Product and Process Innovation in the U.S. from 1980-2015 and Over the Firm's Life Cycle

Colin Davison*

October 17, 2025

Click Here for Current Version

Abstract

This paper uses machine learning techniques to classify patent claims as product or process innovations. I document that the overall process share of innovation was on a secular decline from 1980-2015. I find that the process share is low at the beginning of a firm's product life cycle, peaks in the middle before plateauing at an intermediate level at the end of the life cycle. Firm size is positively associated with the process share of innovation.

Keywords: Product Innovation, Process Innovation, R&D, Product Life Cycle
JEL: C81, L25, O31, O32, O33, Y10

Acknowledgements: I would like to thank William Evans, Kirk Doran, Robert C. Johnson, and Jeffrey Campbell for their advice and feedback.
Data Availability Statement: The data supporting this article can be found at https://doi.org/10.5281/zenodo.8364951.

*The College of Wooster

1 Introduction

Firms engage in product innovation by introducing new product varieties. They also create process innovations by altering the assembly of their products. Understanding what factors incentivize a firm to choose product or process innovation along with the implications of the firm's choice are topics which have received substantial attention. For example, Klepper 1996 and Utterback and Abernathy 1975 develop models of the firm life cycle describing how the relative amounts of product and process innovation evolve over the product life cycle. In addition, Cohen and Klepper 1996 theoretically argue that the incentive to engage in process innovation is increasing in the scale of a firm's production. Despite this, the empirical examinations on how the life cycle and firm size impact product and process innovation have been sparse and mixed (Damanpour 2010). One explanation for the lack of study and consensus is a dearth of measurement of product and process innovation.

This study aims to fill these gaps by introducing a novel dataset, the Economically Based Product Process Patent Dataset (EPP)[1], which classifies all the independent publication claims of patents granted between 1980-2015 to publicly traded U.S. manufacturing firms as product or process innovations (Davison 2023). This data provides an alternative measurement of product and process patenting based on an economically inspired notion of product and process innovation as opposed to using the technically inspired definitions provided by the U.S. Patent and Trademark (USPTO) office. To create the classification, I hand classify approximately 40,000 publication claims as product or process innovations using a consistent and economically based definition of product and process innovation. Using these hand classified claims, I train machine learning (ML) classifiers to predict whether a claim is a product or process innovation based on features present in the claim's text. The ML classifiers are able to predict whether a claim is a product or process innovation with 85% balanced accuracy.

With this data, I describe the time trends in the process share of innovation. I find that the process share of innovation has been on secular decline since 1980, falling from 26.3% in 1980 to 11.7% in 2015. From 1980-2000 the process share fell from 26.3% to 19.1% with the decline being driven by between industry reallocation as patenting moved from industries with high process shares to lower process shares. From 2000-2015 the process share fell from 19.1% to 11.7% and is primarily accounted for by within industry declines in the process share.

Next, I examine how the process share of innovation varies over a firm's life cycle. I find that firms start out with low process innovation intensity as they focus on developing their product(s). Next, they enter a stage of process optimization where their process share of innovation peaks. Finally, they enter a stage of intermediate process intensity as their product(s) mature and then ultimately decline. The results are consistent with the life cycle theories of Klepper 1996 and Utterback and Abernathy 1975. Further, I find a positive association between firm size and process innovation intensity which supports the view that the returns to process innovation are increasing in the output of the firm (Cohen and Klepper 1996).

2 Theoretical Framework

2.1 Measuring Product and Process Innovation

2.1.1 Survey Based Measures

This paper contributes to the literature concerned with measuring product and process innovation. One strategy for measuring process innovation has been through surveys which ask firms whether they have introduced a process innovation in a given period of time (Mairesse et al. 2005; Griffith et al. 2006; Arundel et al. 2007; Antonucci and Pianta 2002; Classen et al. 2014; Demmel et al. 2017; Diéguez-Soto et al. 2018; Sastre 2015; Haneda and Ito 2018; Van Beveren and Vandenbussche 2010). These surveys are not generally conducted in the United States and limited in their ability to distinguish both the quantity and quality of process innovation that has occurred. Rammer 2023 improves on these survey based measures by introducing targeted questions to the German Community Innovation Surveys which seek to quantify and disentangle cost reducing and quality improving process innovations. While Rammer 2023 and other surveys provide valuable information in the German and European contexts respectively, the data described in this paper measures product and process innovation in the US context over a long period of time, something that is not currently possible using survey methods. Further, the EPP data is publicly available to researchers as it is not based on confidential surveys.

