \gamma' \Delta \text{Controls}_p + \varepsilon_p, \tag{3}$$

which includes fixed effects to denote the membership of product markets, p, to the relevant knowledge market k.

2.1 Increased Effective Inventors Concentration across Patent Classes and NAICS sectors

First, I compute Gini coefficients of effective inventors, α_i (shifted to be nonnegative) across patent classes. This is reported in Figure 1. Second, I use the subset of patents classified by Zolas (which ends in 2016) that I can match to PatentsView. This is shown in Figure 2. In both cases the coefficient increased by about 10% from 1978.

Figure 1: Effective inventor Gini coefficients at CPC-4 level

58_Users_andreamanera_Dropbox__MIT__Research_Pr___pet

Figure 2: Effective inventor Gini coefficients at NAICS 4-digit level

59_Users_andreamanera_Dropbox__MIT__Research_Pr___-sc

2.2 Markets with Growing Concentration Increased Their Inventor Share

I compute inventor markets as described above, on the set of all patents that are assigned a NAICS code by Zolas et al's (2016) procedure. The share of effective knowledge-market inventors employed in each product market in year t is then given by the sum of inventor fixed effects from regression 1 for inventors patenting in NAICS sector, p, over their knowledge-market total:

$$\text{Share}^{k}_{p,t} \equiv \frac{\sum_{i \in p(t)} \alpha_i}{\sum_{i \in k(t)} \alpha_i}.$$

Here, the notation $i \in p(t)$ indicates that inventor i patents in product market p at time t, while $i \in k(t)$ denotes that inventor i patents in knowledge market k at time t. Two notes are in order. First, by construction, each product market, p, belongs to only one knowledge market, k. Second, inventors can patent in several product markets in each year. In this case, I attribute the inventor fixed-effect α_i to all product and knowledge markets to which the inventor contributes. The total effective inventor share is defined analogously as

$$\text{Share}_{p,t} \equiv \frac{\sum_{i \in p(t)} \alpha_i}{\sum_{i} \alpha_i}.$$

In this section, I present three sets of results for each specification, which differs in the estimation sample to account for extreme observations. In regression tables, “Full Sample” refers to the sample of observations with non-missing observations for all the variables included. I propose two sample selections to rule out that outliers drive the baseline results. “Trim Outliers” refer to a sample which trims the most extreme observations for the outcome and the independent variable separately. I trim the observations that fall beyond three standard deviations from the sample average of each variable, and that are most likely to drive the results estimated using the full sample.[5] “Mahalanobis 5%” denotes the sample where I trim the 5% extreme observations based on the Mahalanobis distance of pairs of observations from the data centroid. Since this procedure is based on the joint distribution of the outcome and independent variable, the sample varies in each regression.[6]

[5] I justify the choices for each variable in detail in my replication code using the empirical kernel density and detailed tabulations.

[6] I like this but it does not always produce the results one would expect. I am thinking of implementing a version where I trim the residualized variables. Or experiment with difference distances.

Table 1 presents the results of regression 3 where the outcome variable is the change in knowledge-market inventor share, and the independent variable is the change in the lower bound of the Herfindal-Hirschman Index, or the index as reported by the Economic Census. The Economic Census reports HHI indices only for a subset of NAICS 4-digit sectors, but includes concentration ratios for a much wider set. These concentration ratios are used by Keil (2017) to construct a lower bound on the HHI, which I employ as my main measure of concentration due to its wider availability. The results in Table 1 highlight a strongly significant positive correlation between the change in HHI and the change in the share of effective inventors accruing to each NAICS sector. Note that this regression is only partially driven by the contemporaneous correlation between the two variables. As discussed above, the share of effective inventors is average over the five years *starting* in the Economic Census year, while the concentration measures refer to the Economic Census year only.

