FEDERAL RESERVE BANK *of* ATLANTA WORKING PAPER SERIES

Patents to Products: Product Innovation and Firm Dynamics

David Argente, Salomé Baslandze, Douglas Hanley, and Sara Moreira

Working Paper 2020-4
April 2020

Abstract: We study the relationship between patents and actual product innovation in the market, and how this relationship varies with firms' market share. We use textual analysis to create a new data set that links patents to products of firms in the consumer goods sector. We find that patent filings are positively associated with subsequent product innovation by firms, but at least half of product innovation and growth comes from firms that never patent. We also find that market leaders use patents differently from followers. Market leaders have lower product innovation rates, though they rely on patents more. Patents of market leaders relate to higher future sales above and beyond their effect on product innovation, and these patents are associated with declining product introduction on the part of competitors, which is consistent with the notion that market leaders use their patents to limit competition. We then use a model to analyze the firms' patenting and product innovation decisions. We show that the private value of a patent is particularly high for large firms as patents protect large market shares of existing products.

JEL classification: O3, O4

Key words: product innovation, patents, creative destruction, growth, productivity, patent value

https://doi.org/10.29338/wp2020-04

The authors are grateful to Antonin Bergeaud, Pete Klenow, and Jesse Perla for very helpful discussions. They also thank Ufuk Akcigit, Fernando Alvarez, Toni Braun, Benjamin F. Jones, Claudio Michelacci, Juan Rubio-Ramirez, and numerous seminar and conference participants at Harvard Business School, the University of Austin, Chicago Fed, Atlanta Fed, Bank of Portugal, EIEF, EPFL, Northwestern, Rochester, the University of Lugano, and the University of Milan. They also thank participants in the CEPR Symposium, NBER SI Innovation, NBER SI Economic Growth, NBER Productivity Group, SED (Mexico), and Midwest Macro. Researcher(s)' own analyses were calculated (or derived) based in part on data from The Nielsen Company (US) LLC and marketing databases provided through the Nielsen data sets at the Kilts Center for Marketing Data Center at the University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researcher(s) and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. The views expressed here are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the authors' responsibility.

Please address questions regarding content to: David Argente, Department of Economics, Pennsylvania State University, 606 Kern Building, University Park, PA 16801, dargente@psu.edu; Salomé Baslandze, Research Department, Federal Reserve Bank of Atlanta and CEPR, 1000 Peachtree Street NE, Atlanta, GA 30309-4470, salome.baslandze@atl.frb.org; Douglas Hanley, Department of Economics, University of Pittsburgh, 4510 W. Posvar Hall, 230 S. Bouquet Street, Pittsburgh, PA 15260, doughanley@pitt.edu; or Sara Moreira, Kellogg School of Management, Northwestern University, 2211 Campus Drive, Evanston, IL 60208, sara.moreira@kellogg.northwestern.edu.

1 Introduction

Product innovation – the introduction of new and improved products to the market – is a key contributor to economic growth and a central element of endogenous growth models (Romer, 1990; Aghion and Howitt, 1992). However, the paucity of detailed data about the introduction and quality of new products has led researchers to use other metrics to measure innovation. As a result, patents have emerged as the primary metric of innovation, especially after comprehensive data sets with information about their timing and characteristics were made readily available (Griliches, 1981).

Many great inventions like solar panels and liquid-crystal displays (LCD) have been patented. Yet no patents have been filed for other inventions that have transformed our lives in important ways like the World Wide Web and the magnetic strip behind modern-day credit cards. In other cases, firms file patents that never turn into new products in the market.[1] These examples suggest that the relationship between patents and innovation is complex and that patents are a crude measure of innovation. Patents are also a protective tool that firms can use to preempt competitors from entering their product market space. This protection is especially advantageous for large market leaders because of greater incentives to defend their existing lead (Gilbert and Newbery, 1982; Jaffe and Lerner, 2004). While the protective role of patents is important in shaping firm dynamics and industry competition, we lack large-scale systematic evidence on the nature of the relationship between patents and product innovation.

In this paper we use textual analysis to create a unique data set that links patents to products of firms in the consumer goods sector. We use these data to study the relationship between patents and actual product innovation, and how this relationship varies with firms' market share. After documenting a set of new facts, we complement our empirical analysis with a theoretical model of the patenting and product innovation decisions of firms.

Our key empirical findings can be summarized as follows:

- Fact 1: More than half of product innovation comes from firms that do not patent.
- Fact 2: On average, patents are positively associated with subsequent product innovation by firms.

[1] In fact, in recent years, patenting activity has skyrocketed whereas innovation and productivity growth have not (Bloom, Jones, Van Reenen and Webb, 2017).

- Fact 3: Larger firms have lower product innovation rates (quantity and quality), but file more patents for each new product.
- Fact 4: Patenting by larger firms is strongly associated with an increase in sales above and beyond the patents' effect on product innovation.
- Fact 5: Patenting by larger firms is associated with a decline in product innovation by competing firms.

There are two main challenges in studying the relationship between patents and product innovation. First, while patent data are broadly available, measures of product innovation in the market are rarely available at large scale. To address this challenge, we use comprehensive data for firms and products in the consumer goods sector collected from Nielsen-Kilts point-of-sale systems in retail locations. This data set includes detailed information about the characteristics of each consumer-goods product sold from 2006 to 2015, along with sales and price information. We exploit this rich data set to construct measures of product innovation. Our simplest measure is the number of new products (barcodes) introduced at the firm and product category level in a given year. Since many new products represent only minor innovations relative to existing products, we also construct measures of the quality-adjusted number of new products. We infer quality improvements by tracking the new attributes (e.g. formula, style, content) that a product brings to the market and by exploiting variation in product prices and sales.

The second challenge involves linking product innovations to their respective patents. We address this challenge by developing two distinct matching procedures. The first procedure maps each firm's patents to its full product portfolio using the names of firms in the patent and product data sets. This results in a yearly firm-level data set (Match 1). The matching procedure in this step is simple and parsimonious, but it is too coarse. Most firms in our data are active in several product categories and could be patenting products in some product categories and not in others, so we need a more granular procedure to match patents with products. We leverage the richness of the information about product and patent characteristics in our data and use modern methods from the field of natural language processing and information retrieval to link firms' patents with sets of its products (Manning et al., 2008). For this match, we first define product categories – sets of similar products – by applying clustering analysis to the short product descriptions included in the Nielsen data extended with text from Wikipedia articles about the products. We then analyze the text of patent applications and assign each patent to the product category with which it has the highest text similarity.[2] This classification of firms' products and patents into various product categories of the firm results in our benchmark patent-to-products data set at the yearly firm × product category level (Match 2).

The resulting granular data set tracks patents and products for firms in the consumer goods sector. The patenting intensities and product introduction rates of these firms are, on average, comparable to those of other manufacturing sectors. Out of 35 thousand firms covered in our data set, 15% applied for a patent at least once (9% applied during the period covered by Nielsen). This value is in line with that of the manufacturing sector and is substantially higher than that of other sectors in the economy (Graham et al., 2018). The consumer-goods sector also covers a wide range of product categories with distinct patenting intensities. The share of patenting firms varies from zero in some food categories to more than half in printers or water purification products.

We begin our analysis by documenting that never-patenting firms account for a large share of product innovation. Over our sample period, they introduced more than 54% of new products and more than 65% quality-adjusted new products. These shares are larger if we rely on the patent-to-products link at the firm × category level. These statistics are corroborated by similar statistics about sources of growth in the sector. We decompose the 10-year sales growth of the sector into growth coming from patenting and non-patenting firm × categories, and find that although non-patenting firms are smaller, they account for 58% of growth in the sector.

Nonetheless, we find that patenting is positively associated with product innovation both at the extensive margin – when firms switch to patenting – and at the intensive margin. Firms introduce more and better-quality products around the time of a patent application, with the largest correlation in the year following the application. Exploiting our matched firm × category-level data over time, which allows us to control for product category-specific trends and firm-category specific effects, we find that the elasticity of product innovation to the number of patents filed a year before ranges from 0.02 to 0.04. We find similar patterns when we focus our attention on granted patents or on patents that receive many forward citations, but not when we consider non-granted patents or uncited patents. This evidence suggests that commonly used measures of the quality of patents are informative about product innovation rates. These elasticities are instructive in the context of various

[2] Younge and Kuhn (2016) and Kelly et al. (2018) use similar techniques when evaluating textual similarities between patents.

policies meant to encourage innovation.[3] The evaluation of these policies often relies on the estimated elasticity of patents to R&D inputs. However, by and large, patents are not the main policy target of such policies – innovation is. Hence, to study how policies encouraging R&D affect product innovation, for instance, one needs to take into account not only the R&D-to-patents elasticity, but also the patents-to-product innovation elasticity.

The importance of patents for protecting firms' products and deterring competition draws attention to the question of how different firms use this strategic role of patenting. Gilbert and Newbery (1982) and Blundell et al. (1999) suggest, for instance, that market leaders have greater incentives to use preemptive patenting to protect their market lead. Survey results from Cohen et al. (2000) report that the motives behind large firms' patenting often go beyond the direct commercialization of patented innovations and extend to strategic deterrance of rivals. For this reason, a major focus of our paper is to understand how the relationship between patents and product innovation changes with a firm's market leadership.

Using variation across firms within product categories, we estimate that firms at the bottom quintile of the size distribution in a given year, as measured by total sales in a product category, introduce one new product for every five existing products in their portfolio, on average. Firms at the top quintile of the size distribution, on the other hand, introduce one new product for every seven existing products in their portfolio. Though larger firms' innovation rates are lower, they are patenting more intensively. But the patent filings of larger firms have significantly weaker association with their product introduction. Moreover, the average quality improvements of new products decline more steeply with firm size than the rate of product introduction does. Overall, these empirical patterns indicate that the disconnect between patent-based measures of innovation and firms' actual product innovation in the market is bigger for firms with large market shares.

Our results suggest that the main role of patents for these market leaders is to constrain product innovation of competitors and thereby protect sales of their existing products. First, we find that patents filed by market leaders carry a larger revenue premium, even after controlling for the quantity and quality of new products these firms introduce. To the contrary, for smaller firms the revenue premium is fully accounted for by product innovation associated with these patents. Second, we show that patent filings by market leaders are associated with a decline in competitors' product introduction in shared product categories.

[3] For example, R&D tax incentives and subsidies as in Dechezlepretre et al. (2016) or Akcigit et al. (2016a).

The same is not true if we consider patent filings of smaller firms.[4]

In addition to the empirical analysis, we consider the patenting and product innovation decisions of firms in a model. The model builds on quality-ladder models that feature creative destruction (Aghion and Howitt, 1992), but it allows for separation between the decision to innovate and the decision to patent – a distinction we can discipline with the data set we have constructed. In the model, both innovation and patenting are costly activities. Introducing higher quality products increases a firm's profit, while patenting decreases the firm's chances of being displaced by entrants. The model can replicate key empirical facts from our data. Larger firms (market leaders) shift their product innovation towards protective strategies, meaning an increase in the number of patents by large firms restricts competition and innovation and does not translate into higher consumer welfare. We use the model to provide a back of the envelope calculation for the private value of a patent and to decompose it into its protective and productive components.[5] The productive component represents the option value of incorporating a patented idea into higher-quality products in order to gain additional profits. The protective component represents the gains for the firm from impeding creative destruction by competitors. After calibrating the model to our data, we estimate that 43% of the average patent value is accounted by the protective component of the patent, and this share increases substantially with firm size.

Our new data set that combines information on product innovation and patents contributes to our understanding of the usefulness of patent statistics for measuring innovation. In the absence of direct measures of innovation, the literature has relied on indirect inference approaches using data about employment growth (Garcia-Macia et al., 2019) or valuing innovation from patent statistics themselves (e.g., Akcigit and Kerr, 2018). Other researchers have looked at innovations that occur outside of the patent system by examining the number of new books on technical topics (Alexopoulos, 2011) or innovations featured at World Fairs between 1851 and 1915 (Moser, 2012). While we document an overall positive relationship between patents and product innovation, we highlight that the usefulness of patent metrics in inferring innovation significantly hinges on the size of the firms that own the patents.

[4] Various studies have analyzed the effect of patenting on follow-on innovation by other firms: for example, see Williams (2013), Heller and Eisenberg (1998), Sampat and Williams (2019) for biomedical research; Cockburn and J. MacGarvie (2011) for the software industry; and Lampe and Moser (2015) for a more general discussion.

[5] This decomposition is possible because we directly observe both the sales from products linked to patents as well as the behavior by competitors. The previous approach in the literature to infer the (total) monetary value of a patent using surveys, samples of patent sales, or patent renewals is discussed in Section 6.

The findings in this paper contribute to our understanding of firms' growth strategies. Recent studies have shown that large firms rely on other protective strategies such as acquiring potential competitors (Cunningham et al., 2018) or forging political connections (Akcigit et al., 2018) as they slow down on innovation (Akcigit and Kerr, 2018; Cavenaile and Roldan, 2019). We suggest that patenting is yet another protective tool that firms substitute for actual product innovation as they grow.

Additionally, our findings regarding the patenting and innovation decisions of firms can potentially speak to several puzzling macroeconomic trends in recent data: patenting is soaring, but productivity growth is stagnating (Gordon, 2016; Bloom et al., 2017); large firms funnel more resources into intangible capital – including intellectual property, but this is manifested in the increasing dominance of those firms instead of perceptible improvements in aggregate innovation in the economy (Crouzet and Eberly, 2019). Our results show that the incentives of large incumbents to direct their efforts towards productive rather than protective strategies may be limited, which is particularly relevant as more economic activities are reallocated towards firms with a large degree of market power (De Loecker and Eeckhout, 2017; Autor et al., 2017; Gutiérrez and Philippon, 2017; Akcigit and Ates, 2019).

The rest of the paper is organized as follows: in Section 2, we present a description of the data sets along with our data matching procedures. We also discuss the validation exercises and present summary statistics. In Section 4, we explore the relationship between patents and product innovation. Section 5 explores the role of firm size. Section 6 presents the theoretical framework and the patent value calculation. Section 7 concludes.

2 Patent and Product Data

2.1 Overview

We face two main challenges in our study of the relationship between patents and product innovation. First, while data about patents are broadly available, information about the introduction of new products is rarely available at large scale. Second, we need to create a link between patents and new products. This section overviews the empirical strategies we use to address these challenges.

We construct a data set about product introduction by beginning with product-scanner data that cover the product portfolio of firms in the consumer goods sector between 2006 and 2015. This data set allows us to identify new products by their barcodes and to observe their characteristics in detail from which we can compute various measures of innovation for a large representative sector. We draw patent information from the United States Patent and Trademark Office (USPTO). The combination of these two data sets gives us information about patents and product innovations covering a large sector of the economy.

To address the second challenge of linking patents to products, we develop several matching procedures. We begin by using the names of the firms in the patent and product data sets to produce a mapping between firms' patent portfolios and their respective products. We refer to this firm level data set as Match 1. This matching procedure is simple and parsimonious, but is too coarse to allow us to connect patents with specific products.

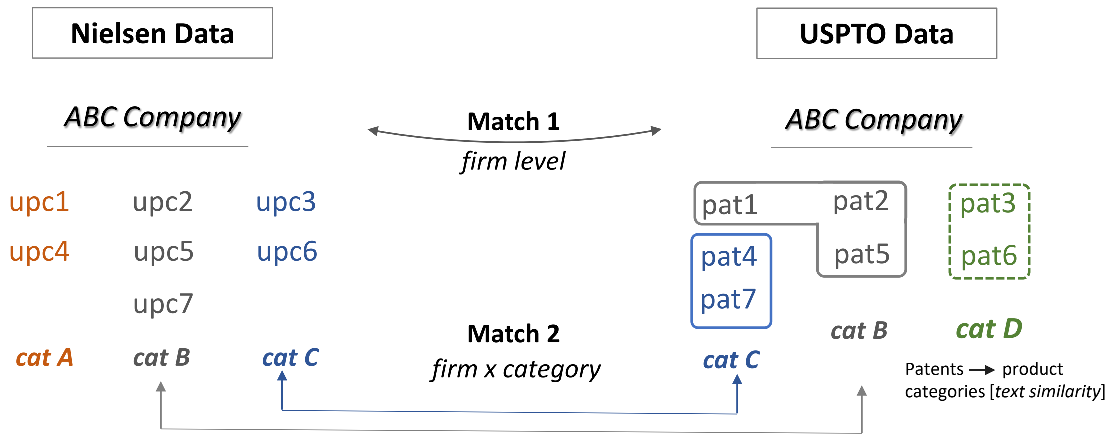
In turn, our second matching procedure leverages the richness of product and patent characteristics using methods from the natural language processing literature to create systematic links between sets of patents and sets of products within a firm. A patent may generate no products or multiple products, and a product may have benefited from multiple patents or from none at all. Therefore, forcing a one-to-one matching between a specific narrowly defined product and a specific patent is neither possible nor desirable.

Hence, our approach is to first define product categories as sets of similar products, which are identified using clustering analysis of product descriptions extended with Wikipedia-based dictionaries. We then assign each specific patent to the product category with which it has the highest text similarity. This classification of a firm's products and patents into the various product categories offered by that firm yields our benchmark patent-to-products data set, which we will refer to as Match 2. Figure 1 illustrates our data schematically, and our matching algorithms are described in detail below.

2.2 Data

Product Data.—Our primary source of product information is the scanner data set from Nielsen Retail Measurement Services (RMS), provided by the Kilts-Nielsen Data Center at the University of Chicago Booth School of Business. This data is collected from point-of-sale systems in grocery, drug, and general-merchandise stores. The original data set consists of more than one million distinct products identified by Universal Product Codes (UPCs), which are scanned at the point of sale. Each UPC consists of 12 numerical digits that are uniquely assigned to each product, and we use these to identify products. UPCs carry information about the brand and a rich set of product attributes like its size, packaging, formula, and flavor.

Figure 1: Product and Patent Data Sets



Notes: This diagram exemplifies the construction of the two data sets linking products and patents. In this example, under Match 1, all products of a firm with name "ABC Company" match to all the patents with assignee name "ABC Company". Under Match 2, upc2, upc5, and upc7 match to pat1, pat2, and pat5 under product category B; upc3 and upc6 match to pat4 and pat7 under product category C; upc1 and upc4 of category A do not match to any patents of the firm; pat3 and pat6 of category D do not match to any products of the firm.

The data focuses on the consumer product goods (CPG) sector, which accounts for 14% of the total consumption of goods in the U.S.[6] The Nielsen RMS dataset covers about 40% of the CPG sector, and nearly covers the universe of firms and product introductions in the sector (Argente et al., 2020). Our sample period covers the years from 2006-2015, and combines all sales, quantities, and prices at the national and annual levels. We use the panel structure of each UPC to measure its entry year. Appendix A.1 provides more detail about types of products and store coverage represented in our sample.

Patent Data.—Our main source of data for patent analysis is the USPTO data on the universe of published patent applications, granted or not. We use the original bulk data files provided by USPTO's Bulk Data Storage System for our analysis. Our sample initially contains information on more than 7 million patent applications filed by more than 500 thousand patent assignees in the years 1975-2017. For each patent, we use information about the patent application year, patent status (granted, pending, or abandoned), patent technology classifications, forward patent citations received, the number of claims on a patent, and whether it is a utility or design patent. For our textual analysis of patent documents, we extract patent titles, the text of patent abstracts, the text of corresponding patent classification titles, claims text, and the titles of citing patents. Appendix A.2 gives more detail

[6]This sector includes non-durables (also known as consumer packaged goods) and semi-durable goods. It excludes consumer durables, producer intermediates, and producer capital.

about our sample and the variables we use.

2.3 Matching Firms

In our firm level data set (Match 1), we match patents to products at the firm level using the firm names in both patent and product data sets. To match firms to patents, we obtain the firm names for each product using UPCs and data set from the GS1 US, which is the single official source of UPCs. This data set links barcodes with the names of firms that sell the product. For the patent data, we begin with the assignee name(s) of each patent. This name is typically the original assignee of the patent and may not represent the current user of the patent because of sales or company reorganizations. We combine the USPTO patent re-assignment data with Thomson Reuters Mergers & Acquisition data to re-assign each patent to its most current holder. This step relies on the assumption that when a firm acquires (or merges with) another firm, the new firm will own all patents that the firms owned before the acquisition (merger). The details of these steps are described in Appendix A.1 and A.2.

A firm's name could be formatted or abbreviated differently in the product and patent data sets, or it may even be misspelled, which presents a challenge in joining information from the two data sets. We developed a name-cleaning algorithm to clean and standardize the firm names to overcome this challenge. This procedure builds on and extends cleaning algorithms from the NBER Patent Data Project (Hall et al., 2001) and Akcigit et al. (2016b) and is described in detail in Appendix A.3.

2.4 Matching Patents to Product Categories

In this section, we describe the details of the algorithm we used to build Match 2 (Firm $\times$ Category level). The algorithm has three crucial steps. The first step creates product categories at a level of aggregation such that they collect distinct and sufficiently large sets of similar products that are meaningfully related to a distinct set of patents. This step yields a set of product categories, a vector of terms used to describe each product category, and a mapping of products into categories. In the second step, similarity scores between patents and product categories are computed. We use various text descriptions to build a vector of terms that describe each patent. We then compute similarity scores between each patent and every product category. These scores represent the overlap between the texts in patents and the text associated with each product category. The final step of our patent-product matching algorithm consists in using the similarity scores and information about the production of the respective patenting firms to generate a procedure that systematically classifies each patent into a product category.

Defining Product Categories.—We define product categories by exploring the product classification scheme used by Nielsen. In the original data, each product is classified into one of 1,070 detailed product modules. These product modules are further aggregated into a set of 114 product groups, and those are further aggregated into ten departments. For example, "disposable cups" and "disposable dishes" are two distinct product modules that are part of the group "paper products" which is part of the department "non-food grocery". Nielsen's modules aggregate products that are close in their technological characteristics. However, there are some sets of distinct modules that have very similar products. At the same time, many Nielsen's groups include products that are quite distinct. For example, "disposable cups", "disposable dishes", "pre-moistened towelettes" and "paper napkins" are all part of the group "paper products", but only "disposable cups" and "disposable dishes" are technically similar. Hence, we seek an intermediate categorization of products – more aggregated than modules and less aggregated than groups – to be able to meaningfully associate patents to a well-defined set of products.

To this end, we apply a clustering procedure to aggregate the Nielsen modules into distinct product categories. Each module is paired with a vector of descriptive words that are weighted by their importance. We expand short module descriptions from Nielsen data with the text of hand-collected Wikipedia articles to get to the comprehensive description of product content of the modules. The resulting vectors collect all the words from the Wikipedia and Nielsen texts, after applying standard parsing and lemmatizing algorithms. When building vectors of words, one must appropriately weight terms by their importance. We use the leading approach in textural analysis – the "term-frequency-inverse-document-frequency" sublinear transformation – that accounts for both the frequency with which a term appears describing a module and how commonly it is used to describe other modules.

