# 1 Data description

In this section, I describe my main data sources and the procedure to build the dataset employed in my empirical analysis. Subsection 1.1 lists the sources of the raw data. Subsection 1.2 focuses on the definition of inventor productivity measures and knowledge markets, which I identify through realized inventor flows across sectors. Subsection 1.3 briefly describes other data construction steps that are discussed in more detail in Appendix **??**.

## 1.1 Data Sources

My empirical analysis relies on the variation of concentration across product markets, as defined by 4-digit NAICS sectors, and the impact of these shifts on the allocation of inventors with specific competences across these sectors, and their effect on inventors' productivity. I use USPTO patent data to measure inventor productivity and establish the set of product markets that share similar inventors, and US Economic Census data to obtain concentration and productivity growth measures. Finally, I also use a dataset of product market regulations, Mercatus RegData 4.0, to conduct an instrumental variable analysis, as well as NBER-CES that I employ to obtain estimates of the Lerner Index that I employ in the calibration of my theoretical model.

My primary source is given by USPTO patent data from PatentsView. This dataset contains disambiguated patent, inventor and assignee identifiers, as well as Cooperative Patent Classification (CPC) classes for each of the patents deposited in the period 1975-2021. I employ these data to construct inventor flows across different sectors, which I build employing the ALP classification of 1976-2016 patents into NAICS sectors of application developed by **?**. Since this classification is constructed using the PATSTAT dataset, I rely on the crosswalk built by Gianluca Tarasconi to match these two sources.[1] This leaves me with around one third of all the patents registered between 1975 and 2021, due to the restriction of the time frame to 1976-2016 as well as an incomplete match between PATSTAT and PatentsView. Patent data also provides my with citation data, which I use to build self-citations as well as truncation-corrected forward citations and patent generality, following the procedure in **?** and **?**. I restrict my attention to utility patents, since I am interested in patents with a technological content, and not just design improvements.

My main data source for concentration and sales data is given by the US Economic Census (EC), which reports sales shares for the top 4, 8, 20, and 50 firms, Herfindal-Hirschman Index, sales and number of companies in various NAICS 4-digit sectors at a 5 year frequency. I restrict my attention to the period between 1997 and 2012 for three main reasons. First, as I show below, this period saw a substantial increase in the concentration of inventors in specific technology classes. Second, the start

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

of this period coincides with an acceleration in the growth of market concentration and markups (see, e.g., **?**). Third, 1997 marks the adoption of the NAICS classification, ensuring a consistent definition of product markets throughout the period I analyze, without the need to rely on crosswalk with the pre-existing SIC classification. As my baseline concentration measure, I rely on the HHI lower bound constructed by **?**.[2] This choice is dictated by the fact that the Economic Census reports the HHI only for a subset of industries, which would severely limit my sample. The method proposed by **?** obviates to this issue by constructing the implied lower bound of the HHI implied by top sales shares reported in the Economic Census, which are available for a much wider set of industries than the HHI.[3] While my estimates are robust to using the EC-reported HHI, this choice allows me to obtain more power for my findings as well as generalize them.

The Economic Census also provides me with sector-level growth in output per worker, which constitutes my main measure of productivity growth in the sectors I analyze. I choose this measure instead of multi-factor productivity since the latter is available only for a limited set of sectors, mostly in manufacturing. I deflate sales using NAICS-specific price indices from the BLS.

I employ two additional data sources in the empirical analysis and in the calibration of my model. First, I obtain sector-specific counts of regulations for various NAICS 4-digit sector from the Mercatus RegData 4.0 dataset, which I employ to conduct an instrumental variable analysis, strengthening the causal interpretation of my results.[4] Second, I use NBER-CES data to produce estimates of the Lerner Index following **?**.

Matching my various data sources results in a dataset of 157 NAICS 4-digit sectors over the period 1997-2012, out of a total of 304 business sectors, for which I have data on concentration and that I merge to my estimates of knowledge markets.

## 1.2 Effective Inventors and Knowledge Markets

In order to identify the effect of product market concentration on the allocation of inventors, I first have to establish which set of sectors different categories of inventors can actually choose from. Indeed, pooling all sectors together would inevitably attenuate the response of the inventor distribution to changes in the features of individual product markets, as this design would correspond to the assumption that all inventors could potentially be employed by any sector. Since inventors have limited employment opportunities due to their specific skills set, this assumption would incorrectly suggest that, in a majority of cases, sector characteristics have no impact on the allocation of researchers. That is, the mismeasurement of labor market for inventors would produce an attenuation bias in my

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

[3] As detailed in **?** this measure is very strongly correlated with the HHI reported by the Economic Census when this is available, with a correlation of around 0.93.

[4] Available at https://www.quantgov.org/bulk-download.

coefficient of interest.

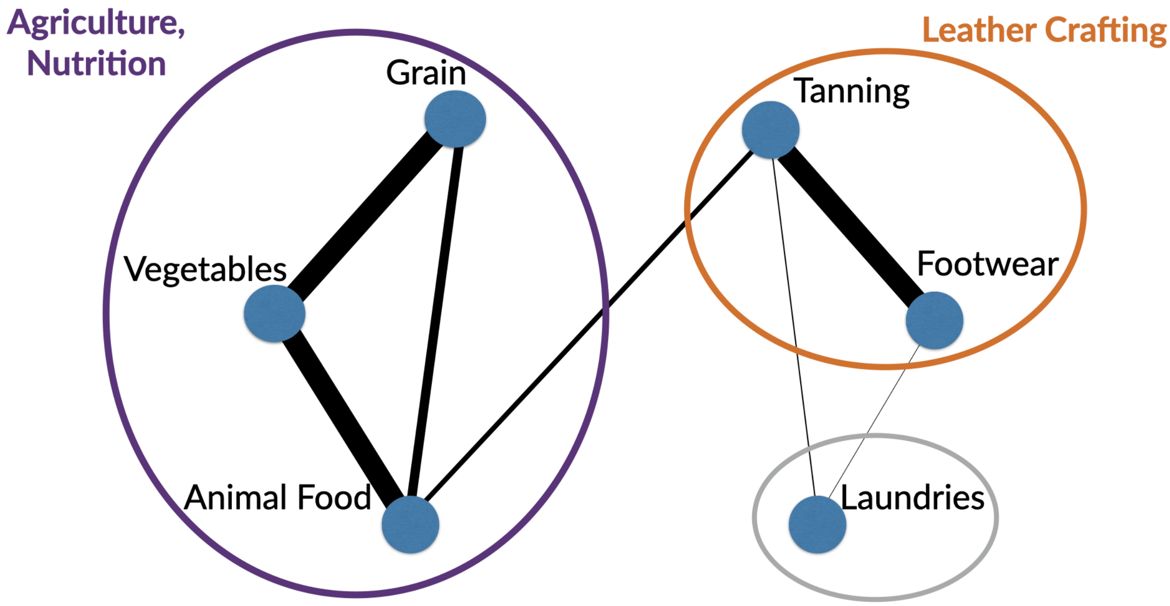
The main aim of this section is correctly grouping product markets that share the same *required knowledge to innovate*, and therefore compete for the same R&D inputs, namely inventors. In practice, I construct transition of inventors across different product markets, and use the resulting network of flows to identify sectors that routinely exchange researchers through a Louvain community-detection algorithm (**?**).