Two important notes on the scale of the variables are in order. First, here and in all following tables and graphs, all variables which refer to shares or growth rates are reported in percentage points for ease of interpretation. Therefore, for example the coefficient in Column (1) of Table 1 should be interpreted as saying that an increase in one unit of the HHI index leads to an increase in the share of the relevant knowledge market of 27.25 pp. Second, HHI indices are instead constructed to range between 0 and 1. In particular, the HHI lower bound has sales-weighted an average of about .03, and a standard deviation of .032 in 2012. According to Table 1, a standard deviation increase in this measure is associated to a 0.87 pp increase in the share of inventors accruing to the relevant NAICS sector. This In comparison, the sales-weighted average share of inventors in 2012 is 1.18%, with a standard deviation of 1.82%, so the estimated effect of a one standard deviation increase in concentration corresponds to about half a standard deviation increase in the share of inventors in the relevant market. Clearly, the estimates using the HHI lower bound tend to be noisier as this is a constructed, and therefore imprecise, measure of concentration. However, the number of available observations is much larger than the actual HHI, so I use the HHI lower bound in most analyses below.

While suggestive, the correlation presented above is far from ideal, as it neglects two fundamental components. First, it does not include controls for the size of the sectors or firms, which could have a confounding and mechanical effect on the share of scientists in a specific sector. Second, it estimates the correlation both across and within knowledge markets. In Table 2, I address these two limitations by restricting the analysis within knowledge markets, and controlling for two measures of size. In the upper panel of Table 2, I use the change in the logarithm of real sales as a measure of the size of each sector, while in the lower panel I present the results when average sales per firm are included as a control. The inclusion of sales per firm is motivated by the fact that there might be significant barriers to entry to R&D, easier to overcome for larger firms, mechanically linking concentration and inventor hiring. Since the Economic Census reports the number of companies only for a subset of firms, the sample used in the lower panel is smaller than the upper panel. The results in Table 2 confirm the positive relation between the change in inventor shares and concentration, and are largely unchanged relative to the estimates in Table 1, suggesting that the correlation does not arise mechanically from factors related to firm or sector size. In particular, these findings imply that sectors with increasing concentration have attracted a rising share of scientists above what would be implied by their expansion in overall sales as well as average firm size.

Figure 3 depicts graphically the residualized observations underlying the estimated coefficients in Columns (2) and (6) of Table 2, Panel (a). The upper panel portrays changes of knowledge-market inventor shares over the change in the HHI lower bound, after partialling out fixed effects for the relevant knowledge market and changes in log real sales, with the marker size proportional to the regression weight. Although the sample displays some observations that appear extreme, the bulk of observations—and especially weighted observations—falls on the regression lines, mitigating the concerns that a few outliers might drive the results. In any event, I explore the robustness of the results to the exclusion of non-residualized observations, both manually and defining extreme observations based on the Mahalanobis distance. Importantly, this exercise reveals that the observations that appear extreme in the residualized scatter are not unusual when considering the marginal or joint distribution of non-residualized outcome and independent variable. The bottom panel of Figure 3 reports the binned scatter plot corresponding to the sample where the 5% extreme observations according to the Mahalanobis distance have been removed, and confirms that the positive relation between concentration and inventor shares is not driven by a few extreme observations. In particular, the corresponding regression results in Table 2(a), Column (6), show that the estimated coefficient is significant at a 5% confidence level. The results presented in this section are robust to using the raw number of inventors to compute the share of researchers captured by each product market. TODO: Add Appendix.

Table 1: Regressions of Change in 4-digit Knowledge Market Share over Change in HHI Measures, Long-Differences, 1997-2012

	Ch. 4d K.M. Eff. Inv. Share (%)					
	(1)	(2)	(3)	(4)	(5)	(6)
Ch. HHI lower bound	27.293*		27.183*		27.326*	
	(11.569)		(11.941)		(11.620)	
Ch. HHI		22.399***		22.399***		22.350***
		(6.345)		(6.345)		(6.343)
4D Knowledge Market						
Sample	Full Sample	Full Sample	Trim Outliers	Trim Outliers	Mahalanobis 5%	Mahalanobis 5%
Weight	Sales	Sales	Sales	Sales	Sales	Sales
Observations	157	80	155	80	150	71

Note: Regressions weighted by sales in 2012; Robust standard errors in parentheses; Symbols denote significance levels $(+\ p<0.1,^{*}\ p<0.05,^{**}\ p<.01,^{***}\ p<.001)$; Checkmarks indicate the inclusion of fixed effects. This Tables presents the results of specifications (2), when the outcome is the share of effective inventors of sector p over total inventors in knowledge market k, and the independent variable is the change in the lower bound of the Herfindal-Hirschman Index for product market p, as implied by Economic Census concentration ratios, or the HHI index reported in the Economic Census. "Full Sample", "Trim Outliers" and "Mahalanobis 5%" refer to the samples described in the main text.

