Patents to Products: Innovation, Product Creation, and Firm Growth*

David Argente[†] Penn State

Salomé Baslandze[‡] EIEF & CEPR

Douglas Hanley[§] U. of Pittsburgh

Sara Moreira[¶] Northwestern

This version: July 5, 2019

[Link to the latest version]

Abstract

How do patents relate to product innovation? To study this question, we construct a new patent-to-product dataset combining patent data with the detailed product- and firm-level data for the consumer goods sector. Using textual analysis of patent documents together with product descriptions, we link specific patents to finely defined product categories within firms and time periods. Our findings indicate that there is a substantial amount of product innovation that comes from firms that have never patented. Nevertheless, for patenting firms, standard patent-based metrics of innovation are correlated to product innovation defined based on both quantity and quality of new products. We find that market leaders use patents differently than followers. In particular, patents of large firms have a weaker association with the quality and quantity of product innovations. Nevertheless, consistent with the notion that patents being used to limit competition, we find that patents of larger firms are associated with higher future revenues even after accounting for the introduction of new products associated with those patents. Motivated by these empirical patterns, we develop a theoretical framework and use it to decompose the value of a patent. We show that the private value of a patent increases as firms become market leaders. This increase is mostly driven by an increasing value derived from protective patenting as opposed to productive patenting.

JEL Classification Numbers: O3, O4

Keywords: Product innovation, patents, productivity, patent value, protective patents.

*We are grateful to Ufuk Akcigit, Fernando Alvarez, Ariel Burstein, Benjamin F. Jones, Claudio Michelacci, and seminar participants at the Bank of Portugal, EIEF (Pizzanomics), EPFL, NBER Productivity, Northwestern, SED (Mexico), Rochester, University of Lugano, and University of Milan. Researcher(s) own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets 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.

[†]Email: dargente@psu.edu. Address: 403 Kern Building, University Park, PA 16801.

[‡]Email: salome.baslandze@eief.it. Address: Via Sallustiana, 62, 00187 Roma.

[§]Email: doughanley@pitt.edu. Address: 230 South Bouquet Street Pittsburgh, PA 15260.

[¶]Email: sara.moreira@kellogg.northwestern.edu. Address: 2211 Campus Drive, Evanston, IL 60208.

I Introduction

For decades, economists and policymakers have seen innovation as a key contributor to productivity gains and economic growth. Descriptions of the process of innovation suggest that the introduction of new and better products account for a large fraction of the value creation from innovating activities (Romer (1990), Aghion and Howitt (1992)). However, the lack of detailed data on the quantity and quality of new products has led researchers to use other metrics to value innovation. In fact, patents have emerged as the primary measure to value innovation, especially since comprehensive datasets containing information about the timing and characteristics of patents became easily available (Griliches (1981, 1990)).

Yet, patents are crude measures of whether innovative activities turn into new and better products. Many firms opt to not patent their product innovations and in many instances firms obtain patents but ultimately decide to not develop them into new commercial products. It is also well known that patents are often used by firms – especially among large market leaders – to protect and defend their current products from competitors (*protective role*), rather than to obtain incremental profits from the commercialization of its own product inventions (*productive role*)(Gilbert and Newbery (1982), Blundell et al. (1999), Cohen et al. (2000)).

In this paper, we (i) evaluate how patent-based metrics of innovation relate to actual product innovation and (ii) explore variation in the mapping of patents to products across the distribution of firm size to better understand the underlying motives that drive firms to use patents. To answer these questions, we build a novel large-scale dataset that combines information on product introductions with information on patents. This dataset covers firms in the consumer goods sector and combines highly detailed scanner point-of-sales data during the 2006-2015 period (Nielsen) with patent information from the United States Patent Office (USPTO). We link these two datasets using name matching algorithms to single out patent applications filed by firms included in the Nielsen data. These algorithms allow us to build a unique firm-level dataset that follows the evolution of the portfolio of patents and products of each firm over time. Moreover, because many firms are multi-product with distinct product offerings, it is also necessary to construct a closer link between patents and products. By leveraging modern techniques in the field of information retrieval (Manning et al. (2008)) and document classification, we match each patent to finely defined product categories and create a granular dataset that tracks the patents and products of each firm in each product category over time.

The combination of the Nielsen and USPTO datasets allows us to conduct an investigation of the relation between patents and products. Product innovation refers to the introduction of new and improved products in the market.[1] Using the Nielsen data, we directly observe all product innovations in the sector. In this paper, we consider products at their finest level of aggregation, the barcode, and we define product innovation not only in terms of the quantity of new products that each firm introduces in the market over a period of time, but also in terms of the quality of those new products, based on the share of new attributes of the products, weighted by their respective importance. Thus, because Nielsen data has rich information on products, product attributes, and rspective market performance, we have the ability to study the relation of patents to several dimensions and types of product innovation.

Our main analysis consists of two parts. First, we study the relationship between patent-based metrics of innovation and product introduction. We establish the following empirical regularities:

Fact 1: A large amount of product innovation comes from firms that do not patent.

Fact 2: Patenting is positively associated with product innovation.

We find that 85% of the approximately 30,000 firms covered in our dataset never applied for a patent. Among the firms that have at least one patent, about 37% did not make any patent application in the period 2006–2015. Compared to other sectors, firms in the consumer goods sector are relatively patent-intensive.[2] Yet, firms that never filed a patent account for more than 50% of new products introduced between 2007 and 2015.

Next, 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 (the correlation is largest in the year following the application). A ten percent increase in patent applications in $t-1$ is associated with a 0.2-0.4% increase in product introduction in t. We find similar patterns when we focus our attention on patents that are granted or on patents that receive a lot of forward citations, suggesting that commonly used measures of the quality of patents are also associated with product innovation rates. We observe these results using within-firm and within-firm category variation, which implies that the

[1] Throughout the paper, we will use "product innovation" and "product introduction" interchangeably

[2] Graham et al. (2018) documents that less than 1% of firms in the U.S. economy are granted a patents between 2000 and 2011, and 6% of firms in the manufacturing sector have at least a patent in the same period.

results are driven not by changes in the composition of firms in the economy but rather by the relation between the patenting and product innovation decisions of firms. Overall, our interpretation of these correlations is that firms come up with some ideas for new products and they want to protect them from being copied by competitors by applying for patents. Simultaneously they develop those ideas into consumer products, which we observe in our measures of product introduction.

Having established a significant association between patents and products, we now turn to the central question of the paper, which is to evaluate how the mapping from patents to products varies across firm size and to understand the circumstances that explain such differences in the relation between patents and products across the size distribution. Empirically, we find the following empirical regularities:

Fact 3: Larger firms have lower product innovation rate (quantity and quality), but higher patents per new product.

Fact 4: Patents relate to higher future sales beyond their effect through product introduction, especially for large firms.

Fact 5: Patents of larger firms are associated with declines in competitors' product introduction.

Using variation across firms within product categories we estimate that, on average, firms in the bottom quintile of the size distribution, as measured by total sales in that product category, introduce a new product for each five products in their portfolio every year, whereas firms in the top quintile of the distribution of size introduce a new product for every seven products on their product portfolio. Moreover, we find that, conditional on introducing new products, the decline in average quality of new products across the distribution of size is even steeper. Yet, larger firms patent relatively more per product introduced in the market.

One potential reason large firms patent more per product is that they have greater incentives to patent. Although patents should carry technological novelty, they also include a protective component and hence can be used strategically by firms. Indeed, Blundell et al. (1999) argue that market leaders have greater incentives to use preemptive patenting to protect their market lead. Consistent with this conjecture, we find that patents have a positive effect on future sales, even after conditioning on the level of product introduction. Furthermore, we find that such impact is only large and significant among large firms.

If this incremental revenue premium that large firms derive from patenting activities reflects the greater protection that the existing products of those firms receive, we should observe that patents discourage competitors from introducing new products in protected product categories. We test this in the data by estimating the elasticity of patents (conditional on product introduction) on future product introduction by competitors, and our results indicate that the elasticity changes with firms size, and thus patents of larger firms are associated with declines in product introduction by competitors.

We build a simple theoretical framework that allows us to match the empirical patterns in the data and to capture the main determinants of the product and patenting decisions. The model builds on the quality-ladder model setup (e.g. Aghion and Howitt (1992)). Innovations come from a technology leader trying to prolong its lead or from potential entrants aiming to become the new leader. There is an incumbent firm that obtains a costless blueprint/idea and makes a once in a lifetime decision regarding product commercialization and patenting of the idea. If the firm decides to introduce a new product, it incurs the cost of product development and commercialization and gains extra revenue from higher-quality products. Simultaneously, the firm also decides if it will patent the blueprint. Patenting involves costs and grants the firm extra protection against being replaced by an entrant. The model has three basic ingredients: product upgrade exhibits decreasing returns, the probability of creative destruction depends on patent protection, and patenting as well as product introduction are costly activities. The model predicts that as firms become market leaders, they shift their innovation towards protective strategies – thus relying on patenting to limit competition instead of product innovation and commercialization.

We use the model to write a simple back of the envelope calculation for the private value of a patent and decompose it into its protective and productive components. We refer to the protective component of patent value as the value that patent brings to a firm by limiting creative destruction, holding fixed the technology of a firm. We refer to the productive component of patent value as the value that patents brings to a firm that commercializes the innovation embedded in the patent, holding creative destruction fixed. After parameterizing the model to our data, we conduct a simple back of the envelope calculation of patent value to the firms. Patent value is estimated to be around $65,000 and it increases with firm's size in the market. The share of the patent value coming from the protective component of a patent is on average 43% and increases largely with firm size.

The paper is related to recent work that attempts to reconcile the trend toward greater

investments in research and patenting and their limited effects on overall innovation and growth (J.Gordon (2016), Bloom et al. (2017)). We suggest that another explanation for this trend may well be that large incumbents rely on protective and defensive strategies instead of innovation to maintain their market leadership. In the context of an environment where economic activities reallocate towards high-market power firms (De Loecker and Eeckhout (2017), Autor et al. (2017)), incentives of large incumbents to direct their innovation effort towards productive rather than protective strategies may be very limited.

The remainder of this paper is organized as follows. In section II, we present the description of the different datasets we use and the procedure to link them along with our text analysis algorithm. We also discuss the validation exercises we performed and present the summary statistics. In section III, we explore the relationship between patenting and product innovation. Section IV presents results on the relationship between patenting, innovation, and deterrence over firm's size. Section V presents the theoretical framework and the patent value calculation. Section VI concludes.

Related Literature

Patents represent the most common metric for innovation in the literature. Firms with larger number of patent filings are considered to be more innovative. Patent filings are standardized, comparable across firms and countries, and they contain rich information on characteristic of invention – hence, providing the most systematic and large-scale dataset proxying innovation. However, in the absence of independent measures of innovation (for example, as in Alexopoulos (2011) or Moser (2012) considering alternative measures, but for aggregate innovation), it is hard to evaluate how good the patents are as proxy for innovation.[3]

There are at least two main challenges with this proxy. First, not all innovations are patented (Moser, 2012); and second, patents by their nature – in addition to their role to reflect certain technological novelties – give the right to exclude others from using same or similar technology (Hall and Harhoff, 2012). These negative competitive spillovers from patenting have been long recognized in the literature (Lanjouw (1998),Lanjouw and Schankerman (2001), Jaffe AB (2004), Bloom et al. (2013)). Since most of innovation is cumulative in

[3]As Boldrin and Levine (2013) put it, "there is no empirical evidence that they [patents] serve to increase innovation and productivity, unless productivity is identified with the number of patents awarded – which, as evidence shows, has no correlation with measured productivity".

nature, this feature of the patents may prevent others to "stand on the shoulders of the giants" (Scotchmer (1991), Furman and Stern (2011)) and may induce large social costs. For example, various studies have studied the effect of patenting on follow-on innovation (see Williams (2013), Heller and Eisenberg (1998),Sampat and Williams (2019) for the papers on biomedical research and Cockburn and J. MacGarvie (2011) for software industry, and Lampe and Moser (2015) for more general discussion).

In this paper, the unique match of patents to products in the CPG sector allows us to partly address these challenges. First, by directly measuring product innovation from the product-level data, we examine how much product innovation is associated with patent introductions. Second, we try to decompose the private value of a patent that comes from its technological component – patent representing novelty, and its strategic component – the ability to protect market from competitors. Importantly, we provide the evidence that these different components that lead to the revenue premium for the firm crucially depend on firm's current market lead. Patents by larger incumbents contain more strategic component – these patents translate less into actual product innovation, but translate more into competitor's deterrance. Hence, we contribute to the literature that tries to estimate patent value. For example, earlier studies (Hall and Harhoff (2012)) tried to infer monetary value of a patent using patent renewals (Schankerman and Pakes, 1986), the direct survey questions (Gambardella et al. (n.d.), Harhoff et al. (2003)), market value estimations (Hall et al. (2005), Toivanen et al. (2002), Seru et al. (2017)), and patent sales (Abrams et al., 2013). In our case, we *directly* observe revenues from products linked to patents, and we directly observe the behavior of competitors – how they react to patents through product introduction.

Our results indicate that patents may help market leaders to further strengthen their position without necessarily introducing a lot of high-quality subsequent products. In their classical paper, Gilbert and Newbery (1982) propose the possibility of preemptive patenting that may contribute to the persistence of monopoly. Empirically, Blundell et al. (1999) show that the impact of patenting on market value is larger for market leaders arguing that partly it comes from pre-emptive patenting. Indeed, large firms may rely on various protective or defensive strategies – like firm acquisitions (Cunningham et al., 2018) or connections with politicians (Akcigit et al. (2018)) – as they slow down on innovation (Cavenaile and Roldan (2019),Akcigit and Kerr (2018)). In this paper, we look at changes in firm incentives from productive patenting to protective patenting, hence contributing to our understanding of firm's growth strategy as a function of their market lead.

By better understanding these growth strategies we can potentially understand the link between increasing market concentration, markups and dominance of large firms on one hand (Rossi-Hansberg et al. (2018), Gutierrez and Philippon (2017), De Loecker and Eeckhout (2018), De Loecker and Eeckhout (2017), Autor et al. (2017)) and a declining growth and business dynamism (Decker et al., 2016) on the other.

II Constructing the Patent-Product Dataset

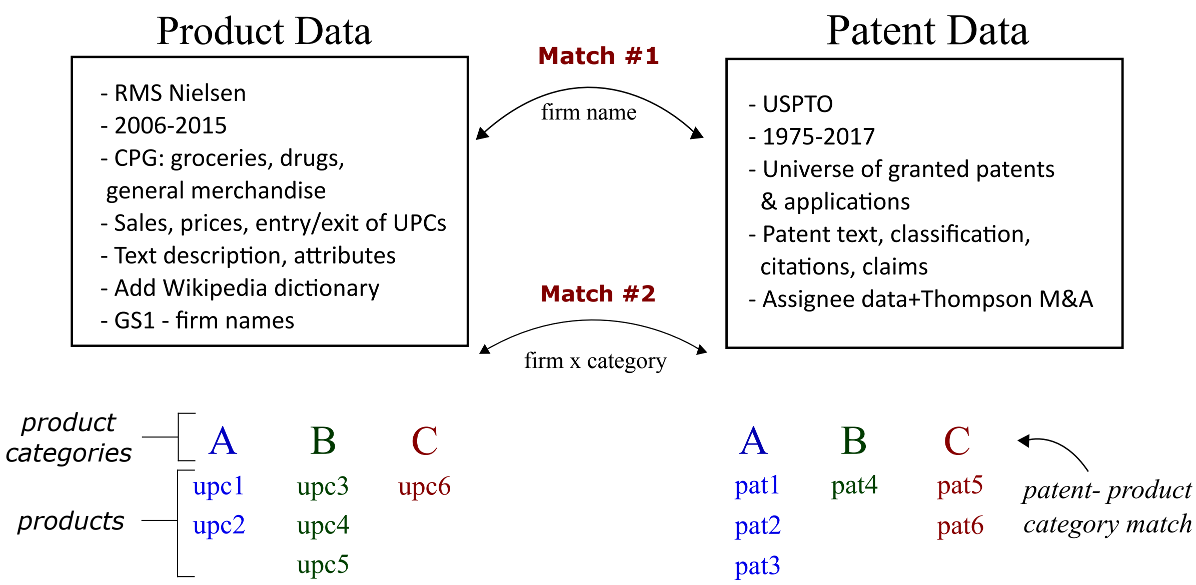
In this paper, we study the relationship between new products and patents. There are two important challenges associated with this: identification of product portfolio of firms and the match of products and patents.

We tackle the first challenge by making use of scanner data and the richness of the data on the product portfolio available for firms in the consumer goods sector (CPG) between 2006 and 2015. Figure 1 schematically represents our data construction. For each firm in the CPG sector, we identify their barcodes and classify them into different *product categories*. The advantage of this dataset is that we can observe new products in this sector. These product introductions may represent small or big innovations on the market. Hence, using detailed information on each product's attributes, we build proxies for quality changes in new products. At the firm and firm × product category levels, we then define *product innovation* based on quantity and quality of new products introduced.[4]

We address the second challenge by matching all United States Patent and Trademark Office (USPTO) patents of CPG firms to their product portfolios. Our *Match 1 – at the firm level* – simply matches patents to products based on name matching algorithms in product and patent datasets. However, many firms produce products in multiple categories that may be very heterogeneous in their patenting intensity, as well as their product introduction rates (Argente et al. (2018)). Therefore, we need to filter out this heterogeneity and establish a closer link between patents and products. Hence, our preferred *Match 2 – at the firm × product category level* – goes a step beyond by classifying firms' patents into product categories applying modern techniques in textual analysis to patent documents and to product descriptors extended with Wikipedia-based dictionaries. This is the bulk of our data analysis and provides a closer link between patents and products. As a result, at the firm and firm

[4]We describe product-level data from Nielsen Retail Measurement Services (RMS) in Section II.1. Section II.1.2 defines product innovation based on quantity and quality measures of products.

Figure 1: Product and Patents Dataset



× product category levels, on the one hand, we observe product innovation from CPG data and, on the other hand, various patenting measures, such as patent and quality-adjusted patent counts.[5]

II.1 Product Data

II.1.1 Datasets

We rely primarily on the Nielsen Retail Measurement Services (RMS) scanner dataset that is provided by the Kilts-Nielsen Data Center at the University of Chicago Booth School of Business. Nielsen data is generated by point-of-sale systems in retail stores. Each individual store reports weekly sales and quantities of every UPC code that had any sales volume during that week. The original data consist of more than one million distinct ***products*** identified by the finest level of aggregation – 12-digit Universal Product Code (UPC) that uniquely identify specific goods available in grocery and drug stores. Each barcode contains

[5]Section II.2 introduces our patent data from USPTO. Section II.3.1 contains the bulk of our analysis on matching patents to products. In Section II.3.2, we consider extensive validation exercises for our matching algorithm. Section II.4 summarizes our final dataset from a two-way match of patents to products: at the firm and at firm × product category levels.

information on the brand, size, packaging, and a rich set of product attributes.

The main advantage of this dataset is its size and coverage. Overall, the RMS consists of more than 100 billion unique observations at the week × store × UPC level. Our sample period covers the period 2006-2015. The dataset comprises around 12 billion transactions worth per year, $220 billion of dollars on average. Over our sample period, 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 MSAs and 2,500 counties. As a result, the data provide good coverage of the universe of products and firms in this sector. In comparison to other scanner datasets collected at the store level, the RMS covers a much wider range of products and stores. In comparison to scanner datasets collected at the household level, the RMS also has a wider range of products because it reflects the universe of transactions for the categories it covers, as opposed to the purchases of a sample of households.

