Patents to Products: Product Innovation and Firm Dynamics*

David Argente, Penn State
Salomé Baslandze, FRB Atlanta & CEPR
Douglas Hanley, U. of Pittsburgh
Sara Moreira, Northwestern

March 15, 2023

Abstract

We study the prevalence and implications of firms' strategic practices to patent but not commercialize new ideas. We develop a model illustrating that market leaders have higher incentives to use these practices, and we create a novel data set linking patents to products to evaluate the model's implications. We find that the degree of association between patent and product introduction is positive, on average, but smaller for market leaders. Patents of market leaders generate larger revenue premiums and greater deterrence of future product introduction by competitors, which is also consistent with leaders' strategic use of patents. Using the model, we quantify that 62% of patents do not lead to product introductions. These patents imply a 2.5% reduction in the rate of creative destruction. Our counterfactuals under different patent regimes indicate that the possibility of obtaining patent protection without product commercialization also reduces market leaders' innovation efforts, potentially reversing the benefits of the patent system as a whole.

*Emails: dargente@psu.edu; salome.baslandze@atl.frb.org; doughanley@pitt.edu; sara.moreira@kellogg.northwestern.edu.
We are grateful to Antonin Bergeaud, Laurent Frésard, Pete Klenow, and Jesse Perla for very helpful discussions. We also thank Ufuk Akcigit, Fernando Alvarez, Toni Braun, Hugo Hopenhayn, Benjamin F. Jones, Claudio Michelacci, Juan Rubio-Ramirez and numerous seminar and conference participants. 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 Data sets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researcher(s) and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. The views expressed in this paper do not necessarily represent the views of the Federal Reserve System or the Federal Reserve Bank of Atlanta. First draft: March 2019.

1 Introduction

There is a growing literature focused on the increased market concentration seen across many U.S. industries and the role that large firms might play in this trend. One strand considers the possibility that large firms have achieved greater technological dominance through innovation, and that this allows them to expand their market share (e.g., Autor, Dorn, Katz, Patterson and Van Reenen, 2020; Crouzet and Eberly, 2019). Another strand of thought posits that the increasing dominance of large firms is associated with market imperfections that are deleterious to economic growth (e.g., Gutiérrez and Philippon, 2017; De Loecker, Eeckhout and Unger, 2020). These market imperfections may result from firms using various methods to deter potential competitors, such as lobbying for political favoritism (Gutiérrez and Philippon, 2019; Akcigit, Baslandze and Lotti, 2023) or acquiring innovative startups only to terminate their operations (Cunningham, Ederer and Ma, 2021). This paper proposes and tests another type of strategy used by large firms: using patents to deter competition while not actually innovating or advancing the state of technology in the market.

The patent system's main purpose is to promote innovation by incentivizing inventors to invest in costly research projects. Today, however, the patent system is viewed with increasing skepticism, and both policy-makers and academic scholars are pointing out its deficiencies.[1] Some of those deficiencies are associated with the perverse incentives for large firms to maintain their market power by patenting ideas before their potential competitors, which leads to patents not for innovation, but solely to deter other firms from coming up with future innovations (Gilbert and Newbery, 1982). While these patenting deficiencies are often discussed, it is unclear how prevalent they are in practice and by how much they hinder innovation. In this paper, we develop a theoretical framework and build a unique patents-to-products data set to identify the prevalence and implications of patenting without product innovation.

We first build a simple theoretical model that illustrates how the relationship between product introduction and patenting decisions differs for firms with different market positions. The model provides a set of testable predictions and identifies relevant data moments that guide our empirical strategy. We formalize the occurrence of strategic patents—defined as patents that deter innovation without leading to product introductions—and show which data moments can identify strategic patenting in the data.

[1] See a survey of the literature by Bryan and Williams (2021), as well as Jaffe and Lerner (2004), Boldrin and Levine (2013), and Cohen, Nelson and Walsh (2000). Some of the popular press articles are: "The Experts: Does the Patent System Encourage Innovation?" (WSJ, 2013) and "Save Americas Patent System" (NYT, 2022).

The framework builds on the quality-ladder model of innovation through creative destruction introduced in Aghion and Howitt (1992) and adds an additional distinction between a firm's decision to innovate and their decision to patent. In the model, product innovation takes the form of quality upgrades to products in the market.[2] These innovations come either from the incumbent's efforts to prolong its technological lead or from market entrants' attempts to become new leaders. Once an incumbent firm obtains an idea for a new innovation, it makes a one-time decision about whether to commercialize the idea as a product and/or patent it. Both patenting and new product commercialization are costly activities. If the incumbent decides to commercialize the idea, it gains additional profits from introducing higher-quality product into the market. Patenting the idea grants the incumbent additional protection against creative destruction by potential entrants. The model shows that patenting and product introduction are complementary activities: the higher the return from product introduction, the higher the benefit of protecting new products with patents; the stronger the protection from patents, the more profitable it is to introduce a new product. These results formalize why we should observe, in practice, a positive relationship between patent filings and product innovation.

More importantly, the model shows that larger incumbents optimally rely more on strategic patenting because the incentives for patenting increase in the incumbents' size while the incentives for product introduction decrease. Large incumbents forego many new ideas that smaller firms would find worth commercializing to avoid cannibalizing their existing rents. The incentives for patenting go in the opposite direction: larger incumbents want to protect their existing rents relatively more and file patents even if they do not intend to commercialize the product. Putting these results together, the model predicts that the strength of the relationship between product commercialization and patents declines with the incumbent's size, reflecting the increase in the probability of strategic patenting to protect existing products.[3]

Our model leads to a set of empirical moments on product introduction and patenting that help identify the presence of strategic patents and their implications. To study these empirical moments, we need detailed data on firms' patents and their linkage to product introductions, in addition to measures of firm revenue and innovations by firms' competitors within a product category. Assembling such a data set presents three empirical challenges. First, while patent data are broadly available, measures of product innovation in the market are rarely available at large scale. To address this challenge, we use comprehensive data

[2]Throughout the paper, we will interchangeably use "product innovation", "product introduction", and "product commercialization".

[3]We show that if the economy did not allow for patenting without product introduction, the strength of the relationship between patenting and product introduction would, in fact, increase with firm size.

on firms and products sold in the consumer goods sector from 2006 to 2015 collected by Nielsen-Kilts from point-of-sale systems in retail locations. This data set includes detailed information about the characteristics of each consumer-goods product; notably, it includes information on product attributes (e.g. formula, style, content), well-measured prices, and sales. We exploit this rich data set to construct measures of product innovation. Our simplest measure is the number of new products (barcodes) introduced at the firm and product category level in a given year. Since many new products represent only minor innovations relative to existing products, we also construct measures of the quality-adjusted number of new products. We measure quality improvements by tracking the new attributes that a product brings to the market and by exploiting variation in product prices and sales.

The second challenge lies in linking product innovations to their respective patents. We start by using firm names in the patent and product data sets to map each firm's patents to its full product portfolio. After this parsimonious matching procedure, we can simultaneously measure patenting and product introduction at the firm level over time. However, we need a more granular procedure to match patents with products within firms. Establishing this closer link between products and patents is particularly important for firms operating in multiple product categories or filing patents that are not related to products in our data. In addition, this link allows us to exploit variation within-firm and across time in our analysis. We leverage the richness of the information about product and patent characteristics in our data and use modern methods from the fields of natural language processing and information retrieval to link each firm's patents with the products it sells (Manning, Raghavan and Schütze, 2008). We define product categories—sets of similar products—by applying clustering analysis to the short product descriptions included in the Nielsen data enriched with text from Wikipedia articles about the products. We then analyze the texts of patent applications and assign each patent to the product category with which it has the highest text-similarity score.[4] This categorization of firms' products and patents leads to our benchmark patent-to-products data set, which includes variation at the firm-category-year level.

Finally, it is important to accurately measure product commercialization and patenting decisions by all firms competing in the same relevant product market. By using the scanner data covering hundreds of consumer goods products (e.g. lamps, batteries, yogurts, diapers) with near-universal coverage and our unique firm-category-year match, we can identify relevant competitors in narrowly defined product categories, determine their market shares, and measure their activities.

[4] Younge and Kuhn (2016), Kelly, Papanikolaou, Seru and Taddy (2021), and Webb (2019) use similar techniques when analyzing patent documents.

The resulting granular data set tracks patents and products for firms in the consumer goods sector. Although our empirical results come only from this sector, the patenting intensities and product introduction rates of such firms are, on average, comparable to those in other manufacturing sectors. Out of 35 thousand firms covered in our data set, 15% applied for a patent at least once (9% applied during the period covered by Nielsen). This patenting rate is in line with that of the manufacturing sector and is substantially higher than that of other sectors in the economy (Graham, Grim, Islam, Marco and Miranda, 2018). The consumer goods sector also covers a wide range of product categories with distinct patenting intensities, allowing us to show the heterogeneity of our results by product types. Over our sample period and across all product categories, never-patenting firms introduce more than 54% of new products and 65% of quality-adjusted new products.

We begin our empirical analysis by exploring the relationship between patents and product introduction across all firms, independent of their market share. Our theoretical framework shows that product introduction and patenting are complementary activities; hence we should observe a positive relationship between these activities in the data. To capture this relationship in the data, it is key that we can directly associate specific patents with specific products. We use variation within firm $\times$ product category over time and estimate how filing new patents relates to introducing new products, controlling for product category-specific trends and firm-category fixed effects that capture, for example, heterogeneous propensities to innovate and file patents. We find that a 10% increase in the number of new patent filings is associated with a 0.4% increase in product introduction in the next year. We observe higher elasticities between product introduction and filings of higher-quality patents, such as granted patents and those with more forward citations.

Next, we evaluate how patenting and innovation decisions change with firm size. Using variation across firms within product categories, we estimate that firms at the bottom quintile of the size distribution in a given year, as measured by total sales within a product category, introduce one new product for every five existing products in their portfolio, on average. Firms at the top quintile of the size distribution, on the other hand, introduce one new product for every seven existing products in their portfolio. Though larger firms' innovation rates are lower, they file more patents. We show that the patent filings of larger firms have a significantly weaker association with product introduction.

Through the lens of our model, the weakening relationship between patenting and product innovation for large firms reflects the higher occurrence of strategic patenting, which is not directed toward product innovation but aims to deter the competitors of market leaders. We find additional evidence consistent with this idea in the data, namely that patents filed by market leaders carry a larger revenue premium above and beyond what would be pre-

dicted by the quantity and quality of new products introduced by these firms. In contrast, the patent revenue premium of small firms is fully accounted for by the product introduction associated with these patents. This empirical regularity is consistent with the model prediction that large firms enjoy higher revenue premiums from patents because patents reduce competitive pressure. We also show that patent filings by market leaders are associated with a decline in competitors' product introduction in shared product categories. The same is not true if we consider smaller firms' patent filings. Finally, we rule out alternative hypotheses that might explain the weakening patents-to-innovation relationship, such as market leaders having longer time lags between patent filings and the introduction of products in the market (for example, because of higher experimentation) or their use of patents for licensing purposes. We also show that patents of market leaders often exhibit lower quality in terms of novelty and their scientific impact: these patents have fewer follow-on citations, a higher self-citations share, exhibit higher textual similarity with preceding patents, and are more often used in litigation.

We then provide back-of-the-envelope calculations for the frequency of strategic patenting and its aggregate implications using the structure of our model. We calibrate the model parameters and compare the benchmark economy to a counterfactual one, where firms can file patents but strategic patents without product introduction are not allowed. Our quantification implies that 61.6% of the patent filings of an average firm and 83.5% of the filings of the largest firms at the top of the sales distribution are strategic. This share of strategic patents reduces aggregate creative destruction by 2.5% relative to the counterfactual economy with no strategic patenting.

Finally, we analyze further counterfactuals to understand the potential benefits and costs of different patent policies. We compare economies with different patent regimes: no patents, no strategic patents , and a benchmark regime with no restrictions on the nature of patents. Comparing the no-patents regime to the no-strategic-patents regime illustrates the innovation-enhancing role of patent rights in the economy (Bryan and Williams, 2021): incumbents introduce new products at a much higher rate in a no-strategic-patents regime because their new products can be protected by patents. However, if the economy also allows for strategic patents, two important forces operate. On the one hand, creative destruction and reallocation fall relative to the no-strategic-patents regime, but, more interestingly, incumbents' product introduction rate falls too. In an economy where incumbents can rely on strategic patenting to maintain their market shares, the incentives for additional product introduction are lower. The comparison of these three patent regimes indicates that while, in principle, patent protection can incentivize ex-ante product innovation by firms, the possibility of obtaining patent protection without product commercialization can reverse

the benefits of the patent system for the economy as a whole.

Related Literature – Our findings regarding the patenting and innovation decisions of firms can speak to several puzzling macroeconomic trends in recent data: patenting is soaring, but productivity growth is stagnating (Gordon, 2016; Bloom, Jones, Van Reenen and Webb, 2020); large firms funnel more resources into intangible capital—including intellectual property—but these expenditures are manifested in the increasing dominance of those firms instead of perceptible improvements in aggregate innovation in the economy (Crouzet and Eberly, 2019). Our results show that large incumbents have limited incentives to direct their efforts towards productive rather than strategic patenting, which is particularly relevant as more economic activity is reallocated towards firms with a large degree of market power (De Loecker, Eeckhout and Unger, 2020; Autor, Dorn, Katz, Patterson and Van Reenen, 2020; Gutiérrez and Philippon, 2017) and business dynamism declines (Decker, Haltiwanger, Jarmin and Miranda, 2016). Recently, Akcigit and Ates (2023, 2021) have argued that the decline in knowledge diffusion from market leaders to laggards has largely contributed to the slowing business dynamism in the United States. Our study lends support to this idea by providing direct evidence on strategic patenting by market leaders and showing that these patenting practices lead to a quantitatively sizable decline in aggregate reallocation and innovation.

Our results also contribute to our understanding of firms' growth strategies. Recent studies have shown that large firms rely on other protective strategies such as acquiring potential competitors (Cunningham, Ederer and Ma, 2021) or forging political connections (Akcigit, Baslandze and Lotti, 2023) as their innovative activity slows down (Akcigit and Kerr, 2018; Cavenaile and Roldan-Blanco, 2021).[5] We show both theoretically and empirically that patenting is yet another protective tool that firms substitute for actual product innovation as they grow.

Given the importance of strategic patenting for market leaders, there is surprisingly little empirical evidence on the differential use of patents by firms with different market positions. Relevant papers that have studied patents and the associated follow-on innovation include: Williams (2013) and Sampat and Williams (2019) for human genes; Cockburn and J. MacGarvie (2011) for software products; and Lampe and Moser (2015) for follow-on patenting with patent pools. While these papers have not considered their effects by firm size, Galasso and Schankerman (2015) examined 1,357 Federal Circuit patents and showed that invalidating patent rights of large patentees led to more follow-on citations to the focal patents by small patentees. In our data, we observe direct measures of product

[5] See Baslandze (2021) for a review.

innovation in the market for all firms in the consumer goods sector and show that patenting by market leaders is related to lower product commercialization by competitors. To the best of our knowledge, this paper is the first to study the relationship between patent filings and product innovations over the full firm size distribution using direct information on the commercialization of products.

Finally, our novel data set sheds light on the usefulness of patent statistics for measuring innovation. In the absence of direct measures of innovation, the literature has relied on indirect inference approaches using data about employment growth (Garcia-Macia, Hsieh and Klenow, 2019) or valuing innovation from patent statistics themselves (e.g., Akcigit and Kerr, 2018). Other researchers have looked at innovations that occur outside of the patent system by examining the number of new books on technical topics (Alexopoulos, 2011) or innovations featured at World Fairs between 1851 and 1915 (Moser, 2012). By linking patents to specific products within firms—usually an unobservable relationship—our data enable us to document that the usefulness of patent metrics in inferring innovation depends significantly on the market position of the firms that own the patents.

The rest of the paper is organized as follows: we start from a theoretical framework in Section 2; we then describe the data sets, our data-matching procedures, and extensive validation checks in Section 3; the main empirical results are presented in Section 4; Section 5 provides model-based estimates of the frequency and aggregate implications of strategic patenting and evaluates various counterfactual patent regimes; and Section 6 concludes.

2 Conceptual Framework

We build a simple Schumpeterian model of endogenous innovation and creative destruction (e.g. Aghion and Howitt, 1992) that illustrates the relationship between product introduction and patenting decisions for firms with different market positions. The model provides a set of empirical predictions and identifies relevant data moments that guide our empirical strategy.

2.1 Setup

Production – Consider a partial equilibrium framework that depicts the innovation process in a single product category. There are J potential producers, and aggregate output is produced using a combination of their quality-weighted varieties:

$$Y = \frac{1}{1-\beta}\left[\sum_{j=1}^{J} q_j^{\frac{\alpha}{1-\beta}} y_j\right]^{1-\beta}, \tag{1}$$

where y_j denotes the quantity, and q_j is the quality level of variety j. This specification implies that products from different producers are perfect substitutes after adjusting for their qualities. The parameter α captures the consumer's satiation with respect to additional quality. Labor is the only factor of production. Producers use labor to make intermediates by hiring workers at the common wage w. Output of variety j is then given by $y_j = l_j$, where l_j is the amount of labor used to produce variety j. We assume that an overhead cost of production ϵ must be paid before choosing prices and output. Since producers' marginal costs are the same and qualities are different, under Bertrand competition, even a small overhead cost allows the highest-quality firm to act as a the sole producer.[6]

The incumbent producer maximizes profits by choosing the price of its product subject to demand from (1),[7] which delivers the following equilibrium values for output (y), sales (R), and profits (Π):

$$y = \frac{1-\beta}{\beta}\frac{\pi}{w}q^{\gamma}, \quad R = \frac{\pi}{\beta}q^{\gamma}, \quad \Pi = \pi q^{\gamma}, \tag{2}$$

where $\pi \equiv \beta\left(\frac{1-\beta}{w}\right)^{\frac{1-\beta}{\beta}}$ and $\gamma \equiv \frac{\alpha}{\beta}$. Hence, an incumbent with a higher-quality product is larger and generates higher sales and profits. We assume $0 < \alpha < \beta$ (meaning $0 < \gamma < 1$) so that the marginal quantity, sales, and profits decrease with quality. Later in the paper, we show that this is the relevant empirical case, so for brevity we adopt this parametric assumption now.

Product Introduction and Patenting – Incumbent firms can improve upon their existing products and file patents. We consider a one-time decision of product introduction and patenting for an incumbent with quality q who exogenously obtains an idea of size λ.[8] Product introduction is not certain. The firm chooses the probability of product introduction z_m by incurring marketing and product commercialization cost $\frac{c_m z_m^2}{2}$. If product introduction is successful, the incumbent brings a new product with a higher quality $q + \lambda$

[6] This assumption, common in this class of models, simplifies the setup. Alternatively, we could work with limit pricing where the firm with the highest quality still captures the entire market, but the price is determined by the price of the second highest-quality producer.

[7] Price of Y is normalized to one.

[8] For simplicity, we consider a one-shot decision; an idea is either used or disappears afterwards. A dynamic problem of patenting and product introduction would bring similar qualitative predictions at the expense of tracking the evolution of a firm's position both in the product and patent spaces.

to the market and earns higher profits.

Simultaneously, the firm also chooses the probability of patenting, z_p, and incurs patenting cost $\frac{c_p z_p^2}{2}$. This cost of obtaining patent protection on the idea λ can be thought of as a combination of research, filing fees, legal fees, and potential patent enforcement costs. A patent grants the firm additional protection against being replaced by a competitor (more details below). If the incumbent successfully obtains a patent, even if the idea is not commercialized, the highest-available quality in the economy becomes $q + \lambda$ since, by patenting, the firm makes the idea public (Hegde, Herkenhoff and Zhu, 2022). As a result, the highest-available quality in the economy could be different than that commercialized by the firm and available to consumers. In this sense, firms' activities in the product and patent spaces are separated. New product introduction does not necessarily imply a patenting activity, and obtaining a patent does not necessarily imply that the firm introduces a new product—an important departure from standard models of innovation and growth.

Creative Destruction – Incumbents can be replaced by competing entrants through creative destruction. For that to happen, entrants need to introduce a higher-quality product and clear the legal barrier created by incumbents' patents, if any.

Entrants arrive at an exogenous rate p at each instant and draw an innovation of step size λ^e from a uniform distribution on $(0, 1)$ relative to the highest available quality. They build on "the shoulders of giants" and learn from products available in the market as well as from existing patents. Hence, the highest available quality on which entrants are building corresponds to $q + \lambda$ if the incumbent does product introduction and/or patents, and to q if incumbent does neither of these two activities.

Normally, an entrant would win the market if it introduced a higher-quality product than that offered to consumers by the incumbent. However, this is not necessarily the case when the incumbent also holds legal protection from patents. Patenting protects the quality level of incumbents $q + \lambda$ by creating a "wall" of height ε $(0 < \varepsilon < 1)$ that entrants need to overcome to enter the market. The parameter ε captures the idea that entrants need to come up with an innovation sufficiently different from what has been patented before, which can depend on the strength of intellectual property protection as well as the scope of the patent. As a result, the probability of creative destruction is p if the incumbent does not patent and is $p(1 - \varepsilon)$ if the incumbent patents (Appendix A.1 provides the proof). This implies that, in contrast to standard models of creative destruction, not all product quality improvements by entrants will find their way to the market. The separation between the patent space and the product space introduces the possibility that a higher-quality product developed by an entrant does not get introduced to the market because it is blocked by an

incumbent's patents.

Value Function – Consider an incumbent with initial product quality q.[9] Denote its (gross) value when it introduces a new product and patents the idea by V^{11}, the value of only introducing a new product by V^{10}, the value of only patenting by V^{01}, and the value of neither introducing a new product nor patenting the idea by V^{00}. Then we obtain

$$\begin{aligned} V^{11}(q) &= \frac{\pi(q+\lambda)^{\gamma}}{r+p(1-\varepsilon)}, \qquad V^{10}(q) = \frac{\pi(q+\lambda)^{\gamma}}{r+p}, \\ V^{01}(q) &= \frac{\pi q^{\gamma}}{r+p(1-\varepsilon)}, \qquad V^{00}(q) = \frac{\pi q^{\gamma}}{r+p}, \end{aligned} \tag{3}$$

where r is interest rate. The value of the incumbent firm with existing quality q is then an expectation over these values, net of product introduction and patenting costs:

$$\begin{aligned} \mathbb{V}(q) = \max_{z_m, z_p} \Big\{ & z_m z_p V^{11}(q) + z_m(1-z_p)V^{10}(q) + (1-z_m)z_p V^{01}(q) \\ & + (1-z_m)(1-z_p)V^{00}(q) - \frac{c_m z_m^2}{2} - \frac{c_p z_p^2}{2} \Big\}. \end{aligned} \tag{4}$$

A key feature of this economy is that incumbent firms can engage in *strategic patenting*, defined as filing patents without product introduction. This possibility of patenting without product introduction is captured by the availability of option $V^{01}(q)$.