Table 2: Regressions of Change in 4-digit Knowledge Market Share over Change in HHI Lower Bound, Long-Differences, 1997-2012

(a) Controlling for Change in Log Real Sales

	Ch. 4d K.M. Eff. Inv. Share (%)					
	(1)	(2)	(3)	(4)	(5)	(6)
Ch. HHI lower bound	26.093*	22.509*	25.904*	22.716*	26.111*	22.554*
	(10.696)	(10.848)	(11.124)	(10.948)	(10.725)	(11.019)
Ch. Log Real Sales	0.914**	0.548*	0.881**	0.539*	0.918**	0.562*
	(0.278)	(0.243)	(0.275)	(0.242)	(0.283)	(0.261)
4D Knowledge Market FE		✓		✓		✓
Sample	Full Sample	Full Sample	Trim Outliers	Trim Outliers	Mahalanobis 5%	Mahalanobis 5%
Weight	Sales	Sales	Sales	Sales	Sales	Sales
Observations	157	153	155	152	150	139

(b) Controlling for Change in Log Real Sales per Company

	Ch. 4d K.M. Eff. Inv. Share (%)					
	(1)	(2)	(3)	(4)	(5)	(6)
Ch. HHI lower bound	35.230**	20.783+	35.230**	20.783+	35.154**	22.854*
	(12.759)	(10.615)	(12.759)	(10.615)	(12.647)	(11.197)
Ch. Log Real Sales per company	0.175	-0.040	0.175	-0.040	0.300	-0.055
	(0.382)	(0.253)	(0.382)	(0.253)	(0.460)	(0.346)
4D Knowledge Market FE		✓		✓		✓
Sample	Full Sample	Full Sample	Trim Outliers	Trim Outliers	Mahalanobis 5%	Mahalanobis 5%
Weight	Sales	Sales	Sales	Sales	Sales	Sales
Observations	81	79	81	79	75	67

Note: Regressions weighted by sales in 2012; Robust standard errors in parentheses; Symbols denote significance levels $\left(+\ p<0.1,^{*}\ p<0.05,^{**}\ p<.01,^{***}\ p<.001\right)$; Checkmarks indicate the inclusion of fixed effects. This Tables presents the results of specifications (2) and (3), when the outcome is the share of effective inventors of sector p over total inventors in knowledge market k, and the independent variable is the change in the lower bound of the Herfindal-Hirschman Index for product market p, as implied by Census concentration ratios. “Full Sample”, “Trim Outliers” and “Mahalanobis 5%” refer to the samples described in the main text.

Figure 3: Residualized Scatter Plots Corresponding to Selected Columns in Table 2, Panel (a)

(a) Raw Scatter Plot, Specification in Column (2)

72_Users_andreamanera_Dropbox__MIT__Research_Pr___ng-for-scientists_draft_graphs_raw_Dk4_h

(b) Binned Scatter Plot, Specification in Column (6)

2.2.1 IV Results

In this subsection, I present IV results that suggest that the relation between concentration and inventor shares is causal. Indeed, more concentration might just be the result of an increase in technological entry barriers, established by incumbents through an increase in their R&D inventors. In this scenario, the causality would flow from increased inventor shares to higher concentration. Above, I tried to mitigate this concern using the the average share of inventors following the Economic Census years to which the HHI refers as my outcome variable. However, reverse causality could still be present if the autocorrelation of inventor shares is sufficiently high. This motivates me to produce 2SLS estimates that instrument the change in the HHI lower bound with changes in product market restrictions, as measured by the Mercatus dataset RegData 4.0. Theoretically, an increase in restrictions should raise barriers to entry in the affected product markets, thus leading to higher concentration. As discussed below, this proves to be the case empirically, making a case for the validity of restrictions as an instrument for concentration. A violation of the exclusion restriction requires a causal connection between product market regulations and the share of inventors hired by each sector, which acts independently of product market concentration. A possibility in this sense is the increase in the number of inventors required to fulfill product market restrictions, if such regulations specifically affect technologies currently in use in the industry. However, this effect should be both large and persistent to be captured by my measure of inventor shares. Further, while RegData certainly include product restrictions, there are also a number of regulatory burdens that are not related to technological components, like reporting obligations and other legal burdens. In addition, while product restrictions might certainly induce a change in the direction of innovation, there is no a priori reason to believe that the scale of innovation activity should also increase. These considerations lead me to believe that the exclusion restriction is not highly likely to be violated.