In the Nielsen data, each product is classified into one of 1,070 low-level product modules. These product modules are further aggregated into a set of 114 product groups. For the purposes of our analysis, we would like to use a product classification scheme that groups together products that are close in their technological characteristics. Of course, the precise definition of "close" is somewhat subjective. Nonetheless, through experimentation with different classification schemes, we have found that the module -level specification is too fine: products across multiple modules are often quite similar (e.g. "Detergents – light duty" vs "Detergents – heavy duty") and it would not be possible to accurately match a patent to a specific module. On the other hand, the group-level classification is too coarse, since it often groups very distinct products. Thus there is a need for an intermediate level of aggregation, between modules and groups, which we generically call ***product category***. Relying on textual description of product modules in Nielsen data, we use text vectorization approach and the state-of-the art clustering techniques to classify similar products into same product categories.[6] The end result of this procedure is 400 clusters covering all CPG goods.

Our dataset combines all sales at the national and annual levels. Hence, for each product in a year, we define its sales as the total sales across all stores and weeks in the year. Likewise,

[6]Extensive discussion of our text analysis techniques comes in Section II.3.1. Details on clustering techniques are given in Appendix A.1.

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.[7]

Finally, we match each UPC to the parent company owning the product. We link products to firms using information obtained from GS1 US – the single official source of UPCs.[8] As a result of this data construction, we have a panel data at the ***firm level*** and ***firm*** $\times$ ***product category level*** with a detailed product portfolio information in 2006-2015.

II.1.2 Product Innovation

We define ***product innovation*** of a firm in a product category based on the amount and nature of products introduced by the firm. An important part of this exercise is the identification of entries. We define entry as the first quarter of sales of a product.[9]

We define two measures of product innovation. Our first measure is simply the count of new UPC's in a year by a firm in that product category:

$$\mathrm{N}^c_{jt} = \sum_{i=1}^{T^c_{jt}} \mathbb{1}[i \text{ is entrant}], \tag{1}$$

where j is firm, t is time, and c is product category. T^c_{jt} is the set of products of a firm in c at t. Throughout the paper we refer to this measure as **product introduction**. Our second measure adjusts for differences in newness/novelty across new products introduced:

$$\text{q-N}^c_{jt} = \sum_{i=1}^{T^c_{jt}} \text{Newness}_i \times \mathbb{1}[i \text{ is entry}]. \tag{2}$$

We refer to this measure as **quality-adjusted product introduction**.

We identify product *newness* by using detailed product attributes of each barcode. Like in Argente and Yeh (2017), we construct a newness index that proxies for a quality improvement in a new product. Our index uses detailed information about the observable characteristics of each UPC provided in the Nielsen RMS dataset and counts the number of new and unique attributes a product has at the time of its introduction relative to all of

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

[8]Argente, Lee and Moreira (2018) provide more details on this data.

[9]We cannot determine entry for products that are already active in the first year of the sample, 2006.

the other products ever sold within the same product module, weighted by importance of each attribute. We define a product i in product module m as a vector of characteristics $V_i^m = [v_{i1}^m, v_{i2}^m, .., v_{iK^m}^m]$ where K^m denotes the number of attributes (e.g. color, formula) observed in product module m and v_{ik}^m represents a characteristic within an attribute (e.g. blue, red, green). For example, the product module "pain remedies-headache" consists of $K^{\text{pain remedies - headache}} = 10$ attributes for each barcode: 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), size. Let Ω_t^m contain the set of product characteristics for each product ever sold in product module m at time t, then the *newness index* of a product i in product module m, launched at time t is defined as follows:

$$\text{Newness}_{i(t)}^{(m)} = \sum_{k=1}^{K^m} \omega_k^m \mathbb{1}[v_{ik}^m \notin \Omega_t^m].$$

where ω_k^m represents the module-specific weight given to new characteristics within attribute k. For example, if ω_k^m weights equally each attribute, and a new product within the "pain remedies-headache" enters with a flavor and formula that has never been sold in any store before, its newness index is $(1+1)/K^{\text{soft drinks}} = 2/10$. On average, we observe 7.2 product attributes in each product module.[10] We estimate ω_k^m using hedonic methods in order to be able to quantify the importance of each attribute within a product module. 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_k^m is the average contribution of the characteristics within each attribute to the price normalized so that $\sum_k^{K^m} \omega_k^m = 1$. We then aggregate the newness index to the category level using equal weights. We provide more details on the procedure in the Appendix B.[11]

[10]Comparing the newness index of different products across distinct modules 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 module. The minimum characteristics we observe for each module is 5 and the maximum is 12.

[11]Table A.I in the Appendix shows that the newness measure is correlated with the growth rate of the firm, the share of revenue 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.

II.2 Patent Data

II.2.1 Datasets

Our main data source for patent analysis is USPTO data on the universe of published patent applications, granted or not. We make use of the original bulk data files provided by USPTO's Bulk Data Storage System as well as supplemental data distributed through the PatentsView platform by the USPTO's Office of the Chief Economist. Our data contains information on 10 million patent applications filed by more than 500 thousands assignees for the period 1975-2017. The advantage of using all patent applications, as opposed to just granted applications, is twofold. First, since patents are usually granted with a lag of roughly two years, the more recent years of the sample suffer from severe truncation. Having all patent applications practically alleviates this problem. Second, we now have a larger sample and can differentiate between patents that are granted, pending, or abandoned – this also serves as one of the quality measures of patents as discussed below.[12]

For each patent, we utilize information on the following variables of interest: patent application year, patent status (granted, pending, abandoned), patent technology classifications (IPC), forward patent citations received, and number of claims on a patent. For our textual analysis of patent documents described below, we extract patents' titles, full text of patent abstracts, text of corresponding IPC classifications, and claims text.

To assign patents to firms, we proceed in the following steps. First, since our product-level data assigns products to firms as of 2015 , we need to treat patents in a similar way. Hence, in the first step, we utilize patent assignment dataset together with the Thomson Reuters Mergers & Acquisition data to designate each patent to its most current holder. The details on this step are delegated to Appendix D.1. Second, since firm names in these datasets often have misspellings or various abbreviations, it is challenging to accurately identify from raw data which companies are the same. To overcome this challenge, we develop a company name cleaning algorithm to clean and standardize company names. This procedure builds on and extends cleaning algorithms from the NBER Patent Data Project (Hall et al., 2001) and Akcigit et al. (2016). Details of this procedure are in Appendix D.2.

[12]In fact, adding non-granted patents information increases number of patents in the data from 6 million to 10 million.

II.2.2 Patent-based Measures of Innovation

In our analysis, we construct various measures of firm's patenting based on both count of patents, as well as quality-adjusted count of patents. A simple count of patent applications at the firm or firm $\times$ product category level over time constitutes our basic dynamic measure of firm's patenting.[13] Let P_{jt}^{c} be the set of patent applications of a firm j in category c at time t. Formally, we define

$$\text{Patents}_{jt}^{c} = \mid P_{jt}^{c} \mid,$$

where $\mid \cdot \mid$ denotes the cardinality of a set.

It is well known that patents are very heterogeneous in their quality. Throughout the paper we use different proxies for patent quality, based on whether they were eventually granted and on their citations. Granted patents are perceived as high-quality patents (as opposed to abandoned or pending patents). We define number of patent applications that eventually get granted as:[14]

$$\text{Patents grant}_{jt}^{c} = \sum_{i \in P_{jt}^{c}} \mathbb{1}[i \text{ is granted}].$$

The count of forward citations has traditionally been used as measure of the economic and technological significance of a patent (for earlier contributions, see Pakes (1986), Schankerman and Pakes (1986), Trajtenberg (1990)). We define citations-adjusted patent count as the total number of patents weighted by Cites_i – forward citations received in the first 5 years from the application time.[15]

$$\text{Patent cites}_{jt}^{c} = \sum_{i \in P_{jt}^{c}} \text{citations}_i.$$

[13] Description of the patents-to-product category match comes in the next Section II.3.1.

[14] Abandoned patents account for 16% of all patent applications in our data. Patents may get abandoned before they have been granted or they may get abandoned after the grant if they are not renewed every 4 years.

[15] A 5-year citations measure attempts to reduce truncation issue of citations – the fact that more recent patents have less time to accumulate citations.

II.3 Matching Patents to Product Categories

II.3.1 Algorithm

We leverage modern techniques in the field of information retrieval (Manning et al., 2008) and document classification to build a match between patents and products. After aggregating similar products into product categories (as described in Section II.1) we now match specific patents to the set of similar products grouped into product categories. To do this, we use a text similarity approach. That is, for each patent and product category, we construct a representative document (a set of words), and base our classification on those patent-category pairs yielding the highest similarity.

The first task is to construct these representative documents. Our distinguishing data on the patent side consists primarily of the text of the patent application (or publication), which includes: title, abstract, and list of claims. We also have U.S. and international patent classification codes for each patent, each of which has an associated short text description. The product side is much more limited. Each category has an associated title, and the products within the categories have descriptors, but these are primarily abbreviations which are often hard to interpret.

To get around this limitation of the Nielsen data, we utilize the text of Wikipedia entries as an intermediary. Specifically, for each low-level product classification (1,070 modules), we manually assign a list of one to three Wikipedia entries that most closely represent the module. This task is much more manageable than the patent-category match. We then use the text of these entries to facilitate this full match.[16]

Having constructed these documents, we then convert them into vectors (one for each document) representing word frequencies and calculate numerical similarities between these vectors to generate a match.

These vectors indicate, for each word, how many times it appears in a document. Each document vector is of length M, which is the number of words that we include in our vocabulary.[17] We use a vocabulary consisting of any words that appear in the Wikipedia

[16] In the case where two modules have precisely the same set of associated Wikipedia entries, we aggregate them together. Regardless, the k-means algorithm would trivially group them together as well.

[17] For this exercise, 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. To avoid confusion, we will continue to use "word" rather than "token" to refer to both 1-grams and 2-grams.

entries we consider excluding those in the top 20% of word frequencies (like "the" or "and"). The corpus of documents can be represented by a very sparse matrix c_{ij} of word counts, where $i \in \{1, \ldots, N\} = \mathcal{N}$ represents the document and $j \in \{1, \ldots, K\} = \mathcal{M}$ represents the word.

We also utilize a popular word-based weighting scheme called total-frequency-inverse-document-frequency (tf-idf) (Aizawa, 2003) to account for the fact that more common words tend to be of less importance and vice versa.[18] There are a number of possible forms to use here, but we choose the most commonly used

$$w_j = \log\left(\frac{N+1}{d_j+1}\right) + 1 \quad \text{where} \quad d_j = |\{i \in \mathcal{N} | c_{ij} > 0\}|$$

Thus if a word appears in all documents, it gets 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, which contains far more documents than the product category side.

Constructing representative documents on the patent side consists of simply concatenating all of the available text into one document. On the product module side, we must aggregate the various Wikipedia entries. In this particular setting, we first vectorize each Wikipedia entry, then average these vectors together (in an an ℓ^2-norm-preserving sense) so as not to overweight longer entries. Additionally, we repeat the first 10% of each Wikipedia entry 10 times to emphasize introductory material. Finally, before vectorization, we run each document through a lemmatizer,[19] which reduces words to their root form by removing conjugation.

Finally, we are left with a weighted, ℓ^2-normalized token frequency vector for each document, both on the patent and product side. Specifically, these are defined as

$$f_{ij} = \frac{w_j c_{ij}}{\sqrt{\sum_{j'} (w_j c_{ij'})^2}}$$

Multiplying any two such vectors together yields a similarity metric between two documents. This is guaranteed to be in the range $[0, 1]$ with zero corresponding to zero word overlap and

[18] Similar text analysis techniques were recently used for patent analysis in Younge and Kuhn (2016) and Kelly et al. (2018).

[19] Specifically, the WordNetLemmatizer provided as part of the NLTK (nltk.org) Python module, which utilizes the WordNet lexical database (wordnet.princeton.edu).

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/phrases.

Generating the patent-product match is then a matter of finding, for each patent, which product categories among those where the firm has any products (in any year) are most similar to it in a textual sense.[20] Computing the similarity between each patent and product category generates a similarity matrix

$$s_{k\ell} = \sum_{j \in \mathcal{M}} f_{kj} f_{\ell j}$$

where k corresponds to patents and ℓ corresponds to product categories. Thus the match would be $\ell^*(k) = \arg\max_\ell s_{k\ell}$.[21]

II.3.2 Patent Match Validation

The matching algorithm described above has intuitive appeal and uses modern tools from the big data analysis, but we also go to great lengths to assess the quality of the match using extensive manual checks and external information. Appendix Section E presents extensive discussion of these validation exercises. Here, we list them briefly. First, we manually go over many patent-to-category matches to assert that the match is not poor. Second, we show that by grouping patents into distinct categories, we are indeed carving out well defined neighborhoods in the technological space. For that, we look at actual versus placebo similarity distribution between patents classified in the same product category. Third, we validate the sample of our patent-product matches for Procter & Gamble against virtual patent markings information that P&G reports on its website. Fourth, we examine similarity distributions with different-rank product category matches to compare how strong the algorithm prefers the chosen max-rank category over lower-ranked category matches.

[20] For additional details on how the algorithm filters out non-CPG patents, see Appendix Section D.4.

[21] Our methodology assumes a one-to-one match between patents and product categories. However, one could argue some patents may be more general and relate to multiple categories. We abstract from this possibility for now. First, we believe product category definition encompasses a broad range of products that are still similar such that one patent could plausibly relate to this and only this range of products. Reassuring is that Figure A.8 illustrates that top-match product category has substantially higher similarity that the lower matches. And second, we illustrate our main results also based on the match at the firm level – where patents essentially are allowed to affect all product categories.

II.4 Final Matched Dataset and Summary Statistics

We end up with the following two types of data matches that we will use throughout our analysis; and, for simplicity, will refer to them as "Match 1" and "Match 2".

Match 1: Firm × Year Level The simple match at the firm level uses firm × year level datasets constructed on products (Section II.1) and patents (Section II.2) side and matches them using unique company identifiers obtained from name cleaning as in Appendix Section D.2. Our matched Nielsen-USPTO sample consists of 5,170 firms which can be divided into those firms that issue a patent during our sample period for Nielsen data 2006-2015 (a total of 3,284 firms) and those that issue a patent prior to 2006 (a total of 1,879 firms).

Match 2: Firm × Category × Year Level Our benchmark matched data is at the firm × category × year level and derives from combining firm × category × year level datasets constructed on products (Section II.1) and patents-product category match described in II.3.1. The matched sample has 5,170 firms operating in 400 product categories with more than 250 thousand observations.

II.4.1 Summary Statistics

In this section, we provide descriptive statistics of our basic firm-year dataset on patents and products. Our baseline product data has approximately 35 thousand firms and an average of 300 thousand active products every quarter. To define firm's patenting status, we divide firms into three categories: (i) firms that have never patented, (ii) firms that patented last before 2006 (the beginning of the Nielsen RMS dataset) and (iii) firms that have patents between 2006 and 2015. Table I shows the share of each type of firm in our data. It also shows the share of total products and the share of total revenue accounted by different type firms. The table shows that a large amount of products in the market belong to firms that have never patented. This accounts for approximately 46% of the total revenue generated in the CPG sector. In our matched sample, 9.5% of the firms have patents during our sample period. However, they account for disproportionately large number of products in the data – 34%, and account for almost half of the total revenue in the sector. These firms represent only 2.27% of the total number of U.S. firms filing for a patent in USPTO in 2006-2015. However, their patents represent about 20% of all patents filed – showing the importance of the CPG patenting in the universe of U.S. patents.

Table I: Number of Firms, Products, and Revenue in Nielsen by Patenting Status

	Firm's patenting status		
	No Patents	Patents before '06	Patents in '06-'15
Number of Firms	85.05%	5.44%	9.50%
Total Number of Products	57.35%	8.9%	33.7%
Total Revenue	45.70%	10.81%	43.49%

Notes: The table shows the shares of all firms, total products, and total revenue in Nielsen RMS data accounted by firms with different patenting status. The first column is for firms that have no patents, the second column is for firms that have patents filed before they first appear in Nielsen RMS (before 2006), and the third column is for firms that have patents in our sample period 2006-2015. The total number of firms in RMS Nielsen data is 34,536. The total number of products active every quarter is approximately 300 thousand.

Table II shows more detailed statistics for firms of different patenting status. It shows that patenting firms are larger (in revenue) and they not only introduce more products but these products also generate more revenue. In the CPG sector most product creation occurs in large high-revenue firms (as shown in Figure A.11). More than 60% of new products launched in the sector belong to firms in the top decile. Large firms are also more diversified than firms that never patented or that patented last before 2006. On average, firms that patent have products in more than 4 different product categories. Interestingly, although patenting firms introduce more high-revenue products, a large share of those products are not very novel according to our newness index: on average, firms that never patent introduce fewer, but more novel products.

In which product groups are firms more likely to patent? One advantage of the Nielsen data is that it allows us to explore a wide range of product categories, from perishables to semi-durable products. Interestingly, patenting firms are present in every product group. Figure 5 shows the product groups according to their patenting intensity – the share of products coming from patenting firms. In our sample, the average product group has patenting intensity of 50%. The figure shows that product groups with more durable products have high patenting intensity. Some examples are: deodorants, detergents, shaving needs, cookware, and kitchen gadgets. On the other hand, product groups such as wine, flour, ice, and cheese show low patenting intensity.

To confirm this intuition we construct a measure of the durability of a product based on the Nielsen Consumer Panel. For each product group we count the average number of shopping trips made by households in a given year to purchase products in each product group. Our assumption is that if households take longer to purchase products for a certain category those

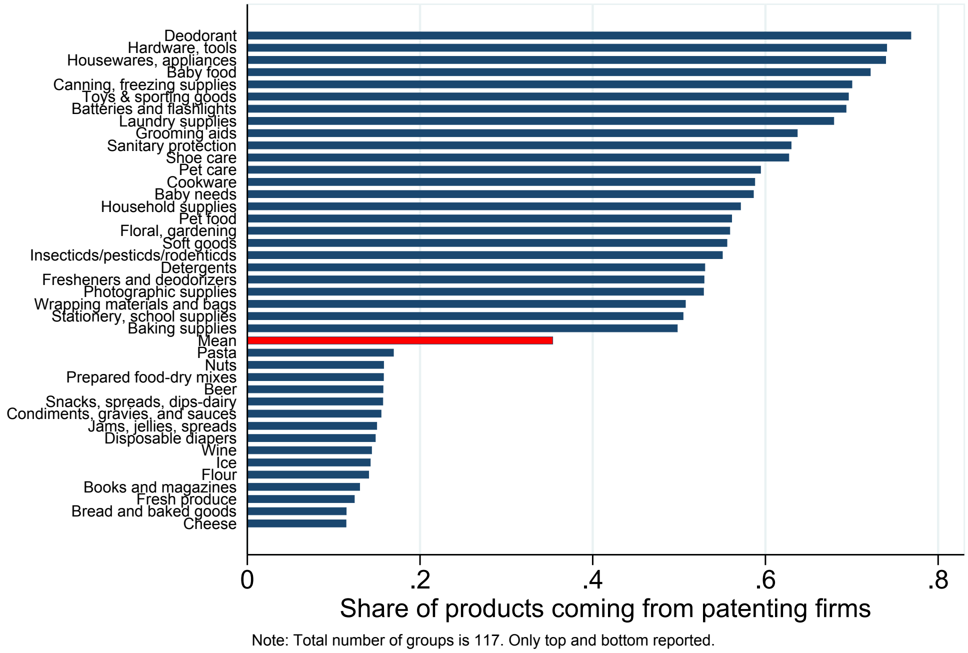
Table II: Summary Statistics by Patenting Status

	Firm's patenting status		
	(1)	(2)	(3)
	No Patents	Patents before '06	Patents in '06-'15
Revenue all products	3708.75	12275.72	27598.79
Revenue new products	243.27	988.62	1670.12
Revenue new products, post entry period	386.62	2064.76	3955.06
Number of products	16.11	35.09	74.09
Number of new products	2.60	6.73	12.91
Product entry rate	0.19	0.17	0.22
Number of product categories	2.36	3.11	4.11
Share of new products lasting more than 4 quarters	0.74	0.70	0.75
Share of new products lasting more than 16 quarters	0.44	0.40	0.42
Average newness of new products	0.13	0.09	0.10
Newness-weighted number of new products	0.57	0.70	1.07
Number of patent applications	0.00	0.00	6.14
Number of granted patent applications	0.00	0.00	4.47
Number of citations-weighted patent applications	0.00	0.00	8.87
Stock of patent applications until year t	0.00	10.88	125.93
Stock of granted patent applications until year t	0.00	10.59	115.39
Number of different technology classes (IPC3) on patents	.	.	5.56
NumFirms	29373	1879	3284
Observations	188118	15285	29030

Notes: The table shows descriptive statistics for a pooled sample of firms for the period 2006-2015 by firms with different patenting status. The first column is for firms that have no patents, the second column is for firms that have patents but before they first appear in Nielsen RMS (before 2006), and the third column is for firms that have patents in our main period 2006-2015. Observations are at the firm × year level.

products must last longer. Thus, we call categories with few trips per year durable categories. Examples of durable categories are sun exposure trackers (1.00), bathroom scales (1.03) and printers (1.03), where the average number of shopping trips per year is in parenthesis. Examples of non-durable categories are refrigerated milk (23.61), cigarettes (19.19) and fresh bread (18.76). Our measure of durability is the inverse of the average number of trips per year in a given product group. Panel (a) in Appendix Figure A.10 shows the relation between the share of patenting firms and our durability measure. The figure shows a clear positive relationship between the fraction of firms patenting in a group and the durability of the products sold in that group. Panel (b) shows that the relationship holds if we focus on the share of new products by patenting firms. These figures suggest that firms are more likely to rely on patenting in sectors where products are more likely to last longer after a household purchase.