Figure 1 presents a graphical illustration that clarifies my definition of knowledge markets as well as the procedure I follow to construct them. Each node in the Figure represents a different NAICS 4-digit sector, which is connected to other sectors by inventor flows, indicated with black lines. The width of each line represents the strength of these flows. The procedure described in detail in the rest of this section, shows how I define these flows and measure their strength. After obtaining these weighted flows, I employ a community detection algorithm to group together sectors that are most closely connected to each other. In the sectors that I selected for this illustration, as one might expect, there are strong flows between grain, vegetable farming and animal food manufacturing, all of which involve knowledge related to agriculture and nutrition, and separately between footwear and leather tanning, which both require knowledge of leather crafting. In this case, my algorithm would identify two knowledge markets, one given by the agriculture and food manufacturing sectors, and one given by leather crafting sectors, leaving the laundry services sector isolated.

This example makes clear the importance of identifying separate knowledge market to analyze the effects of the evolution of product markets on the allocation of inventors. In this figure, where there are routinely almost no links of the laundry sectors to the others, we would not observe a response of the inventor distribution to changes in the concentration of these sectors, since the other inventors in the economy specialize on different skills set and would therefore not be able to transition there. Analogously, we would expect a higher profitability in footwear manufacturing to attract inventors away from leather tanning, but not from vegetable farming.

**Measuring Inventor Transitions** The first step to construct knowledge markets consists in defining transitions of inventors across sectors. I employ the USPTO patent data classified into 4-digit NAICS sectors by **?** to do so. Table 1 exemplifies the structure of the USPTO dataset matched with this NAICS classification. In this dataset, each patent is listed together with all the inventors listed in the registered patents. Each inventor is assigned a disambiguated ID corresponding to the serial number of the first patent the inventor appears. In this example, inventor 00001-1 and 00001-2 both cooperate on the development of patent US00001. The third column in Table 1 shows an example of the **?** classification for NAICS 4-digit industries. The authors conduct a text analysis to classify each patent into the various NAICS sectors of application of that patent. As the table illustrates, this classification is not limited to a single sector per patent, and includes multiple sectors in almost all instances. Importantly, this

Figure 1: Graphical Illustration of Knowledge Markets




Note: This Figure provides a graphical illustration of the definition of knowledge markets as sets of product markets sharing the same required knowledge to innovate. This illustration is based on transitions of inventors across product markets observed in my data and classified in the same knowledge markets, although many other sectors in these markets are excluded for the sake of exposition. In the figure, nodes represent NAICS sectors 1111 (Oilseed and Grain Farming), 1112 (Vegetable and Melon Farming), 3111 (Animal Food Manufacturing), 3161 (Leather and Hide Tanning and Finishing), 3162 (Footwear Manufacturing), and 8123 (Drycleaning and Laundry Services). The edges connecting nodes represent inventor transitions across sectors, while the width of these edges represents the strength of the connection between the two sectors as measured by undirected inventor flows.

classification captures the *technological nature* of the patent and the sectors of application of the knowledge required to develop that patent. While other classifications, like the CPC or the USPC, also describe the technological nature of patents, they do not allow a direct match to sectors of application without arbitrarily assigning sectors to each of these classes.

Given this data structure, I define a transition in two ways. First, I consider inventor transitions *within patents*. That is, I consider that an inventor transition occurs between two sectors if an inventor works on a patent that applies to both of these sectors. The direction of flows does not matter for the definition of knowledge markets, since I am interested in grouping sectors which exchange researchers and not in the specific direction of these exchanges. Therefore in the case of Table 1, I would say that there have been two transitions between sectors 1111 and 1112 in 1980. Another type of transition that I consider is *across patents*. This transition occurs when an inventor works on different patents that apply to different product markets. An example of such transition in Table 1, is the transition between sector 1112 and 3111 by inventor 00001-1. The raw count of transitions of inventors across sectors in each year constitutes the basis of my measure of inventor flows.

Table 1: USPTO Data Structure

| Patent ID | Inventor ID | ? NAICS | Year |
|---|---|---|---|
| US00001 | 00001-1 | 1111 | 1980 |
| US00001 | 00001-1 | 1112 | 1980 |
| US00001 | 00001-2 | 1111 | 1980 |
| US00001 | 00001-2 | 1112 | 1980 |
| US00002 | 00001-1 | 3111 | 1981 |

Note: This Table reports an example of the data structure employed to build knowledge markets. The columns "Patent ID" and "Inventor ID" represent disambiguated patent and inventor identifiers as classified by USPTO PatentsView Data. The column "**?** NAICS" reports an example of patent classification into NAICS 4-digit sectors. The data reported in this table have pure expository aim, and do not represent actual observations in the dataset.

**Weighting Inventor Flows: Effective Inventors** After identifying transitions, I proceed to weigh them by two alternative measures in order to quantify the flow of inventors across sectors. The first measure simply weigh each transition equally, computing inventor flows as the raw count of researchers moving across NAICS. The second measure adjusts for the productivity of individual inventors, since raw counts might overstate or understate the importance of each transition, depending on the size of origin and destination sectors, their technological nature, as well as the ability of each inventor. I therefore define a measure of "effective inventors" that aims to correct for these and other omitted factors. For each inventor, I estimate the fixed effect, $\alpha_i$, in the fully-saturated regression,

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

where $\#\text{Patents}_{cfit}$ denotes the number of patents registered in CPC class, $c$, firm (assignee), $f$, and year $t$, that include inventor $i$. In this regression $\gamma_{cft}$ denotes a of CPC class by firm (assignee) by year fixed effect. I choose to include indicators for CPC classes at one digit, the broadest classification, in order to identify as many fixed effects as possible. The inclusion of CPC, firm, and year controls corrects for specific technological features of the patented technology, the firm environment, as well as the specific year. Further, this specification produces an estimate of inventor productivity that accounts for the number of collaborators on each patent. Clearly, these fixed effects might be inconsistently estimated, and for this reason I check the robustness of all my results, including the construction of knowledge markets, to the use of the raw count of inventors rather than the inventor's productivity captured by the fixed effect, $\alpha_i$. Given this specification, I define an *effective inventor* as one unit of the resulting fixed effects, rescaled to take nonnegative values.