The method ensures that we under-weight common words that appear in many documents as these are less diagnostic of the content of any individual document (Aizawa, 2003). We then aggregate these module vectors into clusters using a popular technique known as k-means clustering (Lloyd, 1982). This procedure allows one to specify the desired number of clusters and yields a clustering assignment that minimizes the within-group vector variance. We use an aggregation of modules into 400 product categories as the baseline, and find that this partition strikes a balance between aggregating very similar products while maximizing the difference between products across categories. Appendices A.4.1 and A.4.2 provide extensive descriptions of methods we have taken from the literature on natural language processing, including the details of clustering, quality assessment, and alternative methods to encourage robustness.

After defining the level of aggregation, we build vectors of terms describing each product category. We use the same methodology that we used to build the vectors of modules, but now we use the titles of the clustered module(s) and all the text from their corresponding Wikipedia articles. We ensure that when a product category aggregates across different modules, we first vectorize each module and then average these vectors together so that we do not overweight longer entries. The final product category vectors are normalized to have unit length.

Patent Vectors and Similarity Scores.—This subsection describes how we measure the amount of overlap between the texts of patent applications and product categories. For the text from patent applications, we use the title, abstract, international patent classification system description, and the titles of cited patents. We create vectors of terms by concatenating all fields into one document, followed by the same parsing and lemmatizing algorithms. As before, we adjust the weights of each term according to the "term-frequency-inverse-document-frequency" sublinear transformation and normalize patent vectors to have unit length.

Finally, we construct a similarity score for each patent p and each product category j by computing the cosine similarity between two normalized vectors, $s_{jp} = f_j \times f_p$. This similarity score is guaranteed to be in the range $[0, 1]$ with zero indicating no word overlap and one indicating that the documents are identical. Appendix A.4.3 provides technical description of this step.

Classifying Patents into Product Categories.—The final step of our patent-products matching algorithm consists in using the similarity scores and determining which are valid matches. We must, however, make some adjustments because we use all patents of each firm with products in the consumer goods sector and some patents may relate to goods outside the consumer goods sector or correspond to more general process/method patents. Hence, we should allow for the possibility that a patent will not be assigned to any product category. After an extensive review of patent texts and a great deal of testing, we identified systematic adjustments to the algorithm that ensure that irrelevant patents remain unmatched with products.

We first adjust the algorithm to include a similarity score threshold. We tested different threshold levels and, in our baseline algorithm, we restrict the set of potential categories for each patent p to the product categories whose similarity score exceeds 0.025. The idea is that patents with low text similarity are unrelated to the product categories that we consider. The implication of this adjustment is that patents whose highest similarity are below that threshold are more likely to be classified as "non-matched".

Second, we use information about the set of product categories sold by the firm. For each patent, we define the set of potential matches, whose elements consist of all product categories in which the patenting firm ever sold a product, according to our product data. Together, these criteria imply that patent p will be classified as unmatched if no product categories satisfy the thresholds and the production conditions. For the patents that have more than one product category satisfying those conditions, we assign the final patent-product category match so that it fits into the product category with the highest similarity score.

Our methodology assumes one product category match for each patent. However, some patents may be more general in nature so that they relate to multiple categories. Our baseline algorithm abstracts from this possibility. However, our procedure to define product categories is designed to ensure that the product categories would encompass a broad range of products that are technically similar such that one patent plausibly relates to this and only this range of products.[7] In Appendix A.4.4, we present the details of this procedure and all the robustness exercises with which we tested our baseline algorithm.

2.5 Match Statistics and Validation

Table 1 provides statistics of the baseline data used in our analysis. The data set Match 1 (Firm level) includes annual data for all 34,665 firms that sold at least one product in our consumer goods sector data (CPG firms). The raw USPTO patent data covers information from 1975 to 2017, but because our product data only covers from 2006 to 2015, our analysis can only consider annual variation for the period 2006-2015. In this shorter period, the USPTO data includes about 3.4 million patent applications in total, and about 500 thousand patent applications filed by CPG firms. The data set Match 2 (Firm $\times$ Category level)

[7]In this sense, the methodology delivers a many-to-many patent-to-product match, where each patent can be matched to multiple products of the firm.

includes 40% of those patent applications. The remaining 60% of patents, while filed by CPG firms, could not be associated with products in the consumer goods sector.

Table 1: Match Statistics

	Period	
	1975-2017	2006-2015
Number of patent applications		
All assignees in USPTO	7,304,072	3,386,208
CPG firms (Match 1)	1,046,030	505,544
CPG firms in product categories (Match 2)	399,684	190,575
Number of firms		
All CPG firms		34,665
CPG with at least a patent applied in 1975-2017		5,209
CPG with a patent applied in 2006-2015		3,266

Notes: Match statistics for the baseline data sets Match 1 (Firm level) and Match 2 (Firm × Category level). Match 1 is described in Section 2.3 and Match 2 is described in Section 2.4.

We perform an extensive set of validation exercises to evaluate the robustness and quality of our match. Appendix A.5 presents details on these validation exercises, while here we focus on summarizing the most important. We use four main types of validation exercises: manual checks, external validations using online-collected data on patent markings, analysis of the robustness of the algorithm-implied similarity scores and placebo tests, and validation of non-matches.

Manual checks.—We manually checked many of the patent-to-products matches and some examples are listed in Table A.I of Appendix. The table lists 100 patent applications by the top-selling firms in the largest product categories according to Nielsen. One can easily see that the patent titles reflect the product categories to which the patents were assigned. For most patents we analyzed, we found that our manual choices of product categories also coincide with the product categories chosen by our matching algorithm using similarity scores.

Virtual patent markings.—We next use virtual patent markings to validate our matches. Using virtual patent markings, firms may give a notice to the public that their product is patented by publishing their products and the patents protecting them online. Website searches showed that very few firms in our data used virtual patent markings, and even when they did, only a selection of products and patents appeared in the markings. Nevertheless, these data give a unique opportunity for an external validation of our matching algorithm.

For Procter & Gamble (P&G) and Kimberly Clark (KC), we manually collected virtual patent markings from the company websites and mapped them to our product categorization. We then validate our patent-product category matches for these firms against this information. Appendix A.5.2 shows that that the algorithm-selected product categories mostly coincide with the patent-product category mapping from virtual markings.

Robustness of the match and placebo tests.—We evaluate robustness of the product category choice by our matching algorithm to potential small perturbations in the algorithm. For the algorithm to be robust against small changes, we should observe that highest-ranked product categories have substantially higher similarity scores with the patents than lower-rank product categories do. Section A.5.3 of Appendix shows this is the case. Next we verify that we are indeed carving out well-defined neighborhoods in the technological space by matching patents into distinct categories. For that, we compare the actual distribution of similarity scores between patents classified in the same product category versus a placebo group of patents drawn at random. Section A.5.4 of Appendix shows that the distribution of similarity scores between pairs of patents within product categories is indeed very different and first order stochastically dominates that of the placebo group.

Validating non-matches.—In our last step of the algorithm for Match 2, multiple criteria are used to allow for the possibility that some patents filed by CPG firms are not associated with any of the consumer-good product categories. A valid "non-match" can arise for two reasons. First, a patent may relate to goods that the firm may be producing outside the CPG sector; second, a patent may be about a general process or method that does not affect the introduction of new products. In the spirit of Hoberg and Phillips (2016), we use information from publicly traded companies' 10K reports to identify firms whose output is mostly in the consumer-goods sector, and we find that only a minority of their patents are classified as "non-match", contrasting with patents held by firms who mostly sell products outside the consumer goods sector. Next, we follow Bena and Simintzi (2017), and use patent claims to create proxies for process-related and product-related patents. We find that the share of "non-matches" is significantly higher among process patents. These exercises, which are presented in Section A.5.5 of Appendix, offer reassurance that our algorithm successfully filters out patents that are not directly related to the products in our data.

3 Measures of Product Innovation and Patenting

3.1 Product Innovation

Our measures of product innovation are based on the number of products that firms introduce to the market and the quality of those products. We use the product data described above to identify the entry dates of products in the market and their respective characteristics and performance. We create separate measures of innovation for the firm-level (Match 1) and firm×category level (Match 2) data. Our first measure is the number of **new products** of firm i (in product category j) in year t:

$$\mathrm{N}_{ijt} \equiv \sum_{u=1}^{T_{ijt}} \mathbb{1}[u \text{ is entrant}],$$

where product u is sold by firm i in product category j, T_{ijt} is the number of products that firm i sells in j as of period t, and $\mathbb{1}[u \text{ is entrant}]$ is an indicator that takes the value of one if u is a new barcode. This measure is simple and parsimonious but does not distinguish major product innovations from innovations that make relatively minor changes to a product's characteristics. Our second set of measures of **quality-adjusted new products**, deals with this potential drawback by explicitly accounting for differences in characteristics across new products:

$$\mathrm{qN}_{ijt} \equiv \sum_{u=1}^{T_{ijt}} \mathrm{q}_u \mathbb{1}[u \text{ is entrant}],$$

where $\mathrm{q}_u \in [0, 1]$ is a measure of quality that we describe below. Together these two metrics allow us to account for differences in both the quantity and quality of product innovation across firms and over time.

Our baseline measure of product quality aims at capturing differences in novelty and economic impact across new products. We build on Argente and Yeh (2017) and use detailed information on product attributes that is available from the product data. Products can then be compared on the basis of characteristics associated with their attributes $\{v_{u,1}, ..., v_{u,A}\}$.[8] We test if each new product has characteristics distinct from those of all existing products available in the market and we compute the quality of a new product as a weighted sum of

[8] For example, "children" and "regular" are two mutually exclusive characteristics associated with the attribute "formula" for "pain remedies-headache" products. Naturally, the number and type of attributes varies across product categories. For example, the product category "pain remedies-headache" includes 10 attributes: brand, flavor, container, style (i.e. children, regular), form, generic, formula (i.e. regular, extra strength, rapid release), type (i.e. aspirin), consumer (i.e. trauma, migraine), and size. On average, we observe that the different product categories include between 5 to 12 attributes. Appendix B gives details.

its novel characteristics across all product attributes:

$$q_u \equiv \sum_{a=1}^{A} \omega_a \mathbb{1}[v_{ua} \text{ is new}].$$

where ω_a are weights that reflect the economic value associated with a particular attribute. We develop a novel approach to estimate weights that capture the importance of each attribute by using "shadow prices" from hedonic pricing regressions (Bresnahan and Gordon, 1996). The underlying assumptions here are that the degree of novelty of a product should be reflected in the price of a product and that the price of a product reflects its embodied characteristics as valued by implicit or shadow prices. A new product has a high novelty score if it has many new characteristics and/or if its characteristics are associated with high implicit prices. We provide details on the properties of this procedure in Appendix B, along with some evidence that the novelty score is strongly associated with the performances of the firm and its products.[9]

We use three alternative measures of new product quality to evaluate the robustness of the empirical patterns we used to relate products to patents. First, we use a simpler version of the quality measure that weighs each attribute equally (quality *q1*). This measure only captures variation in the share of new product characteristics within a product. Second, we develop a measure that is computed much like our baseline measure with the exception that it uses weights that reflect "shadow sales" (quality *q2*). This measure assigns lower quality to new products that are associated with high shadow prices but do not reach many customers. Finally, we use a measure of residual demand taken from Hottman et al. (2016) and Argente et al. (2020) (quality *q3*). This measure does not use information about the degree of novelty of a product and instead captures the relative appeal of new products relative to other products sold by the firm, under some functional-form assumptions. Overall, our baseline measure and these alternative metrics allow us to consider many critical dimensions of the quality of new products, and allow us to assess the robustness of our results.

3.2 Patent Measures

Using an approach similar to how we measured product innovation, we compute measures that allow us to account for differences in the quantity and quality of patent applications

[9]We show that our measure is correlated with the growth rate of the firm, the share of sales generated by new products, and the average duration of new products in the market even after conditioning on the number of products being introduced by the firm (Table A.II in the Appendix).

across firms and over time. Our baseline measure is the number of **patent applications** (P_{it}). Using our patent-product category match, we are also able to measure the number of patent applications filed by firm i in product category j in year t as follows:

$$\mathrm{P}_{ijt} \equiv \sum_{p=1}^{P_{it}} \mathbb{1}[p \text{ is match to } j].$$

Throughout the paper, we use information about whether a patent was granted and information about patent citation counts to compute our measures of patent quality. Patent applications that become **granted patents** (gP_{ijt}) are perceived as high-quality patents because the patent office deemed them novel enough to not be rejected. We compute the number of patent applications that are granted as:[10]

$$\mathrm{gP}_{ijt} \equiv \sum_{p=1}^{P_{it}} \mathbb{1}[p \text{ is granted}] \times \mathbb{1}[p \text{ is match to } j].$$

We also define **patent citations** (cP_{ijt}) as the total number of patents weighted by forward citations received in the first five years since the application was filed:[11]

$$\mathrm{cP}_{ijt} \equiv \sum_{p=1}^{P_{it}} c_p \times \mathbb{1}[p \text{ is match to } j].$$

Measures based on forward citations have traditionally been used to assess the economic and technological significance of a patent (for earlier contributions, see Pakes (1986), Schankerman and Pakes (1986), Trajtenberg (1990)).

3.3 Summary Statistics

Table 2 provides summary statistics about the product- and patent-related variables for the firms in our sample, grouped by their patenting activity. We split firms into three groups: (i) firms that have never filed a patent application, (ii) firms whose last patent application was filed before 2006 (the beginning of the Nielsen RMS data set) and (iii) firms that filed a patent application between 2006 and 2015.

[10] The condition $\mathbb{1}[p \text{ is match to } j]$ is only used for Match 2.

[11] A 5-year citations measure attempts to reduce the truncation issue inherent to citations – the fact that patents filed more recently have had less time to accumulate citations.

Table 2: Summary Statistics by Firm's Patenting Status

	No Patents	Patents before 2006	Patents 2006-2015
Product data			
Number of products	15.49	31.08	78.35
Number of new products (N)	2.58	5.26	13.45
Average quality of new products (q)	0.27	0.20	0.20
Quality-adjusted number of new products (qN)	0.46	0.62	1.48
Product introduction rate (n)	0.19	0.17	0.22
Quality-adjusted product introduction rate (qn)	0.07	0.04	0.06
Sales from all products	2371.59	9392.09	37094.71
Sales from new products	454.74	1811.01	8130.00
Number of product categories	2.36	3.07	5.46
Average quality of new products ($q1$)	0.13	0.10	0.10
Average quality of new products ($q2$)	0.18	0.11	0.12
Average quality of new products ($q3$)	0.06	0.32	0.10
Patent data			
Number of patent applications (P)	0.00	0.00	6.34
Number of granted patent applications (gP)	0.00	0.00	4.57
Number of citations-weighted patent applications (cP)	0.00	0.00	5.88
Stock of patent applications	0.00	11.33	125.36
Stock of granted patent applications	0.00	11.02	107.63
Stock of citations-weighted patent applications	0.00	17.97	215.24
Number of firms	29215	1943	3266
Observations	186934	15803	29052

Notes: The table shows the average of product-based and patent-based variables of the Match 1 data set. The first column groups firms that have no patents; the second column considers firms that have patents, but filed them before they first appear in Nielsen RMS (before 2006); and the third column is for firms that have patents in our focus period of 2006-2015. The statistics regarding product introduction can only be computed for the period 2007-2015 because we cannot determine entries for products first introduced in 2006 (left censored). The statistics for sales are given in thousands of dollars, deflated by the Consumer Price Index for all urban consumers. Information regarding the technology classes of patents is defined using IPC3 data.

Firms in the consumer goods sector are relatively patent-intensive and have product introduction rates comparable to other manufacturing sectors. Table 2 shows that more than 5 thousand firms applied for at least one patent and more than 3 thousand firms filed a patent application during the period 2006-2015. Graham et al. (2018) links Census data to the USPTO and finds that the incidence of patenting is higher in manufacturing than it is in the rest of the economy – 6.3% vs 1% of firms have at least one granted patent application between 2000 and 2011.[12] Table 2 indicates that product introduction rates are on average 20%, which may even be a lower bound to the entry rates in other sectors. Goolsbee and Klenow (2018) use the Adobe Analytics data on online transactions covering multiple prod-

[12]This result is comparable to the figure of 7.6% in our data. Graham et al. (2018)'s patent data includes only granted patents, while our data also includes unsuccessful patent applications. If we count only granted applications, then we would have 2629 patenting firms.

ucts, and report rates even higher for some durable products.[13] Our data covers product categories that exhibit substantial heterogeneity in patenting intensity and entry rates. The share of patenting firms varies from zero in some food categories to a maximum of about 60% in water purification products, and the product introduction rates vary from about 5% in some food categories to a maximum of about 30% in printer supplies.

As expected, patenting firms are larger: they sell more products, operate in more product categories, and have higher sales. Firms that filed patents between 2006 and 2015 account for 61% of sales in our sample. Patenting firms also introduce more products, but this relationship is weaker once we scale the new products and focus on the rates with which new products are introduced instead of the firm's cumulative number of new products. Interestingly, our four different quality measures indicate that the average novelty of new products sold by patenting firms is not higher than that of non-patenting firms, conditional on product introduction.

We also consider the patenting activities across this grouping of firms by patenting activity. Firms with patent applications between 2006 and 2015 file more than six patents per year, on average.[14] Because many patents receive no citations, especially in the first five years, the average number of citation-weighted patent applications, cP_{ijt}, is very similar to the average raw number of patent applications, P_{it}. These firms may hold some design patents, but the majority of patents in our sample are utility patents. Unsurprisingly, the summary statistics show that firms who filed a patent between 2006 and 2015 hold a larger stock of patents than firms who last filed a patent application before 2006. Actively patenting firms hold, on average, 125 patents in stock every year whereas firms that last patented before 2006 hold approximately 11 patents.

4 Relationship Between Product Innovation and Patents

How do patents relate to actual product introduction to the market? How much product innovation in the consumer goods sector is captured by patent-based metrics of innovation? We document the relationship between patents and product innovation using following exercises. First, we show the cross-sectional allocation of product innovation between patenting and non-patenting firms. Second, we consider how a firm's product introduction changes

[13] Goolsbee and Klenow (2018) show that some durable consumer goods (e.g. furniture), not covered in our data set, can have entry rates twice as large as entry rates of non-durables (e.g. food).

[14] Patent statistics are very skewed, and we present averages after winsorizing patent-based variables at top 0.1%.

after it files a first patent application. Third, we quantity the strength of the relationship between changes in the number of patents filed and the amount of product introduction. Finally, we explore the dynamics of these effects. Our findings can be summarized in the following two empirical facts:

Fact 1: More than half of product innovation comes from firms that do not patent.

Fact 2: On average, patents are positively associated with subsequent product innovation by firms.

Product Introduction and Firm's Patenting Status.—We being our analysis of the relationship between patents and new products by exploring cross-sectional variation across firms according to their patenting status. Table 3 shows that in our data, 54% of new products were introduced by firms that never applied for a patent. If we account for the degree of novelty of new products, we estimate that about 65% of quality-adjusted product introduction comes from never-patenting firms. This indicates that, on average, patenting firms introduce more products that make only an incremental improvement over existing products on the market.[15]

Since they rely on the firm-level match, the above statistics implicitly attribute all new products introduced by a patenting firm to some of its patents. However, highly diversified firms might be patenting in one product category, while introducing many products that have no relation to the patents they are filing in other categories. Thus we may be attributing too much product introduction to patents if we rely only on the firm's overall patenting status. This observation exemplifies the importance of establishing a closer link between patents and products using the Match 2 data set. To make these more granular links, we replicate the above exercise but define patenting status at the firm $\times$ category level. As seen from Table 3, firms that never patented in a category are responsible for a greater share of new products introduced in that category.

It is not surprising that a large amount of innovation may not be associated directly with any patents. Even if firms wanted to patent all their new products, some new products represent only small upgrades to existing products, and may not be patentable. Patents are only granted if they exhibit "novelty and non-obviousness", and thus many new products that

[15] This observation holds true regardless of the quality adjustment we use. For example, the share of $q1N$ accounted by never-patenting firms is 65%, and the share of $q2N$ by never-patenting firms is 77%. Our residual quality measure of innovation, $q3$, does not allow us to construct a good counterpart to $q3N$, however as seen from Table 2, $q3$ is not necessarily higher for patenting firms.

Table 3: Share of Product Innovation Accounted for by Patenting Firms

	New Products, N	Quality-adjusted New Products, qN
Match 1		
Firms with patents in 2006-2015	.38	.28
Firms with patents before 2006	.08	.07
Firms with no patents	.54	.65
Match 2		
Firm × category with patents in 2006-2015	.23	.16
Firm × category with patents before 2006	.07	.05
Firm × category with no patents	.71	.79

Notes: the table shows the share of product innovation on the market measured by our two benchmark measures – product introduction (column 1) and quality-adjusted product introduction (column 2) – accounted for by firms and firm×categories with or without patents.

result from very small changes will not be captured by patent metrics. While it is natural that some innovations are not captured by patents, our data offers an unique opportunity to quantify the magnitude of it.

We also evaluate if our measures of product innovation reflect well the sources of growth. Indeed, if we look through the lens of classic innovation-driven growth models, we should expect innovation and growth measures to go hand-in-hand. We conduct simple growth decompositions for our sector to get at this question. We decompose sales growth from 2006 to 2015 into growth that comes from patenting and non-patenting firm × categories as:

$$\underbrace{\text{Growth}_{\text{`}06-\text{`}15}}_{7\%} = \underbrace{\text{Growth}^{\text{Patent}}_{\text{`}06-\text{`}15}}_{4\%} \times \underbrace{s^{\text{Patent}}_{2006}}_{0.72} + \underbrace{\text{Growth}^{\text{No Patent}}_{\text{`}06-\text{`}15}}_{14.4\%} \times \underbrace{s^{\text{No Patent}}_{2006}}_{0.28} \tag{1}$$

where s^{Patent}_{2006} and $s^{\text{No Patent}}_{2006}$ denote sales shares of firm× categories with or without patents, respectively.[16] As with our measures of product innovation, these growth decompositions show that although non-patenting firms are smaller and account for a smaller share of sales in the sector, they contribute more to growth relative to the set of patenting firms – totaling to 58% of the sectoral growth.[17] Hence, the fact that more than a half of the product innovation in the sector is not captured by the patenting status of the firms is corroborated by similar statistics about growth.