Figure 2: Patenting Intensity by Product Groups



Notes: the figure shows the share of products belonging to patenting firms (patenting intensity) for a sample of product groups in the Nielsen data. The figure shows the intensity of the top 25 groups, the average intensity, and the intensity of the bottom 15 groups.

III Patents and Product Innovation

How well do patent-based metrics of innovation capture the actual product innovation in the market? To shed light on this question, we start by documenting that a large amount of product innovation comes from firms that never patent (Section A), followed by exercises that patenting is positively associated with product innovation by the firm (Section B). The positive correlation is corroborated using regression analysis for the behavior of firm's product innovation before and after the first patent application (extensive margin), and for the comovement of new patents and new product introduction (intensive) among patenting firms.

In the section we establish the following stylized facts:

Fact 1: A large amount of product innovation comes from firms that do not patent.

Fact 2: Patenting is positively associated with product innovation.

III.1 Product Introduction and Firm's Patenting Status

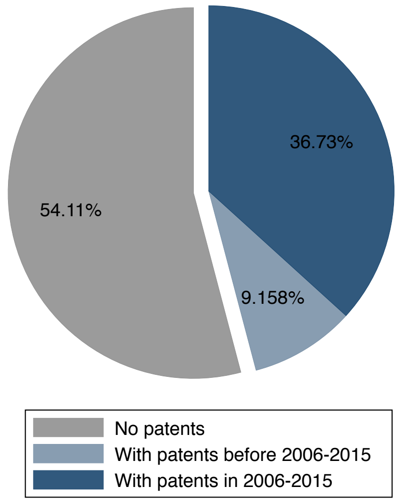
We start the analysis of the relationship between patents and products by exploring cross-sectional variation across firms according to their patenting status. Using the baseline dataset at the firm level (Match 1), we compute the total number of new products introduced in the consumer goods sector in the period 2007-2015 by firms according to their patenting status.[22] In the data, approximately 54% of products were introduced by firms that never applied for a patent (Figure 3). Our interpretation of the magnitude is that a large share of innovation is not captured when using patents as proxies for innovation. One concern with this result may be that firms in the consumer goods sector do not use the patenting system, given that they produce mostly non-durable and semi-durable goods. Nevertheless, 15% of the firms in our dataset use the patent system, which is larger than firms in other sectors: Graham et al. (2018) reports less than 1% of firm in the entire U.S. economy (accounting for 33 % of employment) and 6% in the manufacturing sector.

Another source of concern may be that non-patenting firms may introduce products that are only small upgrades to the existing products, and thus not patentable. A requirement to have a patent granted is "novelty and non-obviousness", and thus many new products that result from very small changes will be captured by the product introduction measure, but not by the patents measure. We evaluate if patenting firms are more likely to add new varieties by computing the total number of "novel" products (as measured by our quality-adjusted measures) in the period 2006-2015 by firms according to their patenting status. We find that non-patenting firms introduce, on average, more novel products in the market as measured by the importance of new attributes of the barcodes (Figure A.9 in Appendix). Thus, while a patent may be perceived as a measure of novelty, there are many firms introducing products in the market that have observable new attributes and that do not patent any of those products.

Another consideration regarding the relevance of the statistic on the share of products introduced by patenting firms is that some of these firms' new products may not be directly supported by their patents. Consider, for example, firms that did not apply for a patent in the period 2006–2015 (Figure 3), but have patents prior to that period. Those firms created 9% of new products in the period 2007-2015, and many of those may be unrelated to their patents. Likewise, even for the firms that applied for at least one patent in the period under analysis, it could be that many of their new products are not the result of innovations re-

[22] Note that since the data starts in 2006, we cannot define new products for 2006.

Figure 3: Share of Product Introduction by Patenting Status



Notes: the figure shows the share of new products by firms' patenting status. It shows the share of entering products launched by firms without patents, with all patents before 2006, and with patents between 2006-2015. In Appendix, Figure A.9 shows the equivalent statistics for the different measures of quality-adjusted product introduction.

flected in their patents.[23] Given these considerations, we interpret the statistic on the share of products introduced by never patenting firms as a lower bound of the amount of product innovation not captured by patent measures on innovation.

III.2 Correlation of Patents and Product Introduction

III.2.1 Extensive Margin of Patenting

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 if there are changes in the firm's product introduction following the first patent application using an event study approach. To evaluate if firm's outcomes change around the time a firm starts patenting, we

[23]This is particularly relevant for highly diversified firms that could be patenting in one product category but introducing products, unrelated to those patents, in another category during the same period. This observation exemplifies the importance of matching specific patents to firm $\times$ product category pairs. For example, using our matched data, we replicated the exercise above, defining patenting status at the level of the product category, and the share of new products introduced by firms that never patented in that sector is substantially larger.

estimate the following specification:

$$\ln Y_{jt} = \sum_{k \neq -1} \beta_k \mathbb{1}\{K_{jt} = k\} + \alpha_j + \gamma_t + u_{jt} \tag{3}$$

where Y_{jt} is the outcome of firm j in year t (e.g. new products), α_j represents firm fixed effects and γ_t are year effects. K_{jt} denotes the number of years at t relative to the first patent application so that β_k for $k < 0$ correspond to pre-trends and $k \geq 0$ to dynamic effects k periods after the first patent application. Since the error term could be serially and cross-sectionally correlated, we use Driscoll and Kraay (1998) standard errors. We consider all firms that appear in the Nielsen data during our period of analysis, but require that firms switching to patenting status are active at least two years before the event in order to be able to estimate pre-trends.

Figure 4 presents the estimated change on number of new products (logs) associated with a switch in status from being a non-patentee to a patentee. The figure shows that, conditional on firm and year effects, there are no pre-trends in the number of new products 3 years before the first patent application of the firm. The dynamic effects after the first patent application indicate an average increase in product introduction of up to 20% percent after the switch to patenting. Firms that become patentees in t=0 (year of the application) increase their product creation on impact, and this effect persists in the following years.[24]

We further evaluate this relationship using alternative definitions of product introduction and patenting status. Table III shows that the estimated change on number of new products (logs) associated with a switch in status from being a non-patentee to a patentee, conditional on firm and time fixed-effects. The results show that the positive correlation is largely driven by high-quality patents (using their granted versus abandoned status as a proxy) as seen from comparing columns (1)-(3).

Following Acemoglu et al. (2019), we tested the strength of the correlation by estimating a dynamic panel model using lags of the dependent variables, to better account for pre-trends in outcome variable and other omitted characteristics that may affect the likelihood of becoming a patentee. The regression consists of adding to the specification one lag of the outcome

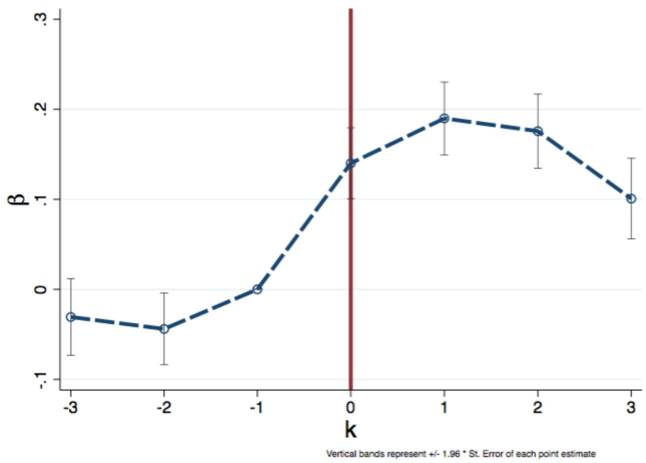
[24]Note that, on average, a patent is granted two to three years after the application filed. Therefore, the persistent effect we observe is not surprising given that some firms may decide to introduce a product at the time of the application and others at the time the patent is granted. Moreover, the persistent effect could be the result of firms deciding to keep filing for patents in the subsequent years after they switch status.

Table III: Product Innovation after the First Patent: Firm-level evidence

	Baseline			Dynamic Panel		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: New Products (Log N)						
After I patent(t)	0.1738**			0.2004***		
	(0.053)			(0.052)		
After I granted patent(t)		0.1605**			0.1013*	
		(0.056)			(0.052)	
After I non-granted patent(t)			0.0696			0.2683**
			(0.127)			(0.128)
Log N(t-1)				0.0004***	0.0004***	0.0004***
				(0.000)	(0.000)	(0.000)
Observations	195,686	195,686	195,686	158,554	158,554	158,554
R-squared	0.897	0.897	0.896	0.899	0.899	0.899
Time	Y	Y	Y	Y	Y	Y
Firm	Y	Y	Y	Y	Y	Y
Panel B: Quality-adjusted New Products (Log q-N)						
After I patent(t)	0.1806*			0.2678**		
	(0.089)			(0.102)		
After I granted patent(t)		0.2668**			0.2583*	
		(0.125)			(0.157)	
After I non-granted patent(t)			-0.1656			-0.0168
			(0.220)			(0.211)
Log q-N(t-1)				0.0013	0.0013	0.0013
				(0.002)	(0.002)	(0.002)
Observations	50,146	50,146	50,146	24,495	24,495	24,495
R-squared	0.836	0.836	0.836	0.800	0.800	0.800
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 ($Log\ q - N$) in a firm × year as a function of a dummy equal to one after the first patent application in a firm × year. Quality of a product is based on our $Newness$ index defined in Section II.1.2. $log\ N$ and $Log\ q - N$ 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).

Figure 4: Event Study: Patenting Status



Notes: The figure plots the estimated coefficients of the impact of switching to patenting status on new products. The observations are at the firm × year level. The coefficients were estimated using an event study approach, as described by equation (4). The sample of firms include all of those that appear in the Nielsen RMS from 2006-2015. The estimates are computed using revenue weights. The vertical lines indicate 95% confidence intervals.

variable – log new products – to control for the dynamics of new products. Columns (4)–(6) of Table III shows that, while the number of new products are serially autocorrelated, the relationship between becoming a patentee and product introduction remains mostly similar.

The exercises above show that we can statistically identify a positive correlation between the timing of patenting and product introduction. One interpretation of this correlation is that firms come up with some ideas for new products and they may want to protect them from being copied by competitors (by applying for a patent). Simultaneously, they develop those ideas into consumer products.[25]

III.2.2 Intensive Margin of Patenting

We now explore how product innovation reacts to changes in intensive margin of patenting as firm accumulates patents (granted and non-granted). As before, we are interested in evaluating the timing of an increase in the number of patents on the product innovation for

[25]Likewise, it may also be the case that patenting gives firms some status in the economy that allows them to create more products in the future. This status can result from consumers' perception of "innovative" firms when firms advertise "patent pending" on their products.

patenting firms. We explore the following specification:

$$\ln Y_{j,t+k} = \beta \ln \text{Patents}_{j,t} + \alpha_j + \gamma_{t+k} + u_{j,t+k} \text{ , } k = -5, ..., 0, ..., 5 \tag{4}$$

where $Y_{j,t+k}$ is the outcome (e.g. number of new products) of firm j in $t+k$, $\text{Patents}_{j,t}$ are the new patents of the firm j in t, α_j are firm fixed-effects, and γ_{t+k} are time fixed-effects. The semi-elasticity of the number of new products to a new patent application filed k years ago is given by the sum of the β coefficients from period 1 to k.

Figure 5 shows the estimated effects. We observe a positive association between new patents and product introduction that spikes one year after the patent application. Unlike the persistent effect we observed after switching to patenting, we do not observe that the accumulation of patents is associated with long-lasting effects. In the Appendix, we consider an alternative measure of product innovation to evaluate if the result is robust (Figure A.12). We find, that the result does not depend on the measure of product innovation used: quality-adjusted measures produce similar patterns. We also explore different patenting measures: granted vs non-granted patents, as well as compare so-called product patents – that presumably should correlate with product introduction more – with process-patents.[26] We find that product introduction co-moves stronger with granted patents and with product patents.

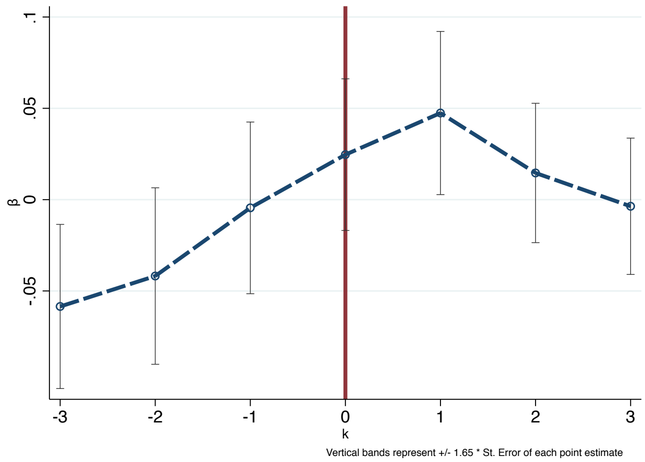
The firm-level analysis may be unsatisfactory to analyze the impact of patents on the introduction of new products, however, if firms operate in multiple unrelated product categories. It could be the case that the patents for one product category are filed during the same period the firm decides to introduce new products in a different category. In this case, our firm-level estimates could capture a spurious correlation between patents and products. Hence, we rely on our Match 2 at the firm $\times$ category level and a apply similar analysis. An advantage of this match is that now we can also condition on product category specific trends (for example, market-wide demand for specific products), as well as firm-category specific effects thus filtering out, for example, firm-specific brand power in specific products.

Table IV presents the estimates of the regression that exploits variation at the firm-category level over time. The columns present results using different explanatory variables – log number of patents, granted patents, and non-granted patents in $t-1$.[27] The standard

[26] We distinguish product from process patents using the information in the patent claims. Details and full description of how we classify patents into product and process patents can be found in Appendix Section D.3.

[27] The results are similar if we use contemporaneous dependent variables instead.

Figure 5: Product Innovation and Patenting Intensity



Notes: The figure plots the estimated coefficients after estimating equation (1) on log number of new products. The graph should be read as follows: Firms that become patentees in t=0, change product creation by β percent in t=-3,...,4.

errors used allow estimates to be both autocorrelated and heteroskedastic. As before, the relationship between patenting and product innovation is positive and mainly driven by higher-quality granted patents. Appendix Tables A.IV and A.V also show that the impact of patenting on new product introduction or quality-adjusted new products is mainly driven by product-related patents as opposed to process-related patents, as defined in Section D.3.[28]

Overall, these exercises show that we can statistically identify a positive correlation between patenting and product introduction, which corroborates that patenting is positively associated with product innovation by the firm.[29]

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

Previous sections show that, although a large amount of product introduction in the market is not related to patenting, conditional on patenting, patents are positively associated with

[28]In Tables A.VIII and A.IX we show that these finding are robust if we consider only granted patents.

[29]The results presented in this section are unaffected if we restrict the sample of firms to those that only operate in the CPG sector, providing evidence that our estimates are not driven by patents filed for activities performed in other sectors. In Appendix A.2 we describe our procedure to identify firms that mostly operate in the CPG industry.

Table IV: Product Innovation and Patenting: Intensive Margin

	Baseline			Dynamic Panel		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: New Products (Log N)						
Patents(t-1)	0.0412***			0.0241**		
	(0.010)			(0.010)		
Patents granted(t-1)		0.0467***			0.0271**	
		(0.011)			(0.011)	
Patents non-granted(t-1)			0.0204			0.0173
			(0.014)			(0.015)
Log N(t-1)				0.0325***	0.0325***	0.0325***
				(0.003)	(0.003)	(0.003)
Observations	412,004	412,004	412,004	365,402	365,402	365,402
R-squared	0.691	0.691	0.691	0.702	0.702	0.702
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y
Panel B: Quality-adjusted New Products (Log q-N)						
Patents(t-1)	0.0188***			0.0109		
	(0.006)			(0.007)		
Patents granted(t)		0.0214***			0.0103	
		(0.007)			(0.007)	
Patents non-granted(t-1)			0.0012			0.0013
			(0.010)			(0.011)
Log q-N(t-1)				-0.0262***	-0.0262***	-0.0262***
				(0.003)	(0.003)	(0.003)
Observations	412,004	412,004	412,004	365,402	365,402	365,402
R-squared	0.558	0.558	0.558	0.575	0.575	0.575
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	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 ($Log\ q - N$) in a firm × category × year in Panel B as a function of log number of patents. $log\ N$ and $Log\ q - N$ use the inverse hyperbolic sine transformation. Quality of a product is based on our *Newness* index defined in Section II.1.2. *Patents* is the natural logarithm of the 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). *Patents*, *Patents granted*, and *Patents non-granted* use the inverse hyperbolic sine transformation.

product innovation. This indicates that patents contain important technological improvements that firms, on average, commercialize in new products. However, by definition, patent has also a protective role: firms can use patents strategically to defend their technology reducing competitive pressure and deterring entry (Cohen et al., 2000). How do firms use these two roles of patenting? In their classical paper, Gilbert and Newbery (1982) suggest that monopolists have high incentives for preemptive patenting. Likewise, Blundell et al. (1999) argue that patents of market leaders may carry large preemptive component. In this section, we shed light on these issues by evaluating how product innovation and patenting varies systematically with firm's market lead (relative sales in the market). In the first part, we establish the following empirical regularities:

Fact 3: Larger firms have lower product innovation rate (quantity and quality), but higher patents per new product.

Fact 4: Patents relate to higher future sales beyond their effect through product introduction, especially for large firms.

Fact 5: Patents of larger firms are associated with declines in competitors' product introduction.