Armed with the results of this estimate, I define *effective inventor flows* between sector $j$ and sector $k$ at time $t$ as:

$$flow_{j\to k,t} = \sum_i \#\{i\text{'s transitions } j \to k \text{ in } t\} \cdot \alpha_i,$$

that is, the sum of transition counts weighted by effective inventors. The total undirected flow between two sectors is then given by the sum of inflows and outflows with ends in one of the two sectors:

$$flow_{jk} = \sum_t \left(flow_{j\to k,t} + flow_{k\to j,t}\right).$$

This flow measure defines a network of inventor transitions across product markets, where the nodes, $j, k$, are given by 4-digit NAICS codes, edges are given by transitions across sectors, and edge weights are defined as a rescaled version of $flow_{jk}$. I use these edge weights as a measure of the strength of the connection between pairs of sectors in the network. Rescaling the flow measure is necessary in order to exclude effects of sector size as well as to avoid double counting of inventors. I describe how I rescale this series in Appendix **??**.

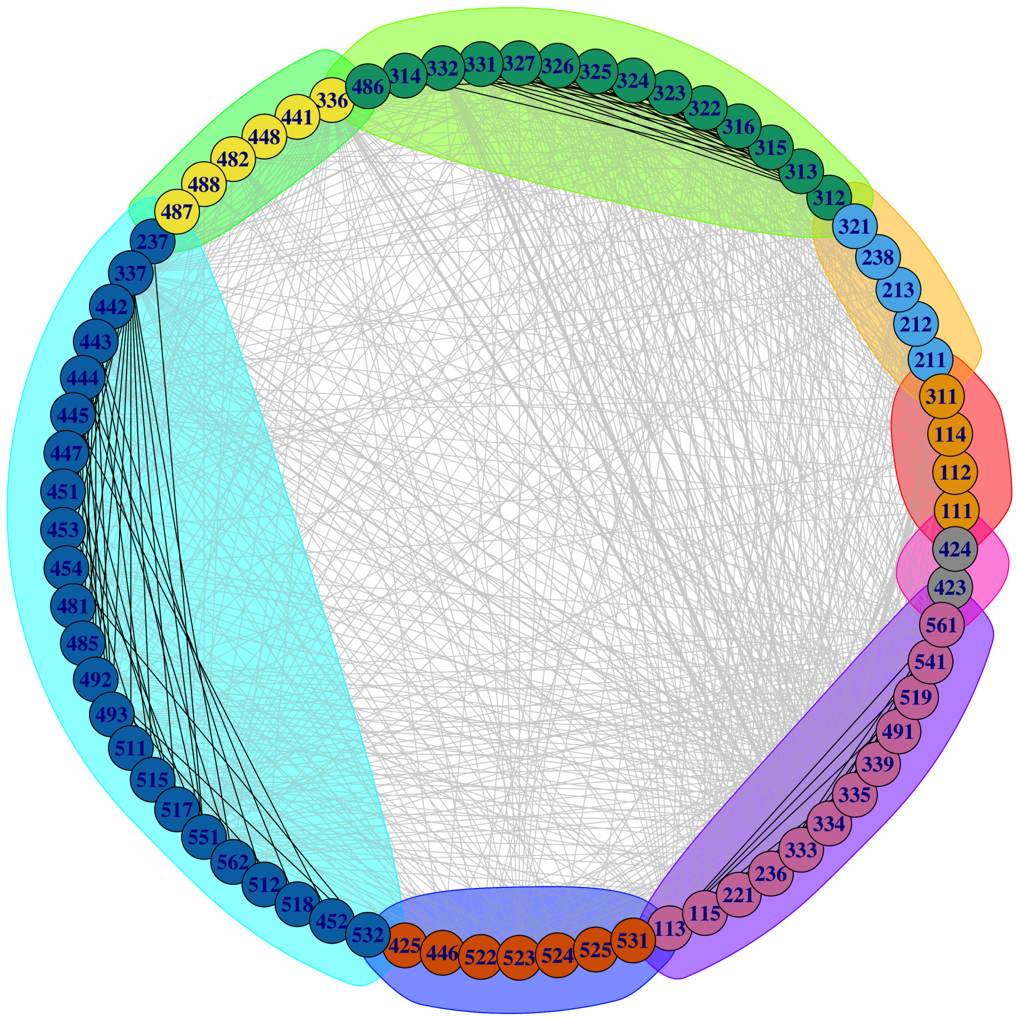
**Community Detection and Resulting Knowledge Markets** I used rescaled the rescaled undirected flow measure as a network edge weight to identify communities through the Louvain algorithm developed by **?**. This procedure maximizes the modularity of the network choosing the number of communities (knowledge market) and the assignment of nodes (NAICS sectors) to communities. Modularity, a commonly used measure of connectedness of networks, measures the distance between the density of links *within* communities versus *between*.

This procedures produces 10 non-singleton sets of NAICS 4-digit sectors that share the same inventors and have non-missing concentration measures. By construction of the community detection algorithm, these knowledge markets do not overlap, so each NAICS 4-digit sector belongs to one and only one knowledge market. Figure 2 displays the result of my procedure applied to NAICS 3-digit sectors. I report this exercise since the 4-digit equivalent would be unreadable. However, the knowledge markets identified by the two exercises are qualitatively similar although they are clearly more numerous in the 4-digit case. In this figure, lines denote inventor transitions, with width proportional to the effective undirected inventor flow between sectors. Nodes represent the NAICS 3-digit sectors reported on each node. Black lines depict flows within knowledge markets, while gray lines represent transitions between communities.