[16]We first write $Rev^{CPG}_t = \sum_j \sum_{i\in\Omega^j_{\text{Patent}}} Rev_{ijt} + \sum_h \sum_{i\in\Omega^j_{\text{No Patent}}} Rev_{ijt}$, where the second sum is across product categories and Ω denotes the set of firms with and without patents in category j; and take the percentage changes in sales to arrive at (1).

[17]These observations are not surprising and should not be limited to our sector: as documented above, non-patenting firms are smaller, and smaller firms tend to contribute less to growth on average, as we know from other studies of the overall economy (Haltiwanger et al., 2013).

First-time Patent Filers.—One important feature of our data is that we observe some firms that change their patenting status in the period of analysis 2006–2015. This allows us to evaluate whether a firm's product introduction tends to change after the firm's first patent application.[18] We do so by estimating the following specification:

$$\log Y_{it} = \beta\, dP_{it} + \alpha_i + \gamma_t + u_{it} \tag{2}$$

where Y_{it} is the outcome of firm i in year t, α_i represents firm fixed effects, and γ_t represents year effects. dP_{it} is an indicator variable that equals 1 after the firm's first patent application. Our goal is to understand if the switch to patenting is associated with increased product innovation, which would be the case if patent-based measures were to approximate well product innovation in the market. To uncover this relationship, we estimate the effects of β relative to firms that are already patenting. These have more similar characteristics and thus are likely a more suitable counterfactual for firms that first apply for a patent than those that never apply.[19]

Table 4 presents the estimated change in our two measures of product innovation associated with a firm's transition from non-patenting to patenting. Conditional on firm and year effects, we find an average increase in product introduction of up to 11% after the switch to patenting. Columns (2) and (3) show that the positive correlation is largely driven by high-quality patent applications, if we take the patent's success with the patent office as a proxy of quality. This result is more pronounced if we study the effect on quality-adjusted product introduction, as shown in columns (4) to (6). These exercises reveal a positive correlation between the timing of patent applications and product innovation. One interpretation of this correlation is that firms come up with ideas for new products and then apply for a patent to protect the idea from being copied by competitors; simultaneously, they develop those ideas into new consumer products.[20] Our findings about the changes in the observed product

[18] Although defining the event of the first patent application at the firm level (as opposed to firm-category) avoids potential cross-category spillovers in patenting and sharpens out definition of the event, we also explore the dynamics of this relationship in detail using Match 2 data later on.

[19] The assumption that this group of firms forms a better control − after accounting for time-invariant differences between firms and common year factors − is supported by the summary statistics presented in Table 2. Nevertheless, we find similar estimates when we test if our results are explained by the contrast with the entire sample of non-switching firms, which includes firms already patenting before the beginning of our sample and those that have not yet patented at the end of our sample (Table A.III in the Appendix).

[20] It may also be the case that patenting gives firms a preferential status in the economy that allows them to create more products in the future. For example, this status may result from consumer perception that firms are more "innovative" when firms advertise "patent pending" on their products.

introduction around the time that a firm first files a patent add to other papers' findings that patenting is associated with other real changes at the firm level, such as increases in stock market prices or the firm's sales and scope (Hall et al., 2005; Balasubramanian and Sivadasan, 2011; Kogan et al., 2017).

Table 4: Product Innovation after First Patent Application

	Log N			Log qN		
	(1)	(2)	(3)	(4)	(5)	(6)
After patent(t)	0.1168**			0.0352		
	(0.045)			(0.020)		
After granted patent(t)		0.1361**			0.0497**	
		(0.048)			(0.018)	
After non-granted patent(t)			-0.0045			-0.0085
			(0.044)			(0.036)
Observations	29,470	29,470	29,470	29,470	29,470	29,470
Time	Y	Y	Y	Y	Y	Y
Firm	Y	Y	Y	Y	Y	Y

Notes: The table shows regressions of log number of new products (Log N) in Panel A and of log quality-adjusted new products (Log qN) of a firm as a function of a dummy equal to one after the first patent application by the firm. Our benchmark quality measure is defined in Section 3.1. The alternative innovation quality measures ($q1, q2, q3$) produce similar results. Both Log N and Log qN use the inverse hyperbolic sine transformation. *After patent* is a dummy equal to one after any patent application; *After granted patent* is a dummy equal to one after a patent application that is granted; and *After non-granted patent* is a dummy equal to one after a patent application that has not been granted (abandoned or pending). The sample includes 596 firms that switch to patenting and those that already patented before the beginning of our sample. Standard errors robust against heteroskedasticity and serial correlation are reported in parentheses.

Elasticity of Product Innovation to Patents.—We next explore how product innovation varies with the changes in intensive margin of patenting exploiting variation in measures of product innovation at the firm $\times$ category level over time. We estimate

$$\log Y_{ijt} = \beta \ \log P_{ijt-1} + \alpha_{ij} + \gamma_{jt} + u_{ijt} \tag{3}$$

where Y_{ijt} is the outcome for firm i in category j in year t and P_{ijt-1} is the log number of patent applications filed by the firm i in category j a year before to allow for a short lag between patent filing and product commercialization. Thanks to the firm $\times$ category level data, we can now control for product category-specific trends (e.g., market-wide demand for specific products), and we can control for firm-category specific effects, thus filtering out, for instance, the effects of firm-specific brand power on the sales of specific products.

We seek to provide reduced-form estimates for the elasticity of product innovation to patents (β). One can think of this elasticity along the lines of the knowledge production function approach (Griliches, 1979). In that approach, the estimated elasticity is for patents (output) with respect to R&D (inputs). However, by and large, patent filings are not the main policy target of various innovation policies (e.g., R&D tax incentives) – innovation in the market is. In our case, patents can be considered as inputs to the production function of the output, which is product innovation. Hence, if one is interested, for instance, in how much a policy that encourages additional R&D spending affects product innovation, one needs to account for an additional elasticity that shows the strength of the relationship between the patents and actual innovation, in addition to the patents-to-R&D elasticity.[21]

Table 5 shows the estimates. The rows present results from using different explanatory variables – the log number of patents, granted patents, and non-granted patents. Conditional on firm-category and category-time fixed effects, we find that the observed elasticities of product introduction and quality-adjusted product introduction to patents are 0.04 and 0.02, respectively. As before, the relationship between patenting and product innovation is mainly driven by higher-quality granted patents. Likewise, Table A.V of Appendix provides similar results for other quality measures of patents – citations and claims. Overall, these exercises show that we can statistically identify a positive correlation between patenting and product introduction, which corroborates that firms' patenting is positively associated with their product innovation.

The estimated elasticity captures the relationship between product introduction and patents associated with products. Not all patents, however, necessarily relate to product improvements: some patents may relate to cost savings from improvements to the firm's general production processes. Nevertheless, our firm $\times$ category data set filters out patents that are not specifically related to product introductions. Hence, to a large extent, our estimates should be driven by product patents rather than process patents. To support this point, we consider the robustness of our results and employ proxies for product-related and process-related patents drawn from Bena and Simintzi (2017); we find that the coefficient on product-related patents is essentially same as our benchmark coefficient, while process-related patents are wholly unrelated to measures of product innovation (Section A.2 and Table A.VI of Appendix).

[21] To underline this point with a simple illustration, consider a standard set-up where $Patents \propto R\&D^{\alpha}$, and α is the knowledge production elasticity estimated in the literature (e.g. Dechezlepretre et al. (2016)). At the same time, $Innovation \propto Patents^{\beta}$. Hence, the effect of additional spending on $R\&D$ for $Innovation$ should, in practice, be inferred from the combined elasticity $\alpha\beta$.

Table 5: Product Innovation and Patenting

	Log N			Log qN		
	(1)	(2)	(3)	(4)	(5)	(6)
Patents(t-1)	0.0380***			0.0189***		
	(0.009)			(0.005)		
Patents granted(t-1)		0.0405***			0.0192***	
		(0.010)			(0.005)	
Patents non-granted(t-1)			0.0234*			0.0082
			(0.013)			(0.007)
Observations	409,641	409,641	409,641	409,641	409,641	409,641
R-squared	0.692	0.692	0.692	0.623	0.623	0.623
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y

Notes: The table shows regressions of the log number of new products (log qN) and of log quality-adjusted new products (log qN) in a firm × category over time as a function of the log number of patents. Our benchmark quality measure is defined in Section 3.1. The alternative innovation-quality measures ($q1, q2, q3$) produce consistent results. *Patents* is the log number of any patent applications in firm × category × year; *Patents granted* is the log number of granted patent applications; and *Patents non-granted* is the log number of patent application that have not been granted (abandoned or pending). The inverse hyperbolic sine transformation is used for logs. Standard errors robust against heteroskedasticity and serial correlation are reported in parentheses.

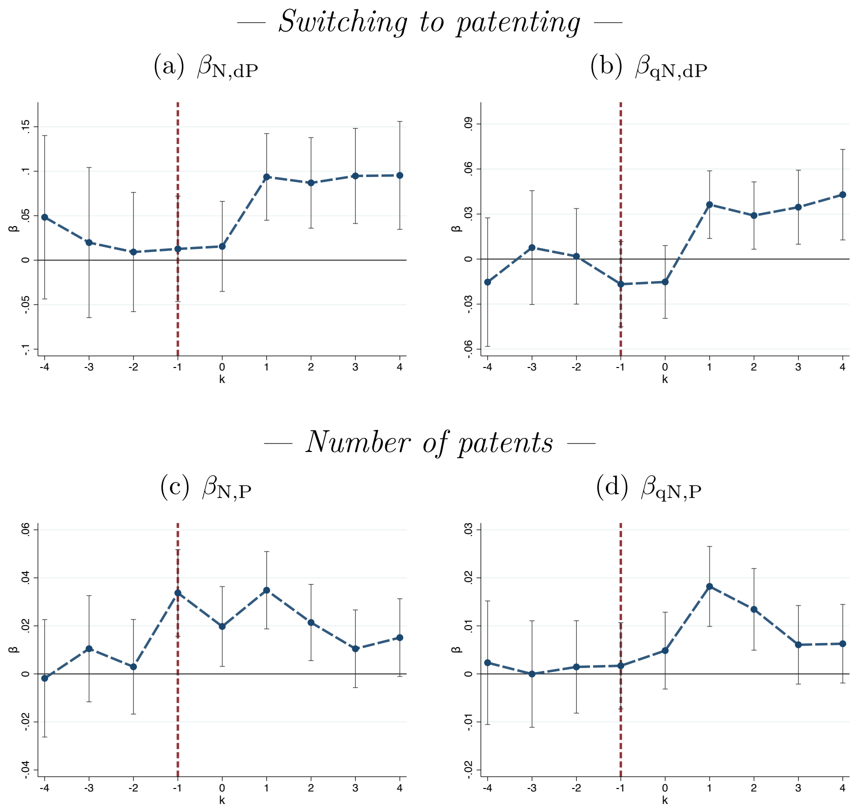
Dynamics of the Effects.—We are now interested in evaluating the timings of the effects captured in (2) and (3). Thus, we study the relationship between patents and product introduction by running the following separate linear regressions using the firm-category level data set:

$$Y_{ijt+k} = \beta_k \mathrm{E}_{ijt} + \alpha_{ij} + \gamma_{jt} + u_{ijt+k} \ , \qquad k = -4, ..., 0, ..., 4 \tag{4}$$

where Y_{ijt+k} is an outcome of firm i in product category j in $t+k$ associated with product introduction and E_{ijt} is either dP_{ijt} (as before, a dummy equal to one after firm starts patenting in category j) or $\log P_{ijt}$, which again denotes the log number of patents filed by firm i in product category j in t. We also include firm-product category and time-product category fixed effects.

Figure 2 plots the estimated coefficients β_k over k. The top panel shows the evolution of N and qN around the time at which the firm starts patenting in a certain product category. The bottom panel is about the intensive margin of patenting, and both are based on patent application years. Consistent with the results above, we find a positive association between patents and product introduction. Our estimates indicate that firms introduce about 10% more products after filing their first patent, with no pre-trends in outcomes before the firm switches to patenting (and 3-4% if we adjust for the novelty of new products– see (a) and (b)). The positive association reaches its maximum magnitude shortly after the first patent is filed in a product category and is fairly persistent thereafter.

Figure 2: Product Innovation and Patenting: Dynamics



Note: The figure plots the estimated coefficients after estimating equation (4) for log product introduction, N, in (a) and (c), and quality-adjusted product introduction, qN, in (b) and (d). Our benchmark quality measure is defined in Section 3.1. The main explanatory variable in (a) and (b) is a dummy equal to one after the firm's first patent in a product category and log number of patent applications in (c) and (d). The inverse hyperbolic sine transformation is used for logs. The vertical bands represent +/- 1.65 $\times$ st. error of each point estimate. Standard errors are clustered at the firm $\times$ category level.

Likewise, our results (see (c) and (d)) exploring the co-movement between patent applications and product introduction indicate that product innovation spikes one year after new patent applications. With an exception for product introduction at $k = -1$ in (c), we do not find a significant relation for k below zero. These dynamic specifications are useful for inferring the long-run elasticity of product introduction to patents, in contrast to the instantaneous elasticities discussed previously. Unlike the results with first-time patent filers, the results for the intensive margin of patenting are not persistent over time, which indicates that filing an extra patent application does not lead to an incremental product introduction in the long run. Under exogeneity assumptions in the context of lineal local projections (Jorda, 2005), the implicit long-run elasticity between patents and product introduction is the sum of the β_k coefficients from $k = 0$ onward. Our results point to an elasticity of about 0.1 for product introduction, and about 0.04 for quality-adjusted product introduction in the four years after the patent is filed. In Figure A.7 of Appendix, we also show that other variables such as the stock of products or sales significantly increase after patents. We also confirm that our results are robust to considering the firm-level data from Match 1.

5 Product Innovation, Patents, and Competition: The Role of Firm Size

The previous sections show that patents are positively associated with product innovation in the market. This associations suggests that patents reflect important technological improvements that firms, on average, commercialize by introducing new products. However, in addition to their role in reflecting certain technological novelties, patents give firms the right to exclude others from using the same or similar technologies (Hall and Harhoff, 2012). These negative competitive spillovers from patenting have long been recognized in the literature (e.g. Lanjouw and Schankerman, 2001; Jaffe and Lerner, 2004; Bloom et al., 2013). Firms can use patents strategically to defend their technologies, reduce competitive pressure, and deter entry (Cohen et al., 2000; Akcigit and Ates, 2019). How do firms use these two roles of patenting? In their classic paper, Gilbert and Newbery (1982) suggest that monopolists have large incentives toward preemptive patenting. Likewise, Blundell et al. (1999) argue that patents held by market leaders may serve a largely preemptive purpose. This section sheds light on these issues by empirically evaluating how product innovation, patenting, and returns on patents vary systematically with a firm's market lead measured by relative sales in the market. We document the following empirical regularities:

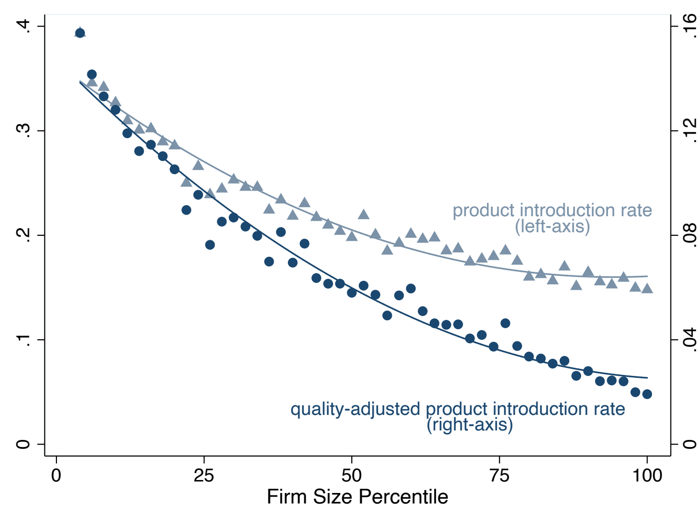
Fact 3: Larger firms have lower product innovation rates (quantity and quality), but file more patents for each new product.

Fact 4: Patenting by larger firms is strongly associated with an increase in sales above and beyond the patents' effect on product innovation.

Fact 5: Patenting by larger firms is associated with a decline in product innovation by competing firms.

Product Innovation and Patenting by Firm Size. — We begin by exploring how product innovation rates vary with firm size. Figure 3 plots the average product introduction rate – the ratio of product introduction to a firm's stock of existing products – for firms across product categories. Larger firms (within product categories) have lower product innovation rates. On average, firms in the top sales quintile have annual innovation rates of about 16%, while firms in the bottom quintile have rates twice as large. Larger firms do not compensate for this decline in the rate of new product introduction with innovations of higher quality. On average, firms in the top sales quintile have quality-adjusted product introduction rates of 3%, while firms in the bottom sales quintile have rates four times larger. The fact that the quality-adjusted introduction rate declines more steeply than the simple product introduction rate indicates that, on average, new products introduced by larger firms represent only incremental improvements over existing products and are thus less novel. Figure A.9 in the Appendix confirms similar patterns using alternative measures of innovation quality based on other novelty metrics and residual demand.

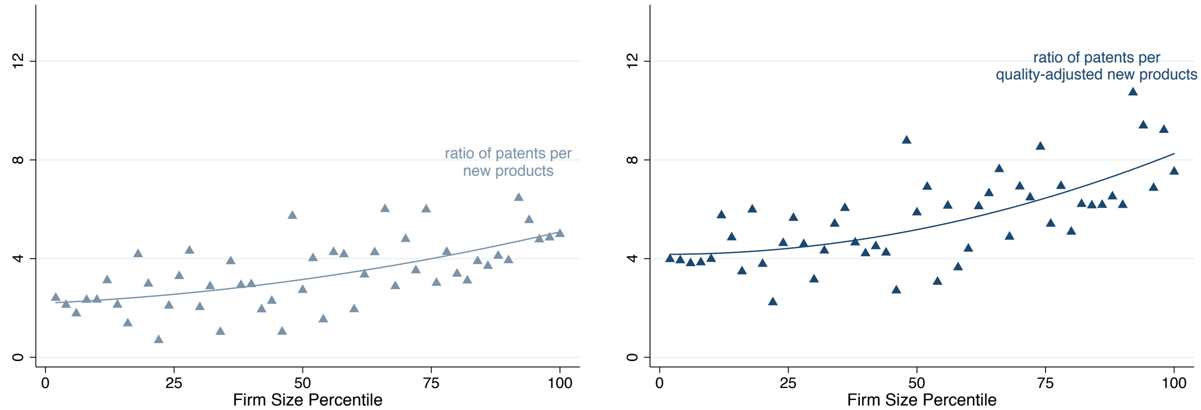
Figure 3: Product Innovation Rate by Firm Size



Notes: This figure plots the relationship between product innovation and the size of the firm, defined by the firm's sales. We use the firm × product category level data for the period 2007–2015, restricting the analysis to observations with sales above $1,000. For each firm × product category, we compute average sales, the average product innovation rate (new products divided by the total number of products sold), and the quality-adjusted product innovation rate (quality-adjusted new products divided by the total number of products sold). Within each product category, we assign firms to 50 bins for average sales and plot the average product innovation rate and the quality-adjusted product innovation rate for each bin. Each dot/triangle plots the averages after weighting each product category by its importance in the whole sector, as measured by the share of sales accounted for by the category.

This fact speaks to the well-known empirical regularity that larger firms grow slower (e.g. Haltiwanger et al., 2013). We show that this slow-down in growth is a reflection of larger firms' slow-down in the introduction of new and higher-quality products. Keeping this relationship between firm size and product innovation in mind, we now explore how patenting activity varies with firm size. Thanks to our data set that matches patents to products, we can simultaneously measure both product innovation and the associated patent applications by firms. As a results, Figure 4 shows that larger firms, on average, file more patents for each new product introduced.[22] Note that this higher intensity of patenting activity relative to the number of new products introduced is not explained by the possibility that larger firms introduce fewer but more novel products: as one can see, after we adjust for the quality of new products, small and large firms' innovation rates diverge even more.

Figure 4: Patents per New Products, by Size



Notes: This figure plots the relationship between the ratio of patent applications per new products and firm size as defined by sales. We use the firm × product category level data set for the period 2007–2015, restricting the analysis to observations with sales above $1,000. For each firm × product category, we compute average sales, the average number of patent applications per new products, and the average number of patent applications per quality-adjusted new products. Within each product category, we assign firms to 50 bins of size based on average sales, and compute the average ratio of patents per new products and average ratio of patents per quality-adjusted new products for each bin. Each triangle plots the averages after weighting different product categories by their importance in the whole sector, as measured by their share of sales. The left figure plots the log ratio of patents per new products (× 1000), and the right figure plots the log ratio of patents per quality-adjusted new products (× 1000).

Motivated by this cross-sectional evidence, we now systematically explore how the elasticity of product innovation to the number of patents varies with firm size. We calculate this elasticity as in equation (3) for both product introduction N and quality-adjusted product introduction qN, after controlling for firm-category and category-time fixed effects. Table 6 reports the elasticities for firms in different size groups. In line with the results discussed above, we estimate an average elasticity of 0.038 (column "All"). The table shows that this elasticity varies substantially across the firm size distribution. Smaller firms in the bottom sales quintile have an elasticity twice as large as that of firms in the top sales quintile (0.059 versus 0.030). In the case of very large market leaders with sales in the top sales decile, we find only a non-significant positive association between patenting and product innovation. These market leaders have the highest rates of patenting, but the patents they file do not seem to translate into new products.

[22] Clearly, if we do not scale our measures of patenting down, the results are even starker: the unconditional probability of patenting and the total number of patents filed by large firms are much higher than they are for small firms (see Figure A.10 of Appendix).

Table 6: Product Innovation and Patenting: by Size

	Log N (t)				Log qN (t)			
	All	Small	Large	Leaders	All	Small	Large	Leaders
Log P(t-1)	0.038***	0.059***	0.030**	0.023	0.019***	0.033***	0.017***	0.015*
	(0.007)	(0.018)	(0.013)	(0.017)	(0.003)	(0.007)	(0.006)	(0.009)
Observations	409,641	61,350	86,953	41,325	409,641	61,350	86,953	41,325
R-squared	0.692	0.463	0.742	0.777	0.623	0.407	0.686	0.732
Time-Category	Y	Y	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y	Y	Y

Notes: The table shows regressions of the log number of new products (log qN) and of log quality-adjusted new products (log qN) in a firm × category over time as a function of the log number of patents. P is the number of patent applications for a firm × category × year. For each firm × product category, we define size based on the average sales over our sample period. The "All" column shows data for all sizes. "Small" column is restricted to the bottom size quintile. "Large" is restricted to the top size quintile. "Leaders" is restricted to the top size decile. The inverse hyperbolic sine transformation is used for logarithms. Standard errors robust against heteroskedasticity and serial correlation are reported in parentheses.