2.2 Equilibrium

After solving the firm's problem in (4), we obtain the optimal product introduction z_m and patenting z_p choices. While the setup is fairly simple, it produces key predictions about the relationship between product introduction and patenting activities. We present all proofs in Appendix A.2.

Proposition 1: Product introduction and patenting activities are complementary: $\frac{\partial z_m}{\partial z_p} > 0$.[10]

[9] Notice that we analyze the problem of a single incumbent in a product category that faces potential competition from entry. Our model implications can be generalized to a richer setting with multiple competing incumbents holding different market shares. Here, instead of comparative statics with respect to firm market share, for simplicity, we chose to analyze the single incumbent problem and consider the comparative statics of product introduction and patenting decisions with respect to firm size (sales), captured in the model by q.

[10] With some abuse of notation, $\frac{\partial z_m}{\partial z_p}$ denotes the degree of complementarity– the cross-derivative of $\mathbb{V}(q)$, which we formalize in Appendix A.2.

This proposition implies that any change to the returns from product introduction or patenting will move both activities in the same direction. For example, if the return from product introduction increases, then the benefit from protecting new products with patents also increases. Likewise, if a firm anticipates stronger protection from patents, it will also find it more profitable to introduce a new product. Note that the latter captures the fundamental reason for the existence of the patent system in the first place—patent protection provides ex-ante innovation incentives. In fact, the following comparative static shows that the complementarity between patenting and product introduction is stronger when patent protection is stronger.

Proposition 1': All else equal, if patent protection is stronger (ε is larger), the complementarity between product introduction and patenting is stronger: $\frac{\partial^2 z_m}{\partial z_p \partial \varepsilon} > 0$.

Next, we move to analyze how this complementarity between product introduction and patenting varies with an incumbent's size, captured by its quality level q. We show that:

Proposition 2: The degree of complementarity between product introduction and patenting is weaker for larger firms: $\frac{\partial^2 z_m}{\partial z_p \partial q} < 0$.

This decline in the degree of complementarity is the impetus for higher strategic patenting by larger firms. In other words, this result stems purely from the fact that firms have the option to patent with no accompanying product introduction and that larger firms have higher incentives to choose this option.[11]

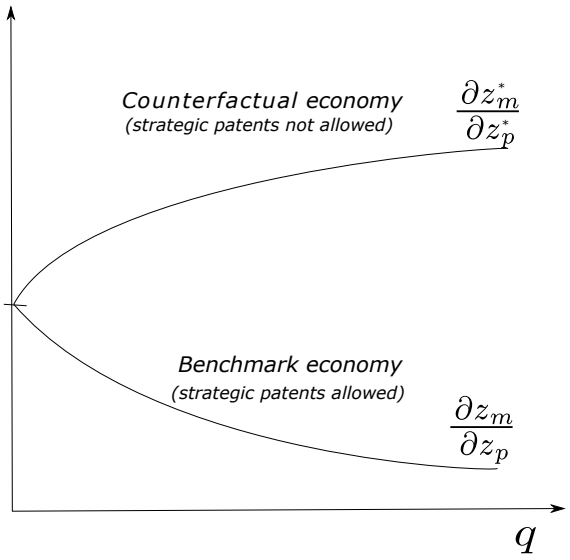
To illustrate this more clearly, let us compare the implied product introduction and patenting incentives of the firms in our economy with those in a counterfactual economy where we do not allow for the existence of strategic patenting. In particular, we consider a *counterfactual economy* where the patent office does not grant patent protection for an idea that is not commercialized in the market.[12] In such a case, the value function becomes:

$$\mathbb{V}^*(q) = \max_{z_m, z_p} \Big\{ z_m z_p V^{11}(q) + z_m(1-z_p)V^{10}(q) + (1-z_m)V^{00}(q) - \frac{c_m z_m^2}{2} - \frac{c_p z_p^2}{2} \Big\}. \tag{5}$$

[11] To keep the analysis tractable and intuitive, our model abstracts from some features such as size-dependent costs and proportional innovation step sizes that might be relevant for this theoretical result. In Appendix Section A.3, we discuss these alternative modeling assumptions and show how this result carries over under new assumptions.

[12] Though commercialization has never been required in the US, until 1880 the USPTO did required applicants to submit a (possibly working) model of the their invention when feasible (Federico, 1936).

Figure 1: Patents and Innovation Complementarity over Firm Size



The difference between $\mathbb{V}^*(q)$ and $\mathbb{V}(q)$ is simply the omission of the $V^{01}(q)$ term (and the residual probability on $V^{00}(q)$). We can show that the result from the previous proposition is reversed in the counterfactual economy:

Proposition 2: If strategic patents are not allowed, the complementarity between product introduction and patenting is stronger for larger firms:* $\frac{\partial^2 z_m^*}{\partial z_p^* \partial q} > 0$.

Figure 1 schematically shows the degree of complementarity between patents and innovation for the benchmark and counterfactual economies, $\frac{\partial z_m}{\partial z_p}$ and $\frac{\partial z_m^*}{\partial z_p^*}$. In both cases, product introduction and patents are positively associated. However, as incumbent's size increases (larger q), the strength of the association between patents and product introduction diverges between the two economies. In the counterfactual economy, where strategic patents are not allowed, the link between product introduction and patents is stronger with firm size because larger firms have more incentives to protect, but at the same time, firms have to introduce new products to benefit from patent protection. This is not the case in the benchmark economy where firms can protect their position solely by filing patents. Hence, it is the availability of the option of patenting without commercialization that generates the decline in the association between patents and product introduction with firm size.

To further evaluate how the incentives for patenting and product introduction vary with firm size, we can conveniently express the gap in the degree of complementarity between the counterfactual and benchmark economies as

$$\Delta \equiv \frac{\partial z_m^*}{\partial z_p^*} - \frac{\partial z_m}{\partial z_p} = \frac{\pi \varepsilon p q^\gamma}{(r + p(1-\varepsilon))(r+p)}. \tag{6}$$

This gap is essentially a revenue premium from strategic patenting, $(V^{01}-V^{00})$. Importantly, this revenue premium varies with firm size

Proposition 3: The revenue premium from strategic patenting is higher for larger firms: $\frac{\partial(V^{01}-V^{00})}{\partial q} > 0$.

Intuitively, larger incumbents have a higher value to protect, hence, they reap a larger return from patenting. Contrary to the incentives for patenting, the incentives for product introduction decline with incumbent's size ($\frac{\partial(V^{10}-V^{00})}{\partial q} < 0$): similar to the well-known *Arrow-replacement effect*, the incremental returns from product introduction decline with firm size.[13] These opposing incentives for product introduction and patenting with firm size explain the higher return from strategic patenting for larger firms.

In the aggregate, the occurrence of strategic patents has negative consequences for creative destruction and reallocation in the economy. The expected creative destruction rate can be expressed as

$$\tau \equiv z_p p(1-\varepsilon) + (1-z_p)p \tag{7}$$

The rate decreases with the patenting rate and because the rate varies with the size of incumbent firms:

Proposition 4. The expected creative destruction rate is lower when incumbents are larger: $\frac{\partial \tau}{\partial q} < 0$.

Intuitively, because larger incumbents rely on patenting more, we find that larger incumbents face a lower risk of creative destruction.[14]

2.3 Main Takeaways and a Pathway to Empirics

We propose a simple extension of the Schumpeterian growth framework (Aghion and Howitt, 1992) that entertains a distinction between a firm's choice to innovate and their choice to patent. The model illustrates how the incentives for patenting and product introduction go in opposite directions as firms grow. Large incumbents give up on many ideas that smaller firms would find worth commercializing to avoid cannibalizing their existing rents. Meanwhile, the incentives for patenting go in the opposite direction—larger firms want to protect

[13]This result requires that $\gamma < 1$ and thus the marginal profits decrease as q increases. We provide an empirical estimate of γ and confirm that it is lower than one. However, instead using decreasing returns to generate a declining relationship between firm size and innovation, one can can also generate it by introducing weaker scalability of R&D technology with increasing size as in Akcigit and Kerr (2018) or an innovation-advertising trade-off as in Cavenaile and Roldan-Blanco (2021).

[14]In the proof, we provide sufficient parametric conditions that guarantee z_p to grow with firm size.

their existing rents relatively more and file patents even if they do not intend to commercialize them. The occurrence of these strategic patents, i.e., patents without associated product introduction, has negative implications for creative destruction and innovation.

Our model leads to a set of empirical moments on product introduction and patenting that help identify the presence of strategic patents and their implications. To start with, Proposition 1 implies that empirically, we should observe a positive association between product introduction and patenting by firms. In turn, Proposition 2 and 2^* tell us that this association naturally gets weaker with the incumbent's size, but only if strategic patenting is allowed. Hence, these propositions highlight that the key moment that helps us detect the existence of firms' strategic patents is the observed empirical association between patenting and product introduction decisions and how it varies with the incumbent's size. Proposition 3 corroborates these findings and shows that the returns to strategic patenting are particularly high for larger incumbents. This implies that while we generally expect incumbents to increase their revenue when they file patents, as patents would reflect product innovations, for larger firms, this patent revenue premium goes beyond product commercialization. Finally, Proposition 4 implies that due to patenting, we should observe lower levels of product introduction by competitors of larger incumbents.

To speak to these empirical predictions from the model, we need to have detailed data on firms' patents linked to product introductions, as well as the revenue and innovation levels of their competitors. The following section describes how we assemble such a data set.

3 Data and Measurement

In this section, we show how we take advantage of rich granular product- and patent-level data and leverage techniques from natural language processing to build a unique data set that allows us to study the relationship between patents and product introduction and ultimately to test the predictions of our theory.

3.1 Patent and Product Data

3.1.1 Overview

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

empirical strategies we use to address these challenges.

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

To address the second challenge of linking patents to products, we start by using firm names recorded in the patent and product data sets to produce a mapping between firms' patent portfolios to their products.[15] This data is too coarse to allow us to connect patents with specific products. Moreover, it does not take into account that some patents are associated with process innovation, or with innovations outside the consumer goods sector that we do not capture with our product introduction measures. Thus, we leverage the richness of product and patent characteristics and use methods from the natural language processing literature to create systematic links between sets of patents and sets of products within a firm.

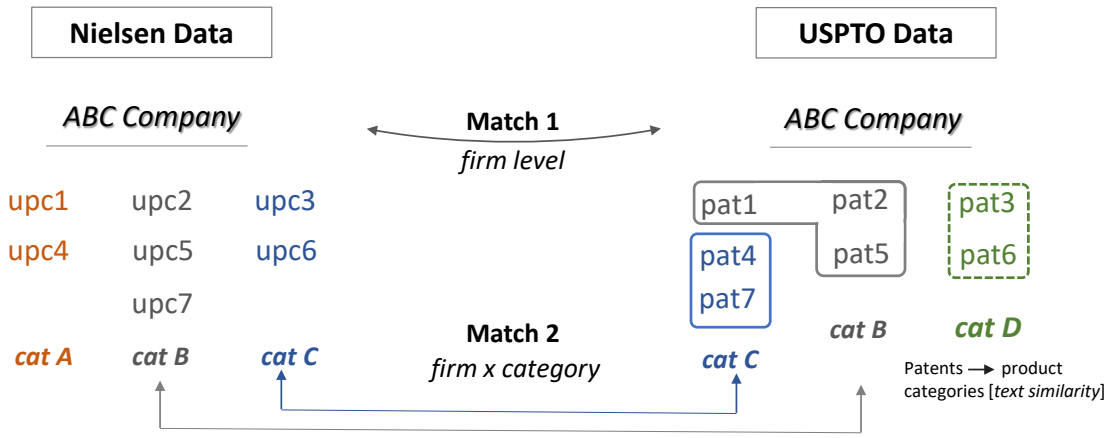
A patent may generate no products or multiple products, and a product may have benefited from multiple patents or from none at all. Therefore, forcing a one-to-one matching between a specific narrowly defined product and a specific patent is neither possible nor desirable. Hence, our approach is to first define product categories as sets of similar products, which are identified using clustering analysis of product descriptions extended with Wikipedia-based dictionaries. We then assign each individual patent to the product category with which it has the highest text similarity within the set of consumer goods covered by the product data. This classification of a firm's products and patents into the various product categories offered by that firm yields our benchmark patent-to-products data set. Figure 2 illustrates our data schematically, and our matching algorithms are described in detail below.

To our knowledge, our algorithm generates a data set that is truly unique. de Rassenfosse (2018) has collected data on about 100 firms with virtual patent markings; and some private companies link patents to products of their clients.[16] However, none of these data sets have

[15]Because this matching procedure is simple and parsimonious, we use this firm-level panel data set in multiple robustness exercises.

[16]Some examples of these companies are FairTech, IPStrategy, Powering ideas, and Intelectual Peritus. Their data sets are confidential and apply only to the portfolio of products of their clients. These services help clients identify their most important patents and the protection they provide to their products portfolio, in addition to helping firms prepare in case of litigation. These companies also use text similarity techniques, using various parts of the patents text and short descriptions of products from trademark data and other

Figure 2: Product and Patent Data Sets



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

rich systematic product-level data for all firms in a sector.

3.1.2 Data Sources

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

The data focus on the consumer product goods (CPG) sector, which accounts for 14% of the total consumption of goods in the U.S. This sector includes food and non-food categories (health and beauty aids, non-food grocery, and general merchandise such as cookware, electronics, gardening, household supplies). Our data cover the years from 2006-2015, and combines all sales, quantities, and prices at the national and annual levels. We use the panel structure of each UPC to measure its entry year. This product data set covers about

sources.

40% of the CPG sector sales, and nearly the universe of firms and new products in the sector. Appendix B.1 provides additional details about the coverage and representativeness of Nielsen RMS to measure product innovation in the consumer goods sector.

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

3.1.3 Matching Firms

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

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

3.1.4 Matching Patents to Product Categories

The algorithm used to build our firm × category panel data set has three key steps. In this section, we describe the details of these steps. The first step creates product categories at a level of aggregation such that they collect distinct and sufficiently large sets of similar products that are meaningfully related to a distinct set of patents. This step yields a set of product categories, a vector of terms used to describe each product category, and a mapping of products into categories. In the second step, similarity scores between patents and product categories are computed. We use various text descriptions to build a vector of terms that describe each patent. We then compute similarity scores between each patent and every product category. These scores represent the overlap between the texts in patents and the text associated with each product category. The final step of our patent-product matching algorithm consists in using the similarity scores and information about the production of the respective patenting firms to classify each patent into a product category and filter out patents not related to CPG products.

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

To this end, we apply a clustering procedure to aggregate the Nielsen modules into distinct product categories. Each module is paired with a vector of descriptive terms (single words and two-word phrases) that are weighted by their importance. We expand short module descriptions from Nielsen data with the text of hand-collected Wikipedia articles to get to a comprehensive description of the product content of the modules. The resulting vectors of descriptive terms collect all the words from the Wikipedia and Nielsen

texts, after applying standard parsing and lemmatizing algorithms. When building term vectors, one must appropriately weight terms by their importance. We use the leading approach in textual analysis—the "term-frequency-inverse-document-frequency" sublinear transformation—that accounts for both the frequency with which a term appears describing a module and how commonly it is used to describe other modules (Aizawa, 2003). This approach ensures that we under-weight common terms that appear in many documents as these are less diagnostic of the content of any individual document.

We then aggregate these module vectors into clusters using a popular technique known as k-means clustering (Lloyd, 1982). This procedure allows one to specify the desired number of clusters and yields a clustering assignment that minimizes the within-cluster term vector variance. As a baseline, we use an aggregation of modules into 400 clusters which we refer to as product categories. We find that this partition strikes a balance between aggregating very similar products while maximizing the difference between products across categories. Appendices B.4.1 and B.4.2 provide extensive descriptions of methods we have taken from the literature on natural language processing, including the details of clustering, quality assessment, and alternative methods to encourage robustness. Appendix E shows the sensitivity of our main findings using the aggregation of product modules defined by Nielsen.

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

Patent Term Vectors and Similarity Scores – We next describe how we measure the amount of overlap between the texts of patent applications and product categories. We create patent term vectors from the patent descriptions using the following fields from the patent applications: the title, abstract, international patent classification description, and the titles of cited patents. We create vectors of terms by concatenating all these fields into one document, followed by the same parsing and lemmatizing algorithms. As before, we adjust the weights of each term according to the "term-frequency-inverse-document-frequency" sublinear transformation and normalize patent vectors to have unit length.

Finally, we construct a similarity score for each patent p and each product category j by computing the cosine similarity between two normalized vectors, $s_{jp} = f_j \times f_p$. This

similarity score is guaranteed to be in the range $[0, 1]$ with zero indicating no word overlap and one indicating that the documents are identical. Appendix B.4.3 provides technical description of this step.

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

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

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

Our methodology assumes one product category match for each patent. However, some patents may be more general in nature so that they relate to multiple categories. Our baseline algorithm abstracts from this possibility. However, our procedure to define product categories is designed to ensure that the product categories would encompass a broad range of products that are technically similar such that one patent plausibly relates to this and only this range of products.[17] In Appendix B.4.4, we present the details of this procedure and all the robustness exercises with which we tested our baseline algorithm. Appendix E

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

Table 1: Match Statistics

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

Notes: Match statistics for the baseline firm-level and firm × category level data sets.

shows the sensitivity of our main findings under higher similarity thresholds.

3.1.5 Match Statistics and Validation

Table 1 provides statistics of the baseline data used in our analysis. Our data includes annual data for all 34,665 firms that sold at least one product in our consumer goods sector data (CPG firms). The raw USPTO patent data cover information from 1975 to 2017, but because our product data only cover years from 2006 to 2015, our analysis can only consider annual variation for the period 2006-2015. In this shorter period, the USPTO data include about 3.4 million patent applications in total, and about 500 thousand patent applications filed by CPG firms. The firm × category data set includes 40% of those patent applications.[18] The remaining 60% of patents, while filed by CPG firms, could not be associated with products in the consumer goods sector.

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

Manual checks – We manually checked many of the patent-to-products matches and some examples are listed in Table A1 in the Appendix. The table lists 100 patent applications by the top-selling firms in the largest product categories according to Nielsen. One

[18] Appendix B.6 illustrates examples of a variety of patent applications and the corresponding products with their respective product entry dates.

can easily see that the patent titles reflect well the product categories to which the patents were assigned. For most patents we analyzed, we found that our manual choices of product categories also coincide with the product categories chosen by our matching algorithm using similarity scores.

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

For Procter & Gamble (P&G) and Kimberly Clark (KC), we manually collected virtual patent markings from the company websites and mapped them to our product categorization. We then validate our patent-product category matches for these firms against this information. Appendix B.5.2 shows that the patent-product category mapping from virtual markings is also selected by our matching algorithm in about 70% of cases.

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

Validating non-matches – In our last step of the algorithm for Match 2, multiple criteria are used to allow for the possibility that some patents filed by CPG firms are not associated with any of the consumer-good product categories. A valid "non-match" can arise for two main reasons. First, a patent may relate to goods that the firm may be producing outside the CPG sector; second, a patent may be about a general process or method that does not affect the introduction of new products. In the spirit of Hoberg and Phillips (2016),

we use information from publicly traded companies' 10K reports to identify firms whose output is mostly in the consumer-goods sector, and we find that only a minority of their patents are classified as "non-match", contrasting with patents held by firms who mostly sell products outside the consumer goods sector. Next, we use alternative procedures to proxy for process patents (completely independently from the algorithm) and compare them with the algorithm's "non-matches". We follow Bena and Simintzi (2017), and use patent claims to create proxies for process-related and product-related patents. We find that the share of "non-matches" is significantly higher among the claims-based measure of process-related patents. These exercises, which are presented in Section B.5.5 in the Appendix, offer reassurance that our algorithm successfully filters out patents that are not directly related to the products in our data.

3.2 Measures of Product Introduction and Patenting

3.2.1 Product introduction

Our measures of product introduction are based on the number of products that firms introduce to the market and the quality improvements in those products. We use the product data described above to identify the entry dates of products in the market and their respective characteristics and performance. We create separate measures of innovation for the firm-level and firm×category level data. Our first measure is the number of **new products** of firm i (in product category j) in year t, as in Broda and Weinstein (2010) and Argente, Lee and Moreira (2018):

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

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

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

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

Our baseline measure of product quality aims at capturing differences in novelty and economic impact across new products. We build on Argente and Yeh (2022) and use detailed information on product attributes that is available from the product data. Products can then be compared on the basis of characteristics associated with their attributes $\{v_{u,1}, ..., v_{u,A}\}$.[19] We test if each new product has characteristics distinct from those of all existing products available in the market, and we compute the quality of a new product as a weighted sum of its novel characteristics across all product attributes:

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

where ω_a are weights that reflect the economic value associated with a particular attribute. We develop a novel approach to estimate weights that capture the importance of each attribute by using "shadow prices" from hedonic pricing regressions (Bresnahan and Gordon, 1996). The underlying assumptions here are that the degree of novelty of a product should be reflected in the price of a product and that the price of a product reflects its embodied characteristics as valued by shadow prices. A new product has a high novelty score if it has many new characteristics and/or if its characteristics are associated with high implicit prices. We provide details on the properties of this procedure in Appendix B.7, along with some evidence that the novelty score is strongly associated with the performance of the firm and its products.[20]

We use three alternative measures of new product quality to evaluate the robustness of our empirical results. First, we use a simpler version of the quality measure that weighs each attribute equally (quality *q1*). This measure only captures variation in the share of new product characteristics contained in a product. Second, we use a weighted quality measure using weights that reflect "shadow sales" (quality *q2*). This measure assigns lower quality to new products that are associated with high shadow prices but do not reach many

[19] For example, "children" and "regular" are two mutually exclusive characteristics associated with the attribute "formula" for "pain remedies-headache" products. Naturally, the number and type of attributes varies across product categories. For example, the product category "pain remedies-headache" includes 10 attributes: brand, flavor, container, style (i.e. children, regular), form, generic, formula (i.e. regular, extra strength, rapid release), type (i.e. aspirin), consumer (i.e. trauma, migraine), and size. On average, we observe that the different product categories include between 5 to 12 attributes. Appendix B.7 gives details.