The results of the 2SLS estimation are presented in the upper panel of Table 3. The specification is the same as in Column (2) of 2, including both knowledge market and sale change fixed effects. The 2SLS estimates confirm the significance of concentration changes for the increase in knowledge market inventor shares. The magnitudes of estimated coefficients are statistically indistinguishable from the ones reported in the baseline regression. The first-stage F clearly indicates that instruments are weak. This is unsurprising since, as detailed above, both the HHI lower bound and the regulation measures are constructed and therefore imprecise. In particular, I had to impute regulations for a large part of the sample using the cosine-similarity between product market restrictions.[7] However, instruments are not irrelevant. The results in the lower panel of Table 3 imply that the first-stage t-statistic for the regression of the change in the HHI lower bound over log-regulations is 2.07, which corresponds to a p-value of 0.041. The reduced form regression of inventor share over log restriction change is equally highly significant. Accordingly, the SW underidentification test rejects the null hypothesis at a 5%

[7] Using only available sectors requires dropping two thirds of the observations. See Appendix **??** for details on data construction.

confidence level. Given the weakness of the instruments, I also report the Anderson-Rubin p-value, which confirms that the coefficient is 5% significant.

2.2.2 Results on Overall Inventor Shares

I also inspect the effect of concentration increases on the share of inventors across all knowledge markets. While the correlation is positive and significant when some outliers are removed, this relation is not robust to the inclusion of all observations or the the alternative trimming procedure provided by the Mahalanobis distance. Results are displayed in Table 4. This shows that the relation highlighted above is not apparent when looking at the overall share of scientists, while it emerges clearly only when looking at economic sectors that actually compete for the same inventors, that is, those that belong to the same knowledge market.

Table 3: IV Regressions of Change in 4-digit Knowledge Market Share over Change in HHI Lower Bound, 2SLS Long-Difference, 1997-2012

(a) 2SLS Results

	Ch. 4d K.M. Eff. Inv. Share (%)	
	(1)	(2)
Ch. HHI lower bound	30.560+	30.096+
	(15.904)	(15.819)
Ch. Log Real Sales	0.544*	0.525*
	(0.244)	(0.247)
4D Knowledge Market FE	✓	✓
Sample	Full Sample	Mahalanobis 5%
Weight	Sales	Sales
Observations	157	150
First-Stage F	4.587229	4.753009
Anderson-Rubin p-value	.0281448	.0321185

(b) First Stage and Reduced Form

	Ch. 4d K.M. Eff. Inv. Share (%)	Ch. HHI lower bound
	(1)	(2)
Ch. Log Restricitions (NAICS 4d)	0.478*	0.016*
	(0.220)	(0.007)
Ch. Log Real Sales	0.539+	-0.000
	(0.274)	(0.005)
4D Knowledge Market FE	✓	✓
Sample	Full Sample	Full Sample
Weight	Sales	Sales
Observations	153	153

Note: Regressions weighted by sales in 2012; Robust standard errors in parentheses; Symbols denote significance levels $\left(+\ p<0.1,^{*}\ p<0.05,^{**}\ p<.01,^{***}\ p<.001\right)$; Checkmarks indicate the inclusion of fixed effects. This Tables presents the results of specifications (2), when the outcome is the share of effective inventors of sector p over total inventors in knowledge market k, and the independent variable is the change in the lower bound of the Herfindal-Hirschman Index for product market p, as implied by Economic Census concentration ratios, instrumented by the change in log-restrictions relevant to the NAICS sector. The lower panel present first-stage and reduced-form relations. “Full Sample” and “Mahalanobis 5%” refer to the samples described in the main text.