The results are similar for the quality-adjusted product introduction, shown in the rightward columns of the table. We further rule out two additional channels that could explain the weak association between patents and product innovation for large firms. First, we do not find evidence that patents held by larger firms are associated wuth product innovation with a longer delay. We test the dynamic specifications of equation (3) and do not find that the elasticity of patents to future product introduction is stronger for larger firms.[23] Second, we test whether patents held by larger firms are related more to process innovation rather than to product innovation (Cohen and Klepper, 1996). Building on Bena and Simintzi (2017), we construct proxies for product-related and process-related patents and we find no systematic pattern between a share of process patents filed by a firm and the firm's size (Figure A.8 in Appendix).[24]

The Role of Patents for Competition by Firm Size.—Our finding that patents are more weakly associated with product innovation for larger firms is by itself important from the perspective of measuring innovation in the market: for market leaders, patents may be misleading proxies for the actual innovation that drives productivity growth. At the same time, this finding may be indicative of market leaders using patents as protective strategy tools. Firms may grow sales not just by introducing new and innovative products, but by accumulating patents that reduce competition and help the firm to capture market share

[23]Indeed, recall that in Section 4, we saw that the increase in product innovation after a new patent filing was short-term.

[24]In addition, if cost reductions due to process innovations are reflected in lower subsequent prices, we can test whether the price changes of larger firms react to patents more. However, we do not find such relationship in the data.

from its competitors.[25] To address this possibility, we take advantage of the information on product sales to understand the revenue premiums from patents for firms of various sizes. We use the following specification:

$$\Delta \log \text{Sales}_{ijt} = \psi \log P_{ijt-1} + \rho \log N_{ijt} + \theta_{ij} + \gamma_{jt} + \varepsilon_{ijt} \tag{5}$$

where the dependent variable is the logarithm of the change in sales at time t, $\log P_{ijt-1}$ is the total number of patent applications until time $t-1$, and $\log N_{ijt}$ is the number of new products introduced at t (we also use the quality-adjusted product introduction $\log qN_{ijt}$). Our coefficient of interest is ψ, which measures the elasticity of sales growth to patents after controlling for the effect that patents may have on sales through increased product innovation.

Table 7 shows the results for all firms and for firms grouped according to size. Overall, we find a positive significant relationship between patents and future growth in sales even after controlling for product innovation (columns "All"). This finding indicates that firms enjoy some additional value from holding a patent beyond the patent's value through new product offerings. Interestingly, this effect is highly heterogeneous across firm sizes. For firms in the bottom sales quintile (columns "Small") there is no statistical association between patents and sales growth after we control for product introduction. However, for firms in the top quintile (columns "Large"), we find that an increase in total patent applications has a significant positive association with sales growth above and beyond its effect through product introduction. This effect is even larger (elasticity of 0.1) when we further restrict the analysis to firms in the top size decile (columns "Leader"). Note also that the direct impact of product innovation on sales growth (coefficients on $N(t)$ and $qN(t)$) decreases as firms increase in size. Hence, by splitting the sample into small and large firms, we learn that while both patents and new products are associated with increased future sales, the conditional impact of new products is more important for smaller firms, while the impact of patents is important for larger firms.

This additional revenue premium that larger firms draw from a patent may likely operate through patents' effect on competition: if patents discourage competitors from introducing new products, patent holders will benefit by serving a larger market. Thus, we now investigate whether patents by market leaders are associated with declining product introduction

[25] The accumulation of patents often creates a web of overlapping intellectual rights which make it difficult for competitors to approach the market leader's technology domain and to leapfrog them. See, for example, Shapiro (2000) for the discussion of patent thickets.

Table 7: Patenting and Sales Growth

	Δ Log Sales (t)				Δ Log Sales (t)			
	All	Small	Large	Leaders	All	Small	Large	Leaders
Log P(t-1)	0.061***	-0.081	0.072***	0.099***	0.073***	-0.101	0.089***	0.111***
	(0.016)	(0.077)	(0.017)	(0.019)	(0.016)	(0.077)	(0.018)	(0.019)
Log N(t)	0.265***	0.316***	0.214***	0.160***				
	(0.003)	(0.011)	(0.003)	(0.004)				
Log qN(t)					0.406***	0.581***	0.310***	0.215***
					(0.006)	(0.029)	(0.007)	(0.007)
Observations	296,320	40,666	131,804	65,680	296,320	40,666	131,804	65,680
R-squared	0.291	0.377	0.294	0.296	0.275	0.368	0.277	0.281
Time-Category	Y	Y	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y	Y	Y

Notes: The table presents estimated outcomes of changes in log sales at the firm × category level as a function of the log number of patents (P) and log number of new products (log qN) and log quality-adjusted new products (log qN), by size groups. We use the firm × product category data set for the period 2007–2015, restricting the analysis to observations with sales above $1,000. For each firm × product category, we define size based on average sales over the sample period. "All" column uses data for all sizes. "Small" column is restricted to the lowest size quintile. "Large" is restricted to the top size quintile. "Leaders" is restricted to the top size decile. The inverse hyperbolic sine transformation is used for logarithms. Standard errors robust against heteroskedasticity and serial correlation are reported in parentheses.

on the part of their competitors, who we will refer to, for simplicity, as market followers. We identify the market leader in each category as the firm with the highest sales in that category and the followers as the remaining firms operating in that market.[26] Then for each year t and market j, we compute the total number of new products introduced by the leader N^{L}_{jt} and by its followers N^{F}_{jt} in t, and we compute the total numbers of patent applications introduced by the leader P^{L}_{jt} and by its followers P^{F}_{jt} until t. We evaluate how product innovation by followers responds to patenting (and product innovation) of the leaders as follows:

$$\log N^{\mathrm{F}}_{jt} = \eta^{\mathrm{F}} \log P^{\mathrm{L}}_{jt-1} + \alpha^{\mathrm{F}} \log N^{\mathrm{L}}_{jt-1} + \theta^{\mathrm{F}}_{j} + \gamma^{\mathrm{F}}_{t} + \varepsilon^{\mathrm{F}}_{jt}, \tag{6}$$

where η^{F} is our coefficient of interest, measuring the association of patents of leaders with the product introduction by followers. We control for $\ln N^{\mathrm{L}}_{jt-1}$ to ensure that the relationship between leaders' patents and followers' product introduction is not driven by possible direct interactions between the leader's and followers' product offerings (such as learning from new products on the market).[27] We also include both time- and product-category-fixed effects to control for time trends and differences in the intensities of patenting and product innovation

[26] To have a static firm-level measure, we define leaders as of 2006, which is the first year of our data. However, the results are not sensitive to a different choice, like using average sales over all years. Moreover, we consider alternative definitions of market leaders (e.g. top decile) and the results are robust.

[27] We also use quality-adjusted new products in all of these regressions and the results are similar.

across product categories. Likewise, we estimate a symmetric regression that measures how leaders' innovation is affected by followers' patenting:

$$\log N^{\mathrm{L}}_{jt} = \eta^{\mathrm{L}} \log P^{\mathrm{F}}_{jt-1} + \alpha^{\mathrm{L}} \log N^{\mathrm{F}}_{jt-1} + \theta^{\mathrm{L}}_{j} + \gamma^{\mathrm{L}}_{t} + \varepsilon^{\mathrm{L}}_{jt} \tag{7}$$

These regressions help us test if the relation between patents of competitors and product introduction is affected by whether we focus on leaders or followers.

Table 8 presents the estimated coefficients. Column 1 shows that product introduction by followers is negatively correlated with the size of the leader's patent portfolio. This means that the product categories in which the leader is intensifying patenting over time are also the product categories in which followers are reducing the introduction of new products. In column 2, we also control for total sales of the market to account for potential shifts over time in the importance of different types of products. In turn, columns 3 and 4 show that product innovation by leaders is not related to the followers' patenting activity. Hence, while patents can be thought of as a protective tool used to hinder competition, our results indicate that this hypothesis is likely to apply when patents are in the hands of large market leaders.

Table 8: Patenting of Market Leaders and Followers

	Followers Log N^{F}			**Leaders** Log N^{L}	
	(1)	(2)		(3)	(4)
Leaders			**Followers**		
Log P^{L} (t-1)	-0.071***	-0.059***	Log P^{F} (t-1)	-0.015	-0.012
	(0.007)	(0.007)		(0.047)	(0.044)
Log N^{L} (t-1)	0.010***	0.005*	Log N^{F} (t-1)	0.215*	0.185*
	(0.002)	(0.002)		(0.112)	(0.094)
Observations	3,192	3,192	Observations	3,188	3,188
Category	Y	Y	Category	Y	Y
Time	Y	Y	Time	Y	Y
Controls	N	Y	Controls	N	Y

Notes: The table shows the relationship between the patents of leaders (followers) and the product introduction of followers (leaders). The leader is defined as the firm with the highest sales in a given category in 2006; the followers are defined as the rest of the firms in the categories. In columns (1) and (2), the dependent variable is the log number of products introduced by followers at time t, and the independent variables are the log number of patent applications by leaders until time $t-1$ and the log number of new products introduced by the leader at time $t-1$. In columns (3) and (4), the dependent variable is the log number of products introduced by leaders at time t, and the independent variables are the log number of patent applications filed by followers until time $t-1$ and the log number of new products introduced by the followers at time $t-1$. Columns (2) and (4) also control for total sales in the category-time. The inverse hyperbolic sine transformation is used for logarithms.

We view the results of this section as providing a consistent story for firms' use of their growth strategies. Firms may use both productive and protective strategies to grow sales.

We document that as firms grow, they rely less on productive strategies that encourage the introduction of new and improved products in the market. At the same time, as firms grow, they rely increasingly on protective strategies such as patenting, which is associated with declines in competitors' product introduction and the resulting higher sales for larger firms.

6 Conceptual Framework

In this section, we offer a simple theoretical framework that illustrates the relationship between innovation, patenting, and creative destruction. Our goal is twofold. First, the framework is meant to build intuition about the incentives for patenting, consistent with empirical patterns documented in the previous sections. Second, we use the model to perform a simple back-of-the-envelope calculation of the private value of a patent, and we decompose this private value into its protective versus productive components. The productive component is the option value of implementing the patented idea into higher-quality products in the market, thereby increasing profits. The protective component is the value that the firm gains by impeding creative destruction.

Our framework builds on the quality-ladder model of innovation with creative destruction (e.g. Aghion and Howitt (1992)). In this model, product innovation takes the form of upgrades to the quality of products on the market. These innovations come from either the incumbent leader trying to prolong its lead or from market entrants aiming to become the new leaders. We consider an exercise in which the incumbent firm obtains an idea for an innovation and makes a once-in-a-lifetime decision about commercializing that idea into a product and/or patenting it. If the firm decides to commercialize the idea, it will gain additional profits when it introduces higher-quality products to the market. In turn, patenting the idea grants the firm extra protection against creative destruction from entrants. The model has three basic ingredients: profits exhibit decreasing returns in quality, the probability of creative destruction depends on patent protection, and both patenting and product innovation are costly activities. This simple framework can easily rationalize the main empirical facts we have documented. For the same idea, smaller firms decide to commercialize ideas into better-quality products, mid-size firms will do both product innovation and patenting, and very large firms will file or acquire patents but not upgrade their products on the market. Hence, while larger firms hinder creative destruction more, they are less active in product innovation.[28]

[28]Although the main features of the model - the reduction in product innovation incentives as firms grow

Production.—Consider a partial equilibrium framework that describes innovation in a single sector. There are M potential producers, and aggregate output is a combination of quality-weighted varieties:

$$Y = \frac{1}{1-\beta}\left[\sum_{m=1}^{M} q_m^{\frac{\alpha}{1-\beta}} y_m\right]^{1-\beta}, \quad 0 < \alpha < \beta < 1 \tag{8}$$

where y_m denotes the quantity and q_m is the quality level of variety m. This specification implies that products from different producers are perfect substitutes after adjusting for their qualities. The parameter α captures the consumer's satiation with respect to additional quality. Labor is the only factor of production. Producers use labor to produce output by hiring labor at the common wage rate of w. Output of variety m is then given by $y_m = l_m$, where l_m is labor used to produce variety m. We assume that the overhead cost of production ϵ must be paid before choosing prices and output. Since producers' marginal costs are the same and qualities are different, under Bertrand competition, this overhead cost allows the highest-quality firm to win the market and act as a monopolist.[29]

The monopolist maximizes profits by choosing the price of its product subject to demand from (8),[30] which delivers the following equilibrium objects for output (y), sales (R), and profits (Π), respectively (hereafter, we drop the subscript m):

$$y = \frac{1-\beta}{\beta}\frac{\pi}{w}q^{\gamma}, \qquad R = \frac{\pi}{\beta}q^{\gamma}, \qquad \Pi = \pi q^{\gamma}, \tag{9}$$

where $\pi \equiv \beta\left(\frac{1-\beta}{w}\right)^{\frac{1-\beta}{\beta}}$ and $\gamma \equiv \frac{\alpha}{\beta}$. Hence, firms with higher-quality products are larger, and generate higher sales and profits. Moreover, since $\gamma < 1$ because $0 < \alpha < \beta < 1$, the marginal quantity, sales, and profits decrease with quality.

Dynamic choices: product innovation and patenting.—Now consider the once-in-a-lifetime decision of product quality upgrade and patenting for an incumbent with quality q who exogenously obtains an idea of size λ.[31] If the firm decides to upgrade the quality of

and the increase in protection incentives through patenting - can be easily generalized, the tractability of the model comes at the expense of richer dynamics and a more realistic cost structure.

[29] This assumption simplifies the setup. Alternatively, we would need to work with limit pricing, where the firm with highest quality would still capture the entire market, but the price would be determined by the price of the second highest quality producer.

[30] Price of Y is normalized to one.

[31] For simplicity, we assume that this is a one-time choice. Hence, the idea is either used or disappears afterwards. A more-complete approach with a dynamic decision of patenting and product innovation would bring similar tradeoffs at the expense of tracking the evolution of a firm's position both in the product and

its product from q to $q+\lambda$ to generate higher profits, it also has to pay the costs of product development and commercialization c_m. In this case, $q+\lambda$ becomes the largest available quality in the economy. Simultaneously, the firm can also decide to patent the blueprint at a cost c_p.[32] A patent grants the firm additional protection against being replaced by an entrant (more details below). If the firm decides to patent, even if the idea is not commercialized, the highest-available quality in the economy becomes $q+\lambda$ since, by patenting, the firm makes the idea "public". Note that the highest-available quality in the economy could be different than that commercialized by the firm and available to the consumers. In this sense, firms' activities in the product and patent spaces are separated. Product innovation does not necessarily imply patenting activity, and neither does introducing a patent necessarily imply product innovation.

Entry: creative destruction.—Incumbents can be replaced by entrants through creative destruction. The model includes an exogenous arrival rate of entrants at each instant p. Entrants build on "the shoulders of giants" and can replace incumbents by improving upon the highest-quality product available in the economy. The underlying assumption is that entrants can learn from products available in the market and from patents. Hence, "the shoulders of giants" correspond to $q+\lambda$ unless the incumbent neither upgrades nor patents, in which case the highest available quality is q. Entrants draw innovation of step size λ^e from a uniform distribution on $(0,1)$. Patenting protects the quality level of incumbents $q+\lambda$ by creating a "wall" of height $\varepsilon > 0$ that entrants need to overcome to enter the market. The parameter ε captures the condition that entrants need to come up with an innovation sufficiently different from what has been patented before, which can depend on the strength of intellectual property protections as well as the scope of the patent. Given these assumptions, the probability of creative destruction is p if the incumbent does not patent and is $p(1-\varepsilon)$ if the incumbent patents (Appendix D provides the proof). Notice that, in contrast to standard models of creative destruction, not all product quality improvements by entrants will find their way to the market. The separation between the patent space and the product space introduces the possibility that a better-quality product is not introduced to the market because it is blocked by existing patents.

Value functions and equilibrium.—Let us denote the value of a firm with existing product quality q that both upgrades quality and files patents as V^{11}, the value of only

patent spaces.

[32]We think of c_p as the combination of research, legal filing, and potential patent enforcement costs.

upgrading as V^{10}, the value of only patenting as V^{01}, and the value neither upgrading nor patenting as V^{00}. Then, the value function of the incumbent firm is

$$V(q) = \max\Big\{ \; V^{11}(q) - c_m - c_p, \; V^{10}(q) - c_m, \; V^{01}(q) - c_p \;, \; V^{00}(q) \; \Big\}, \tag{10}$$

where

$$V^{11}(q) = \frac{\pi(q+\lambda)^{\gamma}}{r + p(1-\varepsilon)}, \qquad V^{10}(q) = \frac{\pi(q+\lambda)^{\gamma}}{r+p},$$
$$V^{01}(q) = \frac{\pi q^{\gamma}}{r + p(1-\varepsilon)}, \qquad V^{00}(q) = \frac{\pi q^{\gamma}}{r+p}.$$

Notice that incentives for product innovation decline as firm size increases, while the returns to patenting increase. Because marginal profits decrease as q increases, the incremental returns from product innovation decline with firm size, which is the same intuition that underlies the well-known *Arrow-replacement effect*. This effect describes how larger firms and monopolists find it less profitable to replace themselves: innovations might cannibalize their own rents.[33]

On the other hand, the returns to patenting increase with size as larger firms have a higher value to protect. In fact, we show in Appendix D that under mild conditions on costs, in this economy there exist cutoffs q^* and q^{**} such that a firm only upgrades when $q < q^*$, engages in both product innovation and patenting in the intermediate region $q^* < q < q^{**}$, and only patents when $q > q^{**}$.[34] As a result, the model delivers an equilibrium that rationalizes the main empirical patterns uncovered in the previous sections:

Implication 1: Many firms (below cutoff q^*) develop product innovations without patenting.

Implication 2: On average, patenting and product innovation are positively correlated.

Implication 3: Larger firms develop relatively fewer product upgrades, but file more patents.

The model also speaks to our empirical facts on the relationship between patenting, firm size, and competition. By construction, patents in the model reduce creative destruction. At the same time, larger firms rely on patenting more. Hence, larger firms also face lower risks of creative destruction.

[33] We provide an empirical estimate of γ and confirm that it is lower than one. However, instead of these decreasing returns that generate a declining relationship between size and innovation, one can generate it through other ways, such as by implementing weaker scalability of R&D technology with increasing size as in Akcigit and Kerr (2018) or an innovation-advertising tradeoff as in Cavenaile and Roldan (2019).

[34] The required conditions on the costs c_m and c_p ensure that at least one firm finds it profitable to introduce a product and at least one firm finds it too costly to patent.

Implication 4: Patents deter future entry by competitors; larger firms deter entry more.

The value of a patent.—If a firm in our model were to sell its patent, what would be its price? Patents embed both productive and protective values. Both values come from the underlying technological innovation contained in the patent. Productive value comes from the option value of commercializing an innovation, while protective value comes from the ability of a patent to protect firms' market lead from competitors. We define the (private) value of a patent as the revenue premium that a patented innovation provides:

$$\begin{aligned} Patent\ Value &= V^{11} - V^{00} \\ &= \underbrace{\frac{\pi(q+\lambda)^{\gamma}}{r+p(1-\varepsilon)} - \frac{\pi(q+\lambda)^{\gamma}}{r+p}}_{Protective} + \underbrace{\frac{\pi(q+\lambda)^{\gamma}}{r+p} - \frac{\pi q^{\gamma}}{r+p}}_{Productive} \end{aligned} \tag{11}$$

The total patent value can be decomposed into productive and protective components by adding and subtracting V^{10}.[35] Productive value is the revenue premium from commercializing a product of upgraded quality if we hold creative destruction fixed. This value from product innovation declines as firms grow, since the same amount of innovation brings marginally lower returns. In contrast, protective value, which is the revenue premium from lower creative destruction holding the technology of a firm fixed, increases as firms grow: the use of patent protection is more relevant as the value of the firm increases. Hence, we formulate our final implication of the model:

Implication 5: The revenue premium from patents comes both from product upgrades and protection. The latter becomes more important as firms grow.

We now set the parameters of the model to estimate the average value of a patent for firms in our data. To estimate (11), we need to assign values to π, λ, γ, p, and ε. First, we normalize the average quality within each product category in our data to one. Notice that we do not observe profits, but given (9), we know that sales are proportional to profits such that $\Pi = \frac{\mu-1}{\mu} \times R$, where μ is the markup. The profit of an average firm is then $\frac{\mu-1}{\mu}$ multiplied by the sales of the average firm, which we take to be equal to the average yearly sales of all firms across all product categories (1.36 million USD).[36] We take $\mu = 1.21$ drawing upon

[35] An alternative decomposition would be $\underbrace{\frac{\pi(q+\lambda)^{\gamma}}{r+p(1-\varepsilon)} - \frac{\pi q^{\gamma}}{r+p(1-\varepsilon)}}_{Productive} + \underbrace{\frac{\pi q^{\gamma}}{r+p(1-\varepsilon)} - \frac{\pi q^{\gamma}}{r+p}}_{Protective}$.

[36] All nominal values are deflated to 2015 dollars.

Barkai (2017)'s average estimate of markups in the U.S. economy in 2014. Assigning values to p and $p(1-\varepsilon)$ involves the following considerations. In the model, if firms do not innovate they face a creative destruction rate that leads to the decline of their expected sales. Hence, we infer the values of p from firms' growth in sales when they do not introduce new products in a given year. In our data, the median firm that does not hold any patents suffers a loss in sales when it does not introduce new products: log sales change is equal to -10.3%. This decline in sales is attenuated if a firm holds a patent, thus giving an estimate for ε. The implied values for creative destruction are $p = 0.098$ and $p(1-\varepsilon) = 0.095$.

Lastly, we jointly estimate λ and γ. Intuitively, λ determines the average growth when the firm innovates and γ affects how this growth varies with firm size. Specifically, the model implies the following relationship between firm growth and relative size, conditional on product innovation:

$$\Delta \ln R_t = \gamma ln\Big(1 + \lambda\Big[\frac{R_{t-1}}{\bar{R}_{t-1}}\Big]^{-\frac{1}{\gamma}}\Big)$$

We estimate this relationship with a non-linear least squares regression applied to the sample of firms who introduce new products in that year. We define the relative size of firms as sales divided by the average sales of firms in that year and product category. The resulting estimates are $\gamma = 0.899$ (s.e. 0.364) and $\lambda = 0.024$ (s.e. 0.008).