[20] We show that our measure is correlated with the growth rate of the firm, the share of sales generated by new products, and the average duration of new products in the market even after conditioning on the number of products being introduced by the firm (Table A2 in the Appendix).

customers. Finally, we use a measure of residual demand taken from Hottman et al. (2016) and Argente et al. (2022) (quality $q3$). This measure does not use information about the degree of novelty of a product and instead captures the appeal of new products relative to other products sold in the market, under some functional-form assumptions. Overall, our baseline measure and these alternative metrics allow us to consider many critical dimensions of the quality of new products and to assess the robustness of our results.

3.2.2 Patent Measures

Using an approach similar to how we measured product introduction, we compute measures that allow us to account for differences in the quantity and quality of patent applications across firms and over time. Our baseline measure is the number of **patent applications** (P_{it}). Using our patent-product category match, we are also able to measure the number of patent applications filed by firm i in product category j in year t as follows:

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

Throughout the paper, we use information about whether a patent was granted and information about patent citation counts to compute our measures of patent quality. Patent applications that become **granted patents** (gP_{ijt}) are perceived as high-quality patents because the patent office deemed them novel enough to not be rejected. We compute the number of patent applications that are granted as:[21]

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

We also define **patent citations** (cP_{ijt}) as the total number of patents weighted by forward citations received in the first five years since the application was filed:[22]

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

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

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

[22] A 5-year citations measure attempts to reduce the truncation issue inherent to citations—the fact that patents filed more recently have had less time to accumulate citations (Hall et al., 2001).

3.3 Summary Statistics

Table 2 provides summary statistics about the product- and patent-related variables for the firms in our sample, grouped by their patenting activity. We split firms into three groups: (i) firms that have never filed a patent application, (ii) firms whose last patent application was filed before 2006 (the beginning of the Nielsen RMS data set), and (iii) firms that filed a patent application between 2006 and 2015. The share of patenting firms and product introduction rates in the consumer goods sector are comparable to those of other manufacturing sectors. More than 5 thousand firms (15%) applied for at least one patent and more than 3 thousand firms (9.5%) filed a patent application during the period 2006-2015. For comparison, Graham et al. (2018) links Census data to the USPTO and finds that 6.3% of manufacturing firms have at least one granted patent application between 2000 and 2011.[23] The corresponding number in our data is 7.6%, which is only slightly higher.[24] Table 2 indicates that product introduction rates are on average 20%. While there is no equivalent comprehensive product data for other sectors, Goolsbee and Klenow (2018) use the Adobe Analytics data on online transactions covering multiple products and report product introduction rates that are comparable to those of other non-durable consumer manufacturing sectors.[25]

Firms with patent applications between 2006 and 2015 file more than six patents per year, on average. Because many patents receive no citations, especially in the first five years, the average number of citation-weighted patent applications, cP_{ijt}, is very similar to the average raw number of patent applications, P_{it}. These firms may hold some design patents, but the majority of patents in our sample are utility patents. Unsurprisingly, the summary statistics show that firms who filed a patent between 2006 and 2015 hold a larger stock of patents than firms who last filed a patent application before 2006.

As expected, patenting firms are on average larger: they sell more products, operate in more product categories, and have higher sales. Nevertheless, a large amount of innovation is associated with firms that never used the patent system. Table 3 shows that in our data, 54% of new products were introduced by firms that never applied for a patent. If we account for the degree of novelty of new products, we estimate that about 65% of quality-adjusted product introduction comes from never-patenting firms. This indicates that, on average, patenting firms introduce more products that make only an incremental improvement over

[23] The incidence of patenting in the rest of the economy is lower, at 1%.

[24] Notice that Graham et al. (2018)'s patent data include only granted patents, while our data also include unsuccessful patent applications. If we count only granted applications, we would have 2629 patenting firms.

[25] Goolsbee and Klenow (2018) show that some durable consumer goods (e.g. furniture), not covered in our data set, have entry rates that are larger than those of non-durables (e.g. food).

Table 2: Summary Statistics by Firm's Patenting Status

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

Notes: The table shows the average of product-based and patent-based variables of the Match 1 data set. The first column groups firms that have no patents; the second column considers firms that have patents, but filed them before they first appear in Nielsen RMS (before 2006); and the third column is for firms that have patents in our focus period of 2006-2015. The statistics regarding product introduction can only be computed for the period 2007-2015 because we cannot determine entries for products first introduced in 2006 (left-censored). The statistics for sales are given in thousands of dollars, deflated by the Consumer Price Index for all urban consumers. Patent statistics are very skewed; the table reports averages after winsorizing patent-based variables at the top 0.1%.

existing products on the market.[26] Since they rely on the firm-level match, the above statistics implicitly attribute all new products introduced by a patenting firm to some of its patents. However, highly diversified firms might be patenting in one product category, while introducing many products that have no relation to the patents they are filing in other categories. Thus we may be attributing too much product introduction to patents if we rely only on the firm's overall patenting status. This observation exemplifies the importance of establishing a closer link between patents and products using the Match 2 data set. To make these more granular links, we replicate the above exercise but define patenting status at the firm $\times$ category level. As seen from Table 3, firms that never patented in a category

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

Table 3: Share of New Products Accounted for by Patenting Firms

	New Products, N	Quality-adjusted New Products, qN
Match 1		
Firms with patents in 2006-2015	0.38	0.28
Firms with patents before 2006	0.08	0.07
Firms with no patents	0.54	0.65
Match 2		
Firm × category with patents in 2006-2015	0.23	0.16
Firm × category with patents before 2006	0.07	0.05
Firm × category with no patents	0.71	0.79

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

are responsible for a greater share of new products introduced in that category.

Our data cover product categories that exhibit substantial heterogeneity in entry rates and patenting intensity. In Appendix D we provide some descriptive statistics grouped across food and non-food categories.[27] The two types of product categories have, on average, similar entry rates but distinct patent intensities. The share of patenting firms and the ratio of patents per product is higher for non-food categories such as health and beauty aids (including over-the-counter drugs), non-food groceries, and general merchandise (including cookware, electronics, and various household supplies). It is not surprising that a large fraction of new product introduction is not directly linked with specific patents, especially in the food sector. Even if firms wanted to patent all new products, some products represent only small upgrades to existing ones and thus may not be patentable. Patents are only granted if the idea exhibits "novelty and non-obviousness," and new products that result from incremental changes will likely not be captured by raw patent metrics.

4 Empirical Analysis

In this section, we use our data set with measures of the firms' product introduction and patenting to test the implications of the theoretical framework. We start by studying the average relationship between product introduction and patenting for all firms, followed by the analysis of the key moments that identify strategic patenting by exploring firm size heterogeneity.

[27] Throughout the analysis we mostly use variation within detailed product categories which do not capture heterogeneity in innovation and patenting intensities across types of products. Nevertheless, in Appendix D we provide the results using only food or non-food product categories to ensure that the results are not driven by variation within some specific product categories.

4.1 Relationship Between Product Introduction and Patents

The first implication of the theoretical framework (building from Proposition 1) is that product introduction and patenting are complementary activities. We use our dataset and employ the following baseline specification using variation at the firm × category level over time

$$\log Y_{ijt} = \beta \log P_{ijt-1} + \alpha_{ij} + \gamma_{jt} + u_{ijt} \tag{8}$$

where Y_{ijt} is the measure of product introduction for firm i in category j in year t and P_{ijt-1} is the log number of patent applications filed by the firm i in category j in year $t-1$. The key coefficient is the elasticity β that captures how product introduction activities relate to patenting activities. By using firm × category × year level data, we can control for many potential confounding effects using product category-specific trends (e.g., controls for market-wide demand for specific products) and firm-category specific effects (e.g., controls for the effects of firm-specific market power on the sales of specific products). Importantly, this set of fixed effects also ensures that results are not driven by differences in patentability or coverage across distinct product categories, or firm-specific time-invariant predispositions to apply for patents.

Table A11 shows the estimates of equation (8) for both measures capturing product introduction decisions—new products ($\log N$) and the quality-adjusted new products ($\log qN$). The rows present results from using different explanatory variables—the log number of patents, granted patents, and non-granted patents.[28] The results indicate that there is a significant positive elasticity of product introduction and quality-adjusted product introduction to patent applications, which we interpret as evidence that product introduction and patenting are complementary activities.[29]

Our conceptual framework predicts that the association between product introduction and patenting increases with the degree of protection rendered by the patent. Consistent with this implication, the complementarity between product introduction and patenting is mainly driven by granted patents, which more effectively deter competitors' innovations (Table A11). We also evaluate this implication by exploring heterogeneity in the degree of patent protection across product categories. We find that the association between product introduction and patents is only strong for product categories that are likely to be able to be well protected by patents, such as the non-food categories. Table A7 in Appendix D shows that the association between product introduction and patents is frail for food-related

[28] Appendix Table A3 uses citations and claims as independent variables, while Appendix Table A4 shows the main results at the firm level (using Match 1).

[29] The estimated elasticities between 0.02 and 0.04 are in line with the magnitude of elasticities implied by our model, as discussed in Section 5.

Table 4: Product Introduction and Patenting

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

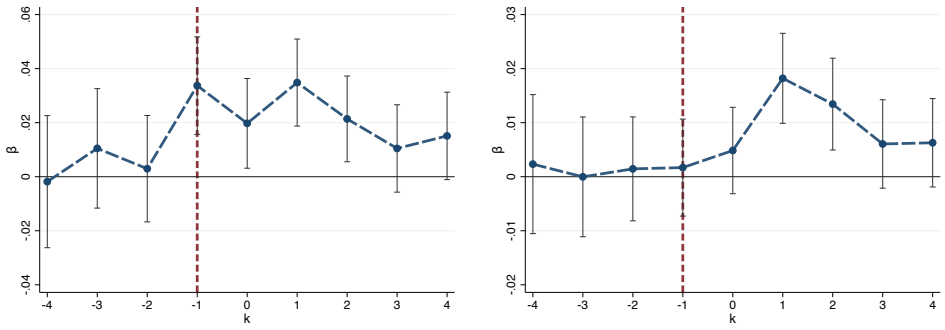
Notes: The table shows regressions of the log number of new products (log qN) and of log quality-adjusted new products (log qN) in a firm × category over time as a function of the log number of patents. Our benchmark quality measure is defined in Section 3.2.1. The alternative innovation-quality measures ($q1, q2, q3$) produce consistent results. *Patents* is the log number of any patent applications in firm × category × year; *Patents granted* is the log number of granted patent applications; *Patents non-granted* is the log number of patent application that have not been granted (abandoned or pending). The inverse hyperbolic sine transformation is used for logs. Standard errors robust against heteroskedasticity and serial correlation are reported in parentheses.

categories and is mostly driven by products in categories such as health and beauty care, non-food grocery, and general merchandise.

Our baseline specification uses one year lag between patent filing and product introduction to account for the fact that it may take longer for firms to develop and commercialize a new product after they apply for patents. We also allow for alternative lags governing the relationship between product introduction and patenting. Figure 3 plots the estimated coefficients for different lags, where k refers to the lag between patenting and product introduction. Consistent with the results above, we find a positive association between patents and product introduction fairly concentrated around one year lag. With an exception for product introduction at $k = -1$, we do not find a significant relationship for k below zero (i.e. where the time of product introduction precedes the time of patent application), and product introduction does not respond significantly after 2 years of the increase in patenting intensity. In Figure A13 in the Appendix, we also show that other variables such as the stock of products or sales significantly increase after patents and that the association is typically concentrated around a one-year lag.

The estimated coefficients of equation (8) capture the relationship between patent applications and product introduction. Not all patents, however, necessarily relate to product improvements: some patents may relate to cost savings from improvements to the firm's general production processes. Our firm × category data set filters out patents that are not specifically related to product introductions. Hence, to a large extent, our estimates should be driven by product patents rather than general process or organizational patents.

Figure 3: Product Introduction and Patenting: Dynamics



(a) $\beta_{\text{N,P}}$ (b) $\beta_{\text{qN,P}}$

Note: The figure plots the estimated coefficients after estimating equation $\log Y_{ijt+k} = \beta_k \log P_{ijt} + \alpha_{ij} + \gamma_{jt} + u_{ijt+k}$, $k = -4, ..., 0, ..., 4$ for log product introduction N in (a) and quality-adjusted product introduction qN in (b). Our benchmark quality measure is defined in Section 3.2.1. The main explanatory variable in (a) and (b) is a dummy equal to one after the firm's first patent in a product category and log number of patent applications in (c) and (d). The inverse hyperbolic sine transformation is used for logs. The vertical bands represent +/- 1.65 × st. error of each point estimate. Standard errors are clustered at the firm × category level.

We find strong evidence supporting this point when we employ independent proxies for product-related and process-related patents drawn from claims texts as in Bena and Simintzi (2017). We find that the coefficient on product-related patents is essentially the same as our benchmark coefficient, while process-related patents are entirely unrelated to measures of product introduction (Section B.2 and Table A5 in the Appendix).

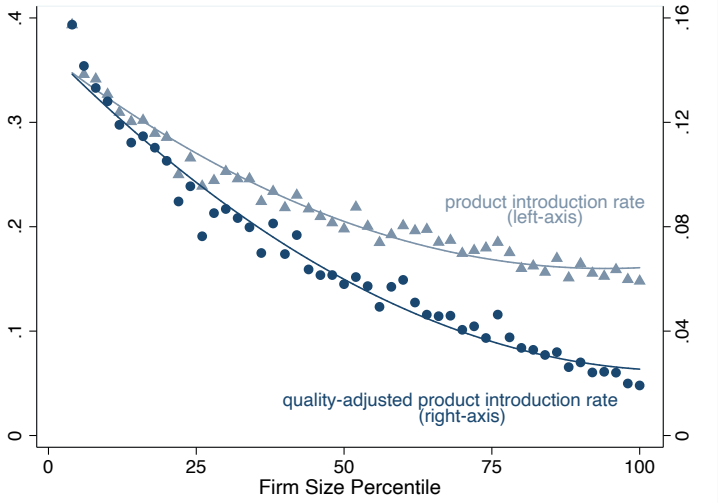
4.2 Strategic Patenting

Our conceptual framework provides a set of empirical moments to test the occurrence of strategic patenting – patents that deter innovation without leading to product introductions. First, we show that the degree of complementary between product introduction and patenting activities reduces with firm size. Through the lens of the model (Propositions 2 and 2*), this is a key moment reflecting the occurrence of strategic patenting. Second, consistent with Proposition 3, we show that patent revenue premium, not accounted for by quantity and quality of product introduction, is higher for larger firms. Finally, consistent with Proposition 4, we show that patents filed by market leaders are associated with a decline in creative destruction, unlike those filed by smaller firms.

4.2.1 Product Introduction and Patenting by Firm Size

We start by exploring the key moments that identifies strategic patenting among market leaders – how the relationship between patenting and product introduction varies with the

Figure 4: Product Introduction Rates by Firm Size



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

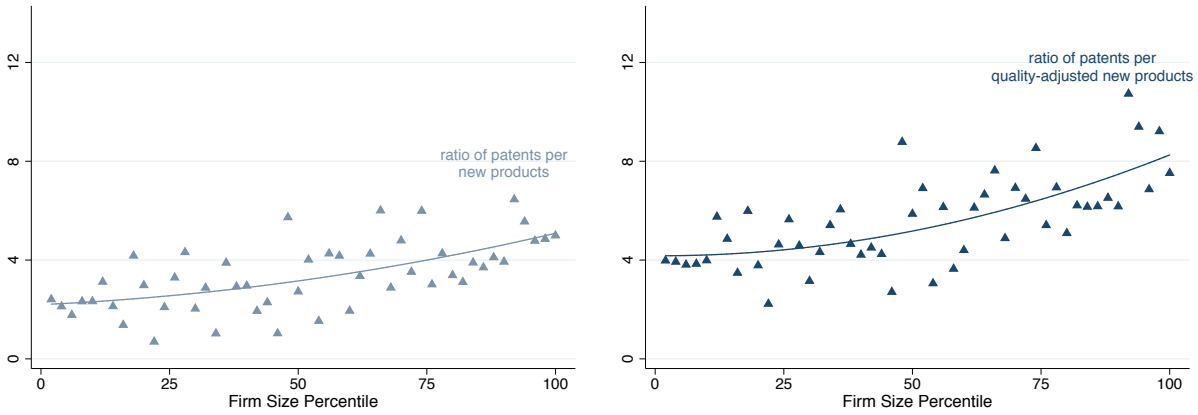
incumbent's size. Figure 4 plots the average product introduction rate—the ratio of product introduction to a firm's stock of existing products—for firms across product categories. Larger firms (within product categories) have lower product innovation rates. On average, firms in the top sales quintile have annual introduction rates of about 16%, while firms in the bottom quintile have rates twice as large. Larger firms do not compensate for this decline in the rate of new product introduction with innovations of higher quality. On average, firms in the top sales quintile have quality-adjusted product introduction rates of 3%, while firms in the bottom sales quintile have rates four times larger. The fact that the quality-adjusted introduction rate declines more steeply than the simple product introduction rate indicates that, on average, new products introduced by larger firms are more likely to represent incremental improvements over existing products and are thus less novel.[30]

Figure 5 shows that larger firms, on average, file more patents for each new product introduced.[31] Note that this higher intensity of patenting activity relative to the number of new products introduced is not explained by the possibility that larger firms introduce fewer but more novel products: as one can see, after we adjust for the quality of new products,

[30]Figure A14 in the Appendix confirms similar patterns using alternative metrics. This result is consistent with previous evidence from patent statistics in Akcigit and Kerr (2018).

[31]If we do not scale our measures of patenting, results are even starker: the unconditional probability of patenting and the total number of patents filed by large firms are much higher than they are for small firms (see Figure A15 in the Appendix).

Figure 5: Patents per New Products, by Size



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

small and large firms' introduction rates diverge even more.

The patterns above are suggestive that the relationship between product introduction and patenting changes with firm size. We now, more formally, explore how the relationship varies with firm size by estimating equation (8) for small and large firms. As before, we control for time × product category and firm × product category fixed effects to ensure that our results are not driven by potential confounders such as differences in patentability across firms and product categories. Table A12 reports the estimated coefficients for firms in different size groups. In line with the results discussed above, we estimate an average coefficient of 0.038 (column "All"). The table shows that the relationship between patents and product innovation weakens with a firm's size: larger firms in the top sales quintile have an elasticity twice as small as that of firms in the bottom sales quintile (0.030 versus 0.059). The results are similar for the quality-adjusted new products shown in the last three columns of the table.

We further evaluate if the relationship between patents and product innovation weakens with firm's size for the different types of product categories independently of their heterogeneity. Figure A17 in Appendix D shows the elasticities by firm size for food and non-food categories. Although the magnitudes of the elasticities vary, we see that the strength of the patents-to-product innovation relationship declines with firm size for both types of products. We interpret this result as suggestive that a weaker association between product introduc-

Table 5: Product Introduction and Patenting: by Size

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

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

tion and patenting activities may prevail for different types of sectors, independently of their degree of patentability or other product characteristics.

4.2.2 Patent Revenue Premium by Firm Size

We next estimate the patent revenue premium that is not accounted for by quantity and quality of product introduction by firms. If market leaders rely more on strategic patenting to reduce competitive pressure, we expect them to enjoy a larger patent revenue premium above and beyond the one explained by the product introduction. To evaluate this in the data, we estimate the relationship between patents and sales growth, conditional on measures of a firm's product innovation. Specifically, we estimate:

$$\Delta \log \text{Sales}_{ijt} = \psi \log P_{ijt-1} + \rho \log N_{ijt} + \theta_{ij} + \gamma_{jt} + \varepsilon_{ijt} \tag{9}$$

where the dependent variable is the logarithm of the change in sales at time t, $\log P_{ijt-1}$ is the total number of patent applications until time $t-1$, and $\log N_{ijt}$ is the number of new products introduced at t (we also use the quality-adjusted product introduction $\log qN_{ijt}$). Our coefficient of interest is ψ, which measures a percent change in sales growth associated with a percent change in patents, conditional on the effect coming from product introduction.

Table A13 shows the results for all firms and for firms grouped according to size. Overall, we find a positive significant relationship between patents and future growth in sales even after controlling for product innovation (columns "All"). This finding suggests that holding an additional patent allows firms to increase their sales even after accounting for the

Table 6: Patenting and Sales Growth

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

Notes: The table presents estimated outcomes of changes in log sales at the firm × category level as a function of the log number of patent applications by until time $t-1$ and the log number of new products introduced at time t (or quality-adjusted new products), by size groups. We use the firm × product category data set for the period 2007–2015, restricting the analysis to observations with sales above $1,000. For each firm × product category, we define size based on average sales over the sample period. The "All" column uses data for all sizes. The "Small" column is restricted to the lowest size quintile. "Large" is restricted to the top size quintile. The inverse hyperbolic sine transformation is used for logarithms.

increase in sales that results from new product offerings. Importantly, this effect is highly heterogeneous across firm sizes. For firms in the bottom sales quintile (columns "Small"), there is no statistical association between patents and sales growth after we control for product introduction. However, for firms in the top quintile (columns "Large"), we find that an increase in total patent applications has a significant positive association with sales growth above and beyond its effect through product introduction. Note also that the direct impact of product innovation on sales growth (coefficients on Log N(t) and Log qN(t)) decreases as firms increase in size. Hence, by splitting the sample into small and large firms, we learn that, while both patents and new products are associated with increased future sales, the conditional impact of new products is more important for smaller firms, while the impact of patents is quantitatively more important for larger firms.

We further explore if this difference in the association between sales and patenting between small and large firms operates through changes in prices and/or changes in quantities sold. Table A6 in Appendix shows the results using specification (9) for prices and quantities. We find that the additional incremental revenue from strategic patenting for large firms occurs both because of higher quantities sold and higher prices. These patterns are consistent with the idea that patents of larger firms are relatively more likely to reduce creative destruction as they discourage competitors from introducing new products, allowing large firms to serve a larger market and charge higher prices.