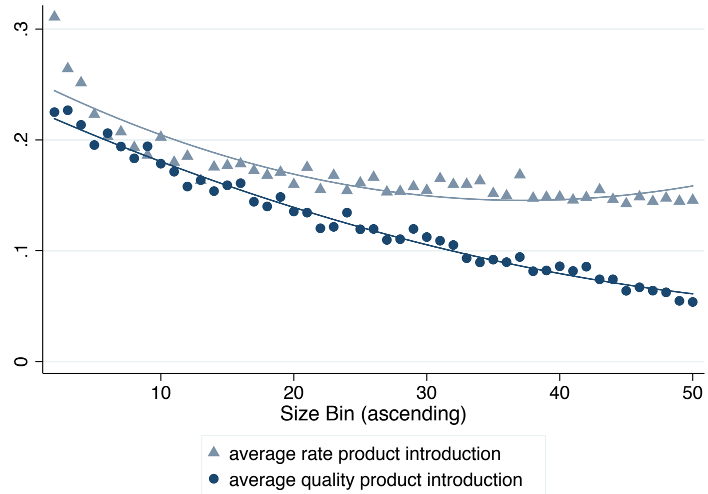
Our analysis shows that product innovation rate declines as firms grow and, at the same time, their patent applications increase. The decline in innovation rates is stronger when we use quality adjusted measures of innovation, which suggests that the use of patenting among larger firms is not explained by larger firms introducing more novel products (as one could expect given that patents are perceived as a measure of novelty).

In this section we establish the following empirical regularities:

IV.1 Product Introduction and Patenting by Firm Size

We begin by exploring how product innovation rates evolve with firm size. In each product category, we rank firms based on their sales and plot the average innovation rate for each size bin. Figure 6 shows that larger firms have lower product innovation rates. The decline in innovation rate is not compensated by innovations of higher quality either; on average, new products of larger firms' have lower levels of newness.

Figure 6: Product Innovation Rate by Size



Notes: This figure plots the relationship between product innovation and size of the firm, defined by sales. We use data on product innovation at the firm x product category for the period 2007–2015. For each product category, we compute the average annual product innovation rate (new products divided by stock of products) and the average quality of the new products (defined by their newness), across 50 bins of size. The figure shows the average after weighting the different product categories by their revenue share.

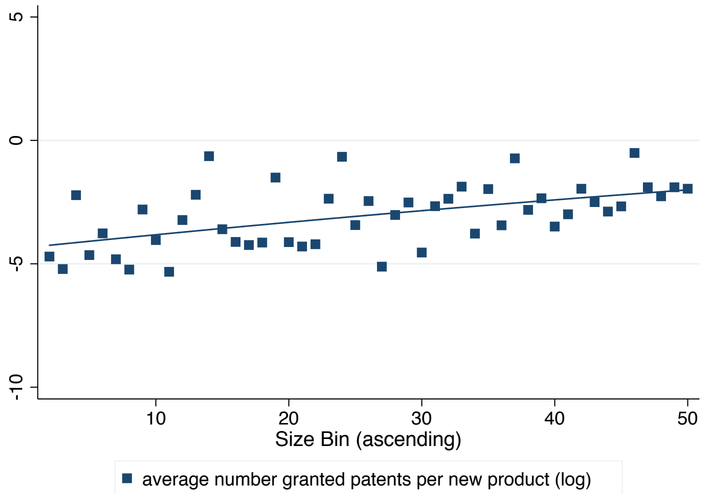
Although product innovation rate declines as firms become market leaders, patenting increases. Figure A.11 shows that patenting disproportionately happens among large firms. Likewise, the number of patent applications increases with size as well (Figure A.13). Putting all together, Figure 7 shows that the ratio of new patents per new products is larger for larger firms.

Motivated by this evidence, we now look at the elasticity of product innovation rates to patenting and firm size using the following specification:

$$E_{fmt} = \alpha + \beta \text{Pat}_{fmt-1} + \gamma \text{Size}_{fmt} + \lambda \text{Pat}_{fmt-1} \times \text{Size}_{fmt} + \eta_{fm} + \theta_{mt} + \epsilon_{jfmt}$$

where E_{fmt} denotes either the entry rate of products launched at time t by firm f in category m, or the average quality of the new products at time t by firm f in category m. Patents_{fmt-1} refers to the number of patent applications the previous year. Size_{fmt} is the natural logarithm of the revenue of firm f in module m at time t. Table V reports our results under several specifications. Patenting is positively related to the rate at which firms introduce new products. Size_{fmt} is mostly negatively associated with innovation rate. Importantly, the interaction between Pat_{fmt-1} and Size_{fmt} indicates that we do not find evidence that patenting increases the product innovation rate of firms as they grow. We find similar

Figure 7: Ratio of Patents per New Product by Size



Notes: This figure plots the relationship between patenting and firm size. We use data on product innovation and patents at the firm x product category for the period 2007–2015. For each product category, we compute the number of new patents per product across 50 bins of size. The figure shows the logs of the average after weighting the different product categories by their revenue share.

results when we focus on the quality of the product introduced by firms, as measured by their average newness. Column (5) shows that the quality of product innovation decreases as firms grow and, while larger firms do relatively more patenting, the interaction term in column (6) shows that this does not translate into higher-quality products.[30]

[30]One could think that this relationship may be driven by firms moving away from product to process innovation (Cohen and Klepper, 1996). We examine this possibility in two ways. Leveraging on our classification of patents into product and process patents (see Appendix Section D.3), we do not observe that firms switch to process patents as they grow. On the other hand, it could be the case that firms, as they grow, focus on process patents to decrease their costs. We examine this question by studying the relationship between process patents and the change in log prices at the firm × category × level. Table A.XI shows that we do not find a relationship between process-related patents and price changes. Table A.XII shows that this finding is robust to interacting process-related patents with the size of the firm.

Table V: Relationship Between Patenting and Product Innovation by Size

	Product Introduction Rate			Quality Product Introduction		
	(1)	(2)	(3)	(4)	(5)	(6)
Patents(t-1)	0.010***		0.011*	0.002		0.018***
	(0.003)		(0.006)	(0.007)		(0.007)
Size(t)		-0.052***	0.041***		-0.112***	-0.027***
		(0.003)	(0.002)		(0.002)	(0.002)
Patents(t-1) x Size(t)			-0.001			-0.012***
			(0.004)			(0.004)
Observations	312,380	334,110	312,380	93,912	112,218	93,912
R-squared	0.386	0.392	0.388	0.510	0.538	0.511
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y

Notes: The table shows the relationship between the rate of product innovation, patenting and firm size. The product innovation rate is the rate of product introduction by firm f in category m at time t. The quality product innovation is the product innovation rate defined by their newness. Patent(t-1) is the natural logarithm of patents applications of product category m by firm f in year t, using the inverse hyperbolic sine transformation. Size(t) is the natural logarithm of the total sales of firm f in product category m at time t (standardized).

IV.2 Patenting and Competition by Firm Size

Large firms patent more on average, yet we find no evidence that these patenting activities translate into more product introduction or higher quality of innovations. Are large firms using their patents to affect competitors rather than to innovate? To answer this question, we first evaluate the relationship between patents and firm revenue. We establish that patents relate to higher future sales beyond their effect through product introduction, especially for large firms.

Table VI shows relationship between patents and firm revenue. Our dependent variable is firm's yearly sales in a product category at time t, while the main explanatory variable is log new patent applications at $t-1$. Overall, there is a positive significant relationship between patents and future revenue, even conditional on product innovation (columns 1 and 2). Hence, there is additional value from holding a patent beyond its value through new

product offerings. By splitting the sample into small and large firms (below and above the median level of firm sales in the product category) in columns (3) and (4), an independent effect of patents after conditioning on product innovation disappears for small firms, but is still significant for large firms. Hence, while both patents and new products are associated with higher future sales, the conditional impact of patents is larger for large firms.

Table VI: Patenting and Revenue

	(1)	(2)	(3)	(4)
Log Revenue				
Log Patent(t-1)	0.0509***	0.0447***	0.0394	0.0528***
	(0.017)	(0.017)	(0.040)	(0.019)
Log N(t)		0.2659***	0.2159***	0.2868***
		(0.004)	(0.006)	(0.005)
Observations	314,815	312,380	147,857	164,434
R-squared	0.906	0.909	0.734	0.846
Time-Category	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y
Sample	All	All	Small	Large

Notes: The table shows regressions of log revenue at the firm $\times$ category $\times$ year level as a function of log number of patents and the log number of products introduced. *Patents* is the natural logarithm of the number of any patent applications in firm $\times$ category $\times$ year using the inverse hyperbolic sine transformation. *Log* $N(t-1)$ is the natural logarithm of products introduced by firm f, using the inverse hyperbolic sine transformation. Columns (4) and (5) split firms according to their size. Column (4) considers firms below the median size (measured in total revenue) within each category. Column (5) considers firms above the median size. The results of the table consider firms $\times$ category combinations with more than \$1000 revenue in a given year.

This additional revenue premium from a patent for larger firms may likely operate through its effect on competition: if patents discourage competitors, patent holders could benefit by serving a larger market. We explore whether the product innovation of competitors in a given market declines when the market leader introduces a product that is patented. We use the following specification:

$$N_{m_{-f}t} = \alpha + \delta N_{fmt-1} + \beta \text{Pat}_{fmt-1} + \gamma \text{Size}_{fmt} + \eta N_{fmt-1} \times \text{Size}_{mt} + \lambda \text{Pat}_{fmt-1} \times \text{Size}_{mt} + \epsilon_{jfmt}$$

where $N_{m_{-f}t}$ is the natural logarithm of the number of products introduced by firms other than firm f in module m at time t, N_{fmt} is the natural logarithm of the products introduced by firm f in module m at time t, Pat_{fmt-1} is the natural logarithm of the patent applications

by firm f in module m at time t.[31] Our coefficient of interest is λ on the interaction between the size of the firm and the number of granted patents, after accounting for the effect of new products on product introduction by others.

Table VII: Patenting and Competitor's Product Introduction

	(1)	(2)	(3)	(4)	(5)
New Products by Other					
Log N(t-1)	-0.00003	-0.00003	0.00045***	0.00046***	-0.01103***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Patent(t-1)		-0.00083		0.00250**	0.00222
		(0.001)		(0.001)	(0.002)
Size(t)			-0.00479***	-0.00465***	-0.00803***
			(0.000)	(0.000)	(0.000)
Patent(t-1) x Size(t)				-0.00296***	-0.00905***
				(0.001)	(0.001)
Observations	264,060	264,060	264,060	264,060	204,166
R-squared	0.99	0.99	0.99	0.99	0.99
Time-Category	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	N
Time-Firm	N	N	N	N	Y

Notes: The table shows the relationship between product introduction of other firms $\hat{f}$ within the product categories m and the size of firm f, the product introduction of firm f , and whether these products are related to a patent. The dependent variable is the natural logarithm of the number of products introduced by firms other than firm f. New(t-1) is the natural logarithm of products introduced by firm f, using the inverse hyperbolic sine transformation. Patent(t-1) is the natural logarithm of patent applications of product category m by firm f applied in year t, using the inverse hyperbolic sine transformation. Patent/New(t-1) is the natural logarithm of granted patents per product. Size(t) is the natural logarithm of the total sales of firm f in product category m at time t (standardized).

Table VII shows our results under several specifications. Columns 1-2 show that both the introduction of new products and patenting are associated with a decline in product innovation rate of competitors within the same product category. Columns 3-5 show that patents launched by larger firms unambiguously relate to the decrease of the rate of product introduction of other firms. Our preferred specification to study whether the behavior of market leaders affects the innovation rate of their competitors is shown in column 5. It shows that, controlling for product category×time and firm×time effects, new products launched by larger firms have a negative effect on the rate at which competitors introduce new products. Protected innovations by larger firms have the opposite effect. In other words, patents seem

[31] We provide results using granted patents and the total number patents of the firm in the Appendix.

to deter innovation of competitors, but only if they belong to larger firms. This is particularly important when larger firms introduce products that are patented. Overall, our findings suggest that the value of a patent for a market leader may come from the deterrence of competitors and that an additional revenue generated by leading incumbents' products comes in from this competitive margin, rather than from protecting and launching high-quality innovative products.

V Conceptual Framework

How do firms decide on patenting and product innovation? Where does the value of a patent come from? On the one hand, a patent is meant to represent certain innovation that will be commercialized (in new and improved products) to generate returns for a firm. On the other hand, patents may be used for other reasons not directly related to innovation. Many patents, for example, are filed purely for protective reasons – strengthening existing patents' position and with the intention to exclude competitors from using a particular invention or building a new one around it (e.g., Cohen et al. (2014), Abrams et al. (2013)). Hence, the value of a patent for firms should capture both, the value from innovation, as well as value from deterring competition. Our results from the previous section suggest that this last type of motive for patenting is likely important for large firms.

In this section, we offer a simple illustrative framework of innovation, patents, and product introduction. Our goal is twofold. First, the framework is meant to build intuition to understand the incentives for patenting, consistent with empirical patterns documented in the previous sections. Second, we use the model to write a simple back of the envelope calculation for the private value of a patent and its decomposition into protective versus productive components. We refer to the *protective* component of patent value as the value that patent brings to a firm by limiting creative destruction, holding fixed the technology of a firm. We refer to the *productive* component of patent value as the value that patent brings to a firm that commercializes the innovation embedded in the patent, holding creative destruction fixed.

Our framework builds on the quality-ladder model setup (e.g. Aghion and Howitt (1992)). Innovations come from a technology leader trying to prolong its lead or from potential entrants aiming to become the new leader. In our simple framework, there is an incumbent firm that obtains a costless blueprint/idea and makes once in a lifetime decision regarding

product commercialization and patenting of the idea. If the firm decides to introduce a new product, it incurs the cost of product development and commercialization to gain extra revenue from higher-quality products. Simultaneously the firm also decides if it will patent the blueprint. Patenting involves costs and grants the firm extra protection against being replaced by an entrant. The model has three basic ingredients: product upgrade exhibits decreasing returns, the probability of creative destruction depends on patent protection, and patenting as well as product introduction are costly activities. The model can rationalize main empirical findings of the previous section. For the same blueprint quality level, small firms will upgrade its products but not patent, middle-sized firms will upgrade and patent, and large firms will patent but not upgrade its products.

V.1 Model

Preferences and technology Consider a partial equilibrium framework describing innovation in a single sector. Suppose that there is a large number M of potential producers and the preferences are characterized by the following function

$$Y = \frac{1}{1-\beta}\left[\sum_{m=1}^{M} q_m^{\frac{\alpha}{1-\beta}} y_m\right]^{1-\beta}, \quad 0 < \alpha < \beta < 1 \tag{5}$$

where y_m denotes the output of a producer m. Different potential producers differ by their qualities q_m. The specification for the preferences implies that products from different producers are perfect substitutes after adjusting for their qualities. The demand elasticity relates to the inverse of β and α captures the consumer's satiation with respect to extra quality. The consumer's first order conditions (taking prices as given) implies that the demand faced by producer m is given by[32]

$$p_m = q_m^{\frac{\alpha}{1-\beta}}\left[\sum_{m=1}^{M} q_m^{\frac{\alpha}{1-\beta}} y_m\right]^{-\beta}. \tag{6}$$

Producers use labor to produce the output hiring labor on the spot market at a common wage rate of w. The production function is simply given by

$$y_m = l_m \tag{7}$$

[32] After normalizing the price index of the sector to 1

and thus all producers have a marginal cost of production w. We assume that producers of different qualities play a two-stage pricing game. In the first stage, producers choose to pay a fee ϵ (which we assume to be arbitrarily small) to enter a price competition. In the second stage, all firms that already paid the fee bid prices. This assumption, and given that products of different producers are perfect substitutes after adjusting for their quality, ensures that only the firm with the highest quality pays the fee and goes to the second stage where price is determined. Therefore, the producing firm charges the unconstrained monopoly price.[33]

The monopolist firm chooses the price of its product by maximizing its profit subject to the demand function, which delivers the following equilibrium objects for output (y), revenue (R), and profits (Π), respectively:[34]

$$
\begin{aligned}
y &= \left(\frac{1-\beta}{w}\right)^{\frac{1}{\beta}} q^{\gamma}, \\
R &= \left(\frac{1-\beta}{w}\right)^{\frac{1-\beta}{\beta}} q^{\gamma} \equiv \frac{\pi}{\beta} q^{\gamma}, \\
\Pi &= \left[\left(\frac{1-\beta}{w}\right)^{\frac{1-\beta}{\beta}} - w\left(\frac{1-\beta}{w}\right)^{\frac{1}{\beta}}\right] q^{\gamma} = \pi q^{\gamma}
\end{aligned} \tag{8}
$$

where $\gamma \equiv \frac{\alpha}{\beta}$. The specifications above make it clear that firms with higher-quality products are larger, and generate higher revenue and profits. Moreover, notice that $\gamma < 1$ because $0 < \alpha < \beta < 1$, and thus the marginal quantity, revenue, and profits decrease with quality.

Incumbent product innovation and patenting In this economy, upon entry the monopolist firm obtains a costless exogenous blueprint/idea of size λ and makes once in a lifetime decision regarding product quality upgrade and patenting.[35] If the firm decides to upgrade quality from q to $q + \lambda$, it needs to incur costs of product development and commercialization c_m. Simultaneously, the firm also decides if it will patent the blueprint and set the patent quality level to $q + \lambda$. To patent, firm needs to incur research and patenting cost c_l. As we will see, a patent grants the firm additional protection against being replaced

[33] 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.

[34] Hereafter, we drop the subscript m.

[35] For simplicity, we assume a one-time choice. Hence, the blueprint is either used or disappears afterwards. A more complete approach with a dynamic choice 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 patent spaces.

by an entrant. Note that the firm can operate in product and patent spaces separately: product innovation does not necessarily imply patenting, and neither introducing a patent necessarily implies product introduction.

Creative destruction In this economy, incumbents can be replaced by entrants through creative destruction. There is exogenous arrival rate p of entrants at each instant. Entrants build on "the shoulders of giants" and can replace incumbents by improving upon the largest available (patent or product) quality. The underlying assumption is that entrants can learn from the products available in the market and from the patents. Hence, "the shoulders of giants" correspond to $q+\lambda$, with the exception of q if incumbent has neither upgraded nor patented. Entrants draw innovation step size λ^e from the 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 beat to enter the market. The parameter ε captures that entrants need to come up with the innovation sufficiently different from what has been patented, as well as the strength of intellectual property protection and broadness of a patent scope.

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 H provides the proof). Notice that unlike in standard models of creative destruction, not all product quality improvements by entrants will find their way to the market. The separation of patent and product spaces introduces a possibility that a better quality product is not being introduced to the consumers because it is blocked by active patents.

Value functions and equilibrium Let us denote the value of a firm who both product and patent upgrade as $V^{1,1}$, the value with product upgrade and no patenting as $V^{1,0}$, the value of no product upgrade with patenting as $V^{0,1}$, and the value with no product upgrade and no patenting as $V^{0,0}$. The value function of the incumbent is given by

$$V(q) = \max\Big\{ V^{11}(q) - c_m - c_l,\ V^{10}(q) - c_m,\ V^{01}(q) - c_l\ ,\ V^{00}(q) \Big\}, \tag{9}$$

where

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

Notice that the incremental gain from product innovation declines, while the returns to patenting increase with firm size. Appendix H shows that in this economy there exist cutoffs q^* and q^{**} such that firms do only product innovation when $q < q^*$ and do only patenting when $q > q^{**}$. This requires only mild conditions (Condition (i) and Condition (ii) from H.2) on costs c_m and c_l that ensure that at least one firm finds it profitable to do product introduction, and at least one firm finds it too costly to engage in research and patenting.

Under the mild conditions above, the model delivers equilibrium that rationalizes the main empirical patterns uncovered in the previous section. Because of diminishing returns in quality, incremental returns from product innovation decline with firm size. The same intuition underlies a well-known *Arrow replacement effect* – larger firms/monopolists find it less profitable to replace themselves: innovations cannibalize their own rents.[36]

Implication 1: Larger firms have lower incentives for product innovation.