2.1.2 Patent Based Measures

In one of the first large scale efforts to use patent data to measure product and process innovation, Scherer 1982 created industry-level product and process innovation data by hand classifying the industry of origin and industry of use for a cross-section of 15,112 patents. With the advent of modern natural language processing, there have been several recent attempts to classify patent claims as product or process innovations (Banholzer et al. 2019; Liu et al. 2020; Bena and Simintzi 2022; Ganglmair et al. 2022). These classifications generally use keyword approaches where they look for the presence of process keywords[2] in the patent claim text which capture the USPTO or technical definition of a process innovation which "define steps, acts, or methods to be performed." For example, Liu et al. 2020 frame process patents as "referring to an *activity* (process, method, or use)" and product patents as "referring to a *physical entity* (produce, device, or apparatus)." This accurately captures the USPTO's technical distinction between product and process innovations but may not correspond to a more economically grounded defintion of product and process innovation.

The methodology of this paper is inspired by the hand classification of Scherer 1982, but differs in its use of ML classifiers to classify patents I have not hand classified. In contrast to the USPTO definition of a product patent, I define a patent claim as a product innovation if it describes a physical object[3] that the firm sells with no discussion about how the object is created. All other claims are process innovations. This method of classification goes beyond distinguishing patents by whether they refer to nouns (products) or verbs (processes), but incorporates reference to the end use of the innovation, that is, whether the firm plans to sell the object or not (Scherer 1982). This avoids misclassification in the situations where a firm either develops an object for internal use (e.g. a firm developing a machine for use in their manufacturing process) or when a firm describes a product in terms of what it can accomplish (e.g. a pharmaceutical company describing a drug as a method for treating a particular disease). Using this economically grounded definition necessitates a difference in classification methodology relative to the currently favored keyword approach, leading to my use of hand classification paired with a rich set of text features to train ML classifiers. Later, I present qualitative evidence that my classification is better able to capture the economic concept of a process innovation relative to the currently available process patent classifications.

2.2 Product and Process Innovation Over Time

The findings in this paper also contribute to our understanding of how patenting and innovation have evolved over long periods of time. Researchers have documented that there has been a decline in scientific research at large corporations (Arora, Belenzon, and Patacconi 2018), an increase in team size (Jones 2009), a decline in research productivity (Bloom et al. 2020), and an increase in patent scope (Marco et al. 2019). This paper documents the secular decline of the process share, highlighting that the innovative efforts of publicly traded U.S. manufacturing firms are increasingly being devoted towards product innovation. Given the differences between product and process innovations in their ability to generate knowledge spillovers, the declining process share of innovation has potential implications for societal welfare (Kraft 1990; Kotabe and Murray 1990; Davison 2022). Further, the decline of the process share indicates an overall shift in how manufacturing firms in the US strategically use innovation. My finding of a secular decline in the process share of innovation stands in contrast to the findings from the data[4] underlying Banholzer et al. 2019, Bena and Simintzi 2022, and Ganglmair et al. 2022 (hereafter BBG), which all indicate a substantial secular increase in the process share of innovation from 1980-2015. The difference in overall time trends is most clearly seen in the "Computers & Communications" technology category where the BBG classifications show a large and increasing process share from 1980-2015 whereas my classification shows a small and declining process share. This discrepancy can be accounted for by a large number of computing patents which use process keywords to describe a product but are clearly not directed towards the manufacturing process for the product (e.g. a semiconductor patent which describes the computer chip as "a method for calculating...").

2.3 Product and Process Innovation Over the Firm's Life Cycle

This paper adds to our understanding of innovation over the life cycle of the firm. Klepper 1996 builds a model of innovation over the product life cycle and predicts that over time producers devote increasing effort to process innovation relative to product innovation. The intuition behind the result is that process innovation scales with firm size in the sense that a given process innovation will be more valuable if it can be applied in the production of a larger quantity of output (Cohen and Klepper 1996). As firms move through the product life cycle, they grow in size, making process innovation increasingly more appealing relative to product innovation.

Utterback and Abernathy 1975 theorize that product innovation will be high at the beginning of the product life cycle when the product is nonstandard and its technical functionality is being refined. As the product becomes standardized and more competition takes place, there is increasing focus on cost minimization and product innovation gives way to process innovation. On the other hand, there is little process innovation at the beginning of the life cycle as the product has not been refined and is being produced in small quantities. As competition increases and the product becomes standardized, firms increasingly focus on process innovation as they seek to lower their costs and take advantage of their growing scale. The focus on process innovation is only justifiable at later stages in the product life cycle when uncertainty about the product is low enough that the risk of creating dedicated processes around a product's design are profitable in expectation (Vernon 1966). In the final stages of the life cycle, when the product is mature, both the product and the production process are highly developed and integrated, making any alterations to the production process costly. In this stage, process innovation falls from its peak but remains above the low levels seen at the beginning of the life cycle.