Three features are worth emphasizing. First, the network is very dense, and transitions across 3-digit as well as 2-digit sectors are pervasive, and differ largely in intensity. This points to the relevance of this classification and its stark difference from what could be obtained by grouping sectors based on broad product markets, which would neglect the linkages across disparate markets, as well as pooling all sectors together, which would neglect the difference in the strength of inventor flows. Second, while the flows between communities might seem more numerous than within communities, this is solely a by-product of the circular layout of the network, whereby flows within close communities on the circle are masked by the nodes. When applying the algorithm to 4-digit sectors, I find that less than a third of flows occur between communities, as expected since the community detection algorithm maximizes the density of within-community linkages. Third, and perhaps most importantly, the classification that I obtain is sensible, grouping together sectors that we might expect to share similar knowledge to innovate. Starting from sector 111 and going counter-clockwise, the knowledge markets in the Figure can be described as follows. The first market, including sector 111, groups sectors involving agricultural production (111, 112 and 114) as well as food manufacturing (311). The second market, starting with 211, includes oil, gas, and mining. The green cluster at the top of the figure groups several heavy manufacturing industries, like chemicals plastics and petroleum products, as well as pipeline transportation (486). The market in yellow is concerned mostly with transportation services and manufacturing as well as motor vehicle dealers. The large blue cluster collects a large number of retail sectors, as well as data processing, telecom and broadcasting services. The remaining three markets include insurance and finance (red cluster), computer, electronics, and machinery manufacturing and professional services (violet), and wholesalers (gray).

As noted above, these markets are identifies using my measure of effective inventors, but I assess the robustness of this procedure to using raw counts of inventor transitions. When applying the procedure to 4-digit NAICS sectors, I find that over 97% of the pairs of NAICS sectors are classified in the same manner using the two measures. That is, 97% of sector pairs belong to the same knowledge market according to both measures, and they do not belong to the same market if they do not according to either measure.

Figure 2: Knowledge Markets Obtained from NAICS 3-digit Sectors




Note: This Figure displays the network of inventor flows between NAICS 3-digit sectors and the knowledge markets resulting from the application of the Louvain community detection algorithm. Lines denote inventor transitions, with width proportional to the effective undirected inventor flow between sectors. Nodes represent the NAICS 3-digit sectors reported on each node. Black lines depict flows within knowledge markets, while gray lines represent transitions between communities.

## 1.3 Other Constructed Measures and Aggregation at Census Frequency

**Patent Citation Measures** For each patent classified by **?**, 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.[5]

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.

I also construct truncation-corrected two forward citation measures and a patent generality measures following the definitions and the procedures described in **?**. The forward citation measures compute the average number of citations received by each firm's patents, giving an indication of the importance of each patent for future technological developments. The correction for truncation is conducted estimating the empirical CDF of the forward citations lag distribution of patents in the relevant CPC 2-digit technology class. The correction is then carried out by dividing the overall amount of forward citations at the latest available date by the inverse of the CDF thus obtained. The procedure suggested by **?** uses only information pertaining to the CPC 2-digit technology class of the cited patent. In addition to this measure, I also conduct an alternative correction that estimates a separate distribution for each citing CPC 2-digit class and sums the corrected forward citations across all citing classes. Patent generality also measures the technological impact of patents, but rather than focusing on citations it examines the scope of application of the patent. In particular, it computes a measure of the dispersion of citation received across different CPC classes. The higher is this dispersion, the wider is the technological applicability of the patent.[6]

**Regulation Data** 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

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

[6] The interested reader should consult **?** for a detailed discussion, and the related appendix for details on the construction of these measures.

texts, and then taking the best match for each document. However, the available data does not include a set of codes for data quality 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 attribute the regulations of the most similar sector otherwise. I chose this threshold by inspecting the similarity associated to various NAICS pairs, and the assignment of regulations to sectors is not highly sensitive to this choice.

**Inventor Distribution Measures** I employ the measure of effective inventors constructed as detailed above to compute measures of researchers' concentration within sectors for each year in my sample. Specifically, I use the PatentsView assignee ID to identify firms that employ specific inventors in each sector, and then compute several measures of the concentration of inventors within sectors. I focus in particular on the top 10% and bottom 50% share of inventors, as well as other commonly employed measures of dispersion like the ratio of the $90^{th}$ quantile to the median. I also compute the Gini coefficient of inventors across CPC classes and NAICS 4-digit, assigning effective inventors to the relevant technology class or NAICS sector, to document the trend in increasing concentration of inventors in specific patent classes and sectors.

**Aggregation at 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), while I use concentration measures for the related census year. 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 Empirical Analysis

This section presents four main findings that apply to the period 1997-2012: (i) effective inventors have become more concentrated across economic sectors; (ii) sectors with increased concentration have attracted a growing share of relevant inventor types; (iii) growth in the share of relevant inventors is negatively correlated with inventor productivity, as measured by forward citations as well as average growth in output per worker divided by effective inventors employed; and (iv) 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.

Results (i) and (ii) are indicative of a growing concentration of inventors, and establish a positive causal link between the growth in product market concentration and the increase in sectors' inventor share. Findings (iii) and (iv) provide evidence in favor of misallocation. Inventors' concentration in less competitive sectors turns out to be inefficient, since there researchers are predominantly employed on projects that do not contribute to the growth of the sector, which I interpret as defensive innovation. This interpretation is supported by the concentration of inventors among larger firms, the decline in forward citations received by patents obtained by these firms, and the decrease in growth per inventor that accompanies the increase in product market concentration.

The rest of this section proceeds as follows. The first subsection presents my empirical framework and variable definition. Remaining sections present in order results (i)-(iv) above. I discuss the causal interpretation of my results an IV specification in Subsection 2.2.

## 2.1 Variable Definitions and Main Specification

The main outcomes of interest throughout the analysis refer to measures of inventors concentration or measures of R&D productivity. For the former, I rely on the definition of effective inventors, $\alpha_i$, provided in Section 1.2, that is productivity-adjusted inventors. For each product market, $p$, I define the share of inventors employed by the sector in year $t$ as:

$$\text{Inventor Share}_{p,t} \equiv \frac{\sum_{p(i,t)=p} \alpha_i}{\sum_{k(i,t)=k} \alpha_i},$$

where the sum at the numerator is taken over all effective inventors that are mentioned in patents registered in product market $p$, while the denominator computes the total effective inventors that belong to the knowledge market. Effective inventors, $\alpha_i$, are also the measure I use to evaluate the dispersion of inventors across sectors and technology classes. My results are robust to computing the inventor share using raw counts of researchers instead of effective inventors.

When analyzing R&D productivity I focus on the three patent-based measures described in Section 1.3, that is, forward citations, share of self-citations, and patent generality. Further, I compute a more direct measure of the productivity of inventors given by the growth in output per worker divided by the number of effective inventors employed by the sector.