Table 4: Regressions of Change in Total Inventors' Share over Change in HHI Lower Bound, Long-Difference, 1997-2012

	Ch. Total Eff. Inv. Share (%)					
	(1)	(2)	(3)	(4)	(5)	(6)
Ch. HHI lower bound	0.297	1.692	1.328*	1.532*	0.271	1.889
	(2.007)	(1.956)	(0.649)	(0.696)	(2.038)	(2.023)
Ch. Log Real Sales	0.460	0.436	0.133**	0.109*	0.464	0.472
	(0.281)	(0.292)	(0.047)	(0.047)	(0.283)	(0.312)
4D Knowledge Market FE		✓		✓		✓
Sample	Full Sample	Full Sample	Trim Outliers	Trim Outliers	Mahalanobis 5%	Mahalanobis 5%
Weight	Sales	Sales	Sales	Sales	Sales	Sales
Observations	157	153	147	143	150	139

Note: Regressions weighted by sales in 2012; Robust standard errors in parentheses; Symbols denote significance levels $\left(+\ p<0.1,^{*}\ p<0.05,^{**}\ p<.01,^{***}\ p<.001\right)$; Checkmarks indicate the inclusion of fixed effects. Please refer to notes in Table 2 for further details.

2.3 Markets with Growing Inventor Shares Experienced a Fall in Inventor Productivity

Table 5 presents the results of running regression (3) when the outcome is the average growth in output per worker per effective inventor. I use growth in annual output per worker provided by the Economic Census and average this measure over the five-year window starting in the EC year, and I proceed analogously to build a measure of average effective inventors over the same period. Inventor productivity is then defined as average output per worker growth divided by average effective inventors. Both the outcome and the dependent variable are measured in percentage points. Table 5 reveals a negative and significant correlation between the increase in the effective inventors' change and inventor productivity. These findings are robust to considering only sectors with positive growth in output per worker over the period 2012-2016. The magnitude of estimated coefficients can be grasped considering the scale of the variables and their changes over the sample period. In particular, the median change in the share of effective inventors over the period was .014%, while the measure of effective inventors has a median of 2018.[8] Using the coefficient in Column (5) to predict the median annual change in labor productivity growth implied by rising inventor concentration amounts to a fall of .15% $(-.005 \times .014\% \times 2018)$. This number increases to . 28% when using the statistics relative to sectors with positive growth in labor productivity only, which accounted for the bulk of the increase in inventor shares. An alternative back-of-the-envelope computation, using the change in product market concentration to predict the change in inventor shares gives even starker results. Using the coefficient in Column (2) of Table 2(a), and given a median change in the HHI of 0.002 yields an increase in the share of effective inventors in concentrating sectors of 0.045%, which implies a fall in average labor productivity implied by misallocation of .45%. While these numbers might appear sizable considering the entirety of the economy, it is worth noting that the sample I have data for includes mainly manufacturing and retail sectors, which experienced a sizable reduction of about 2% in average

[8] Recall that effective inventors in each year are measured as the sum of inventor fixed-effects in each year, and therefore do not represent the simple count of inventors.

annual productivity growth from 1997-2012, driven by a steep decline in output per worker growth in manufacturing. Therefore, the mechanism I propose would explain from 7.5% to 22.5% of the observed decrease in output per worker growth.

Table 5: Regressions of Change in Inventor Productivity over Change in Inventors' Share over Change in 4-digit Knowledge Market Share, Long-Difference, 1997-2012

	Ch. Avg. Output/Worker Growth/Inventor (%)			
	(1)	(2)	(3)	(4)
Ch. 4d K.M. Eff. Inv. Share (%)	-0.007**	-0.005*	-0.007**	-0.005*
	(0.002)	(0.002)	(0.002)	(0.002)
Ch. Log Real Sales		-0.051*		-0.054*
		(0.021)		(0.021)
4D Knowledge Market FE	✓	✓	✓	✓
Sample	Full Sample	Full Sample	Mahalanobis 5%	Mahalanobis 5%
Weight	Sales	Sales	Sales	Sales
Observations	101	101	96	93

Note: Regressions weighted by sales in 2012; Robust standard errors in parentheses; Symbols denote significance levels $\left(+\ p<0.1,^{*}\ p<0.05,^{**}\ p<.01,^{***}\ p<.001\right)$; Checkmarks indicate the inclusion of fixed effects. Please refer to notes in Table 2 for further details. Inventor productivity is measured as the average growth in output per worker over the five years starting in the Economic Census year over the total number of effective inventors in each sector.