In our framework incentives for patenting increase as firms grow – larger firms have higher value to protect, hence rely on patenting more.

Implication 2: Larger firms have higher incentives to protect their innovations by patenting.

Corollary: Elasticity of product innovation to patenting is positive and declines with size.

With the parameter restrictions as in Appendix H, we can also show that in the intermediate range $q^* < q < q^{**}$, firm strictly prefer engaging both in product innovation and patenting.

Finally, the model also speaks to our empirical facts on the relationship between patenting, firm size, and deterrance. 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 creative destruction.

Implication 3: Larger firms deter competition by patenting more than small firms.

Overall, this stylized model gives a simple rationalization to our main empirical regularities on product innovation and patenting. *Implication 1* and *Implication 2* is consistent with the empirical evidence used to support *Fact 3*. Moreover, *Implication 3* is consistent with our empirical evidence regarding the positive association of patents with future sales beyond their effect through product introduction (especially for large firms), and our results that

[36]The Arrow replacement effect is stronger for larger firms because the incremental profit from λ decreases with q. This result is partially a result of the assumption that product upgrade exhibits decreasing returns (consequence of the consumer's satiation with respect to extra quality α).

show that patents of larger firms are associated with declines in product introduction by competitors. Clearly, to generate analytical predictions, the model abstracts from richer dynamics and a more realistic cost structure. However, the main ingredients – both the reduction in product innovation incentives as firms grow and the increase in protection incentives through patenting – are easy to generalize at the expense of analytical tractability. We now use this model to write down the value of a patent and its components.

V.2 The value of a patent

What is the private value of a patent to a firm in our model? If a firm were to price its patent, how much would it value it? By nature, 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 innovation, while protective value comes from the ability of a patent to protect firms' market lead from competitors.

We define the value of a patent as the revenue premium that patented innovation provides:

$$\begin{aligned} Patent\ Value &= V^{11} - V^{00} \\ &= \underbrace{\frac{\pi(q+\lambda)^{\gamma}}{r+p(1-\varepsilon)} - \frac{\pi(q+\lambda)^{\gamma}}{r+p}}_{Protective} + \underbrace{\frac{\pi(q+\lambda)^{\gamma}}{r+p} - \frac{\pi q^{\gamma}}{r+p}}_{Productive} \end{aligned} \tag{10}$$

The second line of this equality decomposes total patent value into productive and protective parts by adding and subtracting V^{10}.[37] Protective value is the revenue premium from lower creative destruction, holding fixed the technology of a firm. Productive value is the revenue premium from commercialization – quality upgrade – holding creative destruction fixed.

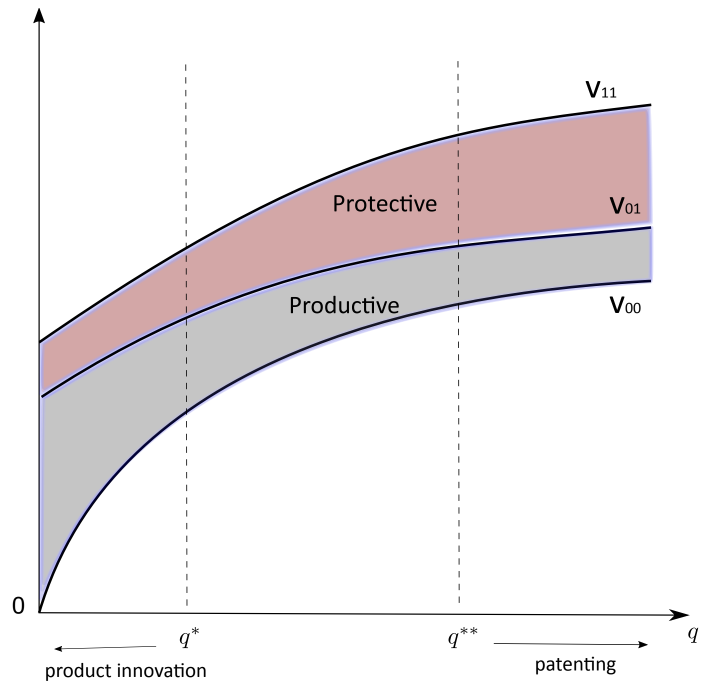
Figure 8 schematically illustrates productive and protective parts of the patent value as a function of q. The grey area – the incremental present value of revenue from higher-quality products on the market – denotes the productive component of patent value. This value, which is simply the value of innovation, declines as firms grow since the same innovation brings marginally lower return to larger firms. To the contrary, the red area – the incremental

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

present value of revenue from lower creative destruction – denotes the protective component of patent value. This value increases as firms grow since patent helps to protect larger value of the firm. Hence, we formulate our final implication of the model:

Implication 4: Patent value to larger firms is higher and is driven more by protective role of patents than productive role.

Figure 8: Firm size and value of a patent



Back of the envelope calculation We now choose parameters of the model to provide a simple back-of-the-envelope calculation of the implied patent value for CPG firms in our data. We need to assign values to $\pi, \lambda, \gamma, p, \varepsilon$ to estimate (10). First, we normalize average quality in each product category in the data to one. Notice that we do not observe profits, but given (8), we know that revenue is 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 $\frac{\mu-1}{\mu}$ times revenue of the average firm, which we take to be equal to the average yearly revenue of firms' across all product categories (1.36 million USD).[38] We take $\mu = 1.21$ based on the average estimate of markups in the U.S. economy in 2014 from Barkai (2017).

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

To assign values to p and $p(1-\varepsilon)$, consider the following. In the model, if firms do not innovate they face creative destruction rate that would lead to the decline of their expected sales in the next period. Hence, we infer the values of p from sales growth[39] of firms when they do not introduce new products in that year. In the data, the median firm that does not hold any patents shrinks when it does not introduce new products: log sales change is equal to -10.3 percent. This decline is attenuated if a firm holds a patent, thus giving us an estimate for ε. The implied values for creative destruction are $p = 0.098$ and $p(1-\varepsilon) = 0.095$.

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

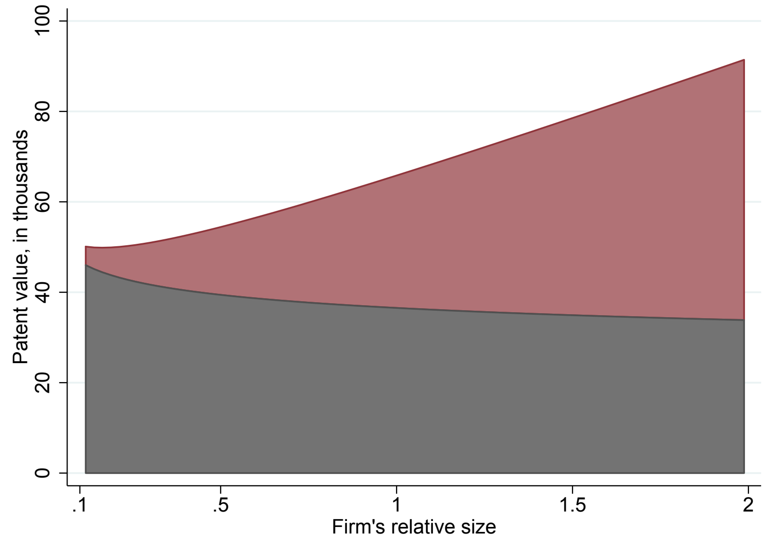
$$\Delta \ln R_t = \gamma ln\Big(1 + \lambda\Big[\frac{R_{t-1}}{\overline{\overline{R_{t-1}}}}\Big]^{-\frac{1}{\gamma}}\Big)$$

We estimate this relationship using non-linear least squares on the sample of firms who introduce new products in that year. We define relative size of firms as firm sales divided by 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 9 plots the resulting patent value over firm's relative size for firms around the mean. Red shaded area depicts the protective value of a patent, while grey area depicts the productive component. The estimated value for the average firm is around \$65,000 and increases with firm size. On average, 43% of this value comes from the protective component. This share varies drastically with firm size. For example, for firms ten times smaller than the average firm in the product category, only 9% of value is protective, while for large firms – twice larger than the average – the protective part goes up 60%. It is worth comparing our estimates of the patent value to the estimates in the literature. Using patent renewal information to infer private value of U.S. patents issued in 1991, Bessen (2008) estimates a mean value of \$121,000 (median \$11,000). Interestingly, consistent with our results, the paper also finds that patent value of smaller firms is lower, while litigated patents are more valuable. Serrano (2010) estimates the average private value of a patent right to be \$90,799 (median \$19,184). Using data from a large non-practicing entity – who presumably hold more valuable patents – Abrams et al. (2013) find that the mean patent value is \$235,723 (median \$47,955).

[39] Because of large outliers, we winsorize growth numbers at 5% and 95%.

Figure 9: Estimated patent value



Though our methodology is very different from the estimation methods in these papers, our estimates are well in the range of these numbers. However, the crucial advantage of our data combined with insights from our model is that we decompose the patent value into its two inherent components – productive and protective. This decomposition may serve as the first step towards better understanding different firms' motives for patenting and the heterogeneous effect of patent system on innovation, competition, and consumer welfare.

VI Conclusion

We construct a new patent-to-product dataset combining information from the U.S. Patent and Trademark Office with detailed product- and firm-level data for the consumer goods sector. Using textual analysis of patent documents together with product descriptions from product data and Wikipedia, we link patents to product categories within firms and time periods. These data allow us to study the patenting behavior of firms that operate in multiple, potentially distant, product categories controlling for category specific trends. Using these data, we find that more than half of the product innovation observed in the data comes from firms that have never patented. Nonetheless, we find that patenting is positively associated with product innovation.

We document substantial cross-sectional heterogeneity in the innovation activities of firms. Larger firms have lower product innovation rates, lower average quality of new products, but higher number of patents per new product. We provide empirical evidence that supports the hypothesis that patents by larger firms are able to generate high revenue by deterring the entry of products to be launched by competitors.

Using a simple framework that builds on quality-ladder models, we decompose the value of a patent considering both the productive and the protective components of a patent. We show that the total private value of a patent is increasing with firm size, because of the value derived from the protective component. We estimate that the average value of a patent in the consumer goods sector is around \$65,000 and increases with firm size. On average, 43% of this value comes from the protective component. We argue that understanding the contribution of the productive and protective components of the patenting, as well as how they vary by firm size has important implications for our understanding of growth, innovation, and intellectual property policy.

References

AB, Lerner Js Jaffe, "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.

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

Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James A Robinson, "Democracy does cause growth," *Journal of Political Economy*, 2019, *127* (1), 000–000.

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

_ **, 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," Working Paper 25136, National Bureau of Economic Research 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**, "How do firms grow? The life cycle of products matters," 2018.

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.

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?," Working Paper 23782, National Bureau of Economic Research 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.

Boldrin, Michele and David K. Levine, "The Case against Patents," *Journal of Economic Perspectives*, February 2013, *27* (1), 3–22.

Cavenaile, Laurent and Pau Roldan, "Advertising, innovation and economic growth," Working Papers 1902, Banco de EspaÃśa;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.

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

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.

Driscoll, John C and Aart C Kraay, "Consistent covariance matrix estimation with spatially dependent panel data," *Review of economics and statistics*, 1998, *80* (4), 549–560.

Furman, Jeffrey L. and Scott Stern, "Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research," *American Economic Review*, August 2011, *101* (5), 1933–63.

Gambardella, Alfonso, Dietmar Harhoff, and Bart Verspagen, "The value of European patents," *European Management Review*, *5* (2), 69–84.

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

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

_ , "Patent Statistics as Economic Indicators: A Survey," NBER Working Papers 3301, National Bureau of Economic Research, Inc March 1990.

Gutierrez, GermÃąn 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," Working Paper 8498, National Bureau of Economic Research 2001.

_ , _ , and _ , "Market Value and Patent Citations," *RAND Journal of Economics*, 2005, *36* (1), 16–38.

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

Harhoff, Dietmar, Frederic M. Scherer, and Katrin Vopel, "Citations, family size, opposition and the value of patent rights," *Research Policy*, 2003, *32* (8), 1343–1363.

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

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

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

Lampe, Ryan and Petra Moser, "Patent Pools, Competition, and InnovationâĂŤEvidence 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.

Lanjouw, Jean Olson, "Patent Protection in the Shadow of Infringement: Simulation Estimations of Patent Value," *The Review of Economic Studies*, 10 1998, *65* (4), 671–710.

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.

_ **and** _ , "Global Market Power," Working Paper 24768, National Bureau of Economic Research June 2018.

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.

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.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Nicholas Trachter, "Diverging Trends in National and Local Concentration," Working Paper 25066, National Bureau of Economic Research September 2018.

Sampat, Bhaven and Heidi L. Williams, "How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome," *American Economic Review*, January 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.

Scotchmer, Suzanne, "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law," *The Journal of Economic Perspectives*, 1991, *5* (1), 29–41.

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

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

Toivanen, Otto, Paul Stoneman, and Derek Bosworth, "Innovation and the Market Value of UK Firms, 1989âĂŞ1995*," *Oxford Bulletin of Economics and Statistics*, 2002, *64* (1), 39–61.

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 Data Appendix

A.1 Intermediate Categorization

As described in Section II.1, we need to develop an intermediate categorization of Nielsen products into product categories that are more aggregated than product modules and more disaggregated than product groups. As discussed in Section II.3.1, we use Wikipedia article texts to match each patent to a particular product category (be it module, group, etc.). The exact details of this procedure are discussed extensively in II.3.1. As part of this analysis, to each product module, we assign a set of Wikipedia articles that describe it. This turns out to be useful for creating new aggregations of modules into product categories. By mapping these documents into a vector space, we are able to use techniques such as k-means clustering to construct an intermediate aggregation. In our baseline analysis, this intermediate aggregation has 400 categories, and we investigate the implications of using numerous alternative clustering techniques.

Using word vector notation, each module is characterized through its associated Wikipedia articles by a word frequency vector. We can aggregate these module vectors into clusters using a popular technique known as k-means clustering. This procedure allows one to specify the desired number of clusters k beforehand and yields a partitioning that minimizes the within-group vector variance (average squared distance from cluster mean). That is, letting x be a given module vector and set S_i be cluster i, we choose our cluster sets S_i so as to minimize

$$\sum_{i=1}^{k} \sum_{x \in S_i} ||x - \mu_i||^2$$

where $\mu_i = \frac{1}{|S_i|} \sum_{i \in S_i} x$.

In our main analysis, we use $k = 400$ for our cluster size. However, we also investigate various other choices for this particular parameter. Additionally, we experiment with various other state-of-the-art clustering techniques such as HDBSCAN. Extensive manual checks confirm that our baseline k-means clustering partitions product modules very well. Regardless of the method used, the advantage of this type of approach is that it will group together precisely those product categories that the patent matching algorithm would have trouble distinguishing between, and vice versa.

To give an example, with this clustering, separate product modules ”Detergents – packaged”,

”Detergents – light duty”, ”Detergents – heavy duty”, ”Laundry treatment aids”, and ”Fabric washes – special” are grouped into one product category. It would be hard for the patent matching algorithm 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.

A.2 Classifying firms into CPG-only firms

Our data provides nearly universal coverage of product portfolios of firms in the CPG sector. However, firms' activities may lie also outside of this sector. To understand how big is the difference between firms' total sales and sales in CPG, we rely on two external data sources: Compustat data on publicly traded firms and National Establishment Time-Series (NETS) Database.

We combine Nielsen data with the Compustat database by matching Nielsen firm names to those in Compustat. We matched around 500 publicly traded companies over our sample period. Our matched sample represents 22% of the total sales in Compustat and 45% of the total revenue in the RMS. Approximately 21% of the total number of products in the data belong to publicly traded firms. We mostly use information contained in the Compustat 10-k. A 10-k is a comprehensive summary report of a firm's performance that must be submitted annually to the Securities and Exchange Commission (on top of the annual report). It includes a section with an overview of the firm's main operations, including its products and services. With this information, we manually classify firms into firms that mostly operate in the CPG industry and firms that operate in CPG and other sectors.

NETS data is provided by Walls & Associates and comprises annual observations on specific lines of business at unique locations over the period 1990-2014. The data allows us to observe and track sales, employment, and industry classification of establishments. After matching Nielsen firm names to those in NETS, we use information on industry of each establishment to classify firms that mostly operate in the CPG industry and firms that operate in CPG and other sectors.

B Newness

In order to quantify the novelty of a product, we follow Argente and Yeh (2017) and construct a *newness index* that uses detailed information about the characteristics of each UPC provided in the Nielsen RMS dataset. The index counts the number of new and unique characteristics a product has at the time of its introduction relative to all of the other products ever sold within the same product module. In contrast to Argente and Yeh (2017), who construct the newness index to capture the novelty of a product from a store's perspective, our measure assigns a higher value to products with more unknown features to the entire market. Moreover, we allow each attribute to have module-specific weights using hedonic methods. The main results of the paper use the module-specific index but our results are similar if we use the equal-weights index.

C Equal-weights

Our equal-weights newness index, counts the number of new characteristics (within each attribute) each new product brings to the market. Figure A.5 shows the average of the equal-weights index for a sample of product groups in our data. Each bar in the figure can be interpreted as the average number of new attributes a new product has out of the total attributes included in our data for each category. Figure A.2 shows the most common attributes with new characteristics for a sample of groups. The figure shows, for example, that the most common attribute of new products included in the newness index for the product group carbonated beverages is brands and for detergents is size. Figure A.3 shows some examples of products with high and low equal-weights newness in our data. For example, the product Asthmanefrin Inhalation Solution - Liquid Refill is part of the group Mediations/Remedies/Health Aids. When it was introduced in the market had 6 new attributes out of 8 that we observe in our data for that product group, including the fact that it was a new brand, launched by a new firm, and that it is a liquid, bronchilator refill. As a results, its equal-weights newness index is 6/8=0.75.

Figure A.1: Newness Index by Product Group

The figure presents the average equal-weights newness index for a sample of groups in our data. In particular, it shows the mean newness index by groups along with the top and bottom groups as ranked by this measure. We compute the newness index for each product using equation II.1.2 and weighting equally each attribute. We average across products and product modules to the group level. We focus on cohorts from 2006Q3 to 2014Q4 and on modules with at least 20 barcodes.