In most specifications, the independent variables are measures of concentration and controls for the size of the sector considered. As discussed in Section 1.1, my baseline measure of concentration is the lower bound of the Herfindal-Hirschman Index constructed by **?** using top sales share reported by the Economic Census for each sector.[7] I label this variable $\underline{\text{HHI}}_{p,t}$, where the line below stands for the lower bound. This choice is motivated by the small number of sectors in my sample that have an HHI index reported by the Economic Census (about 80). Using the lower bound I can extend my analysis to a total of 157 sectors. Due to the high correlation between the two variables, the results are robust to using the Economic Census HHI, as shown in robustness exercises.

I obtain sales variable from the Economic Census, which I deflate using BLS NAICS-specific price indexes. I use sales variables for two purposes. First, real sales in 2012 as the weight in my regressions. Second, I use the logarithm real sales as well as a quartic in real sales in order to control for changes in the size of sectors during my sample period. For the selected subset of sectors that reports the number of establishments, I also explore the robustness of my findings to controlling for sales per establishment.

Given these definitions, my main specification is a sector-level long-difference regression over the period 1997-2012:

$$\Delta\text{Share}_{p,\,2012-1997} = f_k \mathbf{1}\left\{p \in k\right\} \;+\; \beta\Delta\underline{\text{HHI}}_{p,\,2012-1997} \;+\; \gamma'\Delta\text{Size}_{p,\,2012-1997} \;+\; \varepsilon_p, \tag{2}$$

where $\Delta$Sharedenotes the change in the inventors' share of product market $p$, $f_k \mathbf{1}\left\{p \in k\right\}$ is a dummy variable that takes value 1 if the product market, $p$, belongs to knowledge, $k$, $\Delta\underline{\text{HHI}}$ is the change in the HHI lower bound, and $\Delta$Size is a set of controls for the size of sector $p$, which are given by either the change in the logarithm of real sales or real sales per firm, or the change in the terms of a quadratic polynomial in the real sales of the sector.

Regressions are weighted by sector sales in 2012 for the findings which rely on Economic Census sector-level measures, and I estimate robust standard errors in all specifications. When looking at patent measures, I employ the same specification as 2, where I replace the outcome variable with the change in patent productivity and the dependent variable with the change in inventors' share. In this case, since I do not rely on Economic Census measures, I report unweighted regressions.[8] I also discuss the robustness of these findings to adopting the same specification as 2, using the HHI lower bound and weighting by sales.

[7] The expression used to obtain this measure is:

$$\underline{\text{HHI}}_{p,t} = 4\left[\frac{\text{CR4}_{p,t}}{4}\right]^2 + 4\left[\frac{\text{CR8}_{p,t} - \text{CR4}_{p,t}}{4}\right]^2 + 12\left[\frac{\text{CR20}_{p,t} - \text{CR8}_{p,t}}{12}\right]^2 + 30\left[\frac{\text{CR50}_{p,t} - \text{CR20}_{p,t}}{30}\right]^2,$$

where "CR{X}" denotes the concentration ratio, that is the share of sales, of the top X firms. This measure is a lower bound, and coincides with the actual HHI if the sector has less than 50 firms, and sales share are distributed equally in each of the top 0-4, 5-8, 9-20, 21-50 brackets. **?** reports a correlation of $\underline{\text{HHI}}$ with the actual index of 0.93.

[8] As I will show, the change in the inventor share is highly correlated with the change in the HHI, so this specification essentially amounts to a rescaling of the coefficient that would be obtained using the HHI.

## 2.2 Results

### 2.2.1 Inventor Concentration across NAICS Sectors has Increased

Figure 3 reports the time series of inventor concentration across NAICS 4-digit industries for the period 1976-2016, where the **?** series is available. I construct these series using the share of effective inventors (in panel (a)) or raw inventor counts (panel (b)). I use the HHI index of inventor shares accruing to each sector as a measure of concentration. Both panels display an increasing trend in the concentration of inventors starting from the late 1990s. These patterns align closely with the trends reported in **?**, which document a rising share of patents registered by top firms within sectors. Figure 3 extends this findings to the cross-industry allocation of inventors. Quantitatively, the increase in inventor concentration is sizable, corresponding to about a 20% increase in the HHI for the effective inventor measure over the period 1997-2012. Using raw inventor counts increases this number to 30%.

As for the other results presented below, the effective inventor measure and the raw inventor count behave similarly, albeit the series for raw counts is more volatile and exhibits larger changes. This difference in volatility stems from the regression method implied to obtain effective inventors, which accounts for several features through time, firm, and technology class fixed effects.

Figure 3: Herfindal-Hirschman Index of Effective Inventors across NAICS 4-digit Industries, 1976-2016

[every plot/.append style=line width=2pt, ] [width=6.028in, height=4.754in, at=(1.011in,0.642in), scale only axis, unbounded coords=jump, xmin=1978, xmax=2016, xlabel style=font=, xlabel=Year, ymin=0.039, ymax=0.053, yla

Note: This Figure reports the time series of inventor concentration, as measured by the HHI index of inventor shares across NAICS 4-digit sectors. The left panel reports the series constructed using effective inventors as defined in Section 1.2, the right panel uses instead raw inventor counts. Only the NAICS 4-digit sectors with data for all years are included.

### 2.2.2 Markets with Growing Concentration Increased Their Inventor Share

In this section, I present three sets of results for each specification, which differ 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.[9] “Mahalanobis 5%” denotes the sample where I trim the 5% extreme observations based on the Mahalanobis distance of pairs of

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

observations from the data centroid. Since this procedure is based on the joint distribution of the outcome and independent variable, the sample thus obtained varies in each regression.

Table 2 presents the results of regression (2) 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 discussed above, or the index as reported by the Economic Census. The results in Table 2 highlight a strongly significant positive correlation between the change in the 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 2 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 2, 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. 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, allowing me to extend my findings to a larger number of sectors (about double as can be seen from the number of observations in each specification).[10]

While suggestive, the correlation presented above is far from causal, 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 accruing to a specific sector. Second, it estimates the correlation both across and within knowledge markets. In Table 3, 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 3, 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

[10] Regressions using the Economic Census HHI not reported in the main text or the Appendix are available on request.

of firms, the sample used in the lower panel is smaller than the upper panel. The results in Table 3 confirm the positive relation between the change in inventor shares and concentration, and are largely unchanged relative to the estimates in Table 2, 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 4 depicts graphically the residualized observations underlying the estimated coefficients in Columns (2) and (6) of Table 3, 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 4 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 3(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.