2.4 Markets with Growing Inventor Shares Saw an Increase in Self-Citations and in Inventor Concentration

While the findings presented so far establish a connection between the increase in inventor concentration and the fall in inventor productivity, they shed little light on the mechanisms underlying these developments. Table 6 offers a step in this direction, showing that the increase in inventor concentration is positively correlated with the increase in self-citations within the affected sectors. This relation is apparent when restricting attention to the middle range of changes in inventor shares, reported in the Panel (b). Figure 4 clarifies why this restriction is necessary, as the full sample present some extreme observations with little change in self-citations that drive the estimated coefficient towards 0. When restricting to the middle range, which effectively consists in dropping less than 10% extreme observations, a strong positive correlation emerges between change in inventor shares and excess citations. Importantly, the measures of excess self-citations that I construct account for the contribution of active firms to technological advances in the relevant CPC classifications, thus excluding a mechanical increase in self-citations that would result solely from a reduction in R&D activity by competing firms. That is, if a firm contributes 100% of the relevant patents to a filed and cites itself only, its excess self-citations are 0, as explained in Section 1.2.

The rise in self-citations suggests an increasing role of defensive patenting by large incumbents in concentrating sectors. This explanation would also speak to the related finding that sectors that increased their inventor share also saw a *within-sector* increase in inventor concentration. Table 7 shows that the share of effective inventors accruing to top inventor-hiring firms has increased in those sectors that attracted more inventors over the period considered. This finding is consistent across a variety of measures, and Columns (4) and (5) suggests that it is driven by a faster increase of inventors at the top of the distribution more than a transfer from bottom to top firms within the sector.

Table 8 presents regressions of measures of forward citations corrected for truncation, as well as patent generality (both constructed as in Hall et al., 2000) over changes in the inventor market share. The results highlight that sectors increasing their share of inventors have experience a significant fall in forward citations per patent. As for the results in 6, the effects are more pronounced in the middle range of inventor changes. In the restricted sample, the negative correlation between inventor shares and generality is also highly significant.

Table 6: Regressions of Change in Excess Self-Citations over 4-digit Knowledge Market Share, Long-Differences, 1997-2012

(a) Full sample

	Ch. excess self-c. over CPC group		Ch. excess self-c. over CPC subgroup	
	(1)	(2)	(3)	(4)
Ch. 4d K.M. Eff. Inv. Share (%)	0.920	-0.444	0.958+	-0.228
	(0.711)	(1.083)	(0.512)	(0.801)
Ch. Log Real Sales	-1.841	-1.954	-1.456	-1.674
	(1.925)	(1.988)	(1.326)	(1.279)
4D Knowledge Market FE		✓		✓
Sample	Full Sample	Full Sample	Full Sample	Full Sample
Weight				
Observations	157	153	157	153

(b) Full sample, restricting to the middle range of the change in inventor shares (-2% to $+2\%$)

	Ch. excess self-c. over CPC group		Ch. excess self-c. over CPC subgroup	
	(1)	(2)	(3)	(4)
Ch. 4d K.M. Eff. Inv. Share (%)	5.540**	5.244*	4.561***	4.110*
	(1.783)	(2.469)	(1.211)	(1.600)
Ch. Log Real Sales	-2.217	-2.099	-1.780	-1.780
	(1.879)	(1.976)	(1.287)	(1.265)
4D Knowledge Market FE		✓		✓
Sample	Full Sample	Full Sample	Full Sample	Full Sample
Weight				
Observations	145	144	145	144

Note: Unweighted regressions; Robust standard errors in parentheses; Symbols denote significance levels $\left(+\ p<0.1,^{*}\ p<0.05,^{**}\ p<.01,^{***}\ p<.001\right)$; Checkmarks indicate the inclusion of fixed effects. This Tables presents the results of specifications (2) and (3), when the outcome is the change in excess self-citations in sector p over the change in the share of inventors employed in sector p. Upper panel: full sample; Bottom panel: excluding sectors with absolute increase in the inventor share above 2%.