4.2.3 Creative Destruction

Next, we use our data set to investigate whether patents by market leaders are more likely associated with declining product introduction on the part of their competitors, who we will refer to, for simplicity, as market followers. We start by identifying the market leader in each category as the firm with the highest sales in that category and the followers as the remaining firms operating in that market.[32] Then, for each year t and product category j, we compute the total number of new products introduced by the leader N_{jt}^{L} and by its followers N_{jt}^{F} in t, and we compute the total numbers of patent applications introduced by the leader P_{jt}^{L} and by its followers P_{jt}^{F} until t. We evaluate how product introduction by followers responds to patenting (and product introduction) of the leaders using the following specification:

$$\log N_{jt}^{\mathrm{F}} = \eta^{\mathrm{F}} \log P_{jt-1}^{\mathrm{L}} + \alpha^{\mathrm{F}} \log N_{jt-1}^{\mathrm{L}} + \theta_j^{\mathrm{F}} + \gamma_t^{\mathrm{F}} + \varepsilon_{jt}^{\mathrm{F}}, \tag{10}$$

where η^{F} is our coefficient of interest, measuring the association of patents of leaders with the product introduction by followers. We control for $\ln N_{jt-1}^{\mathrm{L}}$ to ensure that the relationship between leaders' patents and followers' product introduction is not driven by possible direct interactions between the leader's and followers' product offerings (such as learning from new products on the market).[33] We also include time- and category-fixed effects to control for time trends and differences in the intensities of patenting and product innovation across product categories. Likewise, we estimate a symmetric regression that estimates the relationship between leaders' innovation and the followers' patenting:

$$\log N_{jt}^{\mathrm{L}} = \eta^{\mathrm{L}} \log P_{jt-1}^{\mathrm{F}} + \alpha^{\mathrm{L}} \log N_{jt-1}^{\mathrm{F}} + \theta_j^{\mathrm{L}} + \gamma_t^{\mathrm{L}} + \varepsilon_{jt}^{\mathrm{L}} \tag{11}$$

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

Table 7 presents the estimated coefficients. Column 1 shows that product introduction by followers is negatively correlated with the size of the leader's patent portfolio. This result suggests that followers reduce the introduction of new products in categories where the leader intensifies its patenting efforts. In column 2, we also control for total sales of the market to account for potential shifts over time in the importance of different types of products. In turn, columns 3 and 4 show that product innovation by leaders is not related to the followers' patenting activity. Hence, while patents can be thought of as a protective tool

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

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

Table 7: Patenting of Market Leaders and Followers

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

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

used to hinder competition in the product market, our results indicate that this hypothesis is likely to apply when patents are in the hands of large market leaders. These findings support that market leaders are more likely to use strategic patenting, and by accumulating patents, they reduce competition and protect their market shares.

4.3 Discussion and Additional Evidence

In this section, we perform a battery of robustness checks to show that our empirical findings are not driven by alternative explanations and examine additional evidence from patent characteristics.

4.3.1 Alternative Explanations

First, we examine whether the relatively weaker association between patent and product introduction of larger firms could be explained by differences in data coverage across firms of different sizes. The patent data set covers the entire portfolio of patents of firms, but our product data set does not cover products outside the consumer goods sector. This could, in turn, result in lower rates of attribution of patents to new products for firms that also produce in other sectors. To allay this concern, we note that our empirical findings are based on our Match 2 algorithm, which filters out patents that are not related to the consumer goods sector. Moreover, we obtain similar results when we use a sample of firms that sell

exclusively CPG products (see Section B.1 for details on the construction of this sample).

Second, we also assess the possibility that our textual analysis of patents artificially weakens the relationship between patents and products of large firms. The potential concern is that the text of patents filed by firms of different sizes may be systematically different and that our matching algorithm could be less effective in ascribing patents filed by larger firms to specific product categories. To better gauge this concern, we study textual characteristics of patents such as patent document length, number of unique words, textual diversity, and relative entropy of patents' word distribution, and we evaluate whether these characteristics vary systematically across firms of different sizes within the same product categories. We do not find systematic differences in these metrics of textual characteristics of patents filed by large and small firms. Furthermore, we also do not observe significant differences in the share of matched patents across firm size (see Figure A18 in Appendix for details). Overall, our exercises indicate that differences in data coverage and in the properties of the matching algorithm seem unlikely to explain the weaker association between patents and product innovation for large firms.

We also study whether the relatively weaker association between patents and product innovation of larger firms could be explained by economic factors other than the strategic use of patents. One possibility is that larger firms shift toward process patents as they grow (Cohen and Klepper, 1996), thereby weakening the relationship between total patents and product introduction for larger firms. Using alternative proxies for product-related and process-related patents constructed in Appendix Section B.2, we find no systematic relationship between the share of this independent measure of process patents in the firm's portfolio and the firm's size (Figure E.1 in the Appendix). Moreover, if cost reductions due to process innovations are reflected in lower subsequent prices, we can test whether future price changes of larger firms react to patents more. However, we do not find such a relationship in the data.[34]

Lastly, we evaluate whether large firms do more experimental research and take more time to commercialize their inventions, which could explain a weaker association between patents and products. Using the dynamic specifications of equation (8), we do not find evidence that patents held by larger firms are associated with product innovation with a longer delay.

[34] A similar concern is that larger firms may file patents not to commercialize products but to license those patents to other firms. We assign patents to the patent holders and do not have information on temporary licensing agreements for all patents (such data do not exist). Moreover, prior work (Fosfuri, 2004; Gambardella, Giuri and Luzzi, 2006) suggests that, if anything, larger firms are less likely to license their patents out.

4.3.2 Additional Evidence from Patent Characteristics

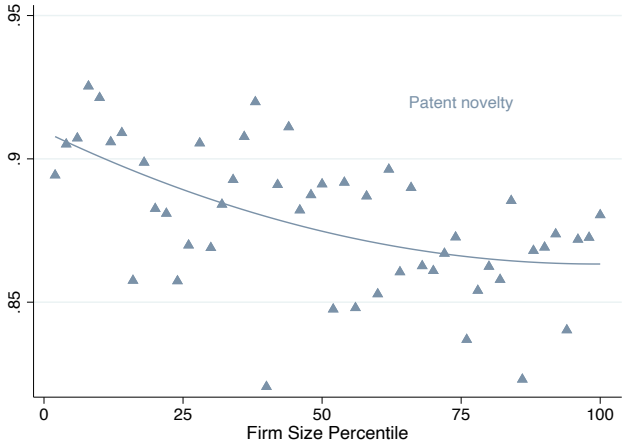
There are multiple accounts of strategic patenting practices in the law and economics innovation literature. For example, the term "sleeping patent" refers to patenting ideas that are not commercialized by the patentee or licensed to another for use (Torrisi, Gambardella, Giuri, Harhoff, Hoisl and Mariani, 2016), the term "patent thickets" refers to filing numerous patents on the same product (Shapiro, 2000; Hall, Graevenitz and Helmers, 2021), and the term "patent evergreening" refers to filing for new patents on secondary features of a particular product as earlier patents expire to extend patent exclusivity past the original twenty-year term, among others (Righi and Simcoe, 2020). There are also some examples from case studies and media reports. A good example for our sector is the patent for P&G Swiffer Wet Jet mops. Instead of patenting the features of the invention, P&G patented the specific functionality of the disposable cloths. The original patent and the more than 80 follow-up patents have made it difficult for competitors to enter the market. Indeed, during our sample period, generic sweeper mops were basically absent. P&G has a market share of approximately 95% in sweeper mops—much larger than that of leaders in other categories, whose share is 40-50% on average. It is also not unusual to find examples of firms with "sleeping" patents that deter entry but do not lead to products on the market. An example is Driscoll's, which controls a third of the U.S. berry market. The company invests heavily in a breeding program to develop new berry varieties which it patents but often does not commercialize. Driscoll's has one of the highest ratios of patents per new product in our data. The company has also recently been involved in several lawsuits to protect its patent portfolio from potential competitors.[35]

While these practices are often discussed, it is not possible to identify in a systematic way for every patent if it does not lead to product innovation but hinders competition. Our benchmark approach to identify strategic patenting relied on using patents-to-products data guided by our model. We now complement our analysis by comparing patents along various dimensions that likely capture the types of patents described above. Note that these measures can only be computed for a smaller set of firms that have multiple patents.

We start by computing measures of patent text similarity to previous patents of the firm to evaluate the average degree of novelty of the firm's patents. Figure 6 plots the average novelty for firms of different sizes. We find that larger firms' patents exhibit lower text-based patent novelty and thus are more similar to their previous patents. Columns 1 and 2 of Table 8 quantify these patterns in the regression analysis. For the cross-section of firms, we look at the average patent characteristics of firms as a function of the leader

[35] *"How Driscoll's Reinvented the Strawberry"*, The New Yorker, August 21, 2017.

Figure 6: Text-Based Patent Novelty by Firm Size



Notes: This figure plots the relationship between patent novelty and the relative size of the firm, defined by the firm's sales. We use the firm × product category level data for the period 2007–2015, restricting the analysis to observations with sales above $1,000. For each firm × product category, we compute average sales and patent novelty, which is a patent-level metric between zero and one. The text-based patent novelty measure is equal to one minus the text similarity between a given patent and its most similar predecessor within a firm (patent text similarity is computed using the same methods outlined in Section B.4.3); Within each product category, we assign firms to 50 bins for average sales and plot patent novelty for each bin. Each dot/triangle plots the averages after weighting each product category by its importance in the whole sector, as measured by the share of sales accounted for by the category.

Table 8: Patent Characteristics: Leaders vs Followers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Text Novelty		Citations		Share Cited Others		Share Litigated	
Leader	-0.042***	-0.033***	-0.538**	-0.675***	-0.062**	-0.076**	0.004	0.008**
	(0.008)	(0.008)	(0.247)	(0.257)	(0.027)	(0.030)	(0.005)	(0.003)
Observations	2,633	1,830	2,179	1,526	1,569	1,095	2,179	1,526
R-squared	0.015	0.405	0.002	0.247	0.003	0.335	0.000	0.094
Controls	N	Y	N	Y	N	Y	N	Y

Notes: The table compares the average patent characteristics of leaders and other firms. "Leader" is a dummy equal to one for the firm with the highest sales in a given category; the followers are defined as the rest of the firms in the categories. *Text Novelty* is a patent-level metric between zero and one. The text-based patent novelty measure is equal to one minus the text similarity between a given patent and its most similar predecessor within a firm (patent text similarity is computed using the same methods outlined in Section B.4.3); *Citations* is the mean number of citations received by patents in the first five years after the application; *Share Cited Others* is the share of forward citations accounted for by citations from other firms different from the patent owner; *Share Litigated* is the share of patents involved in litigation. Data on litigations come from the USPTO Patent Litigation Dataset. Due to truncation concerns, we provide statistics for the patents filed in 2005. Controls include the total number of firms in a category and fixed effects at the firm level.

dummy, without controls (odd columns) and controlling for the total number of firms in the product category and firm fixed effects for multi-category firms (even columns). The estimation indicates that patent text novelty is less likely among market leaders.

We explore several other metrics that are characteristic of strategic patenting practices. Columns 3 and 4 of Table 8 show that large firm's patents accumulate fewer forward citations

Table 9: Model Parameters

Parameter		Identification	Value
r	Interest rate	External calibration	0.04
p	Arrival rate of entrants	Direct data	0.103
ε	Patent protection	Direct data	0.25
λ	Innovation step	Internal calibration	0.09
γ	Elasticity of revenue to quality	Internal calibration	0.88
$\tilde{c}_m$	Cost of commercialization	Internal calibration	2.94
$\tilde{c}_p$	Cost of patenting	Internal calibration	54.99

Notes: Table presents all parameters, and the procedure to parameterize its value. Appendix A.4 provides details.

per patent, hence lead to lower follow-on research. Columns 5 and 6 indicate that the share of forward citations by other firms is lower for market leaders, hence follow-on research is even relatively less likely to occur by potential competitors. Finally, as the last two columns suggest, large firms' patents are somewhat more likely to involve litigation.[36]

5 Frequency and Implications of Strategic Patenting

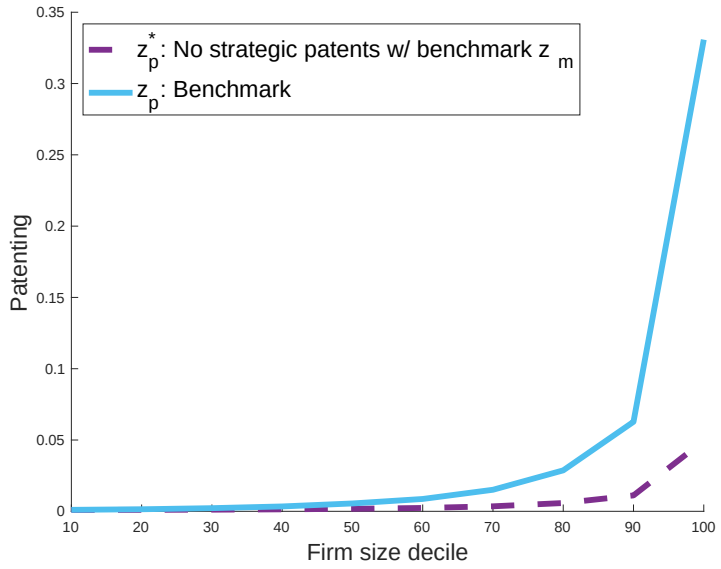
Strategic patenting can have negative consequences for creative destruction and overall innovation in the economy. In this section, we calibrate the model parameters and use the structure of the model to build counterfactuals that allow us to provide rough calculations of the frequency of strategic patenting and its implications for innovation and creative destruction.[37]

We start by parameterizing the model using our data set with measures of product introduction and patenting. The model has seven structural parameters: $\{r, p, \varepsilon, \lambda, \gamma, \tilde{c}_m, \tilde{c}_p\}$. Table 9 reports the calibrated parameters. We set the interest rate to $r = 0.04$, and quantify the parameters governing creative destruction—the arrival rate of entrants (p) and the patent protection step (ε)—using data on sales growth across firms depending on their patenting and product introduction status. In the model, if firms do not innovate or patent, they face a negative expected revenue growth from creative destruction equal to $\log(1-p)$, and in the data, this implies a $p = 0.103$. This decline in sales is attenuated if a firm holds a patent. Conditional on not innovating, the effect of patenting on firm growth is $\log \frac{1-p(1-\varepsilon)}{1-p}$, which we estimate in the data to imply $\varepsilon = 0.25$. For the remaining four parameters, we calibrate the model to match the innovation rate, patents per innovation, and sales growth

[36] Litigation is a rare event – the average share of litigated patents is 0.04, so the coefficient on the leader dummy is economically large.

[37] A comprehensive optimal policy and welfare analysis is beyond the scope of this paper.

Figure 7: Patenting choices with and without strategic patents



Notes: We use the model parameters in Table 9 to compute the benchmark and counterfactual patenting rates. We evaluate the counterfactual z_p^* by holding product introduction rates as in the benchmark economy. Appendix A.4 provides details.

of firms in different firm size percentiles. Intuitively, $\tilde{c}_m$, and $\tilde{c}_p$ affect the levels of innovation and patenting, λ determines the average growth when the firm innovates, and the curvature parameter γ affects how this growth varies with firm size. Despite the stylized nature of our model and few parameters, the model is able to match the innovation rate, patenting intensity, and growth of firms over the size distribution quite well.[38] Appendix A.4 provides more details about the calibration procedure and shows the resulting match between the model and the data.

Using these parameters, we now quantify the frequency of strategic patents and their aggregate impact on creative destruction. We compute the share of strategic patents by comparing the rate of patenting z_p in the benchmark economy (value function in (4)) with the rate of patenting in the counterfactual economy z_p^* (value function in (5)) in which firms can only patent if they simultaneously introduce a new product improvement in the market, while keeping product introduction rates the same as those in the benchmark economy. The gap between z_p^* and z_p captures the additional amount of patenting induced by the possibility of filing patents without product introduction. Figure 7 plots the estimated optimal patenting z_p from the benchmark economy, and the optimal z_p^* from the counterfactual

[38] In addition, model-implied elasticities calculated from equation (12) are consistent in magnitude with elasticities estimated in Table A11.

Table 10: Implications of Strategic Patenting for Innovation: Counterfactual Patent Regimes

	Without Patents	Without Strategic	Benchmark
Innovation			
Incumbents' product introduction (z_m)	0.1582	0.2276	0.1624
Creative destruction (τ)	0.1030	0.1023	0.0999
Patenting			
Incumbents' patenting (z_p)	0	0.0284	0.1190

Notes: We compute the optimal innovation rates using the estimates for the model parameters in Table 9. The table Product introduction, patenting, and creative destruction with large incumbents.

economy, fixing the introduction rate z_m to be the same as that in the benchmark model. We plot these values for firm sizes corresponding to the empirical deciles of the firm size distribution in product categories. Our quantification implies that about 62% of patent filings by an average firm and more than 80% of filings by large firms in the 90th-95th percentiles are strategic. This high share of strategic patents of large firms implies that the creative destruction is 2.5% lower in the benchmark economy with patenting rate z_p relative to the counterfactual economy where incumbents patent at rate z_p^*.

Finally, we quantify the overall implications of strategic patents by comparing the equilibrium innovation in three economies with different patent regimes: the benchmark economy with no restrictions on the nature of patents, the counterfactual with no strategic patents, and a counterfactual without any kind of patents. Table 10 shows the optimal rates of incumbents' product introduction and creative destruction under the three regimes.

The comparison of the first and second columns illustrates the importance of patent protection for providing innovation incentives. While creative destruction is lower relative to the no-patents economy, incumbents choose to introduce new products at a much higher rate because their new products can be protected by patents. Hence, on net, the availability of patent protection has a positive effect on aggregate product introduction in the economy. This comparison is reminiscent of standard arguments for the innovation-enhancing role of patents in the economy (Bryan and Williams, 2021), and in the model, this result comes about from the complementarity between product introduction and patenting (see Proposition 1).

The comparison of the second and third columns shows the additional effect of allowing for strategic patenting. In the benchmark economy, the innovation rate is lower for two reasons. First, in the benchmark economy, the patenting rate is higher, leading to a reduction in the creative destruction rate by 2.4% and hence the lower entry of new products from

entrants.[39] The second channel is less obvious but is quantitatively more important. In the benchmark economy, incumbents' product introduction rate is lower because incumbents can rely on strategic patenting to maintain their market shares, and thus the incentives for additional product introduction are lower.

Overall, the comparison of these three patent regimes indicates that while, in principle, patent protection can incentivize ex-ante product innovation by firms, the possibility of obtaining patent protection without product commercialization can reverse the benefits of the patent system for the economy as a whole.

6 Conclusion

We study the relationship between patents and product innovation by developing a simple framework that separates the decision to upgrade a product and the decision to patent it. Introducing higher-quality products increases a firm's profit, while patenting decreases the firm's chances of being displaced by competitors. Patenting and product innovation are complementary activities because an increase in the effort to patent the idea increases the probability that the firm will maintain its dominance and thus increases the present discounted value of the additional profit from upgrading the product. Crucially, the degree of complementarity is lower for large firms because of the incremental benefits of strategic patents—those that deter competition without leading to product commercialization by the patentee.

The key contribution of our paper is the construction of a unique data set linking patents to products by firms in the consumer goods sector. We rely on textual analysis of patent documents and product descriptions from Wikipedia articles to assign sets of patents to sets of products. We find that patents filed by market leaders are less likely to lead to product introduction. Instead, we find strong evidence that the role of many patents by large firms is to deter future product introduction of competitors and to protect sales of their existing products. Our results indicate that although, on average, patents capture product innovation in the market, patent-based measures of innovation distort the differences between firms of different sizes.

We then perform several counterfactual experiments under different patenting regimes to determine the prevalenceof strategic patenting. We find that most patents by large firms are strategic. The pervasiveness of strategic patents in the economy both hinders innovation

[39]Notice that the difference with the previous exercise in Figure 7 is that now we solve for optimal product introduction and patenting choices in both economies, while before we fixed z_m in the counterfactual economy to be equal to the benchmark value.

by market leaders and decreases creative destruction. Our results illustrate that while, in principle, patent protection can incentivize ex-ante product innovation, the possibility of obtaining patent protection without product commercialization can reverse the benefits of the patent system. They also highlight the possibility that adjustments to the patent system, such as granting patent protection only to patents leading to innovations in the market, can greatly improve creative destruction in the economy.

Lastly, although our estimates point out inefficiencies arising from the existence of strategic patenting, a comprehensive optimal policy and welfare analysis are beyond the scope of this paper. An interesting research agenda going forward is to perform these analyses in a rich structural model, accurately quantifying forces such as the potential knowledge diffusion from patents, the endogenous arrival rate of entrants' ideas as well as the arrival of incumbents' ideas, and the feedback loop between innovation policies and the firm size distribution. These are important considerations that future research should incorporate to quantify the implications of intellectual property rights policies for firm dynamics, aggregate innovation, and growth.

References

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

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

Akcigit, Ufuk and Sina T Ates, "Ten facts on declining business dynamism and lessons from endogenous growth theory," *American Economic Journal: Macroeconomics*, 2021, *13* (1), 257–98.

_ **and** _ , "What Happened to U.S. Business Dynamism?," *Journal of Political Economy*, 2023.

_ **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," *Econometrica*, 2023.

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 Life Cycle, Learning, and Nominal Shocks," *The Review of Economic Studies*, 2022, *89* (6), 2992–3054.

_ , Munseob Lee, and Sara Moreira, "Innovation and Product Reallocation in the Great Recession," *Journal of Monetary Economics*, 2018, *93*, 1–20.

_ , _ , and _ , "The Life Cycle of Products: Evidence and Implications," 2022.

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," *The Quarterly Journal of Economics*, 2020, *135* (2), 645–709.

Baslandze, Salome, "Barriers to Creative Destruction: Large Firms and Nonproductive Strategies," FRB Atlanta Working Paper 2021-23, Federal Reserve Bank of Atlanta September 2021.

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

Bloom, Nicholas, Charles I Jones, John Van Reenen, and Michael Webb, "Are Ideas Getting Harder to Find?," *American Economic Review*, 2020, *110* (4), 1104–44.

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

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

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

Bryan, Kevin A and Heidi L Williams, "Innovation: Market Failures and Public Policies," in "Handbook of Industrial Organization," Vol. 5, Elsevier, 2021, pp. 281–388.

Cavenaile, Laurent and Pau Roldan-Blanco, "Advertising, Innovation, and Economic Growth," *American Economic Journal: Macroeconomics*, July 2021, *13* (3), 251–303.

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

Cohen, Wesley M. and Steven Klepper, "Firm Size and the Nature of Innovation Within Industries: The Case of Process and Product R&D," *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)," NBER Working Paper 2000.