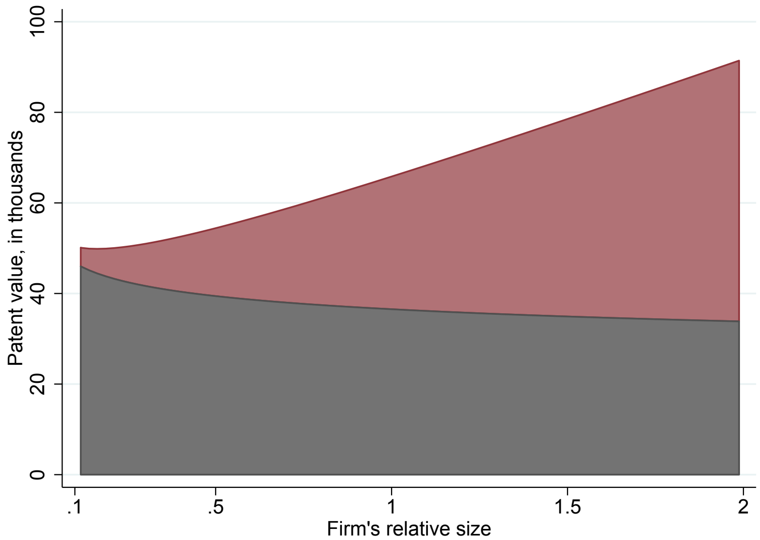
Figure 5 plots the value of a patent against the relative size of firms. The red shaded area depicts the contribution of the protective component of the patent's value, and the gray area depicts the contribution of the value's productive component. For the average firm the value of a patents is around \$65,000.[37] The estimated value increases drastically as firm size increases, mainly due to the contribution of the protective value. For example, for firms ten times smaller than the average firm, only 9% of the value comes from the protective component, while for firms that are twice as large as the average firm, the protective component accounts for 60% of a patent's value.

Our methodology for estimating the value of the patent differs greatly from those used previously in the literature as it relies on the structure of our model, the matched data we constructed between patents and products, and the realized sales of products observed in the data.[38] Nonetheless, we find that our estimates are well in the range of other estimates

[37]Notice that our calculations do not include sales from the stores not covered by Nielsen. To get at the nationwide sales, we can roughly scale our sales twice (see Appendix A.1 for details).

[38]The literature has used various methods to estimate the value of a patent: direct survey questions, inference from observed patent renewals by firms, stock market responses to patent news, as well as direct

Figure 5: Estimated patent value



reported in the literature. Using patent renewal information to infer the private value of U.S. patents issued in 1991, Bessen (2008) estimates a patent's mean value to be $121,000 (median $11,000). Interestingly, consistent with our results, Bessen (2008) also finds that the value of patents held by smaller firms is lower, while litigated patents are more valuable. Serrano (2010) estimates the average private value of holding a patent to be $90,799 (median $19,184). Using data from a large non-practicing entity, which presumably holds mostly valuable patents, Abrams et al. (2013) find that the mean value of a patent is $235,723 (median $47,955).

The advantage of our methodology is that it allows us to decompose the patent value into its two inherent components – productive and protective. The decomposition uncovers the dual role of patenting and the role each component plays for firms of different sizes.

7 Conclusion

Using textual analysis of patent documents and product descriptions, we construct a new patent-to-products data set to study the relationship between patents and product innovation. We find that more than half of the product innovation is not associated with patents. Nonetheless, patent filing is positively associated with subsequent product innovation by firms, on average. We document substantial heterogeneity in this relationship. Patents filed

estimates from patent sales samples. For a comprehensive review, see Hall and Harhoff (2012).

by larger firms reflect less the actual product innovation than other patents do. Instead, we find strong evidence suggesting that the main role of patents for market leaders is to deter product innovation of competitors and protect sales of their existing products. Hence, our results indicate that although on average patents capture product innovation in the market, because the relationship between patents and innovation changes with firm size, patent-based measures distort the differences in actual innovation between firms of different sizes.

Using a simple theoretical framework, we show that for the same patented idea, a larger firm can reap a greater monetary return than a smaller firm can. However, for these large firms, more of this return is derived from the patent's ability to hinder competition than is derived from commercializing a new product using the patented idea. We argue that understanding the contribution of the productive and protective components of patenting and how they vary by firm size has important implications for our understanding of growth, innovation, and intellectual property policy. Specifically, policymakers should pay more attention to the state-dependence of patent and R&D-based policies; these policies should acknowledge that firms' incentives to use intellectual property vary greatly with firm size and market leadership.

In their comprehensive analysis of patent reform in the U.S., Jaffe and Lerner (2004) argue that the seemingly innocent changes in patent policies in the early 1980s significantly affected firms' incentives toward strategic patent filing. In line with this, the survey results published by Cohen et al. (2014) suggest that relative to the early 1980s, large firms now rely somewhat more heavily on patents to protect their sales from competitors. A potential avenue for future research is to understand how large firms' increasing reliance on protective patenting has contributed to the recent trends of increasing dominance of large firms and declining business dynamism (Decker et al., 2016; Akcigit and Ates, 2019).

References

Abrams, David S, Ufuk Akcigit, and Jillian Grennan, "Patent Value and Citations: Creative Destruction or Strategic Disruption?," NBER Working Paper 19647 November 2013.

Aghion, Philippe and Peter Howitt, "A Model of Growth through Creative Destruction," *Econometrica*, 1992, *60* (2), 323–51.

Aizawa, Akiko, "An information theoretic perspective of TF-IDF measures," *Information Processing Management*, 01 2003, *39*, 45–65.

Akcigit, Ufuk and Sina T Ates, "What Happened to U.S. Business Dynamism?," Working Paper 25756, National Bureau of Economic Research April 2019.

_ and William R. Kerr, "Growth through Heterogeneous Innovations," *Journal of Political Economy*, 2018, *126* (4), 1374 – 1443.

_ , Douglas Hanley, and Stefanie Stantcheva, "Optimal Taxation and RD Policies," Working Paper 22908, National Bureau of Economic Research December 2016.

_ , Murat Celik, and Jeremy Greenwood, "Buy, Keep, or Sell: Economic Growth and the Market for Ideas," *Econometrica*, 2016, *84*, 943–984.

_ , Salome Baslandze, and Francesca Lotti, "Connecting to Power: Political Connections, Innovation, and Firm Dynamics," NBER Working Paper 25136 October 2018.

Alexopoulos, Michelle, "Read All about It!! What Happens Following a Technology Shock?," *American Economic Review*, June 2011, *101* (4), 1144–79.

Argente, David and Chen Yeh, "Product's Life-Cycle, Learning, and Monetary Shocks," 2017.

_ , Munseob Lee, and Sara Moreira, "The Life Cycle of Products: Evidence and Implications," 2020.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen, "The Fall of the Labor Share and the Rise of Superstar Firms," Working Paper 23396, National Bureau of Economic Research May 2017.

Balasubramanian, Natarajan and Jagadeesh Sivadasan, "What Happens When Firms Patent? New Evidence from U.S. Economic Census Data," *The Review of Economics and Statistics*, 2011, *93* (1), 126–146.

Barkai, Simcha, "Declining Labor and Capital Shares," *LBS working paper*, 2017.

Bena, Jan and Elena Simintzi, "Globalization of work and innovation: Evidence from doing business in china," Technical Report, Discussion paper 2017.

Bessen, James, "The value of U.S. patents by owner and patent characteristics," *Research Policy*, 2008, *37* (5), 932–945.

Bloom, Nicholas, Charles I Jones, John Van Reenen, and Michael Webb, "Are Ideas Getting Harder to Find?," NBER Working Paper 23782 September 2017.

—, **Mark Schankerman, and John Van Reenen**, "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, 2013, *81* (4).

Blundell, Richard, Rachel Griffith, and John Van Reenen, "Market Share, Market Value and Innovation in a Panel of British Manufacturing Firms," *The Review of Economic Studies*, 07 1999, *66* (3), 529–554.

Bresnahan, Timothy F and Robert J Gordon, "Introduction to" The Economics of New Goods"," in "The Economics of New Goods," University of Chicago Press, 1996, pp. 1–26.

Broda, Christian and David E Weinstein, "Product Creation and Destruction: Evidence and Price Implications," *American Economic Review*, 2010, *100* (3), 691–723.

Campello, Ricardo JGB, Davoud Moulavi, and Jörg Sander, "Density-based clustering based on hierarchical density estimates," in "Pacific-Asia conference on knowledge discovery and data mining" Springer 2013, pp. 160–172.

Cavenaile, Laurent and Pau Roldan, "Advertising, innovation and economic growth," Working Papers 1902, Banco de Espana;Working Papers Homepage February 2019.

Cockburn, Iain and Megan J. MacGarvie, "Entry and Patenting in the Software Industry," *Management Science*, 05 2011, *57*, 915–933.

Cohen, Lauren, Umit Gurun, and Scott Duke Kominers, "Patent Trolls: Evidence from Targeted Firms," Working Paper 20322, National Bureau of Economic Research August 2014.

Cohen, Wesley M. and Steven Klepper, "Firm Size and the Nature of Innovation within Industries: The Case of Process and Product RD," *The Review of Economics and Statistics*, 1996, *78* (2), 232–243.

Cohen, Wesley M, Richard R Nelson, and John P Walsh, "Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not)," Technical Report, National Bureau of Economic Research 2000.

Crouzet, Nicolas and Janice C Eberly, "Understanding Weak Capital Investment: the Role of Market Concentration and Intangibles," NBER Working Paper 25869 May 2019.

Cunningham, Colleen, Song Ma, and Florian Ederer, "Killer Acquisitions," *Academy of Management Proceedings*, 2018, *2018* (1), 11001.

de Rassenfosse, Gaétan, "Notice failure revisited: Evidence on the use of virtual patent marking," Technical Report, National Bureau of Economic Research 2018.

Dechezlepretre, Antoine, Elias Einio, Ralf Martin, Kieu-Trang Nguyen, and John Van Reenen, "Do tax incentives for research increase firm innovation? An RD Design for RD," Working Papers 73, VATT Institute for Economic Research 2016.

Decker, Ryan A., John Haltiwanger, Ron S. Jarmin, and Javier Miranda, "Declining Business Dynamism: What We Know and the Way Forward," *American Economic Review*, May 2016, *106* (5), 203–07.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv preprint arXiv:1810.04805*, 2018.

Feenstra, Robert C, "New product varieties and the measurement of international prices," *The American Economic Review*, 1994, pp. 157–177.

Fellbaum, Christiane, "Princeton university: About WordNet," 2010.

Garcia-Macia, Daniel, ChangâĂŘTai Hsieh, and Peter J. Klenow, "How Destructive is Innovation?," *Econometrica*, 2019, *87* (5), 1507–1541.

Gilbert, Richard J. and David M. G. Newbery, "Preemptive Patenting and the Persistence of Monopoly," *The American Economic Review*, 1982, *72* (3), 514–526.

Goolsbee, Austan D and Peter J Klenow, "Internet rising, prices falling: Measuring inflation in a world of e-commerce," in "Aea papers and proceedings," Vol. 108 2018, pp. 488–92.

Gordon, Robert J., "The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War.," *Journal of Regional Science*, 2016.

Graham, Stuart JH, Cheryl Grim, Tariqul Islam, Alan C Marco, and Javier Miranda, "Business dynamics of innovating firms: Linking US patents with administrative data on workers and firms," *Journal of Economics & Management Strategy*, 2018, *27* (3), 372–402.

Griliches, Zvi, "Issues in Assessing the Contribution of Research and Development to Productivity Growth," *Bell Journal of Economics*, Spring 1979, *10* (1), 92–116.

_ , "Market value, RD, and patents," *Economics Letters*, 1981, *7* (2), 183 – 187.

Gutiérrez, Germán and Thomas Philippon, "Declining Competition and Investment in the U.S.," Working Paper 23583, National Bureau of Economic Research July 2017.

Hall, Bronwyn, Adam Jaffe, and Manuel Trajtenberg, "The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools," NBER Working Paper 8498 2001.

Hall, Bronwyn H., Adam Jaffe, and Manuel Trajtenberg, "Market Value and Patent Citations," *RAND Journal of Economics*, Spring 2005, *36* (1), 16–38.

_ **and Dietmar Harhoff**, "Recent Research on the Economics of Patents," *Annual Review of Economics*, 2012, *4* (1), 541–565.

Haltiwanger, John, Ron Jarmin, and Javier Miranda, "Who Creates Jobs? Small versus Large versus Young," *The Review of Economics and Statistics*, 2013, *95* (2), 347–361.

Heller, Michael A. and Rebecca S. Eisenberg, "Can Patents Deter Innovation? The Anticommons in Biomedical Research," *Science*, 1998, *280* (5364), 698–701.

Hoberg, Gerard and Gordon Phillips, "Text-based network industries and endogenous product differentiation," *Journal of Political Economy*, 2016, *124* (5), 1423–1465.

Hottman, Colin J, Stephen J Redding, and David E Weinstein, "Quantifying the sources of firm heterogeneity," *The Quarterly Journal of Economics*, 2016, *131* (3), 1291–1364.

Jaffe, Adam B. and Josh Lerner, "Innovation and Its Discontents: How Our Broken Patent System is Endangering Innovation and Progress, and What to Do About It.," *Princeton University Press, Princeton, NJ*, 2004.

Joulin, Armand, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov, "Bag of Tricks for Efficient Text Classification," in "Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers" Association for Computational Linguistics April 2017, pp. 427–431.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy, "Measuring Technological Innovation over the Long Run," NBER Working Papers 25266 November 2018.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, "Technological Innovation, Resource Allocation, and Growth*," *The Quarterly Journal of Economics*, 03 2017, *132* (2), 665–712.

Lai, Ronald, Alexander D'Amour, David M. Doolin, Guan-Cheng Li, Ye Sun, Vetle Torvik, Amy Yu, and Lee Fleming, "Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database," *Research Policy*, 2014, *43* (6), 941–955.

Lampe, Ryan and Petra Moser, "Patent Pools, Competition, and Innovation–Evidence from 20 US Industries under the New Deal," *The Journal of Law, Economics, and Organization*, 08 2015, *32* (1), 1–36.

Lanjouw, Jean and Mark Schankerman, "Characteristics of Patent Litigation: A Window on Competition," *RAND Journal of Economics*, 2001, *32* (1), 129–51.

Lloyd, Stuart, "Least squares quantization in PCM," *IEEE transactions on information theory*, 1982, *28* (2), 129–137.

Loecker, Jan De and Jan Eeckhout, "The Rise of Market Power and the Macroeconomic Implications," Working Paper 23687, National Bureau of Economic Research August 2017.

Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze, *Introduction to Information Retrieval*, New York, NY, USA: Cambridge University Press, 2008.

Moser, Petra, "Innovation without Patents: Evidence from World's Fairs," *Journal of Law and Economics*, 2012, *55* (1), 43 – 74.

Öscar Jorda, "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, March 2005, *95* (1), 161–182.

Pakes, Ariel, "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 1986, *54* (4), 755–784.

Romer, Paul M, "Endogenous Technological Change," *Journal of Political Economy*, October 1990, *98* (5), 71–102.

Sampat, Bhaven and Heidi L. Williams, "How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome," *American Economic Review*, 2019, *109* (1), 203–36.

Schankerman, Mark and Ariel Pakes, "Estimates of the Value of Patent Rights in European Countries During the Post-1950 Period," *The Economic Journal*, 1986, *96* (384), 1052–1076.

Serrano, Carlos J., "The dynamics of the transfer and renewal of patents," *The RAND Journal of Economics*, 2010, *41* (4), 686–708.

Shapiro, Carl, "Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting," *Innovation Policy and the Economy*, 2000, *1*, 119–150.

Trajtenberg, Manuel, "A Penny for Your Quotes: Patent Citations and the Value of Innovations," *The RAND Journal of Economics*, 1990, *21* (1), 172–187.

Williams, Heidi L., "Intellectual Property Rights and Innovation: Evidence from the Human Genome," *Journal of Political Economy*, 2013, *121* (1), 1–27.

Younge, Kenneth and Jeffrey M. Kuhn, "Patent-to-Patent Similarity: A Vector Space Model," in "in" 2016.

A Additional Data Information

A.1 Product Data

Data Details.—The main advantage of the RMS data set is its size and coverage. Overall, the RMS data consists of more than 100 billion unique sales observations at the week × store × UPC level. The data set comprises around 12 billion transactions per year which are worth $220 billion dollars on average. Over our sample period, 2006-2015, the total sales across all retail establishments are worth approximately $2 trillion and represent 53% of all sales in grocery stores, 55% in drug stores, 32% in mass merchandisers, 2% in convenience stores, and 1% in liquor stores. A key distinctive feature of this database is that the collection points include more than 40,000 distinct stores from around 90 retail chains, across 371 metropolitan statistical areas (MSAs) and 2,500 counties. As a result, the data provide good coverage of the universe of products and firms in the CPG sector. In comparison to other scanner data sets collected at the store level, Nielsen RMS covers a much wider range of products and stores. In comparison to scanner data sets collected at the household level, Nielsen RMS also has a wider range of products because it reflects the universe of all transactions for the categories it covers, as opposed to the purchases made by a sample of households.

For each product in a year, we define its sales as the total sales across all stores and weeks in the year. Likewise, quantity is defined as total quantities sold across all stores and weeks in the year. Price is defined by the ratio of revenue to quantity, which is equivalent to the quantity-weighted average price.[39] To minimize concerns about potential measurement error caused by Nielsen's treatment of private-label products to protect the identity of the retailers, we exclude all private-label goods from the data.

Nielsen Product Classification system.—The data is organized into 1,070 detailed product modules that are aggregated into 114 product groups. The product groups are then grouped into 10 major departments. The ten major departments are: Health and Beauty Aids, General Merchandise, Dry Grocery (e.g., baby food, canned vegetables), Frozen Foods, Dairy, Deli, Packaged Meat, Fresh Produce, Non-Food Grocery, and Alcohol. For example, a 31-ounce bag of Tide Pods has UPC 037000930389, is produced by Procter & Gamble, and belongs to the product module "Detergent-Packaged" in product group "Detergent," which belongs to the "Non-Food Grocery" department. The product group "Detergent" includes several product modules, including automatic dishwasher compounds, detergents heavy duty liquid, detergents light duty, detergents packaged, dishwasher rinsing aids, and packaged soap.

Defining a Product.—Defining products by their UPCs has some important advantages. First, UPCs are by design unique to every product: changes in any attribute of a good (e.g.

[39] We use the weight and the volume of the product to compute unit values.

forms, sizes, package, formula) result in a new UPC.[40] This offers a unique opportunity for economists to identify products at the finest level of disaggregation. Second, UPCs are so widespread that our data is likely to cover all products sold in the consumer goods sector. Producers have a strong incentive to purchase UPCs for all products that have more than a trivial amount of sales because the codes are inexpensive and they allow sellers to access stores with scanners and internet sales.

Assigning Products to Firms.— Nielsen RMS data does not include information on manufacturing firms. However, products can be linked with firms using information obtained from the GS1 US Data Hub. In order to issue a UPC, firms must first obtain a GS1 company prefix. The prefix is a five- to ten-digit number that identifies firms in their products' UPCs. Argente, Lee and Moreira (2020) provide more details on how to use a subset of the product UPCs to link producers with products.

The GS1 data include the name and address of the firm associated with each prefix, which allows us to append a firm name and location to the UPCs include in the Nielsen-RMS data. A "firm" in the database is defined based on the entity that purchased the barcodes from GS1, which is typically the manufacturer, such as Procter & Gamble.

Classifying firms into CPG-only firms.— Any firm that produces at least one product in the Nielsen RMS data is included in our analysis. We refer to these as CPG firms. However, some of these CPG firms also produce products outside the CPG sector (e.g. Toshiba, Samsung, Whirlpool), while others produce mostly products included in the Nielsen RMS data (e.g. Procter & Gamble, Kimberly Clark, Kraft). Part of our analysis is focused on identifying firms that are solely in the CPG sector. Inspired by Hoberg and Phillips (2016), we use the firm's 10-K reports, which are available from Compustat. The 10-K is a comprehensive summary of a firm's performance that must be submitted annually to the Securities and Exchange Commission, in addition to the annual report. It includes an overview of the firm's main operations, including its products and services. After merging our data with this information from Compustat, we manually classify each publicly traded CPG firm into CPG-only and not CPG-only firms by judging how much of the firmâĂŹs products and services fall outside the CPG sector.

We matched 270 publicly traded companies over our sample period; we classify 23% of them as CPG-only firms.

A.2 Patent Data

Data Details.—Unlike other standard patent data sources such as NBER patent data (Hall et al., 2001) and the data from the Harvard Dataverse Network (Lai et al., 2014), we make use of all patents published in the USPTO, including non-granted patent applications. Using

[40] Firms have strong incentives not to reuse UPCs. Assigning more than one product to a single UPC can interfere with a store's inventory system and pricing policies; it is rare that a meaningful quality change occurs without resulting in a UPC change.

all patent applications, as opposed to just granted applications, offers us two advantages. First, since patents are usually granted with a lag of roughly two years, the more recent years of the sample suffer from severe truncation. Looking at all patent applications alleviates this problem. Second, we can then differentiate between patents that are granted, pending, or abandoned. We use this as one of the patent quality measures, as discussed below. Adding non-granted patent information increases the number of patents in our sample by 1.7 million.

Assigning Patents to Firms.—We begin by selecting all patents that have a valid assignee name.[41] We assign patents to their most recent assignee(s). For this assignment, we use the *current* assignee variable from the USPTO (as of 2017 – our patent data vintage). The current assignee variable is missing for some of the patents included in our sample. In such a case, we start with the name of the original assignee and leverage the USPTO reassignment data to track any change of patent ownership due to a patent sale or firm reorganizations. To further track patent ownership through corporate reorganizations, we rely on Thomson Reuters Mergers & Acquisition data. Our underlying assumption is that patent ownership is transferred to the acquiring firm in case of corporate reorganization. Thomson Reuters M&A provides complete coverage of global mergers and acquisitions activity, including more than 300,000 US-target transactions, since 1970. The data covers mergers of equals, leveraged buyouts, tender offers, reverse takeovers, divestitures, stake purchases, spinoffs , and repurchases. It also provides detailed information about the target, the acquirer, and the terms of the deal. This comprehensiveness is particularly important given that firms that appear both in Nielsen data and USPTO are most likely large firms that undergo many corporate reorganizations.

Product-related and Process-related Patents.—Following Bena and Simintzi (2017), we create proxies for product-related patents and process-related patents based on the formal claims included in patent applications. Patent claims define the scope of a patent's protection and hence represent the essence of a patent application. On average, patents in USPTO have around 15 claims. Some of these are independent claims, while others derive from them. Claim texts are written in technical terms and often have a rigorous semantic structure.