Appendix Table **??** shows estimates using the share of effective inventors of each product market over the total. This amounts to neglecting the fact that inventors flow only across sectors that can employ their skills. In this specification, I find a significant, albeit small, effect of product market concentration on the share of inventors. However, this result only arises when the sample is trimmed to remove outliers. This is unsurprising in light of the discussion in Section 1.2, whereby mismeasuring the labor market for inventors should bias the estimates of inventor mobility towards zero, since many of the sectors would actually not be routinely connected by inventor flows. Finally, this result also conforms with the findings in Table 3, which show that including knowledge-market fixed effects does not alter the coefficients significantly, suggesting that flows across knowledge markets are indeed negligible.

Appendix establishes the robustness of all the findings in this section to the use of raw inventor counts rather than effective inventors to compute both inventor shares and knowledge markets.

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

| | Δ Inventor Share (pp) | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $\Delta \underline{HHI}$ | 27.293* | | 27.183* | | 27.326* | |
| | (11.569) | | (11.941) | | (11.620) | |
| Δ HHI | | 22.399*** | | 22.399*** | | 22.350*** |
| | | (6.345) | | (6.345) | | (6.343) |
| 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 | 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 3: 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

| | Δ Inventor Share (pp) | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $\Delta\underline{HHI}$ | 26.093* | 22.509* | 25.904* | 22.716* | 26.111* | 22.554* |
| | (10.696) | (10.848) | (11.124) | (10.948) | (10.725) | (11.019) |
| Δ log 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) |
| 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 |

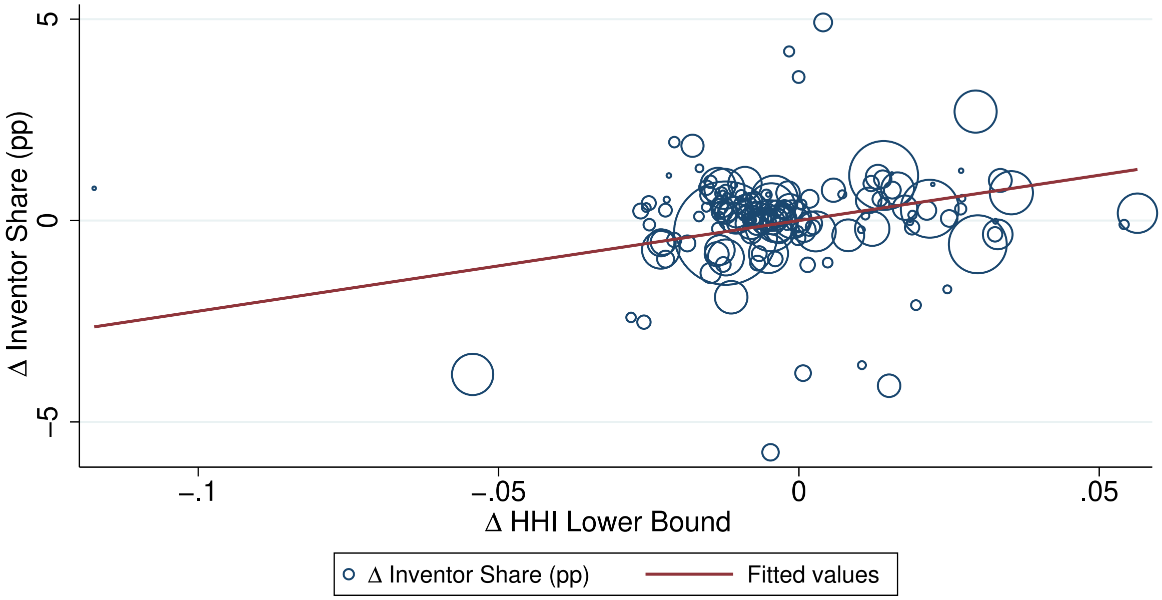
(b) Controlling for Change in Log Real Sales per Company

| | Δ Inventor Share (pp) | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $\Delta\underline{HHI}$ | 35.230** | 20.783+ | 35.230** | 20.783+ | 35.154** | 22.854* |
| | (12.759) | (10.615) | (12.759) | (10.615) | (12.647) | (11.197) |
| Δ log Size | 0.175 | -0.040 | 0.175 | -0.040 | 0.300 | -0.055 |
| | (0.382) | (0.253) | (0.382) | (0.253) | (0.460) | (0.346) |
| 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), 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.

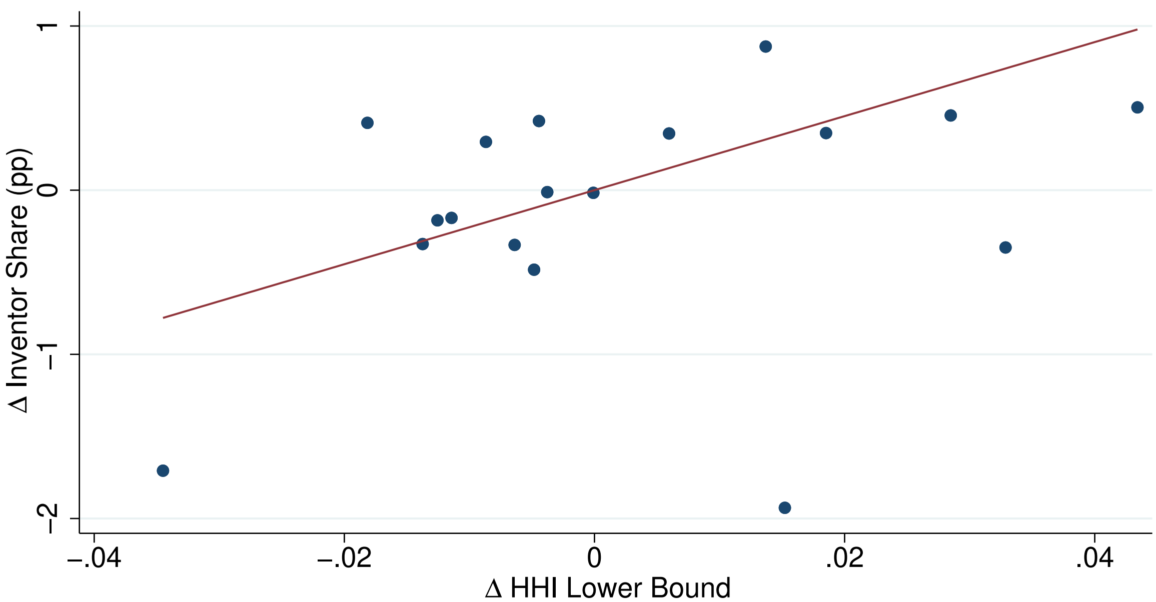
Figure 4: Residualized Scatter Plots Corresponding to Selected Columns in Table 3, Panel (a)

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




Full Sample. Absorbed FE: Knowledge Market. Partialled out: Δ log Sales. Observations weighted by Sales.