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

Cunningham, Colleen, Florian Ederer, and Song Ma, "Killer Acquisitions," *Journal of Political Economy*, 2021, *129* (3), 000–000.

de Rassenfosse, Gaétan, "Notice Failure Revisited: Evidence on the Use of Virtual Patent Marking," NBER Working Paper 2018.

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.

Federico, Pasquale J, "Operation of the Patent Act of 1790," *J. Pat. Off. Soc'y*, 1936, *18*, 237.

Fosfuri, Andrea, "The Licensing Dilemma: Understanding the Determinants of the Rate of Licensing," *Strategic Management Journal*, 03 2004, *27*.

Galasso, Alberto and Mark Schankerman, " Patents and Cumulative Innovation: Causal Evidence from the Courts," *The Quarterly Journal of Economics*, 2 2015, *130* (1), 317–369.

Gambardella, Alfonso, Paola Giuri, and Alessandra Luzzi, "The Market for Patents in Europe," LEM Papers Series 2006/04, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy February 2006.

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

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

Goolsbee, Austan D. and Peter J. Klenow, "Internet Rising, Prices Falling: Measuring Inflation in a World of E-Commerce," *AEA Papers and Proceedings*, May 2018, *108*, 488–92.

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

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

Gutiérrez, Germán and Thomas Philippon, "Declining Competition and Investment in the U.S.," NBER Working Paper 23583 July 2017.

_ **and** _ , "The Failure of Free Entry," NBER Working Paper 26001 2019.

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

Hall, Bronwyn H, Georg von Graevenitz, and Christian Helmers, "Technology Entry in the Presence of Patent Thickets," *Oxford Economic Papers*, 2021, *73* (2), 903–926.

Hegde, Deepak, Kyle Herkenhoff, and Chenqi Zhu, "Patent Publication and Innovation," NBER Working Paper 29770 2022.

Hoberg, Gerard and Gordon Phillips, "Text-based Network Industries and Endogenous Product Differentiation," *Journal of Political Economy*, 2016, *124* (5), 1423–1465.

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

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

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy, "Measuring Technological Innovation over the Long Run," *American Economic Review: Insights*, September 2021, *3* (3), 303–20.

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

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

Loecker, Jan De, Jan Eeckhout, and Gabriel Unger, "The Rise of Market Power and the Macroeconomic Implications," *The Quarterly Journal of Economics*, 2020, *135* (2), 561–644.

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.

Righi, Cesare and Timothy Simcoe, "Patenting Inventions or Inventing Patents? Strategic Use of Continuations at the USPTO," NBER Working Paper 2020.

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

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

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

Torrisi, Salvatore, Alfonso Gambardella, Paola Giuri, Dietmar Harhoff, Karin Hoisl, and Myriam Mariani, "Used, Blocking and Sleeping patents: Empirical Evidence from a Large-scale Inventor Survey," *Research policy*, 2016, *45* (7), 1374–1385.

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

Webb, Michael, "The Impact of Artificial Intelligence on the Labor Market," *Available at SSRN 3482150*, 2019.

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

Younge, Kenneth A and Jeffrey M Kuhn, "Patent-to-patent Similarity: a Vector Space Model," *Available at SSRN 2709238*, 2016.

Patents to Products: Product Innovation and Firm Dynamics

Appendix For Online Publication

A Theoretical Appendix

A.1 Derivation of Rates of Creative Destruction

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

- If the incumbent firm neither patents the idea nor introduces a new product, creative destruction happens at a rate

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

  Hence, any product of higher quality introduced by an entrant will capture the full market.

- If the incumbent firm does not patent but successfully commercializes the product, creative destruction happens at a rate

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

  Again, any product of higher quality introduced by an entrant will capture the market.

- If the incumbent firm patents but does not introduce new products:

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

  Although higher quality products by entrants can still win the market, now entrants' innovation needs to be sufficiently large to also withstand the legal protection from the incumbent's patent.

- Similarly, if the incumbent firm patents and also introduces new products:

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

A.2 Proofs of Propositions

In what follows, to simplify the derivations, we adopt the following notation:

$$(q+\lambda)^{\gamma} - q^{\gamma} \equiv Q_{\lambda},$$
$$\frac{p\varepsilon}{(r+p(1-\varepsilon))(r+p)} \equiv P_{\varepsilon}.$$

Some properties that will be useful for the subsequent proofs are: $P_{\varepsilon} > 0, Q_{\lambda} > 0, \partial Q_{\lambda}/\partial q < 0$ for $\gamma < 1$.

With this notation, some of the main value differences can be written as:

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

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

In addition, we normalize costs such that $\tilde{c}_m \equiv c_m/\pi$ and $\tilde{c}_p \equiv c_p/\pi$.

In the main text, we define the value function $\mathbb{V}(q)$ to be inclusive of the costs $c_m(z_m)$ and $c_p(z_p)$. However, when discussing complementarity, it is useful to characterize things purely in terms of benefits. For that reason we define

$$U(q, z_m, z_p) \equiv z_m z_p V^{11}(q) + z_m(1-z_p)V^{10}(q) + (1-z_m)z_p V^{01}(q) + (1-z_m)(1-z_p)V^{00}(q)$$

We the define complementarity between z_m and z_p to be

$$\frac{\partial z_m}{\partial z_p} \equiv \frac{\partial^2 U}{\partial z_p \partial z_m} > 0$$

In a fairly general setting, this type of benefit-side complementarity will imply that there will be positive co-movement between z_m and z_p in response to some external forcing, such as q. To formalizing this notion, let such a factor be denoted θ and consider the firm's optimization problem

$$\max_{z_m, z_p} = \{U(z_m, z_p, \theta) - c_m(z_m) - c_p(z_p)\}$$

This has associated first order conditions $U_m = c_m'(z_m)$ and $U_p = c_p'(z_p)$. Using the implicit

function theorem we can find the differential effect of a change in θ to be

$$\frac{\partial z_m}{\partial \theta} = \left[\frac{-1}{U_{mm} - c_m''(z_m)}\right]\left(\frac{\partial z_p}{\partial \theta}U_{mp} + U_{m\theta}\right) \equiv A_m\left(\frac{\partial z_p}{\partial \theta}U_{mp} + U_{m\theta}\right)$$
$$\frac{\partial z_p}{\partial \theta} = \left[\frac{-1}{U_{pp} - c_p''(z_p)}\right]\left(\frac{\partial z_m}{\partial \theta}U_{mp} + U_{p\theta}\right) \equiv A_p\left(\frac{\partial z_m}{\partial \theta}U_{mp} + U_{p\theta}\right)$$

where $A_m > 0$ and $A_p > 0$ are abbreviated for simplicity. Solving these two equations we then arrive at

$$\frac{\partial z_m}{\partial \theta} = \frac{A_m A_p U_{p\theta} U_{mp} + U_{m\theta}}{1 - A_m A_p U_{mp}^2} \quad \text{and} \quad \frac{\partial z_p}{\partial \theta} = \frac{A_m A_p U_{m\theta} U_{mp} + U_{p\theta}}{1 - A_m A_p U_{mp}^2}$$

The ratio of these two expressions then yields the co-movement of z_m and z_p, namely

$$\frac{\partial z_m/\partial \theta}{\partial z_p/\partial \theta} = \frac{A_m A_p U_{p\theta} U_{mp} + U_{m\theta}}{A_m A_p U_{m\theta} U_{mp} + U_{p\theta}}$$

So long as $U_{mp} > 0$ and $U_{m\theta}$ and $U_{p\theta}$ are both either positive or negative, we can ensure that this has a positive sign.

Proof of Propositions 1 and 1'.

To show complementarity, we need to show that the value function is supermodular in z_p and z_m. The degree of complementarity between patenting and product introduction is

$$\begin{aligned}\frac{\partial z_m}{\partial z_p} \equiv \frac{\partial^2 U}{\partial z_p \partial z_m} &= V^{11}(q) - V^{01}(q) - V^{10}(q) + V^{00}(q) \\ &= \pi Q_\lambda P_\varepsilon > 0,\end{aligned} \tag{12}$$

which proves Proposition 1. Taking derivative of (12) with respect to ε shows that $\frac{\partial^2 z_m}{\partial z_p \partial \varepsilon} > 0$, proving Proposition 1′.

Proof of Proposition 2.
Taking derivative of (12) with respect to q gives:

$$\frac{\partial^2 z_m}{\partial z_p \partial q} = \pi P_\varepsilon \frac{\partial Q_\lambda}{\partial q} < 0 \tag{13}$$

when $\gamma < 1$.

Proof of Proposition 2.*
The value function in (5) implies the following degree of complementarity:

$$\frac{\partial z_m^*}{\partial z_p^*} = V^{11}(q) - V^{10}(q) = \pi(q+\lambda)^\gamma P_\varepsilon > 0. \tag{14}$$

Taking derivative of (14) with respect to q gives:

$$\frac{\partial^2 z_m^*}{\partial z_p^* \partial q} = \gamma \pi P_\varepsilon (q+\lambda)^{\gamma-1} > 0.$$

Proof of Proposition 3.

$$\frac{\partial (V^{01} - V^{00})}{\partial q} = \pi P_\varepsilon \frac{\partial q^\gamma}{\partial q} > 0.$$

Proof of Proposition 4.

Creative destruction rate is $\tau = z_p(q)p(1-\varepsilon) + (1 - z_p(q))p$. Showing that τ increases with firm size is equivalent to showing that $\frac{\partial z_p}{\partial q} > 0$. Next, we will find the conditions when this holds.

We first derive $z_p(q)$. The first-order conditions from optimizing (4) are

$$\tilde{c}_p z_p = q^\gamma P_\varepsilon + z_m Q_\lambda P_\varepsilon, \tag{15}$$

$$\tilde{c}_m z_m = Q_\lambda \frac{1}{r+p} + z_p Q_\lambda P_\varepsilon. \tag{16}$$

From plugging (15) into equation (16), we get:

$$\begin{aligned} \tilde{c}_p z_p &= q^\gamma P_\varepsilon + \left(Q_\lambda \frac{1}{r+p} + z_p Q_\lambda P_\varepsilon \right) Q_\lambda \frac{P_\varepsilon}{\tilde{c}_m} \\ z_p &= P_\varepsilon \frac{q^\gamma + Q_\lambda^2 \frac{1}{\tilde{c}_m (r+p)}}{\tilde{c}_p - Q_\lambda^2 \frac{P_\varepsilon^2}{\tilde{c}_m}} \end{aligned} \tag{17}$$

Notice that since z_p is the probability, the denominator has to be positive, and the whole ratio is bounded by one. Hence, two conditions are satisfied:

$$\begin{aligned} &\textit{Cond. 1}: && \tilde{c}_p - Q_\lambda^2 \frac{P_\varepsilon^2}{\tilde{c}_m} > 0 \\ &\textit{Cond. 2}: && \tilde{c}_p - Q_\lambda^2 \frac{P_\varepsilon^2}{\tilde{c}_m} > P_\varepsilon \left[q^\gamma + \frac{Q_\lambda^2}{\tilde{c}_m (r+p)} \right] \end{aligned}$$

Now, consider the sign of the derivative:

$$\text{sign}\left[\frac{\partial z_p}{\partial q}\right] = \text{sign}\left[\left(\gamma q^{\gamma-1} + \frac{2Q_\lambda Q'_\lambda}{\tilde{c}_m(r+p)}\right)(\tilde{c}_p - Q_\lambda^2 \frac{P_\varepsilon^2}{c_m}) + \left(q^\gamma + \frac{Q_\lambda^2}{\tilde{c}_m(r+p)}\right) 2Q_\lambda Q'_\lambda \frac{P_\varepsilon^2}{\tilde{c}_m}\right]$$

Now, denoting the above term in the bracket on the right hand side by A and using *Cond. 2*, we get:

$$\begin{aligned} A &> \left(\gamma q^{\gamma-1} + \frac{2Q_\lambda Q'_\lambda}{\tilde{c}_m(r+p)}\right) P_\varepsilon \left[q^\gamma + \frac{Q_\lambda^2}{\tilde{c}_m(r+p)}\right] + \left(q^\gamma + \frac{Q_\lambda^2}{\tilde{c}_m(r+p)}\right) 2Q_\lambda Q'_\lambda \frac{P_\varepsilon^2}{\tilde{c}_m} \\ &= \left[q^\gamma + \frac{Q_\lambda^2}{\tilde{c}_m(r+p)}\right] P_\varepsilon \left(\gamma q^{\gamma-1} + 2Q_\lambda Q'_\lambda (\frac{1}{\tilde{c}_m(r+p)} + \frac{P_\varepsilon}{\tilde{c}_m})\right) \end{aligned}$$

Hence, a sufficient condition for z_p to increase with firm size is $\gamma q^{\gamma-1} + 2Q_\lambda Q'_\lambda (\frac{1}{\tilde{c}_m(r+p)} + \frac{P_\varepsilon}{\tilde{c}_m}) > 0$. Simplifying, we get:

$$\begin{aligned} q^{\gamma-1} &> \frac{2}{\tilde{c}_m(r+p)}[(q+\lambda)^\gamma - q^\gamma][q^{\gamma-1} - (q+\lambda)^{\gamma-1}]\left(1 + \frac{p\varepsilon}{r+p-p\varepsilon}\right), \text{or} \\ &\frac{2}{\tilde{c}_m(r+p)}[(q+\lambda)^\gamma - q^\gamma][1 - (1+\lambda/q)^{\gamma-1}]\left(1 + \frac{p\varepsilon}{r+p-p\varepsilon}\right) < 1 \end{aligned}$$

Notice that $[(q+\lambda)^\gamma - q^\gamma][1-(1+\lambda/q)^{\gamma-1}] \to 0$ when $q \to \infty$, and it is bounded by the maximum value of λ^γ at $q = 0$. So, the left-hand side expression is bounded above by $\frac{2\lambda^\gamma}{\tilde{c}_m(r+p-p\varepsilon)}$. Hence, the following sufficient condition guarantees that $\frac{\partial z_p}{\partial q} > 0$:

$$2\lambda^\gamma < \tilde{c}_m(r+p-p\varepsilon) \tag{18}$$

This condition essentially requires the discounted cost-adjusted return from product introduction to be sufficiently low. This sufficient condition turns out to be numerically quite strict, and in our numerical trials, z_p always increases in q, even if parameters violate the condition (18).

A.3 Alternative Modeling Assumptions

Our model abstracts from various features relevant for the innovation and patenting choices of firms to keep the analysis tractable and intuitive. We discuss some alternative modeling assumptions and show how the main qualitative implication of the model regarding innovation and patenting choices with respect to firm size carry over under these assumptions.

Size-Dependent Cost– Our model assumes the same innovation and patenting cost irrespective of firm size: $\partial c_m/\partial q = 0, \partial c_p/\partial q = 0$. However, the main features of the model

can be easily generalized to include size-dependent costs.[40] Larger firms might be more experienced in patent filings, have dedicated legal lawyers, and have more resources to engage in litigation if need be, effectively leading to lower patenting costs, so that $\partial c_p/\partial q < 0$. In such a case though, large firms' incentives for strategic patent filings would be even bigger than what we find, strengthening the model implications.

Likewise, one could think that because of more effective production process and management, supply chains, and marketing experience, product introduction cost parameter c_m might be lower for larger firms. In this case, as can be seen from equations (12) and (13), Proposition 2 depends on the values of certain parameters. We can show that if the decline in the elasticity of the return to product introduction with respect to size is larger than the decline in the elasticity of the cost parameter c_m: $\frac{\partial Q_\lambda}{\partial q}/Q_\lambda < \frac{\partial \tilde{c}_m}{\partial q}/\tilde{c}_m$, then the same results hold. Alternatively, the relationship between patents and product introduction would be predicted to be stronger with firm size. However, this prediction would contradict what we find empirically.

Proportional Innovation Step Size– The model assumes a constant innovation step size λ irrespective of firm size. Many endogenous growth models generate continual growth by having innovation step sizes that are proportional, rather than absolute. In addition, these models also scale up R&D costs with some term proportional to overall output (such as a wage). We can adopt these assumptions in our setting as well, as a point of reference. We find that both patenting and innovation rates (z_p and z_m) are invariant to firm size, but product introduction and patenting are complements in this setting, too.

Consider that product innovation now results in a size λq innovation, and innovation costs scale up with the base value level V^{00}. We will use the notation $c_k = V^{00}\tilde{c}_k$ and $Q_\lambda = (1+\lambda)^\gamma - 1$. Then we can express the optimal choices of the firm as

$$z_m = \frac{P_\varepsilon(\tilde{c}_p + Q_\lambda^2)}{\tilde{c}_m\tilde{c}_p - Q_\lambda^2 P_\varepsilon^2} \quad \text{and} \quad z_p = \frac{Q_\lambda(\tilde{c}_m + P_\varepsilon^2)}{\tilde{c}_m\tilde{c}_p - Q_\lambda^2 P_\varepsilon^2}$$

These probabilities do not depend on q, though they do require the condition $\tilde{c}_m\tilde{c}_p > Q_\lambda^2 P_\varepsilon^2$ to be well defined. With these, we can do some simple comparative statics. The direct effect of costs is unambiguously negative, meaning $\frac{\partial z_k}{\partial \tilde{c}_k} < 0$. Meanwhile, the cross derivatives can shed some light on the nature of complementarity here. These evaluate to

$$\frac{\partial z_m}{\partial \tilde{c}_p} = -\frac{P_\varepsilon Q_\lambda^2(1+P_\varepsilon)}{(\tilde{c}_m\tilde{c}_p - Q_\lambda^2 P_\varepsilon^2)^2} < 0 \quad \text{and} \quad \frac{\partial z_p}{\partial \tilde{c}_m} = -\frac{Q_\lambda P_\varepsilon^2(1+Q_\lambda)}{(\tilde{c}_m\tilde{c}_p - Q_\lambda^2 P_\varepsilon^2)^2} < 0$$

So from a cost perspective, there is complementarity. From a benefits perspective (Q_k), we can also see complementarity, as we have both $\partial z_m/\partial P_\varepsilon > 0$ and $\partial z_m/\partial Q_\lambda > 0$, as well as $\partial z_p/\partial Q_\lambda > 0$ and $\partial z_p/\partial P_\varepsilon > 0$.

[40] In the literature, there are multiple examples of size-dependent costs, e.g. Breuer, Leuz and Vanhaverbeke (2021).

A.4 Model Quantification and Counterfactuals: Details

Direct data We quantify the parameters governing creative destruction—the arrival rate of entrants (p) and the patent protection step (ε)—using data on sales growth across firms depending on their patenting and product introduction status.

In the model, if firms do not innovate or patent, they face a negative expected sales growth from creative destruction equal to $\log(1-p)$. We use our baseline firm $\times$ category $\times$ year data set for the period 2007-2015 and compute the median log revenue change for firm $\times$ category that did not file a patent nor introduced new products. This statistic is equal to -0.109, implying $p = 1 - \exp(-0.109) = 0.103$.

In the model, conditional on not innovating, the effect of patenting on firm revenue growth is $\log \frac{1-p(1-\varepsilon)}{1-p}$. We use our baseline firm $\times$ category $\times$ year data set for the period 2007-2015 and estimate the coefficient from regressing log revenue change on log patents and firm-category and category-year fixed effects, conditional on no product introduction. The estimated coefficient is 0.0324 (se=0.023), which results in $\varepsilon = 0.25$.[41]

Internal Calibration To jointly estimate λ, γ, $\tilde{c}_m$, and $\tilde{c}_p$, we match the following set of three moments in the data: innovation rate, patents per innovation, and revenue growth of firms in different firm size percentiles.

We use the firm $\times$ product category level data for the period 2007–2015, restricting the analysis to observations with sales above \$1,000 and firms between the 10th and 90th percentiles. For each firm $\times$ product category, we compute innovation rate (new products over its existing products), patents per innovation (log number of patent applications over new products), and revenue growth (change in revenue as in DHS, i.e., $2(y_t - y_{t-1})/(y_t + y_{t-1})$), for innovators relative to non-innovators) and aggregate to the average within a size bin.

To map data percentiles to quality levels in the model, we first normalize the quality q of the average firm in each product category to one. Then, using equation (2), we obtain $q = (\frac{Rev}{\overline{Rev}})^{1/\gamma}$, where we measure Rev as the firm's revenue per products in a product category, and $\overline{Rev}$ denotes the average revenue in the product category. This gives us a mapping between the average normalized revenue in various percentiles to their corresponding levels of q in the model.

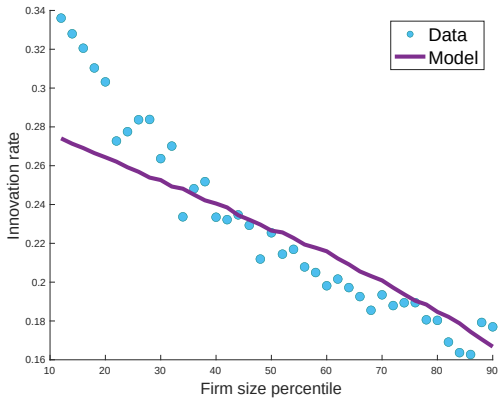
We minimize distance between the model innovation $z_m(q)$, log patent to innovation ratio $\log \frac{z_p(q_i)}{z_m(q_i)}$, and log revenue growth $\Delta \log R$ to their corresponding data moments. Hence, we minimize the following objective function:

$$\min_{\lambda,\gamma,\tilde{c}_m,\tilde{c}_p} \sum_{i=10}^{90} \left[5(z_m(q_i) - m_i^1)^2 + (\log \frac{z_p(q_i)}{z_m(q_i)} - m_i^2)^2 + (\Delta \log R(q_i) - m_i^3)^2\right]^{1/2},$$

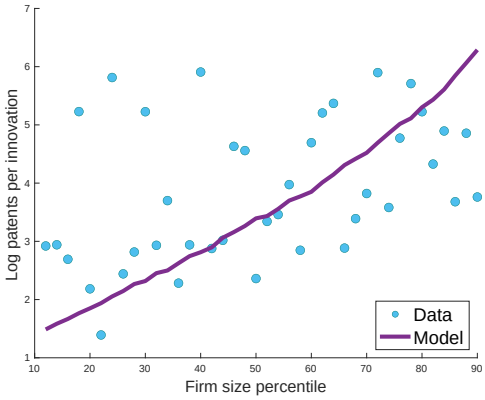
where m^1 is firm's product introduction rate, m^2 is the log number of patent applications over new products, and m^3 is revenue growth. We weight the innovation moment heavier

[41] We calculate a move from zero patents to one patent which implies (with inverse hyperbolic sine transformation for log patents) multiplying the regression coefficient by 0.88. So, $\log \frac{1-p(1-\varepsilon)}{1-p} = 0.88 \times 0.0324$.