The formulaic nature of claims gives us an opportunity to create the following simple classification. We say the claim is a process claim if the claim text starts with "method" phrases ("Method for", "Method of", "Method in", "Method define", and the like) or "process" phrases ("Process for", "Process according", "Process in", and the like). Then, as a baseline, we classify a patent as a process patent if the main (usually, the first) claim of the patent is a process claim. The patent is a product patent if it is either a design patent or a non-process utility patent. In the latter case, claims often start with words like "Apparatus", "Device", and the like). According to this definition, up to 70% of patents are product-related patents. We also tested an alternative definition that defines process patents based on the criteria that the share of process claims is larger than 50%. These two measures are highly correlated

[41] This step eliminates patents assigned to individuals as well as other patents that are missing assignee information, which mostly constitute pending patents.

(0.74) and our results based on the baseline variable are robust to this alternative definition.

A.3 Algorithm of *Match 1*

Firm name Cleaning Algorithm.—We assign each company name to a unique company identifier using the following procedure.

Step 1. In the first step, we run all company names through a name-standardization routine to generate unique company identifiers. Our routine is the following.

(1) After capitalizing all letters, we keep the first part of the company name before the first comma. (2) We remove leading and trailing instances of "THE", we replace different spellings of "AND" words with "&", and replace accented or acute letters with regular ones. (3) We remove special characters. (4) We standardize frequent abbreviations using dictionaries from the NBER Patent Data Project. For example "PUBLIC LIMITED" or "PUBLIC LIABILITY COMPANY" become "PLC";"ASSOCIATES" or "ASSOCIATE" become "ASSOC"; "CENTER" or "CENTRAL" become "CENT". (5) We delete trailing company identifiers. (6) If the resulting string is null, we protect it. (7) We repeat the previous steps on the original company names except for protected strings, for which we now keep the whole string and not just the first portion before the comma. (8) If the string is protected, we remove company identifiers in any place of the string (not just if trailing as in 5). (9) We remove spaces to further decrease misspellings. (10) We assign unique company identifiers based on the cleaned names.

Step 2. In addition to the extensive cleaning in Step 1, we take advantage of a "dictionary" that resulted from a large effort undertaken within the NBER Patent Data Project. After manual checks and searches of various company directories to identify name misspellings and various company reorganizations, the NBER files provide a mapping between patent assignee names and unique company identifiers (*pdpass*). Although this data is based on the assignees of granted patents before 2006, we use this mapping as a "dictionary" that we use in conjunction with our results from Step 1. This helps us leverage both our algorithm from Step 1 and the NBER *pdpass* information, combining the strengths of each method to create new unique company identifiers.

For example, Siemens appears in the data with many different name variations. "SIEMNES AG", "SIEMANS ATKIENGESELLSCHAFT", and "SIEKENS AG" are just a few of such variations that Step 1 does not capture but the NBER files identify as names under the same *pdpass*. In such a case, we use *pdpass* identifiers to group the three firms together. On the other hand, the NBER file does not identify "SIEMENS CORP" "SIEMENS AG" and "SIEMENS" as the same company as the once referenced by the first three name variations above. In such a case, we use the unique identifiers from Step 1 to group these firms together. Finally, after combining information from NBER files with our cleaning after Step 1, we pool all six variations into one new company code.

Our algorithm builds upon proven algorithms from Hall, Jaffe and Trajtenberg (2001) and Akcigit, Celik and Greenwood (2016b). We also applied an extensive number of manual quality checks to our cleaning algorithm. For example, we identified the largest CPG firms, and for each firm we looked up the corresponding set of patents on *Google Patents* to verify that our matching algorithm was obtaining the same patents.

A.4 Algorithms of *Match 2*

A.4.1 *Summary of the Methods of Natural Language Processing*

For convenience, the following section summarizes general methods from natural language processing that we refer to throughout our description of the algorithms below.

i) Parsing Methods

We use 1-grams and 2-grams (single words and two-word phrases) as tokens. In general one could use n-grams, meaning distinct n-length phrases. For the types of documents we are interested in, however, meaningful and irreducible phrases having 3 or more words are quite rare. Also note that we will use the terms "word","term", and "token" interchangeably and these will refer to the set of 1-grams and 2-grams in all cases.

ii) Lemmatizer Methods

We use WordNetLemmatizer provided as part of the NLTK Python module (nltk.org), which utilizes the WordNet lexical database (`wordnet.princeton.edu`), to reduce words to their root forms by removing conjugations like plural suffixes (Fellbaum, 2010). For instance, the word "compounds" would be mapped to "compound".

iii) Word Vector Normalization

Patent (or product category) text documents are first converted into word vectors that indicate, for each term, how many times the term appears in a document. Each document vector is of length $\mathcal{M}$, which is the number of terms that we include in our vocabulary. The corpus of documents can then be represented by a very sparse matrix of term counts with elements c_{km}, where $k \in \{1, \ldots, K\} = \mathcal{K}$ represents the document (patent or a product category) and $m \in \{1, \ldots, M\} = \mathcal{M}$ represents the term.

We then use a word-based weighting scheme called total-frequency-inverse-document-frequency (tf-idf) to account for the fact that more common words tend to be less important and vice versa. A number of possible functional forms could be used here, but we choose the commonly used sublinear form

$$w_m = \log\left(\frac{K+1}{d_m+1}\right) + 1 \quad \text{where} \quad d_m = |\{k \in \mathcal{K} | c_{km} > 0\}|$$

Thus if a word appears in all documents, it is assigned a weight of one, while those appearing in fewer documents get larger weights, and this relationship is sublinear. For our weighting scheme, we use document frequencies from the patent data, as that corpus is considerably larger and less prone to noise.

Finally, we are left with a weighted, ℓ^2-normalized word frequency vector f_k for each document k, both on the patent and product side of our data, with elements

$$f_{km} = \frac{w_m c_{km}}{\sqrt{\sum_{m'} (w_m c_{km'})^2}}$$

A.4.2 *Step 1: Defining Product Categories*

We start by developing an intermediate categorization of Nielsen products into product categories that are more aggregated than product modules but less aggregated than product groups.

Step 1.a - Collect Representative Documents

For each low-level product classification from Nielsen (1,070 modules), we explored different sources of text that might allow us to characterize the modules. First, we studied sources of text within Nielsen. For example, we explored the use of product attributes from each UPC and we found that while informative, some characteristics are shared and not sufficiently different. Second, we explored sources of data outside Nielsen, like dictionaries and various websites. After many manual checks, we decided to use Wikipedia pages.

The main advantage of using Wikipedia entries is that they often include technical descriptions that use words that also appear patents texts and are comprehensive enough to cover all modules. The use of Wikipedia text to encode textual knowledge is already common in the machine learning literature. For instance, two of the most advanced word embeddings currently available, BERT (Google, Devlin et al., 2018) and fastText (Facebook, Joulin et al., 2017), use the entire Wikipedia corpus for training purposes, in a addition to large corpora of text from books and websites. While there are a number of papers in the economics literature that study Wikipedia, we are unaware of any such usage as a direct input into a separate analysis.

For each Wikipedia article, we construct a representative document that includes the title of the module (repeated 10 times), the title of the Wikipedia article (10 times), the entire text (1 time), and the first 10% text of the Wikipedia article (10 times).

Step 1.b - Create Representative Word Vectors

To create the representative word vector for each module, we (i) concatenate all the text; (ii) apply the parsing and lemmatizing algorithms describe above; (iii) exclude terms that appear in more than 80% of documents (to exclude words like "the" and "and"); (iv) and re-weight according to the "term-frequency-inverse-document-frequency" sublinear transformation described above.

Note that for modules that include multiple Wikipedia articles, we first vectorize each Wikipedia entry, then average these vectors together to avoid overweighting longer entries (in an ℓ^2-norm-preserving sense).

Step 1.c - Cluster Analysis

We aggregated these module vectors into clusters using the popular k-means clustering technique. k-means clustering (Lloyd, 1982) is used to find a partitioning of a vector space into clusters of similar vectors. This procedure allows one to specify the desired number of clusters K beforehand and yields a partitioning that minimizes the within-group vector variance, or the average squared distance from the cluster mean.

Letting x be a given module vector and S_i^K be a cluster i of a cluster set S^K, we choose our partitioning S^K so as to minimize

$$\sum_{i=1}^{K} \sum_{x \in S_i^K} ||x - \mu_i||^2, \text{ where } \mu_i = \frac{1}{|S_i^K|} \sum_{x \in S_i^K} x$$

In our main analysis, we use $K = 400$ clusters. This choice is supported by extensive manual checks and experimentation with alternative partitions. We first explore k-means clustering for $K = 100, 200, ..., 900$. We find that our baseline k-means clustering partitions the product space quite well, striking a balance between minimizing the differences of vectors within a cluster while maximizing the differences across clusters.

Additionally, we show that our clustering of the product space is robust. By experimenting with various other state-of-the-art clustering techniques such as HDBSCAN (Campello et al., 2013) – a hierarchical clustering algorithm that does not need substantial tuning – we conclude that many product modules are grouped together independently of the clustering method used.

Finally, the implied clustering also accords well with the external classification scheme from Nielsen. By comparing our partitioning to the original 114 group aggregation from Nielsen (not used an input in our clustering algorithm), we see that products clustered into the same product categories also fall into same groups defined by Nielsen.

The final product clustering groups together precisely capture those product categories that the patent matching algorithm would have trouble distinguishing between, and vice versa. for example, with this clustering, the separate product modules "Detergents – packaged", "Detergents – light duty", "Detergents – heavy duty", "Laundry treatment aids", and "Fabric washes – special" are grouped into one product category. The patent matching algorithm would struggle to accurately map a related patent to one of these modules, especially given that the same patent could plausibly lead to innovations in all of these product modules at the same time.

Step 1.d - Creating Pseudo Product Categories

We create additional pseudo product categories to describe products outside of the consumer goods sector. These pseudo-categories are designed purely to improve the match to consumer products as will be explained below and are not used in our main analysis. We selected a sufficiently large and diverse set of pseudo-categories by experimenting and studying patents held by firms in our sample that produce goods outside of the consumer goods sector. We add 19 of thesepseudo-categories to the existing 400 product categories in the data. Some examples include “computers” and “aviation”. As we did with the original modules, we create word vectors for each pseudo-module based on the associated set of Wikipedia articles the describe it.

Step 1.e - Word Vectors for Product Categories

The final word vector for product categories (including pseudo-product categories) simply combines the titles and word vectors (Step 1.b) of all modules that were clustered together to make a product category (Step 1.c).

A.4.3 *Step 2: Patent Vectors and Similarity Scores*

Step 2.1 - Collect Representative Documents for Patents

We use a variety of text fields to construct patent documents, including the title, abstract, international patent classification system description, and the titles of cited patents. We upweight the title of the patent by a factor of 5 compared to the abstract, because the title has a much higher signal-to-noise ratio than the other patent text fields. Specifically, a patent’s title tends to express the main application of the patent, whereas the abstract, description, and claims contain technical implementation details that are not as relevant for our purposes. For the same reasons we also upweight the patent classification description by a factor of 3.

Step 2.2 - Create Representative Vectors for Patents

To create the representative vector, we: (1) concatenate all the text; (2) apply parsing and lemmatizing algorithms (see description below); (3) exclude terms that appear in more than 80% of documents (excludes words like ”the” and ”and”); (4) and re-weight according to the ”term-frequency-inverse-document-frequency” sublinear transformation (see description above). Constructing representative documents on the patent side consists of simply concatenating all of the available text into one document. For the product categories, we first vectorize each Wikipedia entry, then average these vectors together to avoid overweighting longer entries.

Step 2.3 - Computing Similarity Scores Between Patents and Categories

At this point, we have the normalized word vectors for each product category j, f_{jm}, and the normalized word vectors for each patent p, f_{pm}. Multiplying any two such word vectors together yields the similarity score between two documents:

$$s_{jp} = \sum_{m \in \mathcal{M}} f_{jm} f_{pm},$$

where $\mathcal{M}$, as before, denotes size of a vector, which is the number of terms in the vocabulary. The similarity is guaranteed to lie in the range $[0, 1]$, with zero corresponding to zero word overlap and one corresponding to the case in which the documents are identical (or are multiples of one another). Notice that this vectorization approach (sometimes referred to as "bag of words") ignores any information about the order of words or phrases.

Thus, for each patent, we now have similarity metrics for each product category. The next section describes how we designate the matched product category for each patent.

A.4.4 *Step 3: Classifying Patents into Product Categories*

The final step of our patent-product matching algorithm consists in using the similarity scores to determine which pairs of patents and products are valid matches. Because some patents may correspond to certain general production processes – and not directly to products – or to products outside the consumer goods sector, we allow for the option that a patent is not assigned to any product category, or is a "non-match".

Step 3.1 - Threshold Similarity

We first adjust the algorithm to include a similarity score threshold below which we believe considering the two documents as similar would be too noisy. We tested different threshold levels and, in our baseline algorithm, we restrict the set of potential product categories for each patent p to product categories whose similarity score exceeds 0.025. For those patents that have less than five product categories satisfying this condition, we include the set of product categories that have the five highest similarity scores. For each patent, we denote the set of product categories satisfying these conditions as:

$$\Theta_p = \{j \in \Omega \mid s_{jp} > 0.025 \vee \text{rank}(s_{jp}) \leq 5\} \tag{12}$$

where Ω is the set of all product categories and s_{jp} is the similarity score between patent p and product category j.

Step 3.2 - Production Condition To further improve the match, we leverage firms' production information from Nielsen. For each patent, we define the set of potential matches, G_p, whose elements consist of all product categories in which the patenting firm ever sold a product, according to our product data.

$$G_p = \{j \in \Omega \mid p \text{ is patent of firm } i \;\wedge\; \sum_{t=2006}^{2015} \text{sales}_{ijt} > 0\}, \tag{13}$$

where $sales_{ijt}$ are the sales of firm i in product category j in year t. Note that this condition, will exclude all pseudo-categories and product categories that the firm never produced from the set of potential matches.[42]

[42] This makes it clear that having pseudo-categories helps to filter out many patents of the firms who heavily produce non-CPG products. For example, some firms like Toshiba or Samsung produce small electronics in our data, however they hold large portfolios of patents related to computer hardware or other high-tech technologies that are not relevant for the consumer products sector that we are analyzing. For such patents, the set Θ_p often consists only of pseudo-modules that then are easily filtered out by condition (13).

Step 3.3 - Select Maximum

Together, the criteria above imply that patent p will be classified as a "non-match" if none of its product categories satisfy the thresholds and the production conditions:

$$\Theta_p \wedge G_p = \emptyset$$

For the patents that have at least one product category satisfying those conditions, we assign the final patent-product category match j_p^* to be a product category with the highest similarity score:

$$j_p^* = \max_{j \in \Theta_p \wedge G_p} s_{jp} \tag{14}$$

This defines the matching of a patent p to the set of products grouped in the category j_p^*.

A.5 Robustness and Match Validation

A.5.1 Manual Checks of the Patent-Product Category Matches

We manually checked many patent-to-products matches and Table A.I lists some examples. The top 100 product categories sorted by their revenue and the largest firms selling in those categories are shown. For each firm, we then list an example of the highest-similarity patents in the corresponding product categories and their similarity scores. Comparing the titles of the patents and product categories, we see that product categories selected by our algorithm match the content of the patents well.

Table A.I: Patents with the Highest Similarity Score: Top Selling Firms by Categories

	Company	Product category	Application ID	Title of the Patent	Similarity
1	Philip Morris USA	Cigarette/smoking accessories	13912780	Cigarette and filter sub-assemblies with squeezable flavor capsule and method of manufacture	0.544838
2	Procter & Gamble	Diapers and baby powder	29396475	Absorbent article with a pattern	0.487175
3	Procter & Gamble	Laundry detergent	13905161	Laundry detergent composition	0.387514
4	Nikon	Camera	29385057	Projector equipped digital camera	0.33897
5	General Electric	Lamp	29283361	Lamp	0.427732
6	Coca-Cola USA	Soft drink	13816800	Phytase in ready-to-drink soft drink	0.307128
7	Procter & Gamble	Toilet	13585921	Method of reducing odor	0.191963
8	Procter & Gamble	Paper cup	11897767	Array of paper towel product	0.242879
9	Warner Home Video	Photographic film	10428440	Method of distributing multimedia presentation in different format on optical disc	0.08106
10	Procter & Gamble	Sanitary napkin	29465209	Absorbent article	0.204989
11	L'Oreal USA	Cosmetics	9987885	Anhydrous and water resistant cosmetic composition	0.305982
12	Procter & Gamble	Fabric softener	13070526	Method of making fabric softener	0.41355
13	Kimberly-Clark	Facial tissue	10034881	Method of making a high utility tissue	0.198823
14	Unilever USA	Soap	10320295	Soap wrapper	0.41769
15	L'Oreal USA	Hair coloring	14554789	Hair coloring appliance	0.455061
16	S.C. Johnson & Son	Air freshener	29438208	Dispenser	0.496183
17	Kraft Heinz Foods	Cheese	11618467	Method and system for making extruded portion of cheese	0.596449
18	Nestle Waters North America	Bottle	29434474	Water cooler	0.200115
19	The Hershey Company	Candy	9985948	Confectionary product low fat chocolate and chocolate like product and method for making them	0.282462
20	Procter & Gamble	Hair conditioner	12047712	Tool for separating a hair bundle	0.559868
21	Wm. Wrigley Jr.	Chewing gum	10453862	Method for making coated chewing gum product with a coating including an aldehyde flavor and a dipeptide sweetener	0.578689
22	Kimberly-Clark	Wet wipe	9965645	Wet wipe dispensing	0.506875
23	Procter & Gamble	Razor	29387316	Shaving razor package	0.54803
24	Activision Publishing	PC game	11967969	Video game forward compatibility including software patching	0.347854
25	Frito-Lay	Potato chip	11777839	Method for reducing the oil content of potato chip	0.521346
26	General Mills	Breakfast cereal	29183322	Layered cereal bar having cereal piece included thereon	0.28897
27	Abbott Laboratories	Milk	9910094	Powdered human milk fortifier	0.492503
28	Procter & Gamble	Toothpaste	11240284	Toothpaste dispenser toothpaste dispensing system and kit	0.388327
29	Procter & Gamble	Deodorant	12047430	Deodorant composition and method for making same	0.290906
30	The Minute Maid Company	Juice	12940252	Method of juice production apparatus and system	0.31321
31	Colgate-Palmolive	Toothbrush	11011605	Oral care implement	0.425624
32	Driscoll Strawberry Associates	Fruit	10722055	Strawberry plant named driscoll lanai	0.298149
33	The Duracell Company	Battery charger	10042750	Battery cathode	0.262253

Notes: The table presents information on the top 100 product categories sorted by their revenue. Each row reports the name of the highest-selling firm in a category together with an application ID and title of the firm's patent with the highest similarity score in the corresponding product category. The last column reports a similarity score from matching the patent to the category.

	Company	Product category	Application ID	Title of the Patent	Similarity
34	Alcon Laboratories	Disinfectant	9765234	Conditioning solution for contact lens care	0.362715
35	Pennzoil-Quaker State	Motor oil	10253126	Environmentally friendly lubricant	0.218752
36	Procter & Gamble	Oral hygiene	13150392	Method for whitening teeth	0.361255
37	Abbott Laboratories	Nutrition	10004360	Pediatric formula and method for providing nutrition and improving tolerance	0.108124
38	Anheuser-Busch InBev	Beer	12734356	Process for preparing a fermented beverage	0.419399
39	Procter & Gamble	Shampoo	12040980	Shampoo containing a gel network	0.386299
40	Nabisco Biscuit	Cookie	9761322	Novelty cookie product	0.155735
41	Kraft Heinz Foods	Coffee	13810612	Coffee product and related process	0.497631
42	Royal Appliance Mfg. Co.	Vacuum cleaner	10224483	Vacuum cleaner having hose detachable at nozzle	0.503479
43	Uniden Corp. of America	Mobile phone accessories	10268080	Rotating detachable belt clip	0.052147
44	Lexmark International	Ink cartridge	9766363	Ink cartridge and method for determining ink volume in said ink cartridge	0.505055
45	Gerber Products	Baby food	10295283	Blended baby food	0.24046
46	The Clorox Company	Hard-surface cleaner	12141583	Low residue cleaning solution comprising a c-to-c alkylpolyglucoside and glycerol	0.195491
47	The Clorox Company	Bleach	14724349	Intercalated bleach composition related method of manufacture and use	0.390043
48	L'Oreal USA	Cosmetic mascara	10759614	Two step mascara	0.359273
49	Lifescan	Stool test	10179064	Reagent test strip with alignment notch	0.123588
50	Playtex Products	Tampon	10834386	Tampon assembly having shaped pledget	0.558883
51	Kimberly-Clark	Urinary tract infection	12680575	Management of urinary incontinence in female	0.400734
52	Procter & Gamble	Microfiber	11016522	Rotary spinning process for forming hydroxyl polymercontaining fiber	0.113136
53	Sandisk Corporation	Floppy disk	10772789	Disk acceleration using first and second storage device	0.232516
54	Procter & Gamble	Acne	10633742	hptp-beta a target in treatment of angiogenesis mediated disorder	0.026864
55	Kraft Heinz Foods	Pasta	29220156	Spider shaped pasta	0.643155
56	L'Oreal USA	Eye liner	14368230	Method for delivering cosmetic advice	0.200779
57	Lexmark International	Printer (computing)	11766807	Hand held printer configuration	0.431107
58	Dreyer's Grand Ice Cream	Ice cream	10213212	Apparatus for forming an extruded ice cream dessert with inclusion	0.411786
59	Imation Corp.	Compact cassette	9882669	High speed tape packing	0.240291
60	Conagra Brands	Canning	12814296	Method and apparatus for smoking food product	0.144703
61	Nestle Purina PetCare	Dog food	29212029	Pet food	0.313367
62	Fort James Corporation	Disposable food packaging	29178752	Disposable plate	0.173866
63	L'Oreal USA	Face powder	9847388	Use of fiber in a care composition or a makeup composition to make the skin matte	0.139978
64	Conair Corporation	Hair styling tool	29285527	Curling iron	0.124045
65	Johnson & Johnson	Adhesive bandage	11877794	Adhesive bandage and a process for manufacturing an adhesive bandage	0.229017
66	Unilever USA	Shower gel	10242390	Viscoelastic cleansing gel with micellar surfactant solution	0.121894