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




Mahalanobis 5%. Absorbed FE: Knowledge Market. Partialled out: Δ log Sales. Observations weighted by Sales.

Note: This figure presents residualized scatter plots of the change in the share of effective inventors of sector $p$ over total inventors in knowledge market $k$, over the change in the lower bound of the Herfindal-Hirschman Index for product market $p$, as implied by Census concentration ratios. The upper panel reports the data for the full sample, where both variables are residualized by change in log real sales and knowledge market fixed effects. The size of the markers is proportional to the weight of each observation in the regression (sector sales in 2012). The regression line uses the coefficient on the change in HHI lower bound in Column (2) of Table 3. The lower panel presents a binned scatter plot removing the observations with the highest 5% Mahalanobis distance from the sample centroid. Observations are aggregated using sales weights and the regression line is from Column (6) of Table 3.

**IV Results** I now 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 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 4. The specification is the same as in Column (2) of 3, 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.[11] However, instruments are not irrelevant. The results in the lower panel of Table 4 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%

[11] 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, and the corresponding confidence intervals in brackets.

Taken together, the results presented in this section establish a causal link between the increase in inventor concentration and the shifts in product market concentration across NAICS 4-digit sectors.

### 2.2.3 Sectors that Attracted More Researchers Saw Increasing Top Firms' Inventor Shares and Falling Patent Forward Citations

While the findings presented so far establish a connection between inventor and product market concentration, they do not establish that changes in the distribution of researchers across sectors are inefficient. In particular, it would not be unreasonable to think that more concentrated sectors saw increased entry as a result of the higher rents captured by incumbents. Table 5 shows that the opposite is occurring. Specifically, the share of effective inventors accruing to top inventor-hiring firms has increased in the sectors that attracted more inventors over the period considered, relative to firms with less inventors in the sector. This finding is consistent across a variety of measures displayed in Columns (1) to (6). These findings suggest that inventors have increasingly concentrated among large incumbents, that is, sectors that increased their inventor share also saw a *within-sector* increase in inventor concentration.

Throughout this section, I present results using changes in inventor shares to focus directly on the correlation between inventor transitions and their within-sector distribution. Unless otherwise noted, these findings are robust to using the change in the HHI rather than the inventor share, as should be expected from the strong correlation between these two variables reported in previous tables. For this section, and other patent-based measure, I present robustness results employing this alternative specification in Appendix **??**.

My next finding suggests that this trend might be driven by a rise in defensive innovation, that is R&D aimed at sheltering the incumbents' dominant position and raising barriers to entry. Table 6 shows that inventors' concentration in specific sectors has gone hand in hand with a fall in patents' forward citations in these sectors, a standard measure of the impact of patents and their role in paving the way for further innovations (**?**). The result in Columns (1) and (2) report two different measures of forward citations, that differ in how the series are corrected for truncation. As discussed in Section 1.3, one measure (Column (2)) uses the procedure delineated by **?**, computing the forward citation lag distribution conditioning on the technology class of the cited patent, while the other (Column (1)) also conditions on the technology class of citing patents. Column (3) presents the estimates relative to patent generality, another measure of patent impact, which increases with the scope of application of the patent. The regressions in this table are unweighted since they rely only on patent data, but results are robust to using the HHI as a regressor and weighting by sales. I present results for the full sample,

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

(a) 2SLS Results

| | Δ Inventor Share (pp) | |
|---|---|---|
| | (1) | (2) |
| Δ$\underline{\text{HHI}}$ | 32.426+ | 30.096+ |
| | (16.987) | (15.819) |
| | [4.850, 99.013] | [4.415, 92.104] |
| Δ log Sales | | 0.525* |
| | | (0.247) |
| | | [0.525, 0.525] |
| Knowledge Market FE | ✓ | ✓ |
| Sample | Full Sample | Mahalanobis 5% |
| Weight | Sales | Sales |
| Observations | 157 | 150 |
| First-Stage F | 4.65 | 4.75 |
| Anderson-Rubin p-value | .0298 | .0321 |

(b) First Stage and Reduced Form

| | Δ Inventor Share (pp) | Δ$\underline{\text{HHI}}$ |
|---|---|---|
| | (1) | (2) |
| Δ log Restrictions | 0.478* | 0.016* |
| | (0.220) | (0.007) |
| Δ log Sales | 0.539+ | -0.000 |
| | (0.274) | (0.005) |
| 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. These 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.

as well as restricting to the middle range of changes in inventor shares, which contains more than 90% of the observations. In both samples, I find a highly significant negative relation between changes in inventor shares and the fall in forward citations. The coefficients imply a high semi-elasticity of self citations to changes in the inventor shares, whereby a 1pp increase in the share of inventors leads to an average 0.2-0.5% reduction in forward citations. When dropping extreme observations, I also find a significant decrease in the generality of the patents, indicating that concentrating sectors produce less widely applicable patents. This effect is however not robust to estimating the regression using the HHI as the independent variable.

The fall in forward citations is a first indication of the presence of defensive innovation, aimed primarily at barring entrants from developing new technologies (see, e.g., **?**). In the next section, I show that these patents also appear to produce limited benefits in terms of ensuing productivity growth, as measured by output per worker growth. This provides further support to my interpretation.

Before moving to the results on productivity, I investigate a competing explanation for my findings on output growth. As highlighted by **?** and **?** among others, large incumbents have a strong incentive to focus on improving their own products at the expense of broadly applicable innovation. In the words of the authors, internal and incremental innovation prevails on more radical, external innovations. This mechanism would also imply that an increase in incumbents' share of R&D resources leads to falling innovation productivity. In order to asses the importance of this channel, and in keeping with the analysis in **?**, I use the share of self-citations to measure the extent of internal innovation conducted by firms. Table 7 displays the results pertaining to this measure. All columns use as dependent variable the change in excess log self-citations as defined in Section 1.3. Columns (1) and (2) build excess self-citations correcting for the importance of firms' patents for the CPC group, which reflects the technological classification of the patent. Columns (3) and (4) use the more narrowly-defined CPC subgroups for robustness. Coefficients are mostly non-significant, and turn negative when knowledge market fixed effects are included. Column (3) displays a marginally significant coefficient. However, this result is not robust to using the HHI as regressor and weighting regressions by sales as in the baseline specification. The findings in this table suggest that incremental innovation does changes significantly following changes in sector concentration, reducing the scope for this alternative explanation.