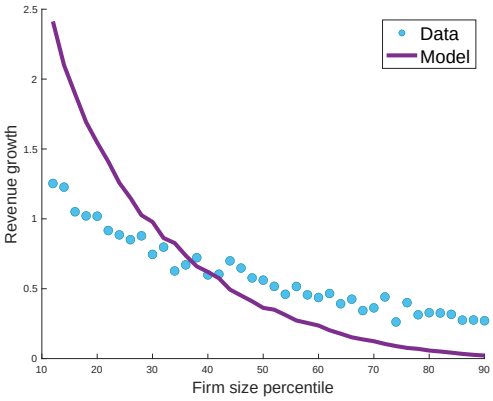
Figure A1: Model vs Data



(a) Innovation



(b) Patents per Innovation



(c) Revenue Growth

than other moments to overrepresent the cleanest moment not influenced by a potential noise in the matching procedure with patents.

Figure A.4 shows the resulting match between data and the model. With only four parameters, we can capture the three important moments over the whole size distribution quite well.

B Additional Data Information

B.1 Product Data

Coverage.— The main advantage of the RMS data set is its size and coverage. Overall, the RMS data consists of more than 100 billion unique sales observations at the week × store × UPC level. The data set comprises around 12 billion transactions per year which are worth $220 billion dollars on average. Over our sample period, 2006-2015, the total sales across all retail establishments are worth approximately $2 trillion and represent 53% of all sales in grocery stores, 55% in drug stores, 32% in mass merchandisers, 2% in convenience stores, and 1% in liquor stores. A key distinctive feature of this database is that the collection points include more than 40,000 distinct stores from around 90 retail chains, across 371 metropolitan statistical areas (MSAs) and 2,500 counties. We keep a balanced set of stores throughout the entire period under the analysis.

Because of its size, the data provides good coverage of the universe of products in the consumer goods sector. Our assessment is based on three considerations. First, comparisons with other scanner data sets reveals that Nielsen RMS covers more product introductions and provides more accurate information on product entry time. Argente et al. (2018) compares Nielsen RMS with other scanner data sets collected at the store level and shows that Nielsen RMS covers a much wider range of products and stores. In comparison to scanner data sets collected at the household level, Nielsen RMS also has a wider range of products because it reflects the universe of all transactions for the categories it covers, as opposed to the purchases made by a sample of households. For example, Nielsen Homescan covers less than 60% of the products the Nielsen RMS covers in a given year.

Second, while the data only covers sales in traditional retail channels and not e-commerce, we do not expect this to substantially affect the total level of innovations in the sector. Between 2000 and 2014, the fraction of all retail sales accounted for by e-commerce went from 0.9 to 6.4 percent, according to figures from the US Census Bureau (Hortaçsu and Syverson, 2015). Thus, during our sample period, online commerce is still a small part of retail activity and will affect innovation numbers by firms that only sell online.

Finally, the data covers sales in food and non-food categories (health and beauty aids, non-food grocery, and general merchandise). However, because the data set has higher coverage of grocery stores, food categories have relatively higher coverage than some general merchandise categories (see, for example, Jaravel (2019) for a thorough comparison of Nielsen RMS and Homescan with the Consumer Expenditure Survey). We assess the impact of this differential coverage of product categories on our measures of product innovation by comparing product introduction rates in our data with those in Nielsen Homescan and other sources (e.g. Goolsbee and Klenow, 2018). We do not find a significant association between sales coverage and the differences in product introductions between data sets across various product categories. Nevertheless, throughout the paper we evaluate the robustness of the results when we keep only products that have high coverage.

Nielsen Product Classification System.— The data is organized into detailed product modules that are aggregated into product groups. The product groups are then grouped into ten major departments. These departments are: Health and Beauty Aids, General

Merchandise, Dry Grocery (e.g., baby food, canned vegetables), Frozen Foods, Dairy, Deli, Packaged Meat, Fresh Produce, Non-Food Grocery, and Alcohol. For example, a 31-ounce bag of Tide Pods has UPC 037000930389, is produced by Procter & Gamble, and belongs to the product module "Detergent-Packaged" in product group "Detergent," which belongs to the "Non-Food Grocery" department. The product group "Detergent" includes several product modules, including automatic dishwasher compounds, detergents heavy duty liquid, detergents light duty, detergents packaged, dishwasher rinsing aids, and packaged soap.

Over time, Nielsen expanded coverage of certain product modules (for instance, some in-store food goods), but we keep a consistent set of product modules that are available throughout the period. This leaves us with 10 departments, 114 product groups and 1,070 modules.

Defining a Product.— Defining products by their UPCs has some important advantages. First, UPCs are by design unique to every product: changes in any attribute of a good (e.g. forms, sizes, package, formula) result in a new UPC.This offers a unique opportunity for economists to identify products at the finest level of disaggregation.

Second, UPCs are so widespread that our data is likely to cover all products sold in the consumer goods sector. Producers have a strong incentive to purchase UPCs for all products that have more than a trivial amount of sales because the codes are inexpensive, and they allow sellers to access stores with scanners and internet sales.

For each product in a year, we define its sales as the total sales across all stores and weeks in the year. Likewise, quantity is defined as total quantities sold across all stores and weeks in the year. Price is defined by the ratio of revenue to quantity, which is equivalent to the quantity-weighted average price.[42] To minimize concerns about potential measurement error caused by Nielsen's treatment of private-label products to protect the identity of the retailers, we exclude all private-label goods from the data.

Assigning Products to Firms.— Nielsen RMS data does not include information on manufacturing firms. However, products can be linked with firms using information obtained from the GS1 US Data Hub. In order to issue a UPC, firms must first obtain a GS1 company prefix. The prefix is a five- to ten-digit number that identifies firms in their products' UPCs. Argente et al. (2018) provide more details on how to use a subset of the product UPCs to link producers with products.

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

Constructing a sample of CPG-only firms.— Any firm that produces at least one product in the Nielsen RMS data is included in our analysis. We refer to these as CPG firms. However, some of these CPG firms also produce products outside the CPG sector (e.g. Toshiba, Samsung, Whirlpool), while others produce mostly products included in the

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

Nielsen RMS data (e.g. Procter & Gamble, Kimberly Clark, Kraft). A part of our analysis is focused on identifying a sample of firms that are solely in the CPG sector. Inspired by Hoberg and Phillips (2016), we use the firm's 10-K reports, which are available from Compustat. The 10-K is a comprehensive summary of a firm's performance that must be submitted annually to the Securities and Exchange Commission, in addition to the annual report. It includes an overview of the firm's main operations, including its products and services. We manually classify each business line reported on the 10-K's into CPG/non-CPG comparing its description with the description of Nielsen modules, and classify each publicly traded CPG firm into CPG-only if the majority of the firm's sales results from CPG business lines. We matched 270 publicly traded companies over our sample period; we classify 23% of them as CPG-only firms.

B.2 Patent Data

Data Details.— Unlike other standard patent data sources such as NBER patent data (Hall et al., 2001) and the data from the Harvard Dataverse Network (Lai et al., 2014), we make use of all patents published in the USPTO, including non-granted patent applications. Using all patent applications, as opposed to just granted applications, offers us two advantages. First, since patents are usually granted with a lag of roughly two years, the more recent years of the sample suffer from severe truncation. Looking at all patent applications alleviates this problem. Second, we can then differentiate between patents that are granted, pending, or abandoned. We use this as one of the patent quality measures, as discussed below. Adding non-granted patent information increases the number of patents in our sample by 1.7 million.

Assigning Patents to Firms.— We begin by selecting all patents that have a valid assignee name.[43] We assign patents to their most recent assignee(s). For this assignment, we use the *current* assignee variable from the USPTO (as of 2017 – our patent data vintage). The current assignee variable is missing for some of the patents included in our sample. In such a case, we start with the name of the original assignee and leverage the USPTO reassignment data to track any change of patent ownership due to a patent sale or firm reorganizations. To further track patent ownership through corporate reorganizations, we rely on Thomson Reuters Mergers & Acquisition data. Our underlying assumption is that patent ownership is transferred to the acquiring firm in case of corporate reorganization. Thomson Reuters M&A provides complete coverage of global mergers and acquisitions activity, including more than 300,000 US-target transactions, since 1970. The data covers mergers of equals, leveraged buyouts, tender offers, reverse takeovers, divestitures, stake purchases, spinoffs , and repurchases. It also provides detailed information about the target, the acquirer, and the terms of the deal. This comprehensiveness is particularly important given that firms that appear both in Nielsen data and USPTO are most likely large firms that undergo many corporate reorganizations.

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

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

The formulaic nature of claims gives us an opportunity to create the following simple classification. We say the claim is a process claim if the claim text starts with "method" phrases ("Method for", "Method of", "Method in", "Method define", and the like) or "process" phrases ("Process for", "Process according", "Process in", and the like). Then, as a baseline, we classify a patent as a process patent if the main (usually, the first) claim of the patent is a process claim. The patent is a product patent if it is either a design patent or a non-process utility patent. In the latter case, claims often start with words like "Apparatus", "Device", and the like. According to this definition, up to 70% of patents are product-related patents. We also tested an alternative definition that defines process patents based on the criteria that the share of process claims is larger than 50%. These two measures are highly correlated (0.74) and our results based on the baseline variable are robust to this alternative definition.

B.3 Algorithm of *Match 1*

Firm Name Cleaning Algorithm.— We assign each company name (from Nielsen or USPTO data) to a unique company identifier using the following procedure.

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

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

Step 2. In addition to the extensive cleaning in Step 1, we take advantage of a "dictionary" that resulted from a large effort undertaken within the NBER Patent Data Project. After manual checks and searches of various company directories to identify name misspellings and various company reorganizations, the NBER files provide a mapping between patent assignee names and unique company identifiers (*pdpass*). Although this data is based

on the assignees of granted patents before 2006, we use this mapping as a "dictionary" that we use in conjunction with our results from Step 1. This helps us leverage both our algorithm from Step 1 and the NBER *pdpass* information, combining the strengths of each method to create new unique company identifiers.

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

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

B.4 Algorithms of *Match 2*

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

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

i) Parsing Methods

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

ii) Lemmatizer Methods

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

iii) Word Vector Normalization

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

We then use a word-based weighting scheme called total-frequency-inverse-document-frequency (tf-idf) to account for the fact that more common words tend to be less important

and vice versa (Aizawa, 2003). A number of possible functional forms could be used here, but we choose the commonly used sublinear form

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

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

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

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

B.4.2 *Step 1: Defining Product Categories*

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

Step 1.a - Collect Representative Documents

For each low-level product classification from Nielsen (1,070 modules), we explored different sources of text that might allow us to characterize the modules. First, we studied sources of text within Nielsen. For example, we explored the use of product attributes from each UPC, and we found that while informative, some characteristics are shared and not sufficiently different. Second, we explored sources of data outside Nielsen, like dictionaries and various websites. After many manual checks, we decided to use Wikipedia pages, and based on module descriptions, we manually selected the closest Wikipedia articles for each product module.[44]

The main advantage of using Wikipedia entries is that they often include technical descriptions that use words that also appear in patent texts and are comprehensive enough to cover all modules. The use of Wikipedia text to encode textual knowledge is already common in the machine learning literature. For instance, two of the most advanced word embeddings currently available, BERT (Google, devlin2018bert, devlin2018bert) and fastText (Facebook, joulin2017bag, joulin2017bag), use the entire Wikipedia corpus for training purposes, in a addition to large corpora of text from books and websites. While there are a number of papers in the economics literature that study Wikipedia, we are unaware of any such usage as a direct input into a separate analysis.

[44]To ensure the best selection of these articles, we cross-checked the results after assigning this task to five independent readers. Some examples of our article selection are: Wikipedia articles titled "Humidifier" and "Dehumidifier" correspond to the "Humidifier and vaporizer appliance" product module; an article "Artificial nails" is assigned to the "False nails and nail decorations" module; articles "Soft drinks" and "Carbonated water" are assigned to the "Soft drinks- carbonated" module.

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

Step 1.b - Create Representative Word Vectors

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

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

Step 1.c - Cluster Analysis

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

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

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

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

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

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

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

these product modules at the same time.

Step 1.d - Creating Pseudo Product Categories

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

Step 1.e - Word Vectors for Product Categories

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

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

Step 2.1 - Collect Representative Documents for Patents

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

Step 2.2 - Create Representative Vectors for Patents

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

The patent corpus is on average shorter than the product category vectors. The average number of words per patent is 263 with standard deviation of 333 (in terms of unique words, we get mean 107 and standard deviation 93). The average number of words is about 7,200 per Wikipedia article, with a standard deviation of 6,500 (in terms of distinct words, the mean is 2,500 and standard deviation of about 2,000). We evaluated if there is a good overlap in the words used on longer documents to insure that there was not too much noise. About 50% of the words seen in our product category vectors show up in the patents somewhere.

Step 2.3 - Computing Similarity Scores Between Patents and Categories

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

vectors together yields the similarity score between two documents:

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

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

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

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

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

Step 3.1 - Threshold Similarity

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

$$\Theta_p = \{j \in \Omega \mid s_{jp} > 0.025 \vee \text{rank}(s_{jp}) \leq 5\} \tag{19}$$

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

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

$$G_p = \{j \in \Omega \mid p \text{ is patent of firm } i \ \wedge \ \sum_{t=2006}^{2015} \text{sales}_{ijt} > 0\}, \tag{20}$$

where $sales_{ijt}$ are the sales of firm i in product category j in year t. Note that this production condition, will exclude all pseudo-categories and product categories that the firm never produced from the set of potential matches.[45]

[45]This makes it clear that having pseudo-categories helps to filter out many patents of the firms who

Step 3.3 - Select the Maximum

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

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

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

$$j_p^* = \max_{j \in \Theta_p \wedge G_p} s_{jp} \tag{21}$$

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

B.5 Robustness and Match Validation

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

We manually checked many patent-to-products matches, and Table A1 lists some examples. The top 100 product categories sorted by their revenue and the largest firms selling in those categories are shown. For each firm, we then list an example of the highest-similarity patents in the corresponding product categories and their similarity scores. Comparing the titles of the patents and product categories, we see that product categories selected by our algorithm match the content of the patents well.

heavily produce non-CPG products. For example, some firms like Toshiba or Samsung produce small electronics in our data, however they hold large portfolios of patents related to computer hardware or other high-tech technologies that are not relevant for the consumer products sector that we are analyzing. For such patents, the set Θ_p often consists only of pseudo-modules that then are easily filtered out by condition (20).

Table A1: Patents with the Highest Similarity Score: Top Selling Firms by Categories

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

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

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

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

B.5.2 External Validation. Virtual Patent Markings

One of the important validation exercises for the patent-to-products match relies on external information. We use information from virtual patent markings which were introduced with the 2011 Leahy-Smith America Invents Act. Under that act, firms may give notice to the public that their product is patented. Recently, de Rassenfosse (2018) provides estimates of the adoption rate of virtual markings and studies factors that account for the likelihood of adoption. Overall, the adoption rate is relatively small and varies systematically with firm size. Indeed, our online searches showed that only a handful of the CPG firms in our sample used virtual patent markings.[46] This means that we cannot use patent markings to match patents to products for all firms in our data set. We can, however, use them as a useful validation exercise to compare the marking's product-patent matches with our algorithm.

To this end, we selected Procter & Gamble (P&G) and Kimberly-Clark (KC) for our validation exercise, as these are among the largest firms in our sample.[47] We start by parsing the product-patent links from the websites. In most cases the markings are associated with brands and not particular products. Hence, an important challenge lies in linking the listed brands on the websites with the brands in Nielsen. We use exact name matches, non-exact name matching, and extensive manual matching to determine the closest Nielsen brand equivalents. We then proceed to identify the product categories that include products of those brands. This parsing process allows us to obtain a mapping between patents and product categories that solely comes from the markings listed by P&G and KC markings.

For each patent, we then compare the matched product categories in our Match 2 data set with the product categories obtained from the virtual markings listed by P&G and KC (311 and 87, respectively).[48] We begin by testing information from the similarity scores. For each patent-product category pair from the virtual markings, we obtain a similarity rank that our algorithm assigns to this product category. For example, when the rank value is one, the product category in the virtual markings corresponds to our algorithm's highest top-1 similarity category. When it is two, the match was very close to the category from the markings, and so on, thus providing a notion of closeness between the algorithm-based and marking-based matches. The first plot in Figure B.5.2 plots the distribution of these ranks. The algorithm-based preferred (highest-similarity) product categories coincide most of the time with the patent-product category mapping we created based on virtual markings. For 69% of patents, and 79% of patents conditional on a match, the virtual marking product categories are ranked as one or two based on similarity scores.[49]

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

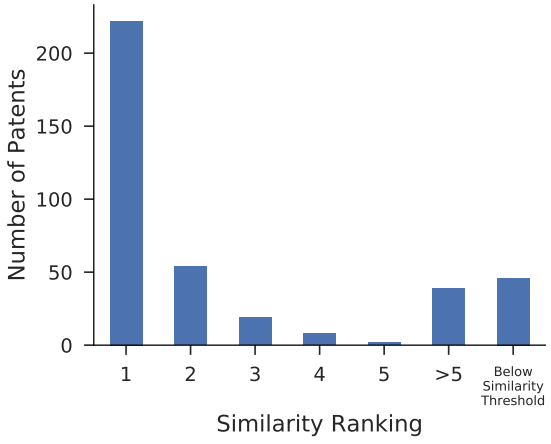
[47] We also found virtual markings are Clorox and Smuckers. However, because the products reported on their websites could not be mapped cleanly to our product categories, we did not analyze them.

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

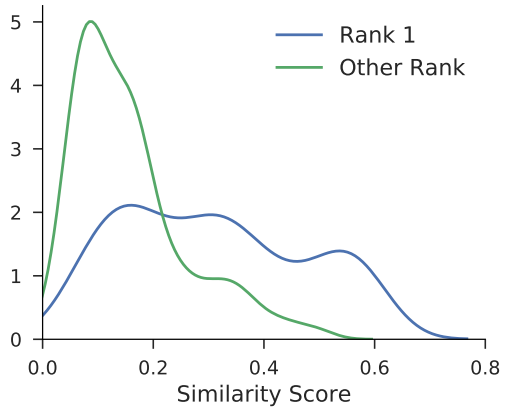
[49] Note that we cannot compare these numbers to 100% given that the ranking is unavoidably affected by some noise that comes from our manual mapping of the product listings on the websites to the notion of product categories in our data.

Figure A2: Virtual Patent Markings. P&G and KC Case Study

Distribution of similarity ranks for virtual markings



Distribution of similarity scores



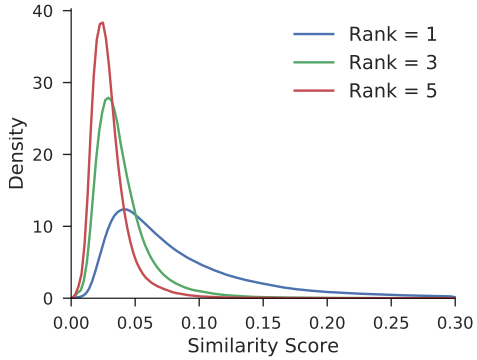
Notes: We use patent markings from P&G and KC. For each patent-product category pair from the virtual markings, we obtain a similarity rank that our algorithm assigned to this product category and show the distribution of ranks in the first graph. When the rank is one, the product category in the virtual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

Another way to visualize the accuracy of the match is to examine the distribution of similarities conditioning on whether the match was rank-1 (coinciding with the category from virtual markings) or a higher rank. If these two distributions were very similar, this would mean that even if the match is accurate, it is not very robust, as small elements of noise or bias could change the results of the match. In fact, as shown in the second plot of Figure B.5.2, these two distributions are quite distinct with the rank-1-match distribution weighted towards the right, meaning the results of the match should be rather robust.

B.5.3 Robustness of the Match. Patent Similarity with Top vs Lower-Rank Categories

As discussed, for our match, we pick product categories which have the highest similarity scores with patents. That is, we first pick the top five categories that have the highest similarity values with patents, and then we assign the top-similarity category conditional on a firm producing a product in that category. However, if the similarity scores for different categories are too close (either because the algorithm is not able to pick up the distinctions between documents or the categories are too finely defined) so that the algorithm cannot clearly differentiate between them, our choice of the top-rank match would not be robust to small perturbations of the algorithm or category clustering. To explore this issue, we plot the distribution of similarity scores of patents with different-rank product categories (Figure B.5.3). The rank-1 category is the category with the highest similarity score for a patent, and so on. We find that top-ranked categories have substantially different (shifted to the right) distributions than slightly lower-ranked categories, thus providing evidence of the robustness of the match. The patents' mean similarity score for rank-1 categories is 3 times higher than the mean similarity score for rank-5 categories.

Figure A3: Similarity Distribution by Rank

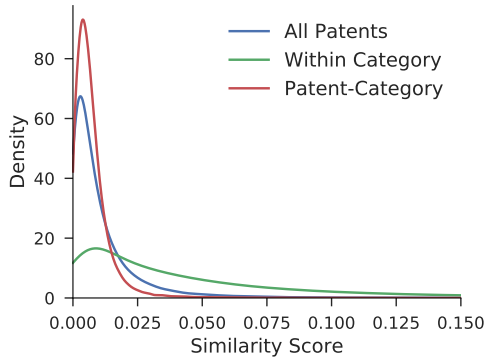


Notes: The figure shows similarity scores distribution of patents for different-rank product categories. Rank-1, Rank-3, and Rank-5 show similarities with categories ranked as the highest, rank-3, and rank-5 similarity categories.

B.5.4 Actual vs Placebo Match of Patents to Product Categories

We next verify that by grouping patents into distinct categories, we are indeed carving out well-defined neighborhoods in the technological space. We again employ word vectors to assess document similarity, but this time between pairs of patent texts. Specifically, we look at the distribution of similarity scores between pairs of patents classified into the same product category and compare this distribution to that of pairs of patents selected at random from the entire set of patents held by CPG firms. The similarity distribution based on this match looks very different from our placebo distribution as seen in Figure B.5.4. The patents' mean similarity score is 5.6 times higher if patents are assigned to the same product categories. In ordinal terms, the median within-category similarity lies at the 93rd percentile in the overall distribution.