	Company	Product category	Application ID	Title of the Patent	Similarity
67	Procter & Gamble	Dishwasher	11348667	Method of cleaning a washing machine or a dishwasher	0.296332
68	Pepsi-Cola North America	Tea	12147245	Coumalic acid to inhibit nonenzymatic browning in tea	0.483404
69	General Mills	Sweet roll	14340046	Method of forming dough composition	0.471588
70	Alcon Laboratories	Eye drop	9919301	Use of certain isoquinolinesulfonyl compound for the treatment of glaucoma and ocular ischemia	0.030695
71	Tyson Foods	Frozen food	13245589	Big poultry cutup method	0.311296
72	Pactiv Corp	Zipper storage bag	10289641	Reclosable bag having tamperevident member removable from the bag along a line of weakness located below the bag zipper	0.224593
73	Lipton	Margarine	9880200	Preparation of a blend of triglyceride	0.317454
74	Handi-Foil Corporation	Kitchen utensil	29418653	Pan with handle	0.167337
75	Hartz Mountain	Pet	10647660	Pet chew and method of providing dental care to pet	0.345158
76	Acco Brands USA	Notebook	11454292	Notebook computer folding ergonomic pad	0.130091
77	Johnson & Johnson	Lotion	12340858	Structured lotion	0.230563
78	Glaxosmithkline	Anti-inflammatory drug	11355808	Use of Immune cell specific conjugate for treatment of inflammatory disease of gastrointestinal tract	0.108521
79	Kraft Heinz Foods	Processed cheese	10207591	Processed cheese made with soy	0.43164
80	Fort James Corporation	Napkin	29215802	Tabletop napkin dispenser	0.263922
81	Omron Healthcare	Sphygmomanometer	29344018	Sphygmomanometer	0.463227
82	General Mills	Cracker (food)	10172401	Advertising quadrate carrier assembly with premium cradle	0.02869
83	BIC USA	Pen	29138586	Writing instrument	0.314765
84	The Libman Company	Mop	29298481	Mop	0.426008
85	Frito-Lay	Snack	10893425	Method and apparatus for layering seasoning	0.12532
86	Fresh Express Incorporated	Salad	29362982	Paper bag with a transparent vertical window for salad ingredient	0.241787
87	Procter & Gamble	Shaving cream	11110034	Shaving system with energy imparting device	0.322912
88	Nestle Purina PetCare	Litter box	29228923	Cat litter box	0.567078
89	Frito-Lay	Corn chip	9998661	Apparatus and method for making stackable tortilla chip	0.15851
90	Elizabeth Arden	Eau de toilette	29414481	Perfume bottle	0.241875
91	Bimbo Bakeries USA	Bread	13618124	Method and system for the preservation and regeneration of pre-baked bread	0.263577
92	E & J Gallo Winery	Wine	10970490	Method and apparatus for managing product planning and marketing	0.215571
93	BIC USA	Lighter	11221295	Multi-mode lighter	0.369379
94	Sara Lee Foods	Sausage	10014160	Split sausage and method and apparatus for producing split sausage	0.520147
95	Frito-Lay	Mixed nuts	11553694	Method for making a cubed nut cluster	0.158285
96	Kiss Nail Products	Manicure	12924589	Artificial nail and method of forming same	0.361627
97	Frito-Lay	Dipping sauce	10109398	Apparatus and method for improving the dimensional quality of direct expanded food product having complex shape	0.1273
98	Kraft Heinz Foods	Bacon	9799985	Bacon chip and patty	0.556922
99	Emerson Radio Corp.	Microwave oven	29149130	Protective cage and radio combination	0.0148
100	Procter & Gamble	Dentures	13043649	Denture adhesive composition	0.467318

A.5.2 External Validation. Virtual Patent Markings

One of our most important validation exercises for the patent-to-products match relies on external information. We use information from virtual patent markings which were introduced with the 2011 Leahy-Smith America Invents Act. Under that act, firms may give notice to the public that their product is patented. Recently, de Rassenfosse (2018) provides estimates of the adoption rate of virtual markings and studies factors that account for the likelihood of adoption. Overall, the adoption rate is relatively small and varies systematically with firm size. Indeed, our online searches showed that only a handful of the CPG firms in our sample used virtual patent markings.[43] This means that we cannot use patent markings to match patents to products for all firms in our data set. We can, however, use them as a useful validation exercise to compare the marking's product-patent matches with our algorithm.

To this end, we selected Procter & Gamble (P&G) and Kimberly-Clark (KC) for our validation exercise, as these are among the largest firms in our sample.[44] We start by parsing the product-patent links from the websites. In most cases the markings are associated with brands and not particular products. Hence, an important challenge lies in linking the listed brands on the websites with the brands in Nielsen. We use exact name matches, non-exact name matching, and extensive manual matching to determine the closest Nielsen brand equivalents. We then proceed to identify the product categories that include products of those brands. This parsing process allows us to obtain a mapping between patents and product categories that solely comes from the markings listed by P&G and KC markings.

For each patent, we lastly compare the matched product categories in our *Match 2* data set with the product categories obtained from the virtual markings listed by P&G and KC (311 and 87, respectively).[45] We begin by testing information from the similarity scores. For each patent-product category pair from the virtual markings, we obtain a similarity rank that our algorithm assigns to this product category. For example, when the rank value is one, the product category in the virtual markings corresponds to our algorithm's highest top-1 similarity category. When it is two, the match was very close to the category from the markings, and so on, thus providing a notion of closeness between the algorithm-based and marking-based matches. The first plot in Figure A.1 plots the distribution of these ranks. The algorithm-based preferred (highest-similarity) product categories coincide most of the time with the patent-product category mapping we created based on virtual markings. 69% of patents and 79% of patents conditional on a match are ranked as one or two based on similarity scores.[46]

[43] Even if firms use virtual patent markings, they report only a selected set of products and just a small fraction of patent portfolio they hold.

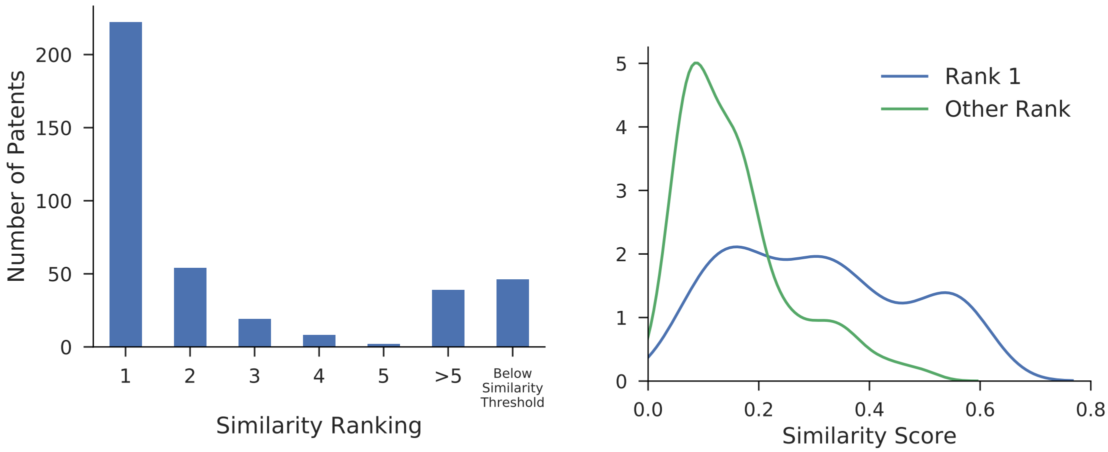
[44] We also found virtual markings are Clorox and Smuckers. However, because the products reported on their websites could not be mapped cleanly to our product categories, we did not analyze them.

[45] P&G and KC hold many more patents that are not included in the virtual markings. We also had to exclude patents listed under brands that we could not cleanly match to the Nielsen data.

[46] Note that we cannot compare these numbers to 100% given that the ranking is unavoidably affected by some noise that comes from our manual mapping of the product listings on the websites to the notion of product categories in our data.

Figure A.1: Virtual Patent Markings. P&G and KC Case Study

Distribution of similarity ranks for virtual markings　　Distribution of similarity scores



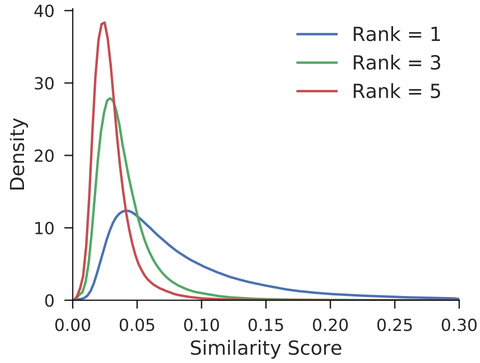
Notes: We use patent markings from P&G and KC. For each patent-product category pair from the virtual markings, we obtain a similarity rank that our algorithm assigned to this product category and show the distribution of ranks in the first graph. When the rank is one, the product category in the virtual marking corresponds to our algorithm's highest top-1 similarity category. The second graph shows the distribution of similarity scores for rank-1 and higher-rank product categories.

Another way to visualize the accuracy of the match is to examine the distribution of similarities conditioning on whether the match was rank-1 (coinciding with the category from virtual markings) or a higher rank. If these two distributions were very similar, this would mean that even if the match is accurate, it is not very robust, as small elements of noise or bias could change the results of the match. In fact, as shown in the second plot of Figure A.1, these two distributions are quite distinct with the rank-1-match distribution weighted towards the right, meaning the results of the match should be rather robust.

A.5.3 Robustness of the Match. Patent Similarity with Top vs Lower-rank Categories.

As discussed, for our match, we pick product categories which have the highest similarity scores with patents. That is, we first pick the top five categories that have the highest similarity values with patents, and then we assign the top-similarity category conditional on a firm producing a product in that category. However, if the similarity scores for different categories are too close (either because the algorithm is not able to pick up the distinctions between documents or the categories are too finely defined) so that the algorithm cannot clearly differentiate between them, our choice of the top-rank match would not be robust to small perturbations of the algorithm or category clustering. To explore this issue, we plot the distribution of similarity scores of patents with different-rank product categories (Figure A.2). The rank-1 category is the category with the highest similarity score for a patent, and so on. We find that top-ranked patents have substantially different (shifted to the right) distributions than slightly lower-ranked patents, thus providing evidence of the robustness of the match. The patents' mean similarity score for rank-1 categories is 3 times higher than the mean similarity score for rank-5 categories.

Figure A.2: Similaritv Distribution bv Rank

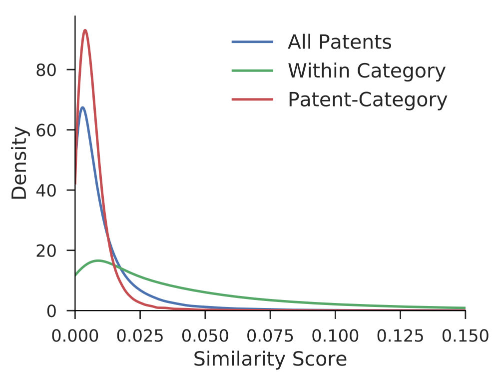


Notes: The figure shows similarity scores distribution of patents for different-rank product categories. Rank-1, Rank-3, and Rank-5 show similarities with categories ranked as the highest, rank-3, and rank-5 similarity categories.

A.5.4 Actual vs placebo match of patents to product categories

We next verify that by grouping patents into distinct categories, we are indeed carving out well-defined neighborhoods in the technological space. We again employ word vectors to assess document similarity, but this time between pairs of patent texts. Specifically, we look at the distribution of similarity scores between pairs of patents classified into the same product category and compare this distribution to that of pairs of patents selected at random from the entire set of patents held by CPG firms. The similarity distribution based on this match looks very different from our placebo distribution as seen in Figure A.3. The patents' mean similarity score is 5.6 times higher if patents are assigned to the same product categories. In ordinal terms, the median within-category similarity lies at the 93rd percentile in the overall distribution.

Figure A.3: Distribution of Pairwise Patent Similarities



Notes: The blue density curve shows the distribution of similarities between pairs of patents classified into the same product category. The green curve shows the distribution of similarities between randomly drawn pairs of patents amongst all those owned by Nielsen firms.

A.5.5 Validating Non-matches. CPG-only Firms and Product-related Patents

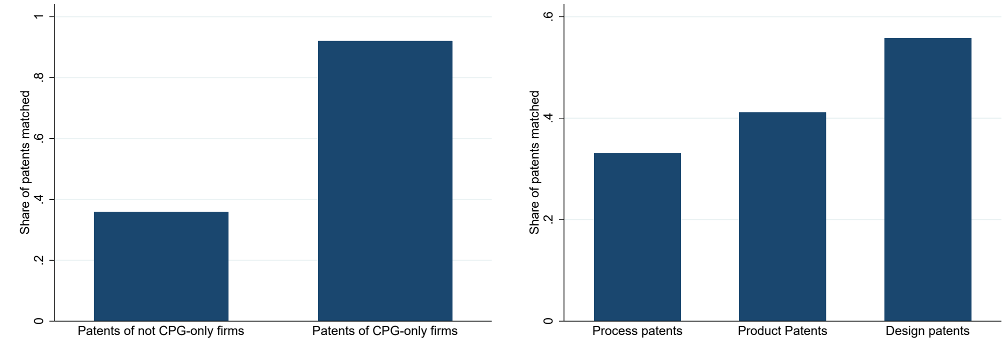
Our *Match 2* of patents at the firm × product category level would ideally filter out patents that are not related to the products in our data. Hence, correct non-matches would arise for the following three reasons. First, a patent may relate to other non-CPG goods that the firm may be producing, which are not covered in our sample; second, a patent may be a general process/method patent that does not relate to the products directly; and third, a patent may just be a by-product of current research that is unrelated to a firm's active product lines, but that in some cases could lead to future product-line expansions. We examine the first and second possibilities.

Panel (a) in Figure A.4 shows the share of patents that match to firms' product categories for a sample of firms that we can accurately identify as CPG-only firms and not CPG-only firms (see Appendix A.1 for details). Indeed, 92% of patents help by CPG-only firms match, while 36% of not CPG-only firms match to our product categories. This result reassures us that our algorithm indeed picks the correct matches. As seen from Panel (b), the similarity scores for CPG-only firm patents are also significantly higher.

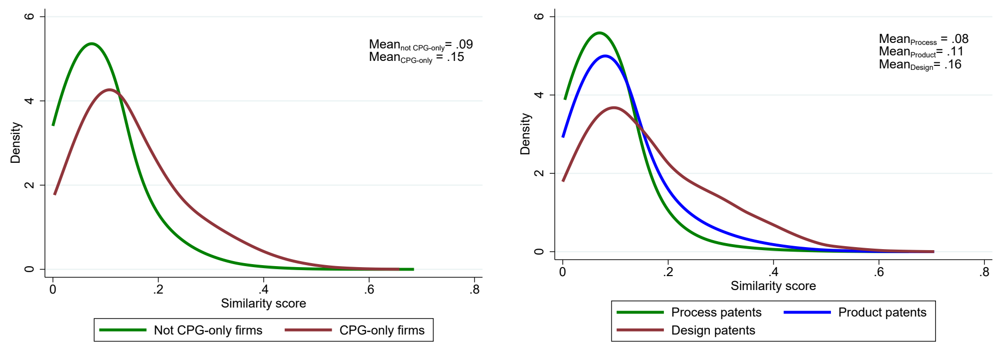
Panel (a) also demonstrates that the share of patents that are matched is higher if the patent is more likely to be directly related to products. Using our proxies for process and product-related patents (see Appendix A.2 for details) and considering design patents as most directly related to products, we plot the share of all process, product, and design patents that are matched. The probability of a match increases along with the likelihood of a patent being related to a product, which is encouraging. Panel (b) also confirms that the similarity scores of product-related patents are much higher than the similarity scores of process patents.

Figure A.4: Match Validation. CPG-only Firms and Product-related Patents

(a) Share of patents matching to firms' product categories



(b) Rank-1 similarity distribution for patents



Notes: Panel (a) shows the share of patents that match with product categories in which firms ever sell a product. The left figure compares patents of the CPG-only and non-CPG-only firms, while the right figure compares process, product-related, and design patents. CPG-only firms and non-CPG-only firms refer to the sample of firms defined in Appendix Section A.1. Process and product-related patents are defined in Appendix Section A.2. Panel (b) displays the similarity score distribution for patents of CPG-only and non-CPG-only firms on the left and of process, product-related, and design patents on the right.

B Measuring Product Innovation

We use four measures of new product quality: a novelty index whose weights are the contributions of each attribute to the product price (baseline q); a novelty index that equally weights each attribute ($q1$); a novelty index whose weights reflect the total revenue accounted by each attribute ($q2$); and a novelty measure that weights each product by its residual demand ($q3$). These measures capture different dimensions of quality. The first three measures (baseline q, $q1$ and $q2$) explicitly capture the novelty of a new product by using information about its attributes. The second type of measure ($q3$) captures any residual demand (or appeal), which can arise from vertical quality differentiation or subjective differences in consumer taste. We next describe the construction of the novelty-based measures and residual demand in detail, followed by a discussion of the descriptive statistics for these measures.

B-1 Novelty-based Measures

Overview. —We define a product u in product category j as a vector of characteristics $V_u^j = \left[v_{u1}^j, v_{u2}^j, .., v_{uA^j}^j\right]$ where A^j denotes the number of attributes (e.g. color, formula, size) observed in product category j and v_{ia}^j represents a characteristic within an attribute (e.g. blue, red, green).[47] Let Ω_t^j contain the set of product characteristics for each product ever sold in product category j at time t, then the *novelty index* of product u in product category j, launched at time t is defined as follows:

$$q_u \equiv \text{Novelty}_{u(t)}^{(j)} = \sum_{a=1}^{A^j} \omega_a^j \mathbb{1}[v_{ua}^j \notin \Omega_t^j].$$

where ω_a^j represents the category-specific weight given to new characteristics within attribute a. The measures q, $q1$ and $q2$ only differ in the way we compute their ω_a^j.

For q, we estimate ω_a^j using hedonic methods in order to be able to quantify the importance of each attribute within a product category. In particular, we estimate a linear-characteristics model using the time-dummy method. We pool data across products and periods and regress prices on a set of product attributes and a sequence of time-dummies. The estimated regression coefficients represent the shadow price for each of the included characteristics. ω_a^j is the average contribution of the characteristics within each attribute to the price normalized so that $\sum_a^{A^j} \omega_a^j = 1$. We then aggregate the newness index to the category level using equal weights. See below for details.

The simplest measure $q1$, simply weights each attribute equally. For example, if a new product within the "pain remedies-headache" category enters the market with a flavor and formula that has never been sold before, its novelty index is $(1+1)/A^{\text{soft drinks}} = 2/10$. Note that comparing the novelty index of different products across distinct categories depends not only on the number of new attributes of each product, but also on the total amount of observable characteristics the Nielsen data provides for each category.

Measure $q2$ is very similar to q. The difference is that instead of using the contribution that each characteristic within an attribute makes to the price as the weight, we use the revenue generated by each characteristic given that we can observe the quantity of products with the same characteristic that were sold. In this case, we also normalize the weights so that all weights within a product module add up to one.

Hedonic Regression Weights. —We estimate product category weights ω_a^j using hedonic methods. We then estimate a linear characteristics model using the time-dummy method. The time-dummy method works by pooling data across products and periods and regressing prices on a set of product characteristics and a sequence of time-dummies. Since the

[47]We refer to product categories for simplicity of notation. Our analysis is conducted first at the product module level (as defined by Nielsen RMS data) and then aggregated at the firm level (*Match 1*) or firm × product category level (*Match 2*).

regression is run over data which is pooled across time periods, any product characteristic which is held by at least one good in some period can be included even if it is not present in all periods. The estimated regression coefficients represent the shadow price for each of the included characteristics. To implement this method, we estimate the following equation by non-negative least squares:

$$p_{ut} = \sum_c \pi^c a_u^c + \lambda_t + \epsilon_{ut} \tag{15}$$

where u denotes the product, c is the characteristic, and t is the time period (years). a_u^c is an indicator that equals one if a given characteristic c is present in product u. Recall that each attribute a (e.g color) has distinct characteristics c (e.g. blue, red). The shadow price of a given characteristics is denoted by π^c. We use non-negative least squares so that the shadow prices are weakly positive. Lastly, λ_t represents time effects.

Using this method, we obtain a correlation of approximately 0.91 between the actual price and $\sum_c \pi^c$.[48] The weight ω_a^j is the average contribution of the characteristics within each attribute to the price normalized so that $\sum_a^{A^j} \omega_a^j = 1$; these are the weights used in our baseline novelty index.

B-2 Residual Demand measure

An alternative way of measuring the degree of product innovation brought by new products to the market is to weight them by their implied quality (or residual demand) using a structural specification of their demand function. To derive an implied quality for each product, we follow Hottman, Redding and Weinstein (2016) and Argente, Lee and Moreira (2020) and use a nested constant elasticity of substitution (CES) utility system that allows the elasticity of substitution between varieties within a firm to differ from the elasticity of substitution between varieties supplied by different firms. The model features oligopolistic competition with a finite number of heterogeneous multi-product firms, where the output of each category is described by a nested CES structure over a finite number of products within a finite number of firms (j is omitted for simplicity of notation)

$$y = \left(\sum_{i=1}^{M} \left(\sum_{u=1}^{N_i} \left(\gamma_{ui} y_{ui}\right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}}$$

where σ is the elasticity of substitution across products within the same firm, η is the elasticity of substitution across firms, and γ_{ui} and y_{ui} are the implied quality and quantity of product u produced by firm i, respectively. Using the first order conditions of the consumer we can write the demand for product u produced by firm i as follows

[48] These dummies for characteristics seem to explain differences in prices well. The variance of linear combination of the fixed effects of the attributes (excluding time fixed-effects) relative to the variance of the prices is 0.827.

$$y_{ui} = (\gamma_{ui})^{\sigma-1} \left(\frac{p_{ui}}{p_i}\right)^{-\sigma} \left(\frac{p_i}{p}\right)^{-\eta} \frac{Y}{p} \quad , \qquad p_i = \left(\sum_{u=1}^{N_i} \left(\frac{p_{ui}}{\gamma_{ui}}\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$$

where the demand for the product depends on the implied quality γ_{ui} and price p_{ui} of the product, as well as the firm's price index p_i, the category's price index p, and the size of the category Y. Conditional on the observing the prices and quantities from the data and obtaining estimates for σ and η, we recover γ_{uijt} as a structural residual that ensures that the model replicates the observed data up to a normalization.[49] We normalize the implied quality so that its geometric mean within each category and time period equals one. The key advantage of this normalization is that we can compare a product's implied quality within the firm and across firms within a category and time period. Using this normalization and equation B, we obtain the product implied quality as:

$$\gamma_{ui} = \left(\frac{s_{ui} \times s_i}{\prod_{u,j} (s_{ui} \times s_i)^{\frac{1}{M}}}\right)^{\frac{1}{\sigma-1}} \left(\frac{s_i}{\prod_{u,j} (s_i)^{\frac{1}{M}}}\right)^{\frac{\sigma-\eta}{(1-\eta)(1-\sigma)}} \left(\frac{p_{ui}}{\prod_{u,j} (p_{ui})^{\frac{1}{M}}}\right)$$

where s_{ui} and s_i are the share of sales of product u and the share of sales of firm i, respectively, and M denotes the total number of products sold in a category. The estimation procedure for σ and η follows Broda and Weinstein (2010) and Feenstra (1994). The estimation has two steps. In the first step, we estimate the elasticity of substitution across products within firms using product shares, product prices, and firm shares using a GMM procedure. The key identification assumption is that demand and supply shocks at the product level are uncorrelated once we control for firm-time specific effects. In the second step, we use these estimates for products to estimate the elasticity of substitution across firms for each category using the procedure developed by Hottman et al. (2016).