While there is substantial empirical interest around how overall innovative inputs and outputs evolve over the firm and industry life cycle (Tavassoli 2015; Shahzad et al. 2022), there has been less work done empirically testing how product and process innovation change over the firm's life cycle, with Utterback and Abernathy 1975 providing a notable exception. Empirically investigating how product and process innovation change over the firm's life cycle is challenging due to the lack of data both distinguishing product and process innovation in firms over time and measuring where firms are in their life cycle. To measure where a firm is specifically located in their own life cycle, studies use proxies such as firm size, age, or cash flow characteristics (Dickinson 2011) which can be problematic as they are indirect measurements of where the firm is located in the life cycle. I address these issues by combining the EPP data with the (Hoberg and Maksimovic 2022) measure of where firms are in the product life cycle. The Hoberg and Maksimovic 2022 improves on the previously listed proxies by using textual analysis of a firm's 10-K filings to document where the firm is at in the life cycle based on the firm's own descriptions of their business.

2.4 Product and Process Innovation and Firm Size

Theoretical examinations of the relative amounts of product and process innovation and firm size have concluded that there should be a positive relationship between the process share of innovation and firm size (Cohen and Klepper 1996). There are several reasons to think this will hold. First, firms apply a given cost reducing process innovation in service of producing each unit of output. As the amount of output a firm produces increases, the firm will be able to apply that process innovation in the production of more output, increasing the returns to the process innovation (Cohen and Klepper 1996). The second argument relates to the firm life cycle discussed previously. In the early stages of the life cycle, a firm is small and tends to focus on product innovation (for reasons discussed previously) while later in the life cycle when a firm is larger they tend to focus on process innovation. The tight connection between firm size and life cycle stage highlights the need to examine these relationships simultaneously. While the theoretical studies conclude that there should be a positive association between firm size and process innovation intensity, the empirical findings are mixed (Pavitt et al. 1987; Scherer 1991; Bertschek 1995; Cohen and Klepper 1996; Fritsch and Meschede 2001) with a meta-study by Damanpour 2010 finding no difference in the strength of the relationship between firm size and product/process innovation. This study seeks to revisit the relationship between firm size and product/process innovation using an economically grounded measurement of product and process innovation.

3 Building Product and Process Patent Data

3.1 Sample Selection

This study constructs its novel classification of patent claims as product or process innovations using the DISCERN database, a match between USPTO patents and COMPUSTAT firms which is described in Arora, Belenzon, and Sheer 2021. The data provide the most comprehensive and accurate match of USPTO patents to publicly traded U.S. firms that is currently available by updating the National Bureau of Economic Research (NBER) patent database (Hall et al. 2001) with all patents granted from 1980 to 2015. To create the data, Arora, Belenzon, and Sheer 2021 match all COMPUSTAT firms ever having conducted R&D to patents on the basis of assignee and firm name. The DISCERN data improve upon Hall et al. 2001 by tracking parent companies and subsidiaries in the COMPUSTAT data, as well as how name changes and mergers and acquisition activity may affect subsidiaries. I aggregate to their identification of parent companies when conducting firm level analysis.

The DISCERN database includes patents of publicly traded companies who are not manufacturing firms. As the distinction between product and process innovation is most salient for manufacturing firms, I limit my sample to the patents of firms of predominantly operate in manufacturing industries over their lifetime. A simple way to define whether a COMPUSTAT firm is a manufacturing firm is to check if their two-digit SIC code falls within 20-39. I depart from this convention primarily because COMPUSTAT industry codes are based on the most current financial statements of the firm. For firms who exist for long periods of time and change their products, the current COMPUSTAT industry classification may not accurately reflect whether the firm primarily operated in manufacturing over its life. Instead, I use COMPUSTAT business segment data to classify a firm as primarily engaged in manufacturing if over 50% of its deflated sales from 1980-2015 are in manufacturing industries. I also remove 38 firms who fit my initial definition of a manufacturing firm, but whose patents don't consistently correspond with their manufacturing activity.[5] Table A.1 shows that my definition of a manufacturing firm closely aligns with the standard SIC classification.[6]

After removing the patents of firms who are not manufacturing firms, I webscrape Google Patents to obtain characteristics of all the remaining patents in my sample. From this I obtain the patent's title, Cooperative Patent Classification (CPC) code, and publication claim text. In order to make the problem of classifying patents as product or process innovations tractable, I limit to firms whose primary 4-digit Standard Industrial Classification (SIC) code belongs in the top 100 patenting SIC codes. This restriction allows me to retain over 95% of the manufacturing patents in the original sample, but it reduces the number of industries from 272 to 100, making the classification problem significantly easier as I am interested in ensuring that I am able to accurately classify patent claims within each industry. The resulting sample includes 1,016,729 patents assigned to 3,092 manufacturing firms, which captures 75% of the DISCERN database. This is consistent with the findings of Autor et al. 2020 who report that more than three-quarters of corporate patents in their sample are granted to manufacturers.