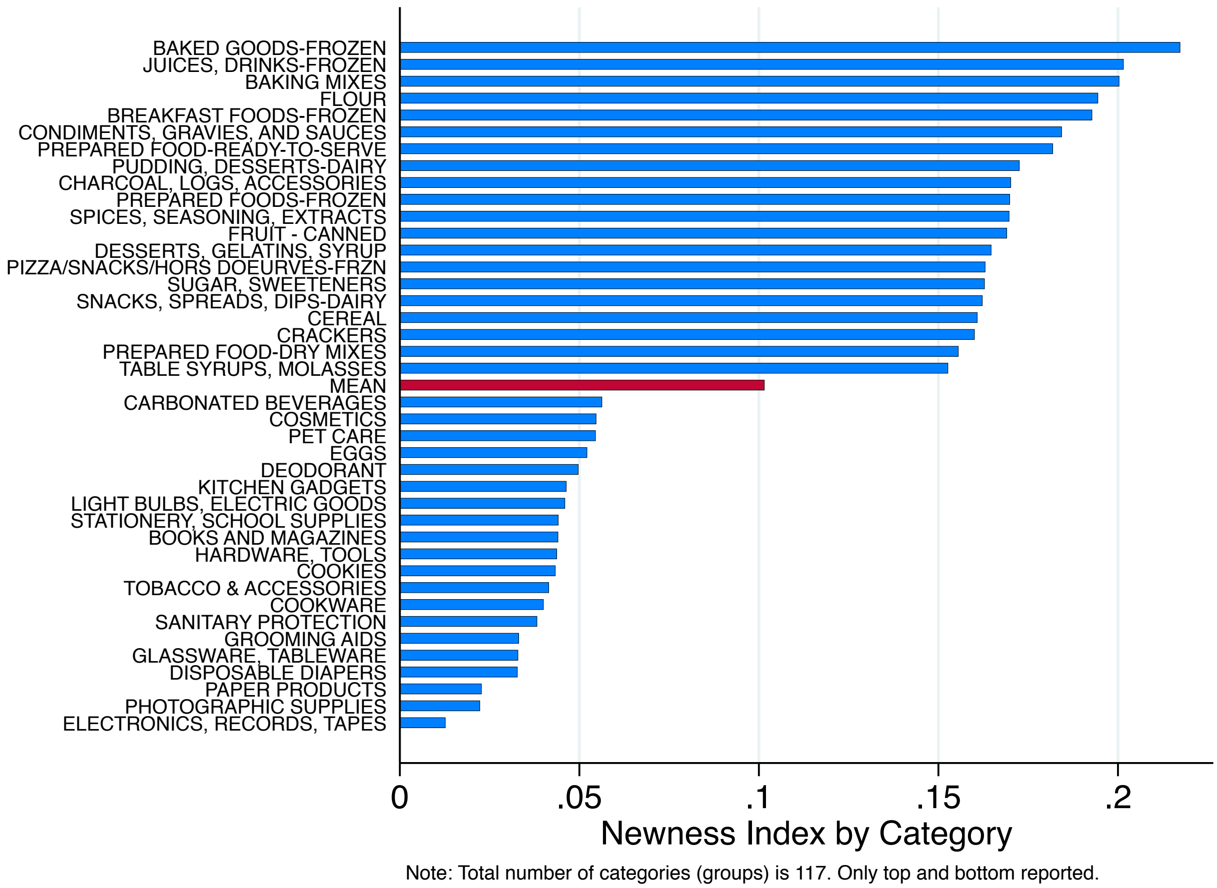


Figure A.2: New Product Characteristics by Product Group

The figure presents the most common new characteristics appearing in new products for a sample of groups: carbonated beverages, cheese, detergents, first aid, hardware tools, and electronics, records and tapes. For each product attribute we construct an indicator if it is the first time the attribute appears in a product. Then we average across products within a product module and then across modules to the group level. We focus on cohorts from 2006Q3 to 2014Q4.

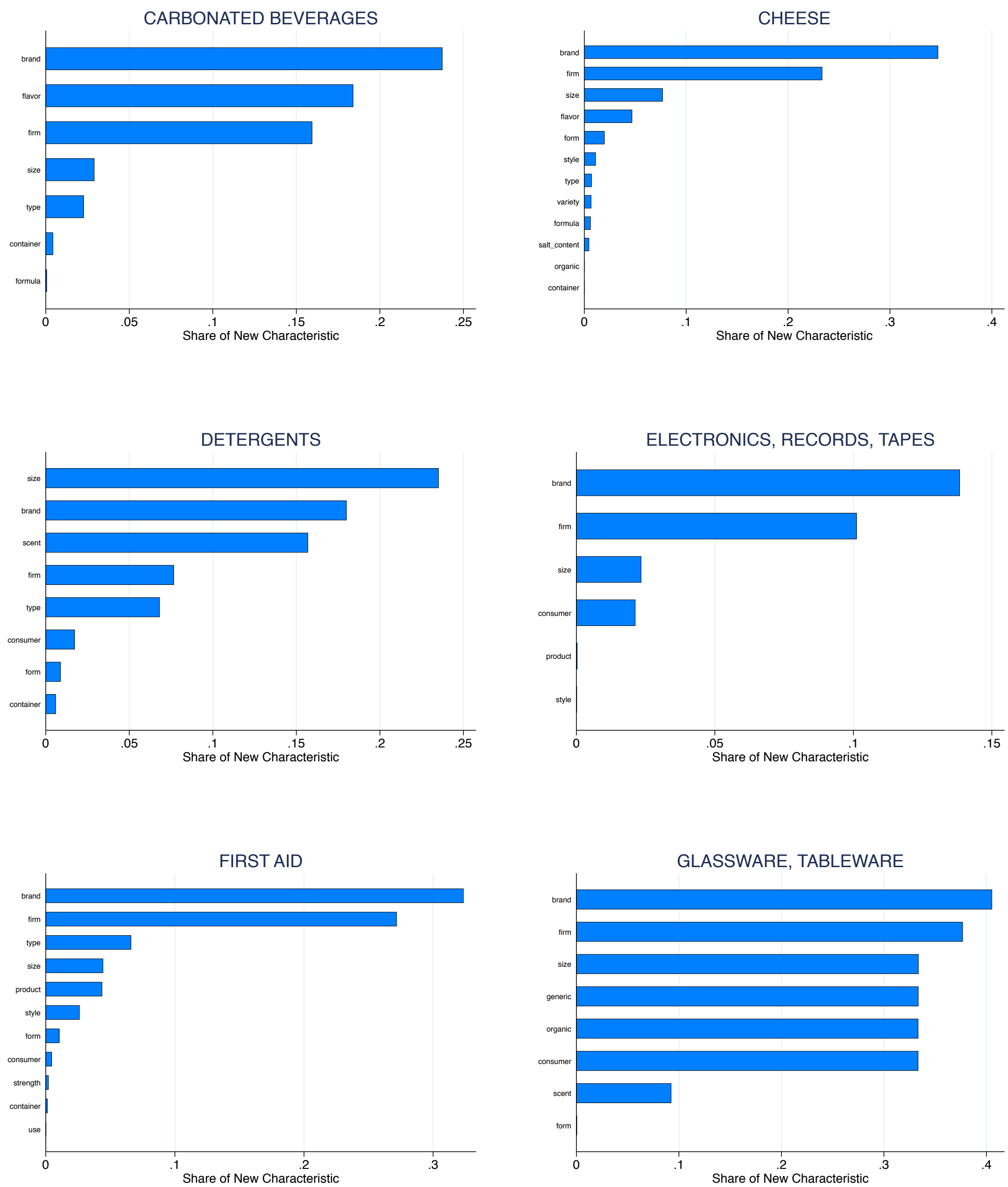
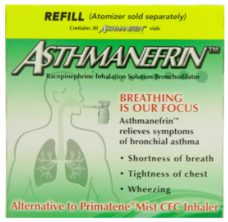


Figure A.3: Newness Index: Examples



Asthmanefrin Inhalation Solution Bronchodilator Asthma Refill Vials
Novelty = 0.75

Kiinde Direct-Pump Adapters for Kiinde Twist Pouch Breast Milk Storage Pouches
Novelty = 0.66

JAWS Hardwood Floor Cleaner Bottle with 2 Refill Pods
Novelty = 0.71

Fire & Flavor Evolution Salt Foot Soak Therapy
Novelty = 0.66

(a) High-Newness products

Swiffer Wetjet Hardwood Floor Mopping and Cleaning Solution Refills

Lavender Vanilla
Novelty = 0.14

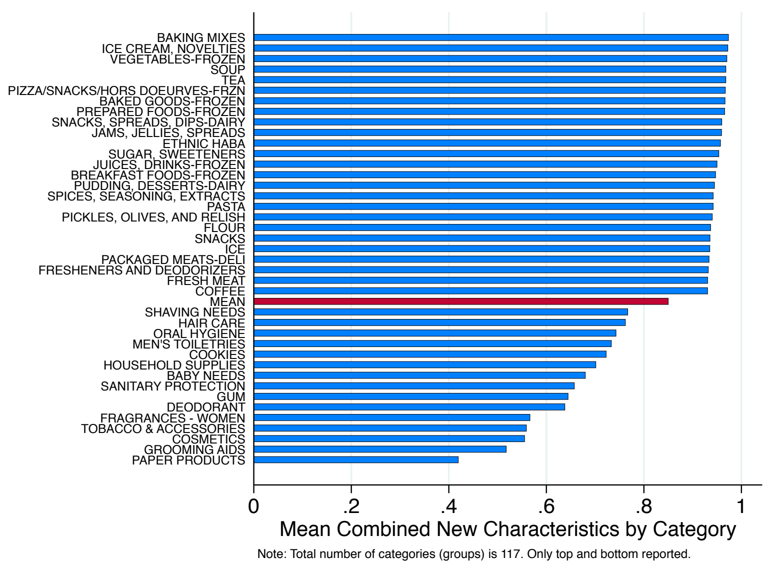
Sweet Citrus
Novelty = 0.14

(b) Low-Newness products

Nielsen collects the most relevant attributes of each product category. Nonetheless, a possible concern of our measure is that it relies heavily on the coverage of attributes of our data. It could be the case that some attributes are omitted from the data and, as a result, our measure of newness could have a downward bias (given that we would assign a value of zero to products whose attributes are not well covered in the data). Given that by definition, changes in any attribute of a good or their potential combinations must result in a new barcode, we compute the share of barcodes in each category for which our data can detect a new combinations of characteristics. Figure A.4 reports the fraction of new products that enter the market with a new combination of product characteristics. Overall, 66% of new products enter the market with a new combination of attributes covered by the Nielsen RMS.

Figure A.4: Combinations by Product Group

Figure A.4 reports the fraction of new products that enter the market with a new combination of product characteristics. We average product modules to the group level. We focus on cohorts from 2006Q3 to 2014Q4 and on modules with at least 20 barcodes.



D Module-specific weights

We estimate the module-specific weights ω_k^m using hedonic methods. We 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 character-

istics and a sequence of time-dummies. Since the regression is run over data which is pooled across time periods, any product characteristic which features in at least some 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:

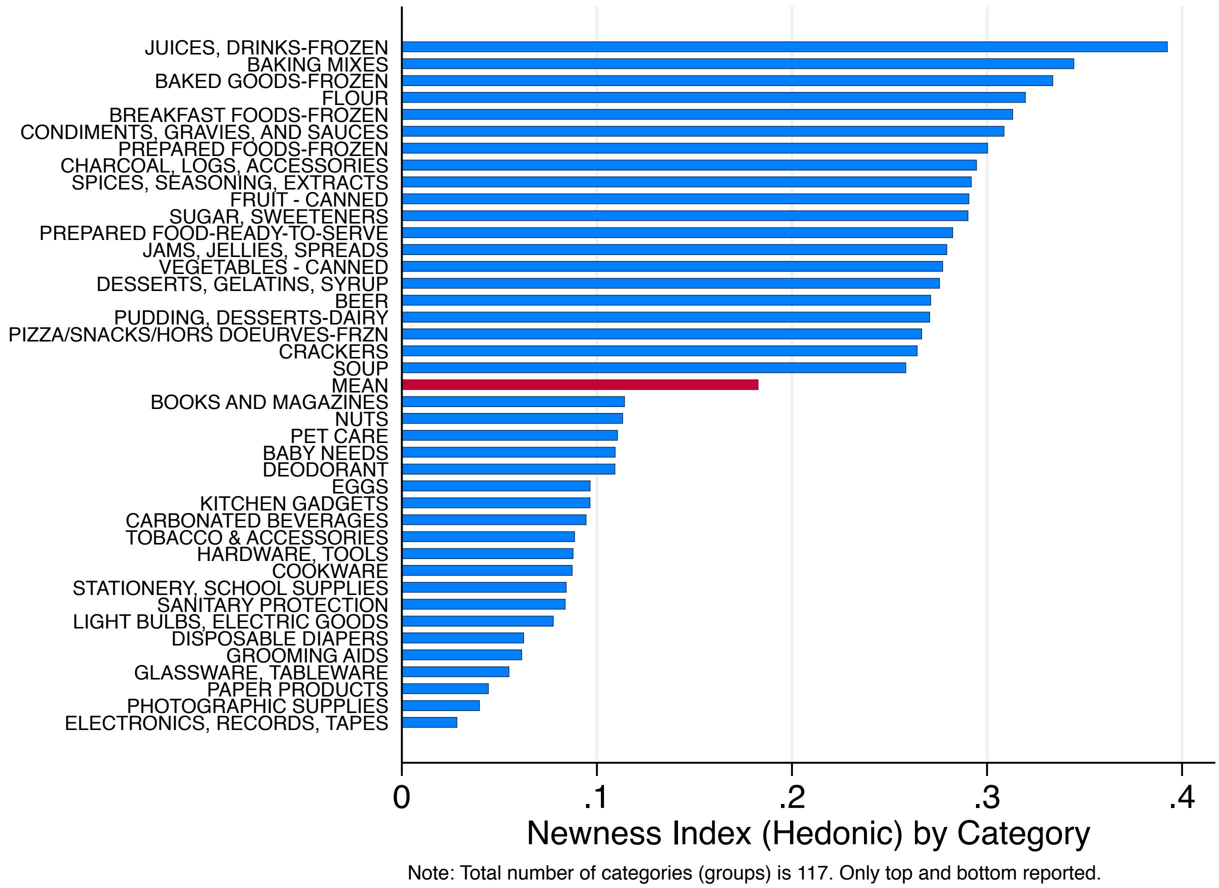
$$p_{it} = \sum_j \pi^j a_i^j + \lambda_t + \epsilon_{it} \tag{11}$$

where i denotes the product, j is the characteristic, and t is the time period (years). a_i^j is an indicator that equals one if a given characteristic j is present in product i. Recall that each attribute k (e.g color) has distinct characteristics j (e.g. blue, red). The shadow price of a given characteristics is denoted by π^j. 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 between the actual price and $\sum_j \pi^j$ of approximately 0.91. ω_k^m is the average contribution of the characteristics within each attribute to the price normalized so that $\sum_k^{Km} \omega_k^m = 1$. Our module-specific weights index has a correlation of 0.93 with the equal-weights index and, conditional on having an equal-weights index larger than zero, the correlation is 0.79.

Figure A.5: Newness Index by Product Group

The figure presents the average newness index for a sample of groups in our data. In particular, it shows the mean newness index by groups along with the top and bottom groups as ranked by this measure. We compute the newness index for each product using equation II.1.2. We average across products and product modules to the group level. We focus on cohorts from 2006Q3 to 2014Q4 and on modules with at least 20 barcodes.



D.1 Patent assignment

To assign patents to their most recent firms, we proceed in the following steps. First, we assign each patent to a *current* assignee(s) (as of 2017 – our patent data vintage). Sometimes this information is missing; in such a case we take the name of an *original* assignee and reassign its patents in case of corporate reorganization. Since most of the time patents get reassigned when a firm is acquired, we track these merger and acquisitions using the

Thomson Reuters Mergers & Acquisition data. 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 of the target, the acquirer, and the deal terms. This is particularly important given that firms that appear both in Nielsen data and USPTO are most likely large firms that undergo many corporate reorganizations.

D.2 Company name cleaning algorithm

We assign each company name to a unique company identifier by using the following procedure that builds on and extends cleaning algorithms from the NBER Patent Data Project and Akcigit et al. (2016).

Step 1. In the first step, we run all company names through name standardization routine and generate unique company identifiers.

1. After capitalizing all letter, we keep the first part of the company name before the first comma.
2. We remove leading and trailing "THE" words; 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 resulting string is null, we protect it.
7. We redo 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 comma.
8. If 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 cleaned names from 9.

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

For example, Siemens has many different variations of its name in the data. ”SIEMNES AG”, ”SIEMANS ATKIENGESELLSCHAFT”, and ”SIEKENS AG” are just 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 will use *pdpass* identifiers to group these three firms together. On the other hand, the NBER file does not identify ”SIEMENS CORP” ”SIEMENS AG” and ”SIEMENS” as the same company and the same as the first three name variations above. In such a case, we use our unique identifiers from Step 1 to group these firms together. As a result, after combining information from NBER files with our cleaning after Step 1, we pool all these six variations into one new company code.

D.3 Product and Process Patents

Similar to Bena and Simintzi (2017), we classify patents into product-related patents and process-related patents based on the claims of patents. Claims of the patents define the scope of patent protection and hence represent the essential part of a patent application. On average, patents in USPTO have around 15 claims: some of them are independent claims, while others derive from them. Claim texts are written in technical terms and often have a rigorous semantic structure.

This gives us an opportunity to create the following simple classification. We say the claim is a process claim if the claim text starts from “method”-phrases (“Method for”, “Method of”, “Method in”, “Method define”, and alike) or “process”-phrases (“Process for”, “Process according”, “Process in”, and alike). Then, as a baseline, we classify a patent into process

patent if the first (which is also considered to be the main) 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 (then its claims often start from words like “Apparatus”, “Device”, and alike). According to this definition, up to 70% of patents are product patents. We have also tried with an alternative definition that defines process patents based on the criteria that share of process claims is larger than 50%. These measures are highly correlated (0.74) and any results based on the baseline variable is robust to this alternative definition.

D.4 Handling non-CPG patents

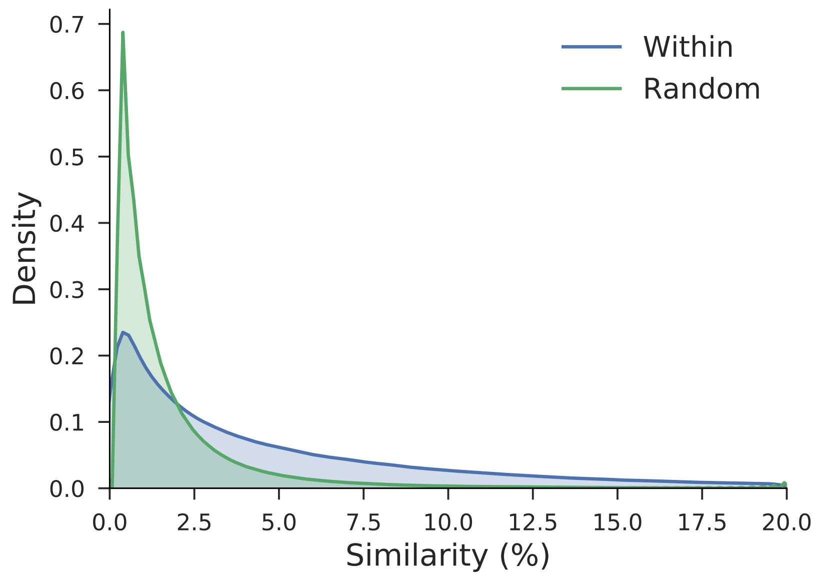
One issue that arises in this setting is that there are numerous patents, particularly those in software and computer hardware, that are not relevant to the consumer products sector that we are analyzing. Most firms operating in these high tech sectors do not appear in the Nielsen data, and hence many patents are filtered out at the firm level. However, some large firms such as Apple and Samsung do make small appearances, and hence we attempt to match their substantial patent portfolios to consumer product categories.

To address this, we introduce additional “pseudo-modules” to draw off these traditional “high-tech” patents, lest they be incorrectly categorized into one of our consumer categories. We have 20 such pseudo-modules, and examples include “computers” and “aviation.” Like the original modules, each pseudo-module also has an associated set of Wikipedia articles that we assign. These categories are designed purely to improve the match to consumer products and are not used in the main analysis that follows.

E Patent Match Validation

Actual vs placebo match. First, we wish to establish that by grouping patents into distinct categories, we are indeed carving out well defined neighborhoods in the technological space. To do this, we again employ word vectors to assess document similarity, but this time between pairs of patents. Specifically, we look at the distribution of the similarity between pairs of patents classified into the same product category and compare this distribution to that for pairs of patents selected at random from the entire dataset. It is reassuring that the similarity distribution based on the match looks very different from our placebo distribution, as seen from Figure A.6.

Figure A.6: Distribution of Patent Similarities



Procter & Gamble Virtual Patent Markings Next we bring in external information on the correct classification of particular patents. To do this, we utilize information from Procter & Gamble (P&G) website on so called virtual patent markings. Firms sometimes mark their products online with active patents to signal to potential competitors that their products are protected by patents. This way, P&G marks some most of their brands with 333 patents (as of March 2019). We started by collecting this information, and manually matched the brands listed online to the brands listed by Nielsen.[40] We then proceeded to identify the product categories that include products of those brands. This allows us to obtain a mapping between products and patents that solely relies on P&G markings, and the mapping between brands and product categories.