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

| | Δ 90/50 Quantile Ratio | Δ Top 10%/Bottom 50% | Δ Top-50/Bottom-50 Share Ratio | Δ Top 10% | Δ Bottom 50% |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Δ Inventor Share (pp) | 0.211+ | 0.243* | 0.314+ | 0.018** | -0.008* |
| | (0.107) | (0.097) | (0.184) | (0.006) | (0.004) |
| Δ log Sales | -0.100 | 0.328 | 0.147 | 0.026 | 0.005 |
| | (0.122) | (0.294) | (0.316) | (0.020) | (0.007) |
| 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 3 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 6: Regressions of Changes in Forward Citation over 4-digit Knowledge Market Share, Long-Differences, 1997-2012

(a) Full sample

| | $\Delta$ log Citations/Patent (CPC) | $\Delta$ log Citations/Patent (Total) | $\Delta$ Patent Generality |
|---|---|---|---|
| | (1) | (2) | (3) |
| $\Delta$ Inventor Share (pp) | -0.197*** | -0.227*** | -0.004 |
| | (0.044) | (0.051) | (0.004) |
| $\Delta$ log Sales | -0.234* | -0.258+ | 0.008 |
| | (0.112) | (0.148) | (0.013) |
| 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%)

| | $\Delta$ log Citations/Patent (CPC) | $\Delta$ log Citations/Patent (Total) | $\Delta$ Patent Generality |
|---|---|---|---|
| | (1) | (2) | (3) |
| $\Delta$ Inventor Share (pp) | -0.545*** | -0.618*** | -0.025* |
| | (0.113) | (0.137) | (0.012) |
| $\Delta$ log Sales | -0.232* | -0.255+ | 0.008 |
| | (0.109) | (0.146) | (0.012) |
| 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. These Tables presents the results of specification (2), 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%.

Table 7: Regressions of Change in Excess Self-Citations over 4-digit Knowledge Market Share, Long-Differences, 1997-2012

| | Δ CPC group self-citations | | Δ CPC subgroup self-citations | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Δ Inventor Share (pp) | 0.920 | -0.444 | 0.958+ | -0.228 |
| | (0.711) | (1.083) | (0.512) | (0.801) |
| Δ log Sales | -1.841 | -1.954 | -1.456 | -1.674 |
| | (1.925) | (1.988) | (1.326) | (1.279) |
| Knowledge Market FE | | ✓ | | ✓ |
| Sample | Full Sample | Full Sample | Full Sample | Full Sample |
| Weight | | | | |
| Observations | 157 | 153 | 157 | 153 |

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), when the outcome is the change in excess self-citations in sector $p$ over the change in the share of inventors employed in sector $p$.

### 2.2.4 Markets with Growing Inventor Shares Experienced a Fall in Inventor Productivity

Table 8 presents the results of running regression (2) 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 Economic Census 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 8 reveals a negative and significant correlation between the increase in the effective inventors' change and inventor productivity, regardless of the independent variable employed and the sample restriction adopted.

The magnitude of estimated coefficients can be grasped considering the scale of the variables and their changes over the sample period. Starting from the upper panel of Table 8, the median change in the share of effective inventors over the period was .014pp, while the measure of effective inventors has a median of 2018.[12] 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 .15pp $(-.005 \times .014pp \times 2018)$. This number increases to . 28pp 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 3(a), and considering a median change in the HHI of 0.002 yields an increase in the share of effective inventors in concentrating sectors of 0.045pp, which implies a fall

[12] 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.

in average labor productivity implied by misallocation of 0.45pp. 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 2.8pp in average 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 around 15% of the observed decrease in output per worker growth in these sectors.

The estimates in the lower panel of Table 8, which uses the HHI instead the change in inventor shares as independent variable, imply even larger growth effects. Using the estimates in Column (2), a median HHI change of 0.02 and a median number of effective inventors of 1421 in sectors with growth in inventor shares implies a $-0.78$pp change in output per worker growth from misallocation, with a confidence interval ranging from $-0.13$ to $-1.45$pp. The midpoint of these estimates would explain 27% of the observed fall in output per worker growth over the sample period, with bounds ranging from around 5% to about 50%.

This last set of results provides further support to the hypothesis that defensive innovation increased in concentrating sectors. Indeed, a central feature of this type of R&D activity is that it does not produce substantial growth, since it is aimed at hampering innovation projects by potential entrants. In this sense, patents are registered in order to prevent others from doing so and implementing inventions that would threaten the incumbents' dominant position. This interpretation also resonates with the finding in **?**, who note that incumbent firms tend to register a large number of patents, but account for a small share of overall innovations; a finding they also interpret as evidence of defensive innovation.

Table 8: Regressions of Changes in Inventor Productivity over Changes in Inventors' Share and HHI, Long-Difference, 1997-2012

(a) Change in Inventors' Share as Independent Variable

| | Δ Growth/Inventor (pp) | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Δ Inventor Share (pp) | -0.007** | -0.005* | -0.007** | -0.005* |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| Δ log Sales | | -0.051* | | -0.054* |
| | | (0.021) | | (0.021) |
| Knowledge Market FE | ✓ | ✓ | ✓ | ✓ |
| Sample | Full Sample | Full Sample | Mahalanobis 5% | Mahalanobis 5% |
| Weight | Sales | Sales | Sales | Sales |
| Observations | 101 | 101 | 96 | 93 |

(b) Change in HHI as Independent Variable

| | Δ Growth/Inventor (pp) | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $\Delta \underline{HHI}$ | -0.332** | -0.292* | -0.332** | -0.290* |
| | (0.113) | (0.123) | (0.114) | (0.126) |
| Δ log Sales | | -0.052* | | -0.053* |
| | | (0.021) | | (0.022) |
| Knowledge Market FE | ✓ | ✓ | ✓ | ✓ |
| Sample | Full Sample | Full Sample | Mahalanobis 5% | Mahalanobis 5% |
| Weight | Sales | Sales | Sales | Sales |
| Observations | 101 | 101 | 98 | 94 |

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 3 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. The upper panel presents estimates when the independent variable is the change in the share of inventors accruing to a sector, while the bottom panel uses the change in the lower bound of the HHI index.