Figure 4: Residualized Scatter Plots Corresponding to Selected Columns in Table 6

(a) Binned Scatter Plot, Full Sample

74_Users_andreamanera_Dropbox__MIT__Research_Pr___g-for-scientists_draft_graphs_bin_self_a

(b) Binned Scatter Plot, Middle Range of the Change in Inventor Shares (−2% to +2%)

Table 7: Regressions of Change in Inventor Distribution Measures over Change in 4-digit Knowledge Market Share, Long-Difference, 1997-2012

	Ch. Inv. 90/50 Quantile Ratio	Ch. Inv. Top-10/Bottom-50 Share Ratio	Ch. Inv. Top-50/Bottom-50 Share Ratio	Ch. Inv. Top 10% Share	Ch. Inv. Bottom 50% Share
	(1)	(2)	(3)	(4)	(5)
Ch. 4d K.M. Eff. Inv. Share (%)	0.211+	0.243*	0.314+	0.018**	-0.008*
	(0.107)	(0.097)	(0.184)	(0.006)	(0.004)
Ch. Log Real Sales	-0.100	0.328	0.147	0.026	0.005
	(0.122)	(0.294)	(0.316)	(0.020)	(0.007)
4D Knowledge Market FE	✓	✓	✓	✓	✓
Sample	Full Sample	Full Sample	Full Sample	Full Sample	Full Sample
Weight	Sales	Sales	Sales	Sales	Sales
Observations	118	118	118	118	118

Note: Regressions weighted by sales in 2012; Robust standard errors in parentheses; Symbols denote significance levels $\left(+\ p<0.1,^{*}\ p<0.05,^{**}\ p<.01,^{***}\ p<.001\right)$; Checkmarks indicate the inclusion of fixed effects. Please refer to notes in Table 2 for further details. Column (1) uses the ratio in the 90 percentile of effective inventors to the median as the outcome variable. Columns (2) and (3) instead present the share ratio, that is the share of effective inventors accruing to the top 10 or 50% relative to the share accruing to the bottom 50% of the distribution within each NAICS sector.

Table 8: Regressions of Changes in Forward Citation over 4-digit Knowledge Market Share, Long-Differences, 1997-2012

(a) Full sample

	Ch. in log citations per patent (CPC2 based)	Ch. in log citations per patent (Total)	Ch. in patent generality
	(1)	(2)	(3)
Ch. 4d K.M. Eff. Inv. Share (%)	-0.197***	-0.227***	-0.004
	(0.044)	(0.051)	(0.004)
Ch. Log Real Sales	-0.234*	-0.258+	0.008
	(0.112)	(0.148)	(0.013)
4D Knowledge Market FE	✓	✓	✓
Sample	Full Sample	Full Sample	Full Sample
Weight			
Observations	153	153	153

(b) Full sample, restricting to the middle range of the change in inventor shares (−2% to +2%)

	Ch. in log citations per patent (CPC2 based)	Ch. in log citations per patent (Total)	Ch. in patent generality
	(1)	(2)	(3)
Ch. 4d K.M. Eff. Inv. Share (%)	-0.545***	-0.618***	-0.025*
	(0.113)	(0.137)	(0.012)
Ch. Log Real Sales	-0.232*	-0.255+	0.008
	(0.109)	(0.146)	(0.012)
4D Knowledge Market FE	✓	✓	✓
Sample	Full Sample	Full Sample	Full Sample
Weight			
Observations	144	144	144

Note: Unweighted regressions; Robust standard errors in parentheses; Symbols denote significance levels $\left(+\ p < 0.1,^{*}\ p < 0.05,^{**}\ p < .01,^{***}\ p < .001\right)$; Checkmarks indicate the inclusion of fixed effects. This Tables presents the results of specification (3), when the outcome is the log-change in forward citations and the change in patent generality in sector p over the change in the share of inventors employed in sector p. Column (1) and (2) presents the results when forward citations are extrapolated the procedure Hall et al. (2000) to avoid truncation bias. A specific cite-lag distribution over 35 years is estimated for each pair of cited and citing CPC2-codes. Column (1) employs the extrapolation scheme by each pair of CPC2 cited and citing sector. Column (2) applies the extrapolation scheme to total citations received by each cited patent. Column (3) presents results on the patent generality measures. All columns exclude self-citations. Upper panel: full sample; Bottom panel: excluding sectors with absolute increase in the inventor share above 2%.