We then compare the results of our match with this validation data. When computing our optimal match, we naturally choose the product category with the highest similarity to a given patent. However, for any category, we can compute its similarity rank for a given patent. When this value is one, that means the match is correct. When it is two, the match was presumably nearly correct, and so on, thus providing a notion of intensity of correctness/incorrectness. Panel (a) of Figure A.7 plots the distribution of such ranks for 1

[40]The challenge is that P&G reports its own classification of brands, hence extensive manual matching is needed to determine closest product module match.

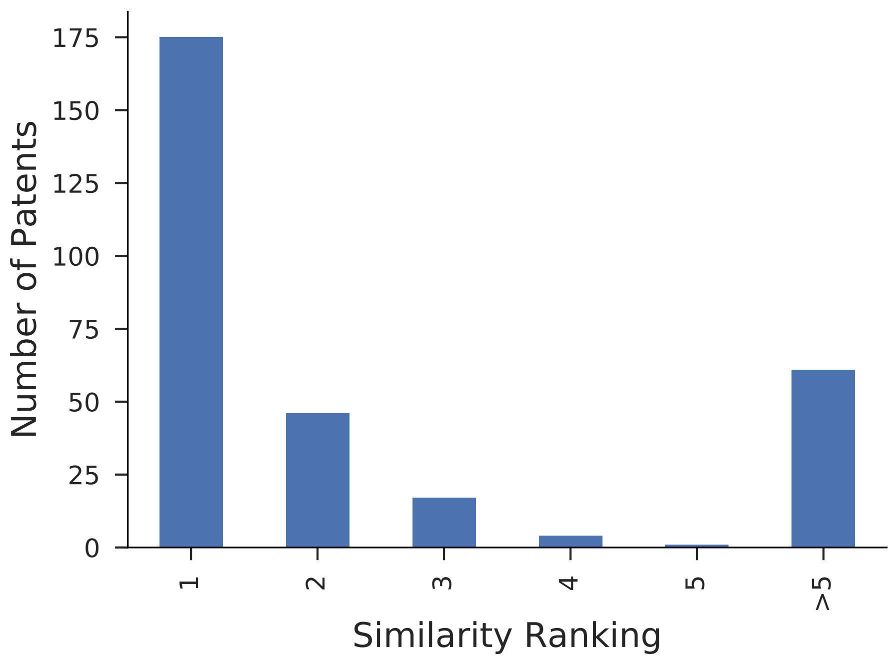
through 5. Here were can see that the match is very often correct or close to correct.

Another way to visualize the accuracy of the match is to look at the distribution of similarities conditioning on whether the match was correct (coinciding with the category from virtual marking) or incorrect. 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 Panel (b) of Figure A.7, these two distributions are quite distinct with a heavier weight of correct-match distribution towards right, meaning the results of the match should be rather robust.

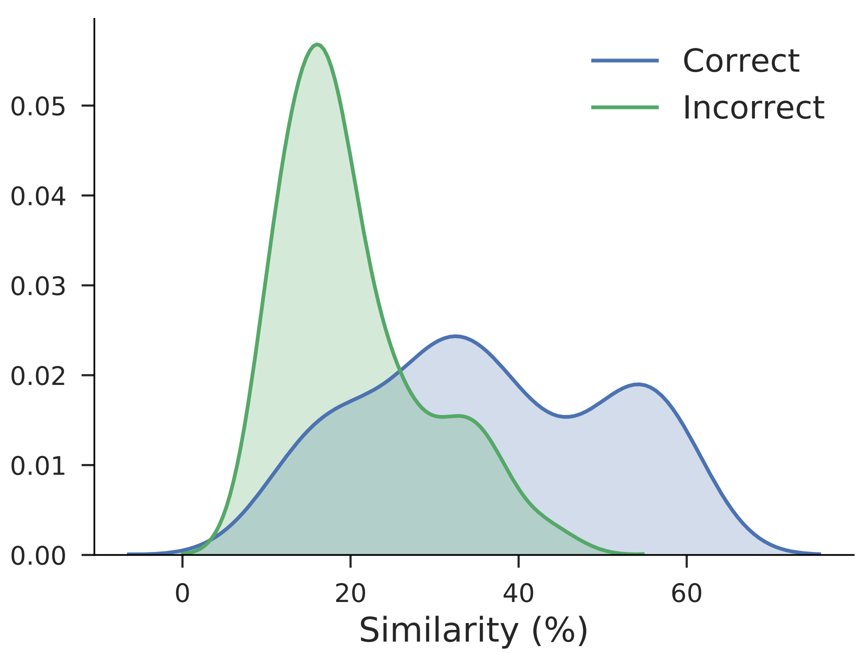
Similarity of top vs non-top ranks. By way of comparison, we can undertake a similar exercise for all patents, not just those owned by P&G. We plot distribution of similarities with different-rank product category matches. Here we again find that top ranked patents have substantially different (and higher) distributions than slightly lower ranked patents, thus providing evidence of robustness of the match. These results are plotted in Figure A.8.

Figure A.7: External validation. P&G case study

(a) Distribution of Patent Ranks



(b) Similarity Distribution by Correctness



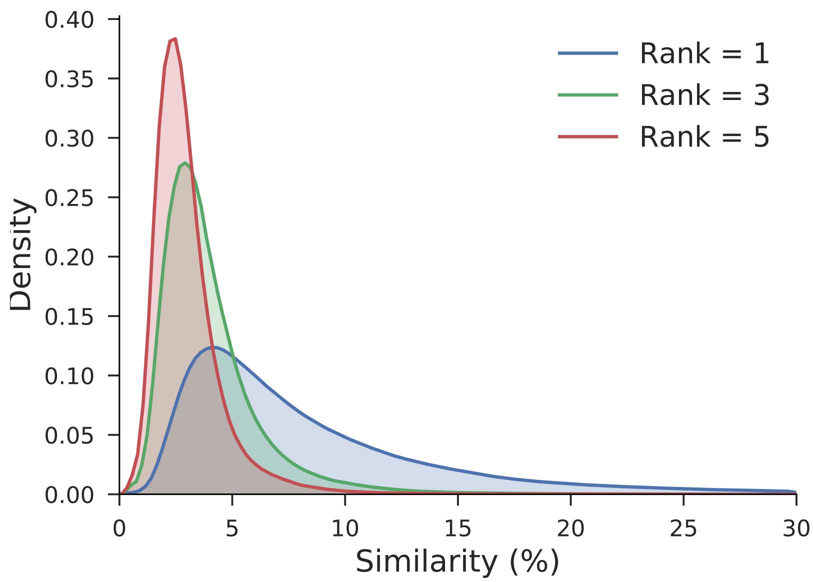
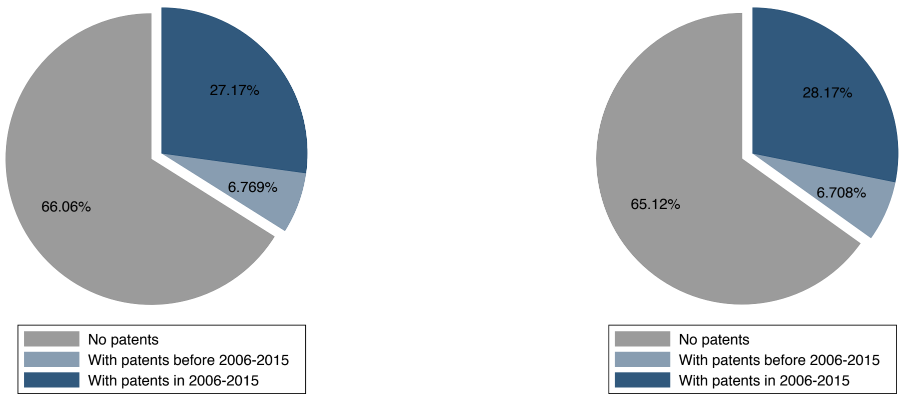


Figure A.8: Similarity Distribution by Rank

F Figures

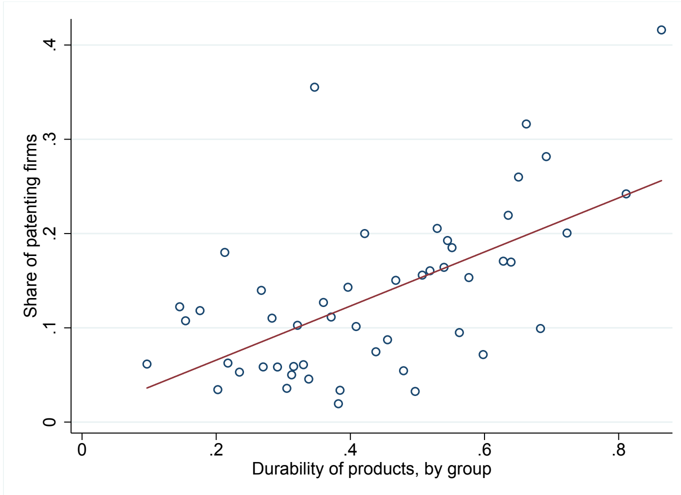
Figure A.9: Share of Product Innovation by Patenting Status: Quality Adjusted New Products



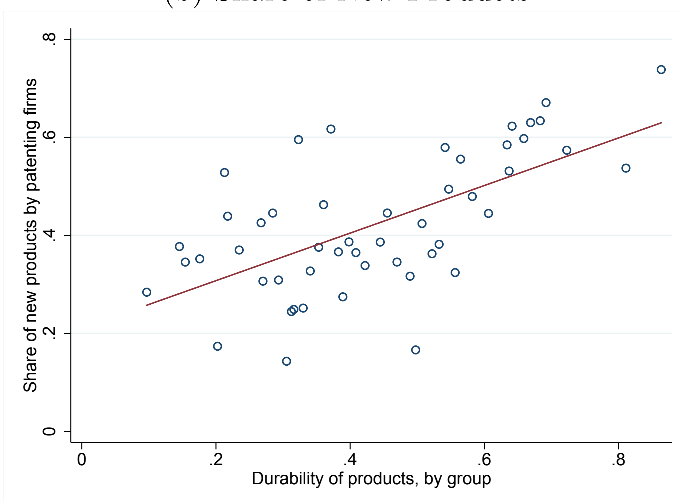
Notes: the figure shows the share of quality(newness)-weighted new products by firms' patenting status. The plots shows quality adjusted unweighted (left) and the price hedonic-weighted (right).

Figure A.10: Product Durability and Firms' Patenting

(a) Share of Patenting Firms

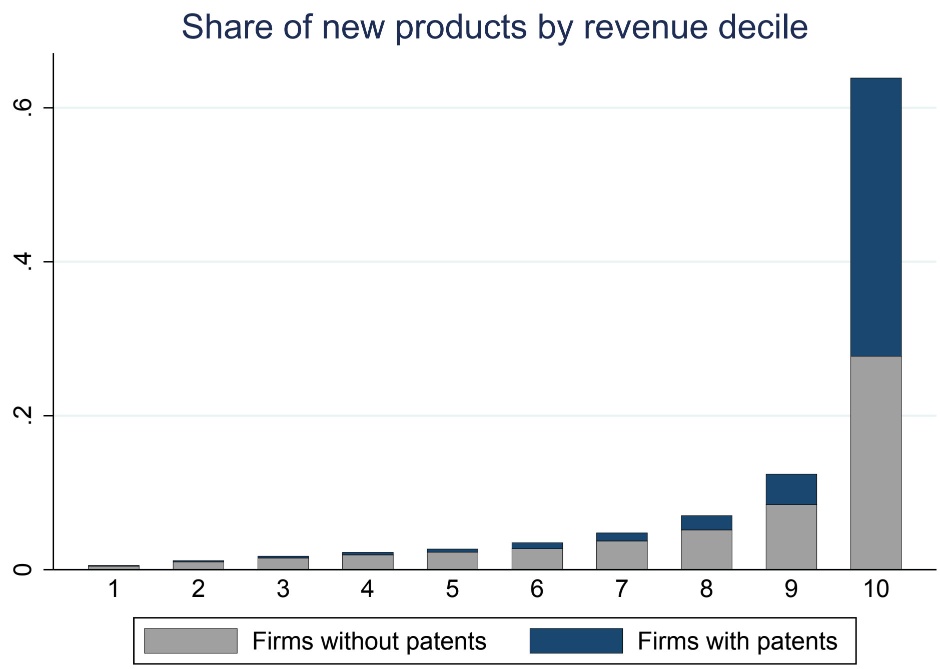


(b) Share of New Products



Note: the figure shows the share of patenting firms and the share of new product launched by patenting firms in each product group of the Nielsen data and their relationship with the durability of the products in each group. In order to approximate the durability of each product group we use the Nielsen Consumer Panel Data and count the average number of shopping trips made by households in a given year to purchase products in each product group. We call categories with few trips per year durable categories. Our measure of durability is the inverse of the average number of trips in a given product group.

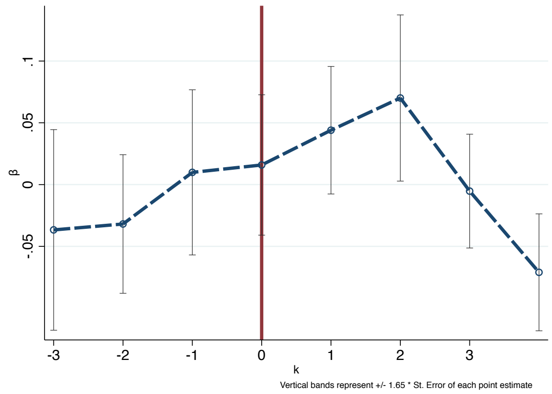
Figure A.11: New Products and Patents by Revenue Decile



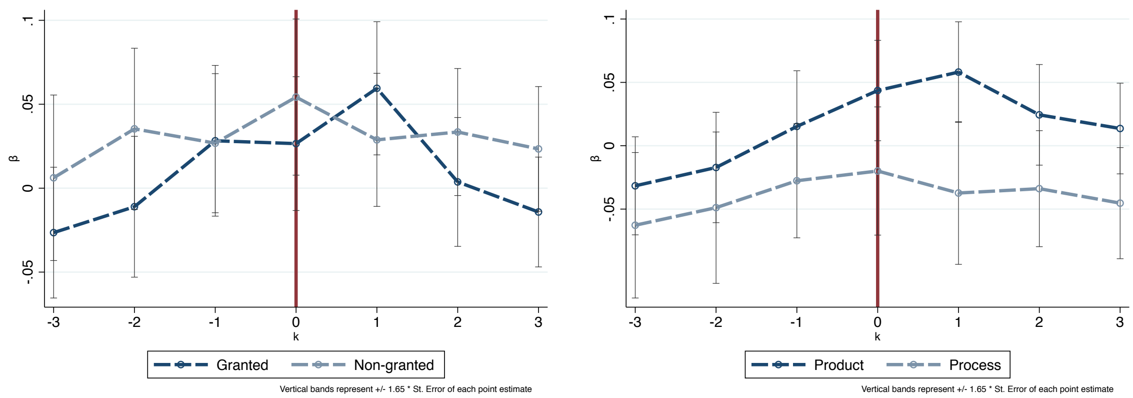
Note: The figure shows the share of new products by revenue decile. The blue part of the bars indicate the share of new products introduced by firms that do not have a patent. The red part of the bars indicate the share of new products that is introduced by firms with patents.

Figure A.12: Product Innovation and Patenting Intensity: alternative measures

Quality-adjusted Product Introduction

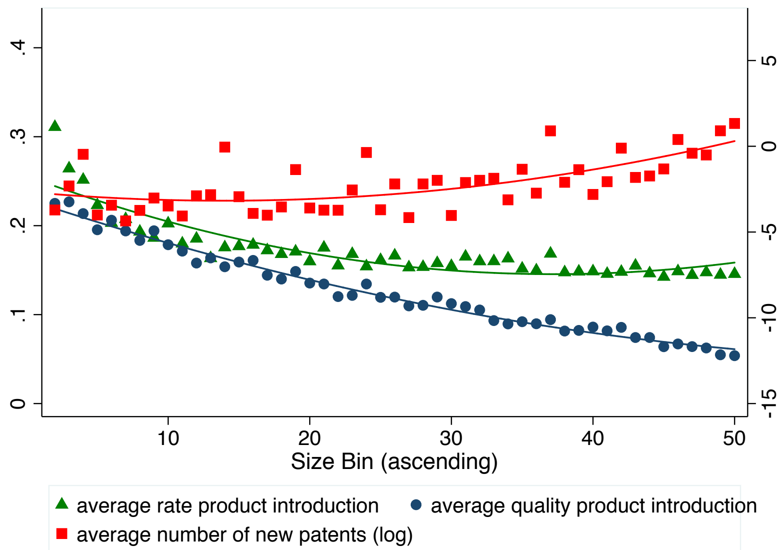


Product Introduction: granted vs. non-granted, product vs. process



Notes: The figure plots the estimated coefficients after estimating equation (1) on log number of new products. The graph should be read as follows: Firms that become patentees in t=0, change product creation by β percent in t=-3,...,4.

Figure A.13: New Products and new Patents by Size



Notes: This figure plots the relationship between product innovation and patenting by firm size. We use data on product innovation and patents at the firm x product category for the period 2007–2015. For each product category, we compute the product entry rate, the average number of new patents, and the average quality of product introduction (newness measure) across 50 bins of size. The figure shows the averages of the three measures after weighting the different product categories by their revenue share.

G Tables

Table A.I: Newness Measure: Correlation with Firm Outcomes

	(1) Growth rate (DH)	(2) Growth rate (New)	(3) Duration 4q	(4) Duration 16q
Newness(t)	0.1476***	0.2773***	0.1172***	0.1118***
	(0.019)	(0.005)	(0.007)	(0.012)
Log N(t)	0.1946***	0.0224***	0.0274***	0.0190***
	(0.004)	(0.001)	(0.002)	(0.003)
Observations	93,290	112,218	97,692	54,148
R-squared	0.383	0.597	0.477	0.569
Time-Category	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y

Notes: The table shows the correlation between our measure of newness 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 the natural logarithm of the total products introduced using the inverse hyperbolic sine transformation.

Table A.II: Product Innovation and Patenting: Intensive Margin (Firm level)

	Baseline			Dynamic Panel		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: New Products (Log N)						
Patents(t-1)	0.0310***			0.0256**		
	(0.009)			(0.010)		
Patents granted(t-1)		0.0303**			0.0296*	
		(0.012)			(0.013)	
Patents non-granted(t-1)			0.0218**			0.0136
			(0.008)			(0.009)
Log N(t-1)				0.0030***	0.0030***	0.0030***
				(0.000)	(0.000)	(0.000)
Observations	178,509	178,509	178,509	158,678	158,678	158,678
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y
Panel B: Quality-adjusted New Products (Log q-N)						
Patents(t-1)	0.0149**			0.0153**		
	(0.005)			(0.006)		
Patents granted(t-1)		0.0160**			0.0180**	
		(0.007)			(0.007)	
Patents non-granted(t-1)			0.0021			-0.0003
			(0.006)			(0.007)
Log N(t-1)				0.0073*	0.0073*	0.0073*
				(0.003)	(0.003)	(0.003)
Observations	178,509	178,509	178,509	158,678	158,678	158,678
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	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 ($Log\ q-N$) in a firm $\times$ year in Panel B as a function of log number of patents. *log N* and $Log\ q-N$ use the inverse hyperbolic sine transformation. Quality of a product is based on our $Newness$ index defined in Section II.1.2. *Patents* is the natural logarithm of the number of any patent applications in firm $\times$ 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). *Patents*, *Patents granted*, and *Patents non-granted* use the inverse hyperbolic sine transformation

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Log N(t)	Log q-N(t)	Log N(t)	Log q-N(t)	Log N(t)	Log N(t)	Log q-N(t)	Log q-N(t)
Citations(t-1)	0.026***	0.014***			0.013*		0.004	
	(0.007)	(0.004)			(0.007)		(0.005)	
Claims(t-1)			0.012***	0.006**		0.005		0.004*
			(0.004)	(0.003)		(0.004)		(0.003)
Log N(t-1)					0.032***	0.032***		
					(0.003)	(0.003)		
Log q-N(t-1)							-0.026***	-0.026***
							(0.003)	(0.003)
Observations	412,004	412,004	411,889	411,889	365,402	365,293	365,402	365,293
R-squared	0.691	0.558	0.691	0.558	0.702	0.702	0.575	0.575
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 log number of new products ($log\ N$) and of log quality-adjusted new products ($Log\ q-N$) in a firm × category × year panel as a function of the log citations and the log claims of a patent. Quality of a product is based on our $Newness$ index defined in Section II.1.2. $log\ N$, $Log\ q-N$, as well as log citations and log claims use the inverse hyperbolic sine transformation.