Figure A4: Distribution of Pairwise Patent Similarities



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

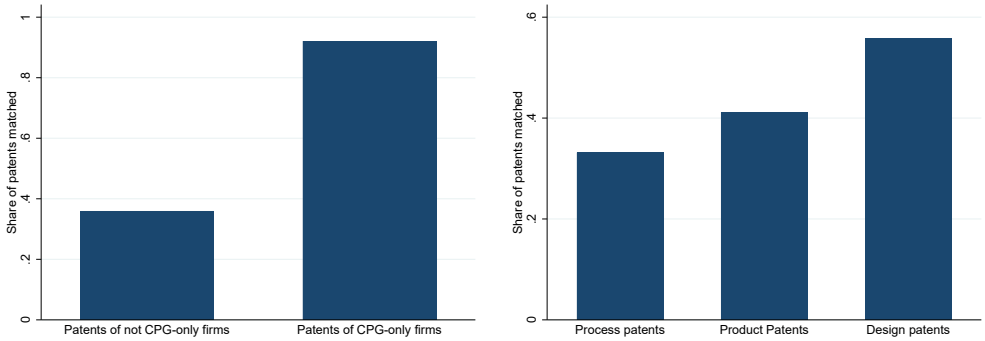
B.5.5 Validating Non-matches. CPG-only Firms and Product-Related Patents

Our Match 2 of patents at the firm × product category level would ideally filter out patents that are not related to the products in our data. Hence, correct non-matches would arise for the following two main reasons. First, a patent may relate to other non-CPG goods that the firm may be producing, which are not covered in our sample; and second, a patent may be a general process/method patent that does not relate to the products directly. We examine these possibilities.

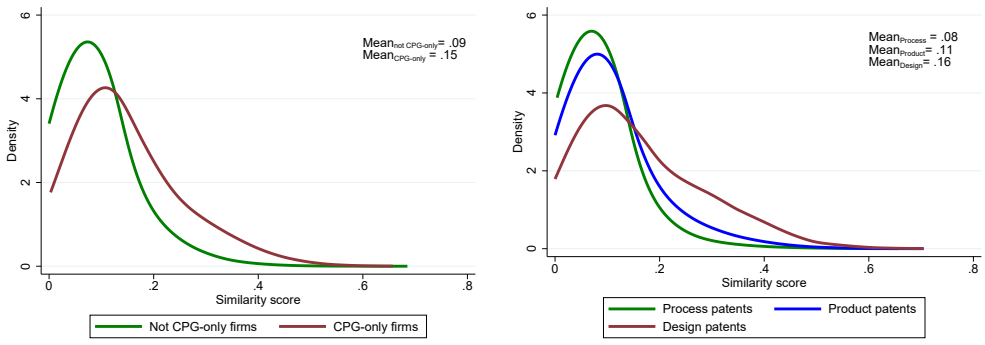
Panel (a) in Figure B.5.5 shows the share of patents that match to firms' product categories for a sample of firms that we can accurately identify as CPG-only firms and not CPG-only firms (see Appendix B.1 for details). Indeed, 92% of patents held by CPG-only firms match, while 36% of patents of not CPG-only firms match to our product categories. This result reassures us that our algorithm indeed picks the correct matches. As seen from Panel (b), the similarity scores for CPG-only firm patents are also significantly higher.

Figure A5: Match Validation. CPG-only Firms and Product-Related Patents

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



(b) Rank-1 similarity distribution for patents



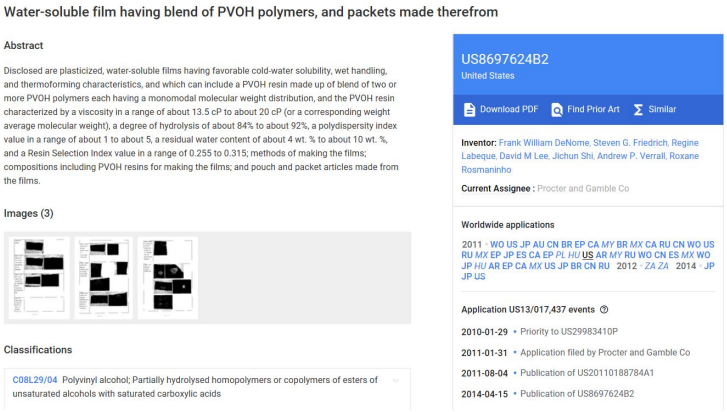
Notes: Panel (a) shows the share of patents that match with product categories in which firms ever sell a product. The left figure compares patents of the CPG-only and non-CPG-only firms, while the right figure compares process, product-related, and design patents. CPG-only firms and non-CPG-only firms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

Panel (a) also demonstrates that the share of patents that are matched is higher if the patent is more likely to be directly related to products. Using our proxies for process- and product-related patents (see Appendix B.2 for details) and considering design patents as

most directly related to products, we plot the share of all process, product, and design patents that are matched. The probability of a match increases along with the likelihood of a patent being related to a product, which is reassuring. Panel (b) also confirms that the similarity scores of product-related patents are much higher than the similarity scores of process patents.

B.6 Patents and Products in CPG. Examples.

Figure A6: Procter & Gamble

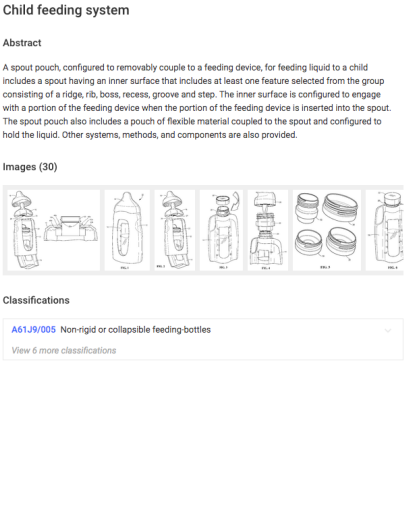


(a) Patent application in 2011



(b) The first Tide Pods in 2012

Figure A7: Example: Kiinde LLC

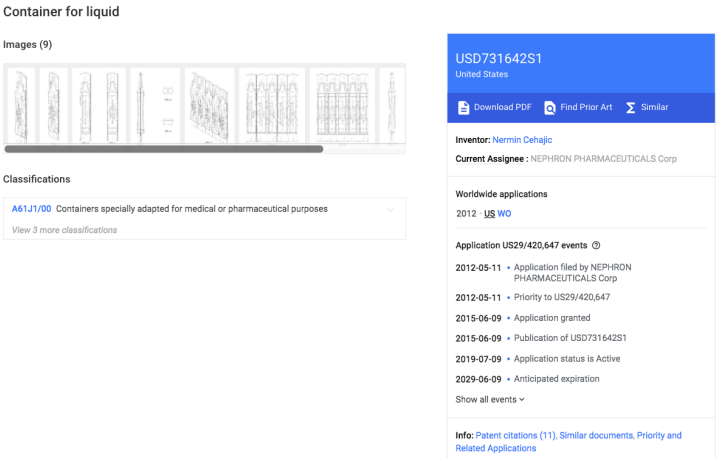


(a) Patent application in 2013



(b) Direct pump adapters introduced in 2014

Figure A8: Example: Nephron Pharma

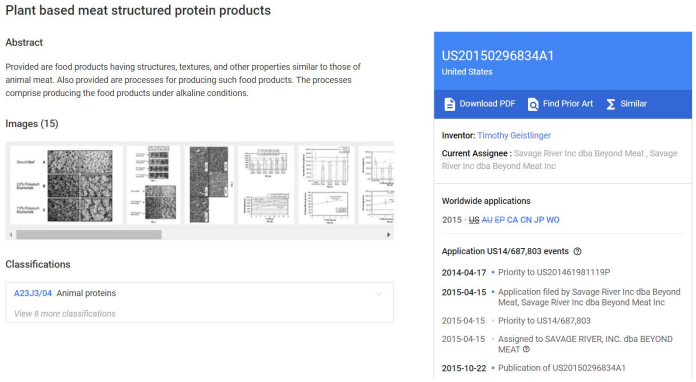


(a) Patent application in 2012



(b) Refill vials in 2012

Figure A9: Example: Beyond Meat Inc.

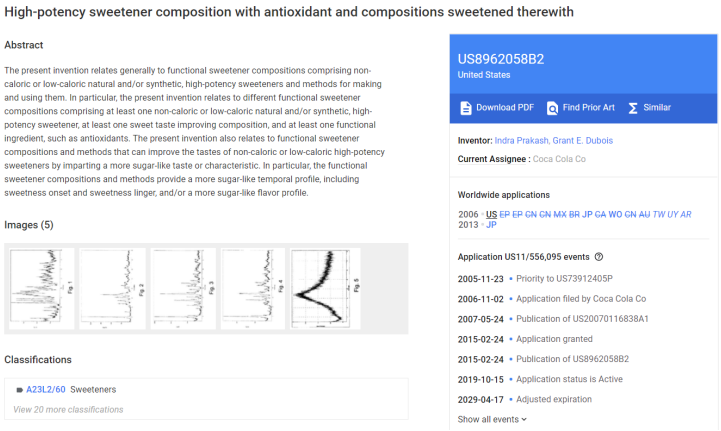


(a) Patent application in 2014



(b) The first simulated beef product in 2014

Figure A10: Example: Coca Cola Inc.



(a) Patent application in 2005



(b) The first sugar-free Coke Zero in 2005

B.7 Measuring Product Innovation

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

B.7.1 Novelty-Based Measures

Overview. — We define a product u in product category j as a vector of characteristics $V_u^j = \left[v_{u1}^j, v_{u2}^j, .., v_{uA^j}^j\right]$, where A^j denotes the number of attributes (e.g. color, formula, size) observed in product category j and v_{ia}^j represents a characteristic within an attribute (e.g. blue, red, green).[50] Let Ω_t^j contain the set of product characteristics for each product ever sold in product category j at time t, then the *novelty index* of product u in product category j, launched at time t is defined as follows:

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

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

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

For q, we estimate ω_a^j using hedonic price regressions in order to be able to quantify the importance of each attribute within a product category. The section below provides the details on the hedonic methods used.

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

Measure $q2$ is very similar to q. We use weights generated by hedonic regressions and scale them by observed quantities to get to the sales-based weights for each attribute. In this case, we also normalize the weights so that all weights within a product category add up to one.

Hedonic Regression Weights. — We estimate product category weights ω_a^j for our measure q using hedonic methods. In particular, 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 characteristics and a sequence of time-dummies. Since the regression is run over data which is pooled across time periods, any product characteristic which is held by at least one product in some period can be included even if it is not present in all periods. The estimated regression coefficients represent the shadow price for each of the included characteristics. To implement this method, we estimate the following equation by non-negative least squares:

$$p_{ut} = \sum_c \pi^c a_u^c + \lambda_t + \epsilon_{ut}, \tag{22}$$

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

Using this method, we obtain a correlation of approximately 0.91 between the actual price and $\sum_c \pi^c$.[51] The weight ω_a^j is the average contribution of the characteristics within each attribute to the price normalized so that $\sum_a^{A^j} \omega_a^j = 1$; these are the weights used in our baseline novelty index.

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

B.7.2 Residual Demand Measure

An alternative way of measuring the degree of product innovation brought by new products to the market is to weight them by their implied quality (or residual demand) using a structural specification of their demand function. To derive an implied quality for each product, we follow Hottman et al. (2016) and Argente et al. (2018) and use a nested constant elasticity of substitution (CES) utility system that allows the elasticity of substitution between varieties within a firm to differ from the elasticity of substitution between varieties supplied by different firms. The model features oligopolistic competition with a finite number of heterogeneous multi-product firms, where the output of each category is described by a nested CES structure over a finite number of products within a finite number of firms (j is omitted for simplicity of notation)

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

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

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

where the demand for the product depends on the implied quality γ_{ui} and price p_{ui} of the product, as well as the firm's price index p_i, the category's price index p, and the size of the category Y. Conditional on observing the prices and quantities from the data and obtaining estimates for σ and η, we recover γ_{uijt} as a structural residual that ensures that the model replicates the observed data up to a normalization.[52] We normalize the implied quality so that its geometric mean within each category and time period equals one. The key advantage of this normalization is that we can compare a product's implied quality within the firm and across firms within a category and time period. Using this normalization and equation B.7.2, we obtain the product implied quality as:

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

where s_{ui} and s_i are the share of sales of product u and the share of sales of firm i, respectively, and M denotes the total number of products sold in a category. The estimation procedure for σ and η follows Broda and Weinstein (2010) and Feenstra (1994). The estimation has two steps. In the first step, we estimate the elasticity of substitution across products within firms using product shares, product prices, and firm shares using a GMM procedure. The key identification assumption is that demand and supply shocks at the

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

product level are uncorrelated once we control for firm-time specific effects. In the second step, we use these estimates for products to estimate the elasticity of substitution across firms for each category using the procedure developed by Hottman et al. (2016). We use the estimates from Argente et al. (2018).

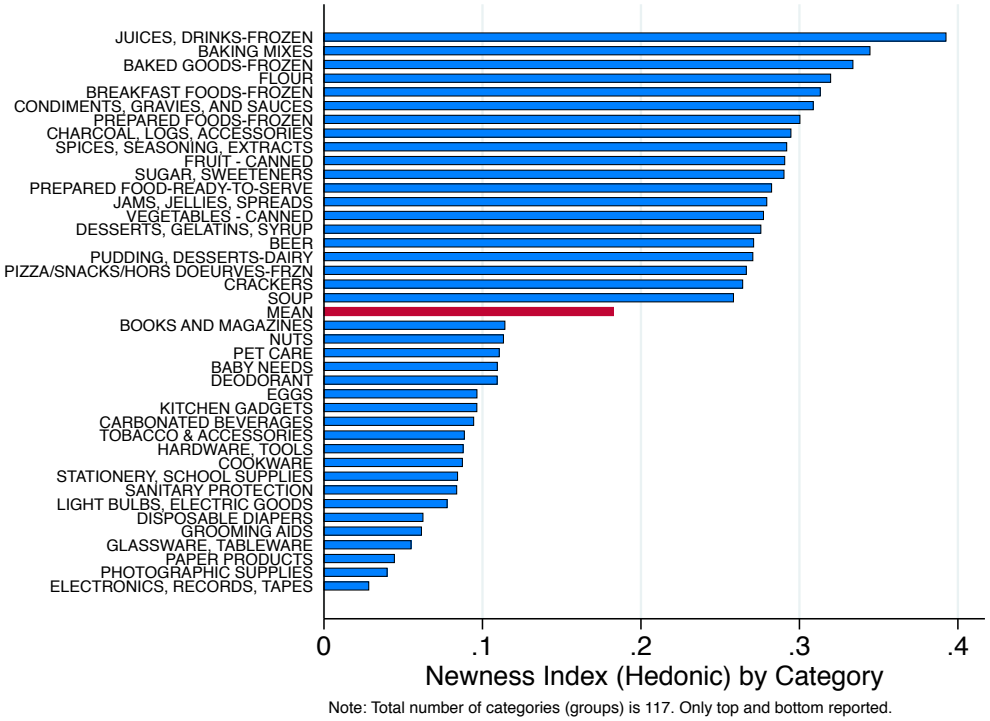
To capture the incremental effect of new products on the residual demand of the firms, our measure of quality improvement $q3$ is the geometric average of the implied quality of the new products relative to the geometric average of all products sold by the firm.

B.7.3 Descriptive Statistics

Novelty Indices Across Product Categories. — Although in our main analyses we only consider within category variation in novelty, Figure A11 shows the degree of heterogeneity in novelty index q across different product categories. The quality measure q has a correlation of 0.93 with the equal-weights measure $q1$. Conditional on having an equal-weights index larger than zero, the correlation is 0.79. "Juices, Drinks-frozen" has a high novelty index mainly due to the prevalence of new brands and new flavors over our sample period. Over our sample period, there are more than 50 new brands and 67 new flavors in this category, which can be explained by recent trends in this category to increase the nutrients, reduce the sugar content, and to create products according to the consumer's lifestyles. The novelty index for "Baking Mixes" and "Flour" can be explained by the surge in home-based baking observed in recent years, which led to more than one thousand new brands in these categories. Only in "Baking Mixes" we observed more than 600 new flavors during our sample period. These categories have also seen significant innovations in packaging. An example is stand-up pouches, which use less plastic, increase the shelf life of products, and reduce the likelihood they are damaged during shipping.

Figure A12 shows some examples of products with high and low equal-weights novelty $q1$ in our data. For example, the product Asthmanefrin Inhalation Solution - Liquid Refill is part of the group Medications/Remedies/Health Aids. When it was introduced in the market, this product had six of the eight attributes that we observe in our data for that product group, it was a new brand, launched by a new firm, it was a liquid, bronchilator refill. As a results, its equal-weights novelty index is 6/8=0.75.

Figure A11: Novelty Index (Baseline q)



Notes: The figure presents the average novelty index for a sample of product groups in our data. In particular, it shows the mean novelty index by groups along with the top and bottom groups as ranked by this measure. We compute the novelty index for each product using equation B.7.1. We average across products and product modules to the category level. We focus on cohorts from 2006Q3 to 2014Q4 and on modules with at least 20 barcodes.

Figure A12: Novelty Index: Examples



(a) High-Novelty Products

(b) Low-Novelty Products

Correlation with Product and Firm Performance. — Our baseline measure of quality q explicitly captures the novelty of a new product by using information about its attributes. This use of product attributes offers important advantages in the context of our paper. Patents are granted on the basis of novelty, and thus using a quality-adjusted

measure of product introduction that explicitly accounts for new features of the product should help to align the notion of innovation on patents and products side. However, new features of the product may not affect the market at all if they are not valued by customers. Our baseline measure q partially accounts for this potential concern by weighting any new characteristic according to its shadow price using hedonic regressions. In addition, Table A2 shows that our baseline measure is correlated with product and firm outcomes, and thus may be capturing some vertical quality differentiation or subjective differences in consumer taste.

Table A2: Novelty Measure: Correlation with Firm Outcomes

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

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

C Additional Empirical Results

Table A3: Product Innovation and Patenting: Citations and Claims

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

Notes: The table shows regressions of the log number of new products (*log N*) and of the log quality-adjusted new products (*Log qN*) in a firm×category over time as a function of log citations- and claims-adjusted number of patents. Our benchmark quality measure is defined in Section 3.2.1. *Citations* is the log number of 5-year citations received by all patents filed in the firm×category×year; *Claims* is the log number of claims on all patents filed in the firm×category×year. The inverse hyperbolic sine transformation is used for logs. Standard errors robust against heteroskedasticity and serial correlation are reported in parentheses.

Table A4: Product Innovation and Patenting (Firm Level)

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

Notes: The table shows regressions of the log number of new products (*log N*) and of log quality-adjusted new products (*Log qN*) in a firm over time as a function of log number of patents. Our benchmark quality measure is defined in Section 3.2.1. *Patents* is the log number of any patent applications in firm × year; *Patents granted* is the log number of granted patent applications; and *Patents non-granted* is the log number of patent applications that have not been granted (abandoned or pending). Standard errors robust against heteroskedasticity and serial correlation are reported in parentheses.

Table A5: Product Innovation and Product & Process-Related Patents

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

Notes: The table shows regressions of the log number of new products (*log N*) and the log quality-adjusted new products (*Log qN*) in a firm×category over time as a function of proxies for product-related and process-related patents. Our benchmark quality measure is defined in Section 3.2.1. *Product patents* is the log number of product-related patents, while *Process patents* is the log number of process-related patents. Proxies for product-related and process-related patents are defined in Section B.2. The inverse hyperbolic sine transformation is used for logs.

Table A6: Patenting and Sales: Role of Price and Quantities

Panel A- Prices

	Log Prices (t)			Log Prices (t)		
	All	Small	Large	All	Small	Large
Log P(t-1)	0.017**	0.010	0.031***	0.018**	0.009	0.031***
	(0.007)	(0.030)	(0.009)	(0.007)	(0.030)	(0.009)
Log N(t)	0.017***	0.024***	0.005***			
	(0.001)	(0.004)	(0.002)			
Log qN(t)				0.021***	0.035***	-0.002
				(0.003)	(0.009)	(0.003)
Observations	520,875	84,041	109,203	520,875	84,041	109,203
R-squared	0.919	0.935	0.951	0.919	0.935	0.951
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y

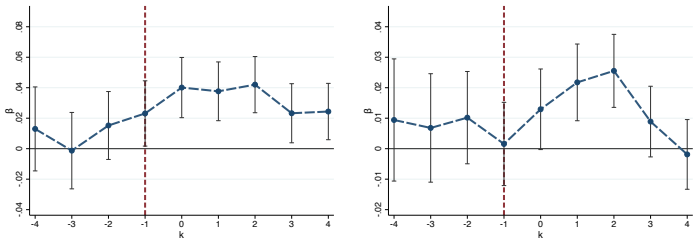
Panel B- Quantities

	Log Quantity (t)			Log Quantity (t)		
	All	Small	Large	All	Small	Large
Log P(t-1)	0.504***	0.167*	0.345***	0.515***	0.159	0.376***
	(0.033)	(0.098)	(0.040)	(0.033)	(0.098)	(0.040)
Log N(t)	0.286***	0.155***	0.366***			
	(0.005)	(0.013)	(0.007)			
Log qN(t)				0.268***	0.179***	0.393***
				(0.012)	(0.028)	(0.014)
Observations	520,875	84,041	109,203	520,875	84,041	109,203
R-squared	0.765	0.788	0.759	0.764	0.788	0.755
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y

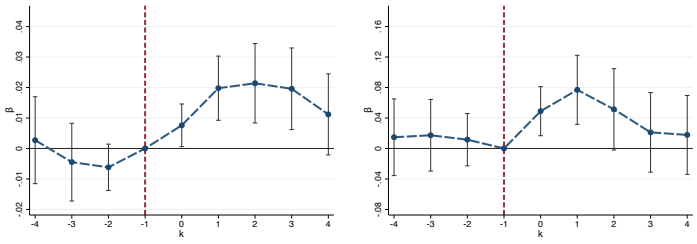
Notes: The table presents estimated outcomes of log prices/quantities at the firm × product module level as a function of the log number of patent applications by until time $t-1$ and the log number of new products introduced at time t (or quality adjusted new products), by size groups. We use prices and quantities at the product module level to have comparable units of measurement. Patent and product innovation measures are defined at the product category level. For each firm × module, we define size based on average sales over the sample period. "All" column uses data for all sizes. "Small" column is restricted to the lowest size quintile. "Large" is restricted to the top size quintile. The inverse hyperbolic sine transformation is used for logarithms.