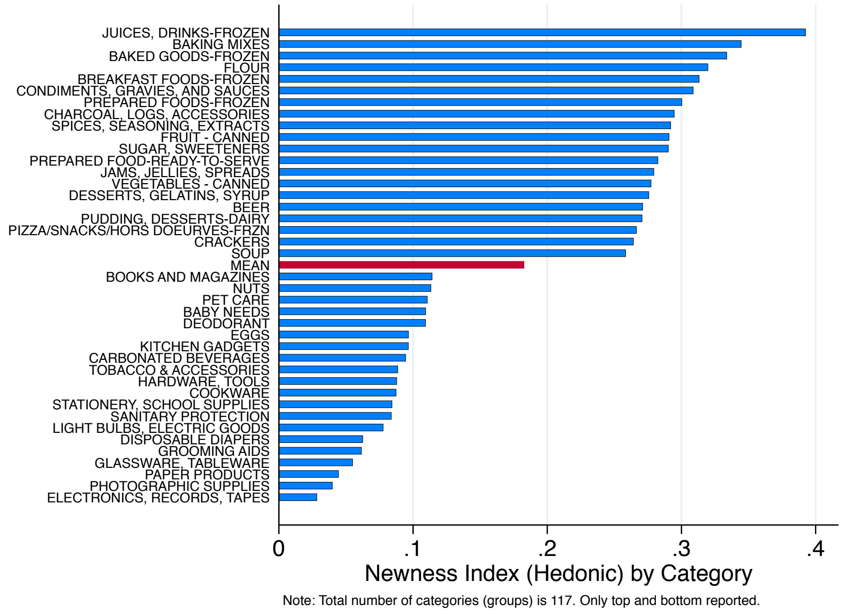
We use the estimates from Argente, Lee and Moreira (2020). To capture the incremental effect of new products on the residual demand of the firms, our measure of novelty is the geometric average of the implied quality of the new products relative to the geometric average of all products sold by the firm.

B-3 Descriptive Statistics

Differences in Novelty Across Products. —Our novelty index, counts the number of new characteristics (e.g. blue, red) within the attributes (e.g. color) of each new product brought to the market. We then weight these attributes according to their importance. Figure A.5 shows the degree of heterogeneity in novelty index q across different product categories. The quality measure q has a correlation of 0.93 with the equal-weights measure $q1$. Conditional on having an equal-weights index larger than zero, the correlation is 0.79.

[49]Normalization is required because the utility function is homogeneous of degree 1 in the implied quality.

Figure A.5: Novelty Index (Baseline q)



Notes: The figure presents the average novelty index for a sample of product groups in our data. In particular, it shows the mean novelty index by groups along with the top and bottom groups as ranked by this measure. We compute the novelty index for each product using equation B. We average across products and product modules to the category level. We focus on cohorts from 2006Q3 to 2014Q4 and on modules with at least 20 barcodes.

Figure A.6 shows some examples of products with high and low equal-weights novelty in our data. For example, the product Asthmanefrin Inhalation Solution - Liquid Refill is part of the group Medications/Remedies/Health Aids. When it was introduced in the market, this product had six of the eight attributes that we observe in our data for that product group, and it was a new brand, launched by a new firm, and it is a liquid, bronchilator refill. As a results, its equal-weights novelty index is 6/8=0.75.

Figure A.6: Novelty Index: Examples

(a) High-Novelty products (b) Low-Novelty products

Correlation with Product and Firm Performance. —Our baseline measure of quality explicitly captures the novelty of a new product by using information about its attributes. This use of product attributes offers important advantages in the context of our paper. Patents are granted on the basis of novelty, and thus using a quality-adjusted measure of product introduction that explicitly accounts for new features of the product may maximize the correlation between metrics of innovation and product introduction that are based on patent text. However, novel features may not affect the market at all if they are not valued by customers. Our baseline partially accounts for this potential source of error by quantifying any new characteristic according to its shadow price using hedonic regressions. Table shows that our baseline measure is correlated with product and firm outcomes, and thus may be capturing some vertical quality differentiation or subjective differences in consumer taste.

Table A.II: Novelty Measure: Correlation with Firm Outcomes

	(1) Growth rate (DH)	(2) Growth rate (New)	(3) Duration 4q	(4) Duration 16q
Novelty(t)	0.1546***	0.3032***	0.1081***	0.0754***
	(0.024)	(0.006)	(0.009)	(0.016)
Log N(t)	0.1953***	0.0245***	0.0287***	0.0203***
	(0.004)	(0.001)	(0.002)	(0.003)
Observations	92,430	111,339	96,942	53,611
R-squared	0.382	0.588	0.476	0.570
Time-Category	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y

Notes: The table shows the correlation between our measure of novelty and several firm outcomes. *Growth rate* (DH) is the revenue growth of the firm estimated as in Davis and Haltiwanger (1992), i.e. $2(y_t - y_{t-1}/(y_t + y_{t-1})$. *Growth rate* (New) is the revenue generated by new products as a share of total revenue in period t. *Duration* $4q$ and *Duration* $16q$ are the share of products introduced a time t that last in the market more than 4 or 16 quarters respectively. *log N* is log number of products introduced using the inverse hyperbolic sine transformation.

C Additional Empirical Results

Table A.III: Product Innovation after First Patent (All Firms)

	Log N			Log qN		
	(1)	(2)	(3)	(4)	(5)	(6)
After I patent(t)	0.0499			0.0101		
	(0.038)			(0.021)		
After I granted patent(t)		0.0745*			0.0254*	
		(0.033)			(0.012)	
After I non-granted patent(t)			-0.0853			-0.0389
			(0.057)			(0.042)
Observations	195,781	195,781	195,781	195,781	195,781	195,781
Time	Y	Y	Y	Y	Y	Y
Firm	Y	Y	Y	Y	Y	Y

Notes: The table shows regressions of the log number of new products (*log N*) in Panel A and the log quality-adjusted new products (*Log qN*) of a firm as a function of a dummy equal to one after the first patent application by the firm. Quality is defined in Section 3.1. *log N* and *Log qN* use the inverse hyperbolic sine transformation. *After I patent* is a dummy equal to one after any patent application; *After I granted patent* is a dummy equal to one after a patent application that is granted; and *After I non-granted patent* is a dummy equal to one after a patent application that has not been granted (abandoned or pending). The sample contains all firms in the data. Standard errors robust against heteroskedasticity and serial correlation are reported in parentheses.

Table A.IV: Product Innovation and Patenting (Firm Level)

	Log N			Log qN		
	(1)	(2)	(3)	(4)	(5)	(6)
Patents(t-1)	0.0310***			0.0149**		
	(0.009)			(0.005)		
Patents granted(t)		0.0303**			0.0160**	
		(0.012)			(0.007)	
Patents non-granted(t-1)			0.0218**			0.0021
			(0.008)			(0.006)
Observations	178,509	178,509	178,509	178,509	178,509	178,509
Time	Y	Y	Y	Y	Y	Y
Firm	Y	Y	Y	Y	Y	Y

Notes: The table shows regressions of the log number of new products (*log N*) and of log quality-adjusted new products (*Log qN*) in a firm over time as a function of log number of patents. Our benchmark quality measure is defined in Section 3.1. *Patents* is the log number of any patent applications in firm × year; *Patents granted* is the log number of granted patent applications; and *Patents non-granted* is the log number of patent applications that have not been granted (abandoned or pending). Standard errors robust against heteroskedasticity and serial correlation are reported in parentheses.

Table A.V: Product Innovation and Patenting: Citations and Claims

	Log N		Log qN	
	(1)	(2)	(3)	(4)
Citations(t-1)	0.0256***		0.0135***	
	(0.006)		(0.003)	
Claims(t-1)		0.0111***		0.0073***
		(0.004)		(0.002)
Observations	409,641	409,210	409,641	409,510
R-squared	0.692	0.692	0.623	0.623
Time-Category	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y

Notes: The table shows regressions of the log number of new products (*log N*) and of the log quality-adjusted new products (*Log qN*) in a firm×category over time as a function of log citations- and claims-adjusted number of patents. Our benchmark quality measure is defined in Section 3.1. *Citations* is the log number of 5-year citations received by all patents filed in the firm×category×year; *Claims* is the log number of claims on all patents filed in the firm×category×year. The inverse hyperbolic sine transformation is used for logs. Standard errors robust against heteroskedasticity and serial correlation are reported in parentheses.

Table A.VI: Product Innovation and Product & Process-Related Patents

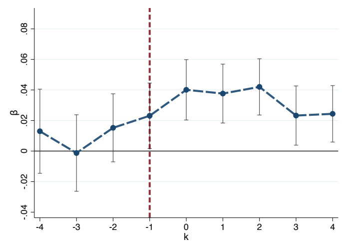
	Log N		Log qN	
	(1)	(2)	(3)	(4)
Product patents(t-1)	0.0402***		0.0185***	
	(0.009)		(0.005)	
Process patents(t-1)		0.0092		0.0030
		(0.016)		(0.009)
Observations	409,510	409,510	409,510	409,510
R-squared	0.692	0.692	0.623	0.623
Time-Category	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y

Notes: The table shows regressions of the log number of new products (*log N*) and the log quality-adjusted new products (*Log qN*) in a firm×category over time as a function of proxies for product-related and process-related patents. Our benchmark quality measure is defined in Section 3.1. *Product patents* is the log number of product-related patents, while *Process patents* is the log number of process-related patents. Proxies for product-related and process-related patents are defined in Section A.2. The inverse hyperbolic sine transformation is used for logs.

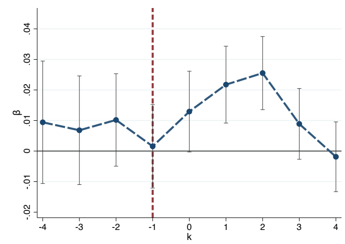
Figure A.7: Product Innovation and Patenting: Dynamics of Other Outcomes

— Firm-level estimates (Match 1) —

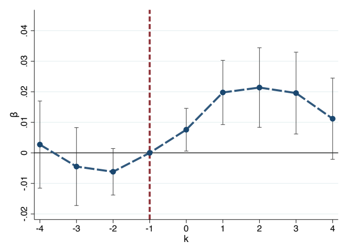
(a) $\beta_{\mathrm{N,P}}$



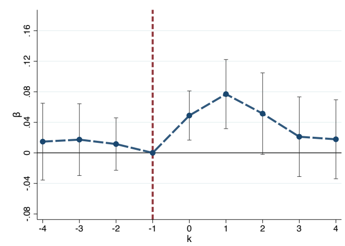
(b) $\beta_{\mathrm{qN,P}}$



(c) $\beta_{\mathrm{T,P}}$

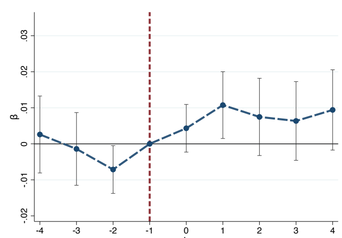


(d) $\beta_{\mathrm{rev,P}}$

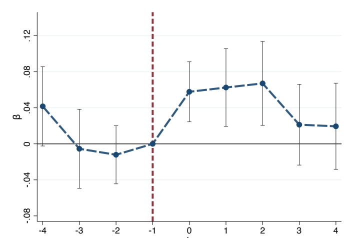


— Firm-category level estimates (Match 2) —

(e) $\beta_{\mathrm{T,P}}$

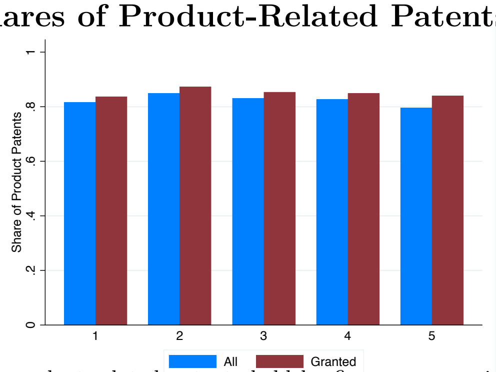


(f) $\beta_{\mathrm{rev,P}}$



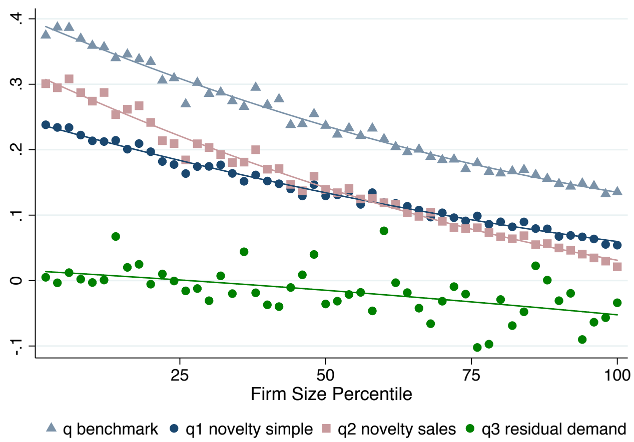
Note: The figure plots the estimated coefficients after estimating equation (4) for the log product introduction, N, in (a), quality-adjusted product introduction, qN, in (b), total number of products, T, in (c) and (e), and yearly revenue in (d) and (f) on log number of patent applications. The top panel uses firm-level data (*Match 1*), and the bottom panel uses firm-product category level data (*Match 2*). The inverse hyperbolic sine transformation is used for logarithms. The vertical bands represent $\pm 1.65\times$ the st. error of each point estimate. Standard errors are clustered at the firm $\times$ category level.

Figure A.8: Shares of Product-Related Patents by Firm Size



Notes: This figure plots the shares of product-related patents held by firms across quintiles of firm size, defined in terms of sales. We classify patents into product-related patents based on the claims of patent documents. Details of this classification can be found in Appendix A.2.

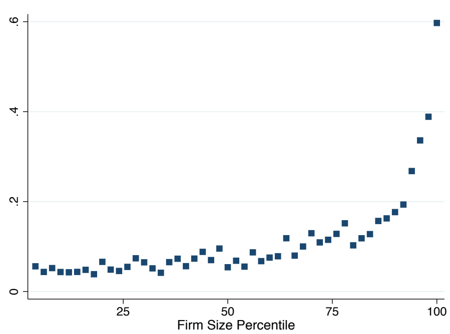
Figure A.9: Product Innovation Rate by Size: Alternative Quality Adjustments



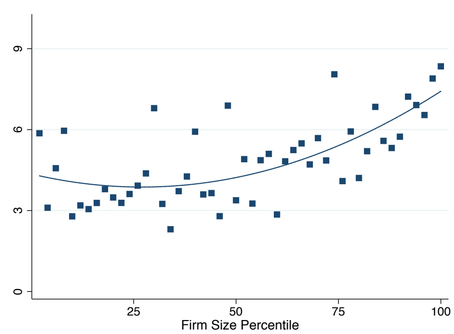
Notes: This figure plots the relationship between product innovation and size of the firm, defined by firm sales. We use the Match 2 data set on product innovation at the firm × product category level for the period 2007–2015, restricting the analysis to observations with sales above \$1,000. For each firm × product category, we compute their average sales and quality-adjusted product entry rates (quality-adjusted new products divided by total number of products) using our benchmark and three alternative quality measures – $q1, q2, q3$. Within each product category, we assign firms to 50 size bins based on their average sales and we plot the average product entry rate and the quality-adjusted product entry rate per bin. Each dot/triangle plots the averages after weighting different product categories by their importance in the whole sector, as measured by their sales share.

Figure A.10: Patenting and Firm Size

(a) Probability of patent application



(b) Number of patent applications (log)



Notes: This figure plots the relationship between patenting and firm size, defined by sales. We use our firm × product category data set covering the period 2007–2015, restricting the analysis to observations with sales above \$1,000. For each firm × product category, we compute the probability of having filed a patent and the average number of patent applications on file. Within each product category, we assign firms to 50 size bins based on their average sales, and we compute the average probability and number of patents ×1000 (log) for each bin. Each dot/triangle plots an averages after weighting different product categories by their importance in the whole sector, as measured by their sales share.

D Model

D.1 Derivation of Rates of Creative Destruction

Depending on the actions of the incumbent, our model delivers the following rates of creative destruction:

- If incumbent does not patent and does not upgrade products

$$p \times \Pr\Big(q + \lambda^e > q\Big) = p,$$

- If incumbent does not patent and does upgrade products

$$p \times \Pr\Big(q + \lambda + \lambda^e > q + \lambda\Big) = p,$$

- If incumbent patents but does not upgrade products

$$p \times \Pr\Big(q + \lambda + \lambda^e > q + \lambda + \varepsilon\Big) = p(1 - \varepsilon),$$

- If incumbent patents and does not upgrade products

$$p \times \Pr\Big(q + \lambda + \lambda^e > q + \lambda + \varepsilon)\Big) = p(1 - \varepsilon).$$

and thus the role of the patent if to reduce the rate of creative destruction.

D.2 Conditions for Equilibrium

We re-write the options that the firm is considering as

$$\max\Big\{\ V^{11}(q) - V^{00}(q) - c_m - c_p,\ V^{10}(q) - V^{00}(q) - c_m,\ V^{01}(q) - V^{00}(q) - c_p\ , 0\Big\},$$

We denote the first term as $O1(q)$, second as $O2(q)$, third as $O3(q)$, and the last as $O4(q)$.

Step 1: We show that, for small firms, introducing a new product with no patenting dominates all the other options: $\exists q^*$ s.t. $\forall q < q^*,\ O2(q) = \max\Big\{O1(q),\ O2(q),\ O3(q),\ O4(q)\Big\}$

First, let us consider the behavior of each function with respect to q.

$O2$ is decreasing in q: $\frac{dO2}{dq} = \frac{\pi\gamma}{p+r}\left(\frac{1}{(q+\lambda)^{1-\gamma}} - \frac{1}{q^{1-\gamma}}\right) < 0, \forall q.$

$O3$ is increasing in q: $\frac{dO3}{dq} = \pi\gamma q^{\gamma-1}\left(\frac{1}{p(1-\varepsilon)+r} - \frac{1}{p+r}\right) > 0, \forall q.$

$O1$ is decreasing in q for small values of q, and $O1$ is increasing in q for large values of q with a minimum at $q = \mathring{q} \equiv \frac{\lambda}{\left(\frac{p+r}{p(1-\varepsilon)+r}\right)^{\frac{1}{1-\gamma}} - 1}$:

$$\frac{dO1}{dq} = \pi\gamma\left(\frac{1}{(q+\lambda)^{1-\gamma}}\frac{1}{p(1-\varepsilon)+r} - \frac{1}{q^{1-\gamma}}\frac{1}{p+r}\right) > 0 \text{ if } q > \mathring{q} \tag{16}$$

Next, notice that $O2(0) = \frac{\pi\lambda^\gamma}{p+r} - c_m$, $O3(0) = -c_p, O1(0) = \frac{\pi\lambda^\gamma}{p(1-\varepsilon)+r} - c_p - c_m, O4(0) = 0$. Consider the following restrictions on the parameters:

Condition (i): $c_m < \frac{\pi\lambda^\gamma}{p+r}$

Condition (ii): $c_p > \pi\lambda^\gamma \frac{p\varepsilon}{(p(1-\varepsilon)+r)(p+r)}$

The first condition simply states that the of cost commercialization is sufficiently low – such that the firm with the lowest-quality products (the firm who marginally gets most out of product innovation) finds it worthwhile to introduce new products. The second condition states that research and patenting costs are too high for the smallest firms. Both conditions are mild and necessary to generate basic patterns in the data – that at least some firms find it worthwhile to innovate with new products, and that smallest firms do not engage in formal intellectual property protection.

Under Conditions (i) and (ii) and given the monotonicity of $O2$ and $O3$ and (16), there exists a threshold level of q^*, such that firms below q^* only do product innovation (option $O2$), while above q^* other options dominate.

Step 2: we show that for sufficiently large firms, patenting while introducing no new products dominates other options: $\exists q^{**}$ s.t. $\forall q > q^{**}$, $O3(q) = \max\left\{O1(q),\ O2(q),\ O3(q),\ O4(q)\right\}$

Let us first compare $O1$ and $O3$.

$$O3(q) > O1(q) \text{ iff } c_m > \frac{\pi(q+\lambda)^\gamma - \pi q^\gamma}{p(1-\varepsilon)+r}$$

Notice that the right-hand side is a decreasing function of q with an asymptote at zero. In words, marginal returns from additional product innovation are so low that they do not cover the cost of commercialization. Hence, for large enough q the inequality is satisfied.[50] This, together with decreasing $O2$, implies that there exists q^{**} such that firms above q^{**} prefer to file patents while introducing no new products (option $O3$).

Step 3: Our next step simply determines the conditions under which $q^* \neq q^{**}$, and when firms engage both in product innovation and patenting in the $q^* < q < q^{**}$ range, which is an empirically relevant case.

Define q_1 and q_2 as the intersection of $O1$ and $O2$ functions and $O2$ and $O3$ functions, respectively. So, $O1(q_1) = O2(q_1)$ and $O2(q_2) = O3(q_2)$. Then we need to show that $q_1 < q_2$,

[50]Also notice that Condition (i) implies that $c_m < \frac{\pi\lambda^\gamma}{p(1-\varepsilon)+r}$, so at $q = 0$, the inequality is not satisfied. Hence, $O3$ is not always preferred over $O1$.

which is equivalent to showing that $O2(q_1) > O2(q_2)$ because $O2$ is a decreasing function. This will be true under *Condition (iii)* that ensures that the parameters satisfy

$$(q_1 + \lambda)^{\gamma} - q_1^{\gamma} > (q_2 + \lambda)^{\gamma} - q_2^{\gamma},$$

In such a case, $q^* = q_1$ and $q^{**} = q_{2.}$