Table A.IV: Product vs Process: Product Introduction

	(1)	(2)	(3)	(4)	(5)	(6)
New Products						
Patents(t-1)	0.0412***			0.0241**		
	(0.010)			(0.010)		
Patents Prod.(t-1)		0.0429***			0.0254**	
		(0.010)			(0.011)	
Patents Proc.(t-1)			0.0124			0.0025
			(0.018)			(0.018)
New(t-1)				0.0325***	0.0325***	0.0325***
				(0.003)	(0.003)	(0.003)
Observations	412,004	411,889	411,889	365,402	365,293	365,293
R-squared	0.691	0.691	0.691	0.702	0.702	0.702
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y

Notes: The table shows regressions of log number of new products ($log\ N$) as a function of log number of patents. The patents are divided into product patents and process patents according to the information detailed in the claim of the patent. *Patents* is the natural logarithm of the number of any patent applications in firm $\times$ category $\times$ year. Both log patents and the log number of new products use the inverse hyperbolic sine transformation.

Table A.V: Product vs Process: Product Introduction (Quality Adjusted)

	(1)	(2)	(3)	(4)	(5)	(6)
New Products (Q-adj.)						
Patents(t-1)	0.0188***			0.0109		
	(0.006)			(0.007)		
Patents Prod.(t-1)		0.0191***			0.0120*	
		(0.007)			(0.007)	
Patents Proc.(t-1)			-0.0003			-0.0055
			(0.013)			(0.014)
New q-adj(t-1)				-0.0262***	-0.0263***	-0.0262***
				(0.003)	(0.003)	(0.003)
Observations	412,004	411,889	411,889	365,402	365,293	365,293
R-squared	0.558	0.558	0.558	0.575	0.575	0.575
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y

Notes: The table shows regressions of log quality-adjusted new products ($Log\ q-N$) as a function of log number of patents. The patents are divided into product patents and process patents according to the information detailed in the claim of the patent. Both log patents and the log number of new products use the inverse hyperbolic sine transformation.

Table A.VI: Patenting and Revenue - Intensive Margin

	(1)	(2)	(3)	(4)	(5)	(6)
Revenue						
Patents(t-1)	0.0509***			0.0442***		
	(0.017)			(0.015)		
Granted(t-1)		0.0619***			0.0535***	
		(0.019)			(0.017)	
Non-granted(t-1)			0.0201			0.0210
			(0.024)			(0.021)
Rev.(t-1)				0.3619***	0.3619***	0.3619***
				(0.003)	(0.003)	(0.003)
Observations	314,815	314,815	314,815	311,578	311,578	311,578
R-squared	0.906	0.906	0.906	0.926	0.926	0.926
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y

Notes: The table shows regressions of log revenue in a firm × category × year as a function of log number of patents. *Patents* is the natural logarithm of the 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). *Patents*, *Patents granted*, and *Patents non-granted* use the inverse hyperbolic sine transformation.

Table A.VII: Product vs Process: Revenue

	(1)	(2)	(3)	(4)	(5)	(6)
Revenue						
Patents(t-1)	0.0652***			0.0418**		
	(0.025)			(0.019)		
Patents Prod.(t-1)		0.0544**			0.0385*	
		(0.026)			(0.020)	
Patents Proc.(t-1)			0.1261***			0.1024***
			(0.043)			(0.029)
Rev.(t-1)				0.6222***	0.6220***	0.6220***
				(0.003)	(0.003)	(0.003)
Observations	415,008	414,893	414,893	410,801	410,692	410,692
R-squared	0.849	0.849	0.849	0.901	0.901	0.901
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y

Notes: The table shows regressions of log revenue as a function of log number of patents. The patents are divided into product patents and process patents according to the information detailed in the claim of the patent. *Patents* is the natural logarithm of the number of any patent applications in firm × category × year. Both log patents and the log number of new products use the inverse hyperbolic sine transformation.

Table A.VIII: Product vs Process: New Products (Granted)

	(1)	(2)	(3)	(4)	(5)	(6)
New Products						
Granted(t-1)	0.0467***			0.0271**		
	(0.011)			(0.011)		
Granted Prod.(t-1)		0.0487***			0.0292**	
		(0.011)			(0.012)	
Granted Proc.(t-1)			0.0158			0.0041
			(0.022)			(0.023)
New(t-1)				0.0325***	0.0325***	0.0325***
				(0.003)	(0.003)	(0.003)
Observations	412,004	411,889	411,889	365,402	365,293	365,293
R-squared	0.691	0.691	0.691	0.702	0.702	0.702
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y

Notes: The table shows regressions of log quality-adjusted new products ($Log\ q - N$) as a function of log number of granted patents. The granted patents are divided into product patents and process patents according to the information detailed in the claim of the patent. *Granted* is the natural logarithm of the number granted patents in firm $\times$ category $\times$ year. Both log granted patents and the log number of new products use the inverse hyperbolic sine transformation.

Table A.IX: Product vs Process: New Products - Quality Adjusted (Granted)

	(1)	(2)	(3)	(4)	(5)	(6)
New Products (Q-adj.)						
Granted(t-1)	0.0214***			0.0103		
	(0.007)			(0.007)		
Granted Prod.(t-1)		0.0208***			0.0109	
		(0.007)			(0.008)	
Granted Proc.(t-1)			0.0065			-0.0043
			(0.015)			(0.016)
New q-adj(t-1)				-0.0262***	-0.0263***	-0.0262***
				(0.003)	(0.003)	(0.003)
Observations	412,004	411,889	411,889	365,402	365,293	365,293
R-squared	0.558	0.558	0.558	0.575	0.575	0.575
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y

Notes: The table shows regressions of log quality-adjusted new products ($Log\ q - N$) as a function of log number of granted patents. The granted patents are divided into product patents and process patents according to the information detailed in the claim of the patent. *Granted* is the natural logarithm of the number of granted patents in firm × category × year. Both log granted patents and the log number of new products use the inverse hyperbolic sine transformation.

Table A.X: Product vs Process: Revenue (Granted)

	(1)	(2)	(3)	(4)	(5)	(6)
Revenue						
Granted(t-1)	0.0826***			0.0566***		
	(0.027)			(0.021)		
Granted Prod.(t-1)		0.0830***			0.0585***	
		(0.028)			(0.021)	
Granted Proc.(t-1)			0.0787			0.0673*
			(0.054)			(0.037)
Rev.(t-1)				0.6222***	0.6220***	0.6220***
				(0.003)	(0.003)	(0.003)
Observations	415,008	414,893	414,893	410,801	410,692	410,692
R-squared	0.849	0.849	0.849	0.901	0.901	0.901
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y

Notes: The table shows regressions of log revenue as a function of log number of granted patents. The granted patents are divided into product patents and process patents according to the information detailed in the claim of the patent. Log granted patents use the inverse hyperbolic sine transformation.

Table A.XI: Product vs Process: Price Changes

	(1)	(2)	(3)	(4)	(5)	(6)
Δ Price						
Patents(t-1)	0.0021					
	(0.005)					
Patents Prod.(t-1)		0.0017				
		(0.005)				
Patents Proc.(t-1)			0.0145**			
			(0.007)			
Granted(t-1)				0.0053		
				(0.005)		
Granted Prod.(t-1)					0.0050	
					(0.005)	
Granted Proc.(t-1)						0.0153*
						(0.009)
Observations	410,577	410,468	410,468	410,577	410,468	410,468
R-squared	0.170	0.170	0.170	0.170	0.170	0.170
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y

Notes: The table shows regressions of change in prices (in logs) as a function of log number of patents both total applications and granted patents. The patents are divided into product patents and process patents according to the information detailed in the claim of the patent. *Patents* is the natural logarithm of the number of any patent applications in firm × category × year and *Granted* is the natural logarithm of granted patents in firm × category × year. Both log patents and log granted patents use the inverse hyperbolic sine transformation.

Table A.XII: Product vs Process: Price Changes by Size

	(1)	(2)	(3)	(4)	(5)	(6)
Δ Price						
Patents(t-1)	-0.0119					
	(0.019)					
Size(t)	0.0088***	0.0088***	0.0088***	0.0088***	0.0088***	0.0088***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Patents(t-1) x Size(t)	0.0010					
	(0.001)					
Patents Prod.(t-1)		0.0002				
		(0.005)				
Patents Proc.(t-1) x Size(t)		0.0006				
		(0.000)				
Patents Proc.(t-1)			0.0127*			
			(0.007)			
Patents Prod.(t-1) x Size(t)			0.0002			
			(0.000)			
Granted(t-1)				-0.0019		
				(0.021)		
Granted(t-1) x Size(t)				0.0005		
				(0.001)		
Granted Prod.(t-1)					-0.0023	
					(0.022)	
Granted Prod.(t-1) x Size(t)					0.0005	
					(0.001)	
Granted Proc.(t-1)						0.0167
						(0.040)
Granted Proc.(t-1) x Size(t)						-0.0001
						(0.002)
Observations	410,577	410,468	410,468	410,577	410,468	410,468
R-squared	0.172	0.172	0.172	0.172	0.172	0.172
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y

Notes: The table shows regressions of change in prices (in logs) as a function of log number of patents both total applications and granted patents. The patents are divided into product patents and process patents according to the information detailed in the claim of the patent. The variable *Size* is the log revenue of the firm × category × year. *Patents* is the natural logarithm of the number of any patent applications in firm × category × year and *Granted* is the natural logarithm of granted patents in firm × category × year. Both log patents and log granted patents use the inverse hyperbolic sine transformation.

Table A.XIII: Deterrence, Patenting, and Market Leadership (Granted)

	(1)	(2)	(3)	(4)	(5)
New Products by Other Firms					
Log N(t-1)	-0.00003	-0.00003	0.00045***	0.00046***	-0.01113***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Patent(t-1)		-0.00132		0.00262**	0.00194
		(0.001)		(0.001)	(0.002)
Size(t)			-0.00479***	-0.00465***	-0.00811***
			(0.000)	(0.000)	(0.000)
Patent(t-1) x Size(t)				-0.00343***	-0.00922***
				(0.001)	(0.001)
Observations	264,060	264,060	264,060	264,060	204,166
R-squared	0.99	0.99	0.99	0.99	0.99
Time-Category	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	N
Time-Firm	N	N	N	N	Y

Notes: The table shows the relationship between product introduction of other firms $\hat{f}$ within the product categories m and the size of firm f, the product introduction of firm f , and whether these products are related to a patent. The dependent variable is the natural logarithm of the number of products introduced by firms other than firm f. New(t-1) is the natural logarithm of products introduced by firm f, using the inverse hyperbolic sine transformation. Patent(t-1) is the natural logarithm of granted patents of product category m by firm f applied in year t, using the inverse hyperbolic sine transformation. Size(t) is the natural logarithm of the total sales of firm f in product category m at time t (standardized).

Table A.XIV: Deterrence, Patenting, and Market Leadership

	(1)	(2)	(3)	(4)
New Products by Other Firms				
Log N(t-1)	-0.00002	-0.00003	0.0004***	0.0004***
	(0.00013)	(0.00013)	(0.0001)	(0.0001)
Cum. Patents(t-1)		0.00494		0.0068***
		(0.00182)		(0.0017)
Size(t)			-0.0047***	-0.0042***
			(0.0002)	(0.0002)
Cum. Patents(t-1) × Size(t)				-0.0023***
				(0.0004)
Observations	264,060	264,060	264,060	264,060
R-squared	0.99	0.99	0.99	0.99
Time-Category	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y

Notes: The table shows the relationship between product introduction of other firms $\hat{f}$ within the product categories m and the size of firm f, the product introduction of firm f , and whether these products are related to a patent. The dependent variable is the natural logarithm of the number of products introduced by firms other than firm f. $Log\ N(t-1)$ is the natural logarithm of products introduced by firm f, using the inverse hyperbolic sine transformation. $Cum\ Patents(t-1)$ is the natural logarithm of cumulative patents of product category m by firm f applied in year t, using the inverse hyperbolic sine transformation. $Size(t)$ is the natural logarithm of the total sales of firm f in product category m at time t (standardized).

H Model

H.1 Derivation of rates of creative destruction

Depending on the actions of the incumbent, we have the following rates of creative destruction:

- If incumbent does not patent and does not do product upgrade

$$\tilde{p}^{00} = p \times \Pr\Big(q + \lambda^e > q\Big) = p,$$

- If incumbent does not patent and does product upgrade

$$\tilde{p}^{10} = p \times \Pr\Big(q + \lambda + \lambda^e > q + \lambda\Big) = p,$$

- If incumbent patents but does not do product upgrade

$$\tilde{p}^{01} = p \times \Pr\Big(q + \lambda + \lambda^e > q + \lambda + \varepsilon\Big) = p(1 - \varepsilon),$$

- If incumbent patents and does not do product upgrade

$$\tilde{p}^{11} = p \times \Pr\Big(q + \lambda + \lambda^e > q + \lambda + \varepsilon)\Big) = p(1 - \varepsilon).$$

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

H.2 Conditions for equilibrium

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

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

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

We define

Option 1

$$O1(q) = \frac{\pi(q+\lambda)^\gamma}{r+p} - \frac{\pi q^\gamma}{r+p} - c_m$$

Option 2

$$O2(q) = \frac{\pi q^\gamma}{r+p(1-\varepsilon)} - \frac{\pi q^\gamma}{r+p} - c_l$$

Option 3

$$O3(q) = \frac{\pi(q+\lambda)^\gamma}{r+p(1-\varepsilon)} - \frac{\pi q^\gamma}{r+p} - c_l - c_m$$

Option 4

$$O4 = 0$$

Step 1: We show that for small firms product introduction with no patenting dominates other options:

$\exists q^*$ s.t. $\forall q < q^*, V^{10}(q) - c_m = \max\left\{\; V^{11}(q) - c_m - c_l,\; V^{10}(q) - c_m,\; V^{01}(q) - c_l\;,\; V^{00}(q)\;\right\}$

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

$O1$ is decreasing in q since its derivative is always negative:

$$\frac{dO1}{dq} = \frac{\pi\gamma}{p+r}\left(\frac{1}{(q+\lambda)^{1-\gamma}} - \frac{1}{q^{1-\gamma}}\right) < 0, \forall q \tag{12}$$

$O2$ is increasing in q:

$$\frac{dO2}{dq} = \pi\gamma q^{\gamma-1}\left(\frac{1}{p(1-\varepsilon)+r} - \frac{1}{p+r}\right) > 0, \forall q \tag{13}$$

$O3$ is decreasing in q for small values of q and increasing in q for large values with a minimum at $q = \mathring{q}$:

$$\begin{aligned} \frac{dO3}{dq} &= \pi\gamma\left(\frac{1}{(q+\lambda)^{1-\gamma}}\frac{1}{p(1-\varepsilon)+r} - \frac{1}{q^{1-\gamma}}\frac{1}{p+r}\right) > 0 \text{ if } q > \mathring{q}, \\ \mathring{q} &\equiv \frac{\lambda}{\left(\frac{p+r}{p(1-\varepsilon)+r}\right)^{\frac{1}{1-\gamma}} - 1} \end{aligned} \tag{14}$$

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

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

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

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

Under Conditions (i) and (ii) and given the monotonicity properties in (12), (13), (14), there exists a threshold level of q^*, such that firms below q^* do product innovation only, while above q^* other options dominate.

Step 2: we show that for large enough firms, patenting with no product introduction dominates other options:

$\exists q^{**}$ s.t. $\forall q > q^{**}, V^{01}(q) - c_m = \max\Big\{\ V^{11}(q) - c_m - c_l,\ V^{10}(q) - c_m,\ V^{01}(q) - c_l\ ,\ V^{00}(q)\ \Big\}$

Let us first compare $O2$ and $O3$.

$$\begin{aligned} O2(q) &= \frac{\pi q^\gamma}{p(1-\varepsilon)+r} - \frac{\pi q^\gamma}{p+r} - c_l > O3(q) = \frac{\pi(q+\lambda)^\gamma}{p(1-\varepsilon)+r} - \frac{\pi q^\gamma}{p+r} - c_l - c_m \text{ iff} \\ c_m &> \frac{\pi(q+\lambda)^\gamma - \pi q^\gamma}{p(1-\varepsilon)+r} \end{aligned}$$

Notice that the right-hand side is a decreasing function[41] with an asymptote at zero. Or in other words, marginal returns from additional product innovation are so low that they do not cover commercialization cost. Hence, for large enough q the inequality is satisfied.[42] This together with (12) implies that there exists q^{**} such that firms above q^{**} prefer doing patenting with no product introduction.

The next step just shows conditions under which $q^* \neq q^{**}$, and when firms engage both in product innovation and patenting in the $q^* < q < q^{**}$ range – an empirically relevant case.

Step 3: We show under which conditions intermediate range with both product introduction and patenting exists.

First, we can make a strong sufficient condition: that $O3 > 0$ for all q. Given this, we show that intersection of $O1$ and $O3$ happens earlier than of $O1$ and $O2$. Say, $O1(q_1) = O3(q_1)$ and $O1(q_2) = O2(q_2)$. Then we need to show that $q_1 < q_2$. In such a case, $q^* = q_1$ and $q^{**} = q_2$.

Because of (12), showing that $q_1 < q_2$ is equivalent to showing $O1(q_1) > O1(q_2)$. We have

$$q_1 = \left(\frac{c_l(r+p)(r+p(1-\varepsilon))}{p\varepsilon\pi}\right)^{\frac{1}{\gamma}} - \lambda \tag{15}$$

and

$$\frac{\pi(q_2+\lambda)^\gamma}{r+p} - c_m = \frac{\pi q_2^\gamma}{r+p(1-\varepsilon)} - c_l \tag{16}$$

Condition (iii) is that parameters satisfy

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

with q_1 and q_2 defined as in (15) and (16).

[41] $\frac{d\left(\frac{(q+\lambda)^\gamma}{\tilde{p}_{11}+r} - \frac{q^\gamma}{\tilde{p}_{10}+r}\right)}{dq} = \frac{d((q+\lambda)^\gamma(\tilde{p}_{10}+r) - q^\gamma(\tilde{p}_{11}+r))}{dq} = \gamma\left[\frac{\tilde{p}_{10}+r}{(q+\lambda)^{1-\gamma}} - \frac{\tilde{p}_{11}+r}{q^{1-\gamma}}\right] < 0$

[42] Also notice that Condition (i) implies that $c_m < \frac{\pi\lambda^\gamma}{\tilde{p}_{11}+r}$, so at $q = 0$, inequality is not satisfied, so $O2$ is not always preferred over $O3$.