Figure A13: Product Innovation and Patenting: Dynamics of Other Outcomes

— Firm-level estimates (Match 1) —

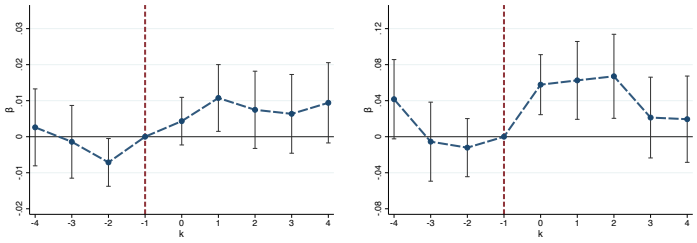


(a) $\beta_{\text{N,P}}$ (b) $\beta_{\text{qN,P}}$



(c) $\beta_{\text{T,P}}$ (d) $\beta_{\text{rev,P}}$

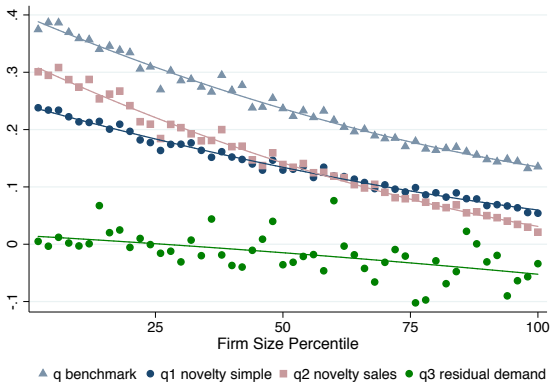
— Firm-category level estimates (Match 2) —



(e) $\beta_{\text{T,P}}$ (f) $\beta_{\text{rev,P}}$

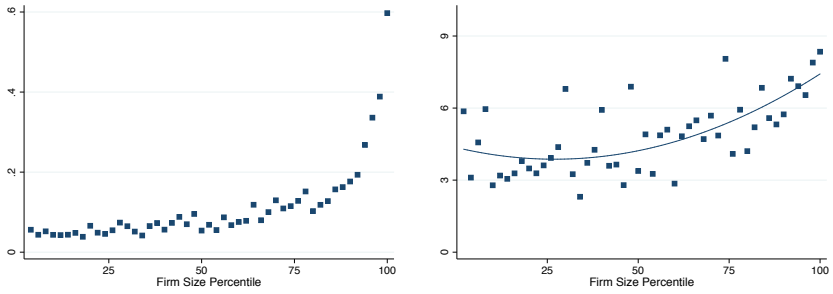
Note: The figure plots the estimated coefficients after estimating equation $\log Y_{ijt+k} = \beta_k \log P_{ijt} + \alpha_{ij} + \gamma_{jt} + u_{ijt+k}$, $k = -4, ..., 0, ..., 4$ for the log product introduction, N, in (a), quality-adjusted product introduction, qN, in (b), total number of products, T, in (c) and (e), and yearly revenue in (d) and (f) on log number of patent applications. The top panel uses firm-level data (Match 1), and the bottom panel uses firm-product category level data (Match 2). The inverse hyperbolic sine transformation is used for logarithms. The vertical bands represent $\pm 1.65\times$ the st. error of each point estimate. Standard errors are clustered at the firm $\times$ category level.

Figure A14: Product Innovation Rate by Size: Alternative Quality Adjustments



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

Figure A15: Patenting and Firm Size



(a) Probability of patent application (b) Number of patent applications (log)

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

D Heterogeneity across Product Categories

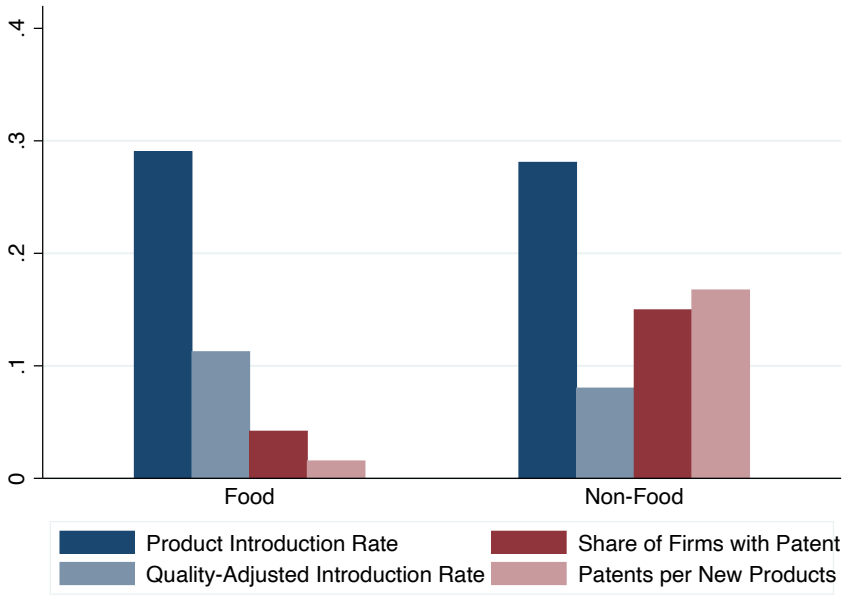
In this section, we explore some of our main results for food and non-food product categories. Food categories include the Nielsen departments dry grocery, frozen foods, dairy, deli, packaged meat, fresh produce, and alcoholic beverages. Non-food categories include the departments health and beauty, non-food grocery, and general merchandise.

Table A7: Product Innovation and Patenting: Food and Non-Food Categories

	Log N			Log qN		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel 1 - Food						
Patents(t-1)	-0.0111			0.0070		
	(0.021)			(0.013)		
Patents granted(t-1)		0.0024			0.0109	
		(0.024)			(0.015)	
Patents non-granted(t-1)			-0.0307			0.0059
			(0.029)			(0.018)
Observations	205,595	205,595	205,595	205,595	205,595	205,595
R-squared	0.650	0.650	0.650	0.617	0.617	0.617
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y
Panel 2 - Non-Food						
Patents(t-1)	0.0413***			0.0199***		
	(0.011)			(0.006)		
Patents granted(t-1)		0.0443***			0.0210***	
		(0.011)			(0.006)	
Patents non-granted(t-1)			0.0175			0.0017
			(0.016)			(0.009)
Observations	104,117	104,117	104,117	104,117	104,117	104,117
R-squared	0.764	0.764	0.764	0.647	0.647	0.647
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y

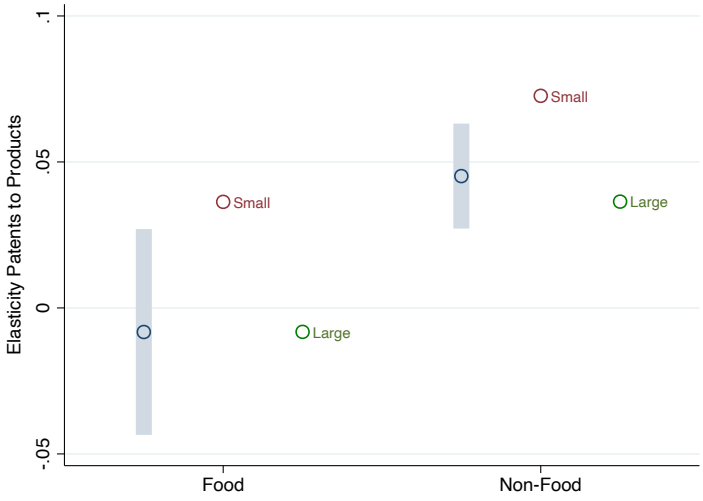
Notes: The table shows regressions of the log number of new products (log qN) and of log quality-adjusted new products (log qN) in a firm × category over time as a function of the log number of patents. Our benchmark quality measure is defined in Section 3.2.1. *Patents* is the log number of any patent applications in firm × category × year; *Patents granted* is the log number of granted patent applications; and *Patents non-granted* is the log number of patent application that have not been granted (abandoned or pending). The inverse hyperbolic sine transformation is used for logs. Standard errors robust against heteroskedasticity and serial correlation are reported in parentheses.

Figure A16: Main Summary Statistics for Food and Non-Food Product Categories



Notes: The figure presents summary statistics for food and non-food product categories in Nielsen. Using Match 2, we compute the average product introduction rate, quality-adjusted introduction rate, number of new products, and patent applications at the firm × category level (patent statistics are winsorized at the top 5%). With these, we compute average product introduction rate, quality-adjusted introduction rate, patents per new products and share of firms with patents at the product category-level. The plot shows these statistics by aggregating them within food and non-food product category (weighting by the total revenue of each product category).

Figure A17: Product Innovation and Patenting by Firm Size

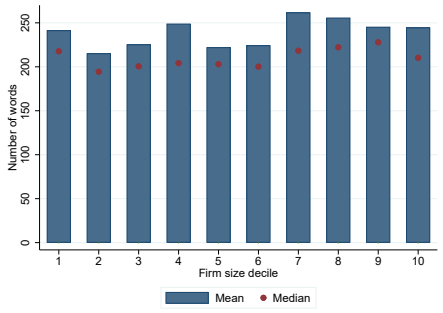


Notes: The figure shows coefficients of the log number of new products (log N) over time as a function of the log number of patents. *Patents* is the log number of any patent applications in firm × category × year. The inverse hyperbolic sine transformation is used for logs. Coefficients in blue include firms of all sizes. "Small" indicates coefficients of firms in the lowest size quintile. "Large" indicates coefficients of firms in the top size quintile.

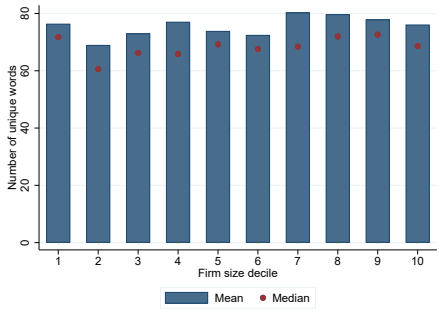
E Robustness to Matching Algorithm

E.1 Statistics on the Matching Algorithm by Firm Size

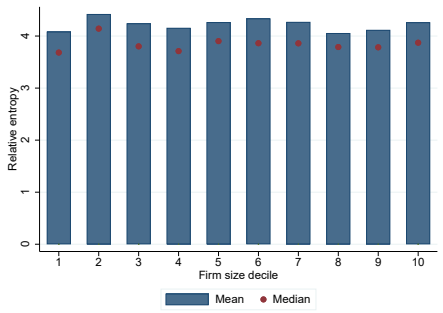
Figure A18: Patent Text and Match Properties by Firm Size Percentile



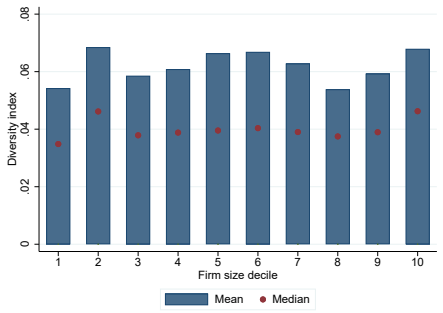
(a) Number of words in patent documents



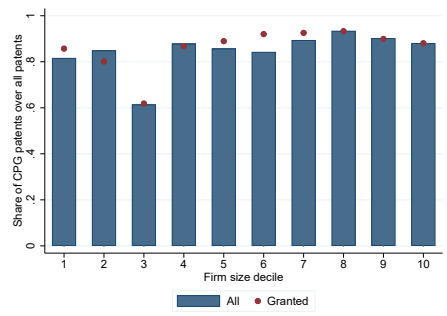
(b) Unique words in patent documents



(c) Relative entropy of patent word dist.



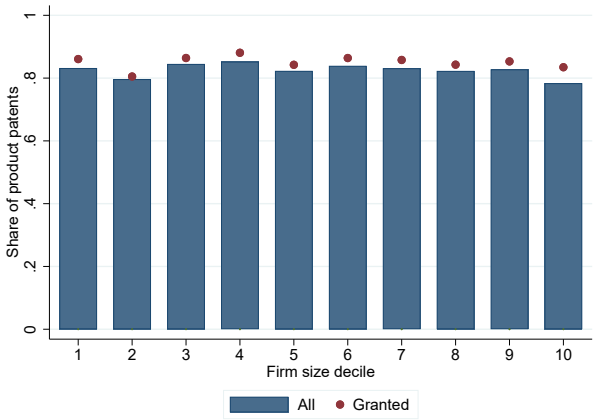
(d) Word diversity index of patent word dist.



(e) Share of matched patents

Notes: The figure plots various text and match characteristics of patents held by firms in different size (sales) deciles. All size deciles are constructed within product categories, except for panel (e) that is based on firm-level deciles. The first panels plot means and medians of the average number of words (a), the number of unique words (b), the relative entropy between the patent's word distribution and the word distribution of all patents (c), and the Simpson's diversity index of the patent's word distribution (d) of firms' patents. Panel (e) looks at the share of matched patents in the firms' whole patent portfolio (that is, the number of patents from Match 2 divided by the number of patents from Match 1) for the sample of CPG-only firms (see Section B.1 for the definition) for which the non-matches are less likely to be due to the firm's operations outside the CPG sector. Panel (f) plots the similarity scores of the matched patents.

Figure A19: Share of Product-Related Patents by Firm Size Percentile



Notes: This figure plots the shares of product-related patents held by firms across deciles of firm size, defined in terms of sales. We classify patents into product-related patents based on the claims of patent documents (Appendix B.2).

E.2 Results with Alternative Matching

In this section, we present our main results under different specifications of our matching algorithm. In our baseline specification, we classify product modules into 400 clusters, which we refer to as product categories. We do so since we believe this partition strikes a balance between aggregating very similar products while maximizing the difference between products across categories. Nonetheless, in this section, we show that our main findings are similar if we use the product classification scheme developed by Nielsen: 1,070 detailed product modules aggregated into a set of 114 broad product groups.

We also present our main results using a higher matching similarity threshold. Recall that patents with low text similarity are deemed unrelated to the product categories that we consider. In our baseline specification, we restrict the set of potential categories for each patent to the product categories whose similarity score exceeds 0.025. This section shows that our main results hold if we use a higher similarity score of 0.05. As before, the implication of this adjustment is that patents whose highest similarity is below that threshold are more likely to be classified as "non-matched".

Table A8: Product Innovation and Patenting (using the Nielsen product group aggregation)

	Log N			Log qN		
	(1)	(2)	(3)	(4)	(5)	(6)
Patents(t-1)	0.0307***			0.0173***		
	(0.010)			(0.006)		
Patents granted(t-1)		0.0366***			0.0191***	
		(0.010)			(0.006)	
Patents non-granted(t-1)			0.0079			0.0025
			(0.014)			(0.008)
Observations	309,718	309,718	309,718	309,718	309,718	309,718
R-squared	0.710	0.710	0.710	0.631	0.631	0.631
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y

Notes: The table shows regressions of the log number of new products (log qN) and of log quality-adjusted new products (log qN) in a firm × category over time as a function of the log number of patents. Our benchmark quality measure is defined in Section 3.2.1. The alternative innovation-quality measures ($q1, q2, q3$) produce consistent results. *Patents* is the log number of any patent applications in firm × category × year; *Patents granted* is the log number of granted patent applications; and *Patents non-granted* is the log number of patent application that have not been granted (abandoned or pending). The inverse hyperbolic sine transformation is used for logs. Standard errors robust against heteroskedasticity and serial correlation are reported in parentheses. We use an aggregation of modules into product groups as defined by Nielsen.

Table A9: Product Innovation and Patenting: by Size (using the Nielsen product group aggregation)

	Log N (t)			Log qN (t)		
	All	Small	Large	All	Small	Large
Log P(t-1)	0.031***	0.033*	0.040***	0.017***	0.037***	0.024***
	(0.008)	(0.020)	(0.014)	(0.004)	(0.008)	(0.008)
Observations	309,718	45,838	66,697	309,718	45,838	66,697
R-squared	0.710	0.468	0.755	0.631	0.419	0.689
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y

Notes: The table shows regressions of the log number of new products (log qN) and of log quality-adjusted new products (log qN) in a firm × category over time as a function of the log number of patents. P is the number of patent applications for a firm × category × year. For each firm × product category, we define size based on the average sales over our sample period. The "All" column shows data for all sizes. The "Small" column is restricted to the bottom size quintile. "Large" is restricted to the top size quintile. The inverse hyperbolic sine transformation is used for logarithms. We use an aggregation of modules into product groups as defined by Nielsen.

Table A10: Patenting and Sales Growth (using the Nielsen product group aggregation)

	Δ Log Sales (t)			Δ Log Sales (t)		
	All	Small	Large	All	Small	Large
Log P(t-1)	0.060***	0.052	0.101***	0.072***	0.068	0.111***
	(0.017)	(0.085)	(0.019)	(0.017)	(0.085)	(0.019)
Log N(t)	0.259***	0.329***	0.146***			
	(0.003)	(0.012)	(0.004)			
Log qN(t)				0.368***	0.540***	0.176***
				(0.006)	(0.029)	(0.007)
Observations	227,183	31,070	51,191	227,183	31,070	51,191
R-squared	0.287	0.326	0.284	0.270	0.315	0.270
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y

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

Table A11: Product Innovation and Patenting (higher similarity threshold)

	Log N			Log qN		
	(1)	(2)	(3)	(4)	(5)	(6)
Patents(t-1)	0.0391***			0.0188***		
	(0.009)			(0.005)		
Patents granted(t-1)		0.0439***			0.0199***	
		(0.010)			(0.005)	
Patents non-granted(t-1)			0.0208			0.0071
			(0.014)			(0.007)
Observations	409,434	409,434	409,434	409,434	409,434	409,434
R-squared	0.692	0.692	0.692	0.623	0.623	0.623
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y

Notes: The table shows regressions of the log number of new products (log qN) and of log quality-adjusted new products (log qN) in a firm × category over time as a function of the log number of patents. Our benchmark quality measure is defined in Section 3.2.1. The alternative innovation-quality measures ($q1, q2, q3$) produce consistent results. *Patents* is the log number of any patent applications in firm × category × year; *Patents granted* is the log number of granted patent applications; and *Patents non-granted* is the log number of patent application that have not been granted (abandoned or pending). The inverse hyperbolic sine transformation is used for logs. Standard errors robust against heteroskedasticity and serial correlation are reported in parentheses. We restrict the set of potential categories for each patent to the product categories whose similarity score exceeds 0.05.

Table A12: Product Innovation and Patenting: by Size (higher similarity threshold)

	Log N (t)			Log qN (t)		
	All	Small	Large	All	Small	Large
log P(t-1)	0.039***	0.062***	0.027**	0.019***	0.032***	0.015**
	(0.007)	(0.019)	(0.013)	(0.003)	(0.007)	(0.007)
Observations	409,434	61,597	86,787	409,434	61,597	86,787
R-squared	0.692	0.461	0.742	0.623	0.407	0.686
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y

Notes: The table shows regressions of the log number of new products (log qN) and of log quality-adjusted new products (log qN) in a firm × category over time as a function of the log number of patents. P is the number of patent applications for a firm × category × year. For each firm × product category, we define size based on the average sales over our sample period. The "All" column shows data for all sizes. "Small" column is restricted to the bottom size quintile. "Large" is restricted to the top size quintile. The inverse hyperbolic sine transformation is used for logarithms. We restrict the set of potential categories for each patent to the product categories whose similarity score exceeds 0.05.

Table A13: Patenting and Sales Growth (higher similarity threshold)

	Δ Log Sales (t)			Δ Log Sales (t)		
	All	Small	Large	All	Small	Large
log P(t-1)	0.061***	-0.132*	0.089***	0.073***	-0.146*	0.100***
	(0.016)	(0.080)	(0.020)	(0.016)	(0.080)	(0.020)
log N(t)	0.265***	0.316***	0.160***			
	(0.003)	(0.011)	(0.004)			
log qN(t)				0.406***	0.581***	0.215***
				(0.006)	(0.029)	(0.007)
Observations	296,320	40,666	65,680	296,320	40,666	65,680
R-squared	0.291	0.377	0.296	0.275	0.368	0.281
Time-Category	Y	Y	Y	Y	Y	Y
Firm-Category	Y	Y	Y	Y	Y	Y

Notes: The table presents estimated outcomes of changes in log sales at the firm × category level as a function of the log number of patents (P) and log number of new products (log qN) and log quality-adjusted new products (log qN), by size groups. We use the firm × product category data set for the period 2007–2015, restricting the analysis to observations with sales above $1,000. For each firm × product category, we define size based on average sales over the sample period. "All" column uses data for all sizes. "Small" column is restricted to the lowest size quintile. "Large" is restricted to the top size quintile. "Leaders" is restricted to the top size decile. The inverse hyperbolic sine transformation is used for logarithms. Standard errors robust against heteroskedasticity and serial correlation are reported in parentheses. We restrict the set of potential categories for each patent to the product categories whose similarity score exceeds 0.05.

Appendix References

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

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

Argente, David, Munseob Lee, and Sara Moreira, "Innovation and Product Reallocation in the Great Recession," *Journal of Monetary Economics*, 2018, *93*, 1–20.

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

Breuer, Matthias, Christian Leuz, and Steven Vanhaverbeke, "Reporting regulation and corporate innovation," NBER Working Paper 2021.

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

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

de Rassenfosse, Gaétan, "Notice Failure Revisited: Evidence on the Use of Virtual Patent Marking," NBER Working Paper 2018.

Feenstra, Robert C, "New Product Varieties and the Measurement of International Prices," *The American Economic Review*, 1994, pp. 157–177.

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

Goolsbee, Austan D. and Peter J. Klenow, "Internet Rising, Prices Falling: Measuring Inflation in a World of E-Commerce," *AEA Papers and Proceedings*, May 2018, *108*, 488–92.

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

Hoberg, Gerard and Gordon Phillips, "Text-based Network Industries and Endogenous Product Differentiation," *Journal of Political Economy*, 2016, *124* (5), 1423–1465.

Hortaçsu, Ali and Chad Syverson, "The Ongoing Evolution of US Retail: A Format Tug-of-war," *Journal of Economic Perspectives*, 2015, *29* (4), 89–112.

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

Jaravel, Xavier, "The Unequal Gains from Product Innovations: Evidence from the U.S. Retail Sector," *The Quarterly Journal of Economics*, 2019, *134* (2), 715–783.

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

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