3.2 Defining Product and Process Innovation

I define a product innovation as an innovation that describes a physical object[7] that a firm sells in the output market with no discussion about how the object is created. All other innovations are defined as process innovations. Scherer 1984 created one of the first large hand classification of product and process innovations from patent data and similarly conceived of process innovations as those that result in "new or improved production processes used internally within the performing company" whereas product innovations create or improve "new products" that are "sold to others."

Other modern classifications of patents into product and process innovations rely on the USPTO[8] definitions of product and process innovations (Banholzer et al. 2019; Liu et al. 2020; Bena and Simintzi 2022; Ganglmair et al. 2022). Overall, the classifications of Banholzer et al. 2019; Liu et al. 2020; Bena and Simintzi 2022; Ganglmair et al. 2022 succeed at categorizing patents according to the patent office's definition of product and process patenting. Liu et al. 2020 reports that in the case of USPTO patents, patent experts agreed with their classification 95% of the time and Ganglmair et al. 2022 reports 98% agreement between their classification and manual classification from hired classifiers. These results indicate that the data provided by BBG provide an accurate categorization of product and process patenting according to the patent office's definition of product and process patenting.

The USPTO defines a product innovation as an invention that is directed to either a machine, manufacture (article created from raw materials) or composition of matter.[9] On the other hand, process innovations define steps, acts, or methods to be performed and include a new use of a known process, machine, manufacture, composition or material. There are two types of process innovations. First, is an innovation that describes the use of an entity to achieve a technical effect. Second, is an innovation that describes a process for the production of a product. In short, the USPTO definitions focus on the distinction between nouns (products) and verbs (processes) without consideration about whether the innovation is intended for internal use or if the innovation is in reference to something that will be sold. My definition incorporates the usage of the innovation as this is the more economically relevant characteristic of the innovation. Particularly, my definition of a product innovation requires that a product innovation be both a noun (physical object) and something that is sold. Banholzer et al. 2019 note the difference between the USPTO definition of product and process claims that BBG use to classify patents and a more economically based definition when they say:

> The language used in the examination guidelines differs from the language used by economists...it is very difficult to determine whether product and process claims as defined by the guidelines can serve the typical functions economists have in mind when talking about product and process inventions.

This classification seeks to overcome these issues and create an economically grounded classification of product and process innovation. To more concretely see how these definitions work in practice, consider three patents granted to Micron Technology, a semiconductor firm specializing in computer memory production. The semiconductor industry is highly innovative, having the most patents of any industry in my data. As an example of a product patent, consider U.S. patent number 6952359, titled: "Static content addressable memory cell" and pictured in Panel (a) of Figure A.1.

This patent is for a content addressable memory cell, a product that Micron sells. The motivation for the patent is described in the text of the patent which says: "There is a...need for an alternative CAM cell design that is relatively small and yet has acceptably low soft-error rates." There is no discussion of how the product is created, making this patent a pure product innovation, meant to address shortcomings in currently available product offerings. According to the USPTO and BBG definitions, this patent would also be classified as a product innovation since it refers to a manufacture. As such, BBG all classify this patent as a pure product innovation.

Now consider Panel (b) of Figure A.1 which depicts U.S. patent number 6051074 assigned to Micron Technology that has the title: "Thermal conditioning apparatus." The description of the patent's CPC classification reads: "Apparatus specially adapted for handling semiconductor or electric solid state devices during manufacture or treatment thereof..." Further the patent goes on to state: "A problem that arises with the prior art...is that when the heating or cooling assemblies must be repaired or replaced, extensive and costly amounts of downtime occur." From the CPC description and the text of the patent, it is clear that this machine is used to more effectively produce semiconductors. According to my definition this invention is a process innovation since it describes a physical object that Micron does not sell but is used to produce physical objects that Micron will sell. I determine whether Micron sells a device such as the one in US6051074 by examining Micron's 10-K documents over the span of many years to see if Micron sells products similar to the object in the patent at any point going forward. Since Micron does not sell semiconductor manufacturing equipment, then this apparatus is classified as a process innovation which is used internally in the manufacturing process. But according to the USPTO and BBG definitions, this patent would be classified as a product innovation since it is directed at a machine/apparatus. Indeed, Banholzer et al. 2019, Bena and Simintzi 2022, and Ganglmair et al. 2022 all classify this patent as a pure product innovation, in accordance with the USPTO definition whereas I classify it as a pure process patent.

But not all inventions are strictly product or process innovations. Consider, U.S. patent number 7271654, which has the title: "Low voltage CMOS differential amplifier" and is shown in the bottom panel of Figure A.1. From the title, it would appear that the patent is for an object that Micron Technology will sell, yet the second sentence of the abstract states that: "there is provided a method of manufacturing a device..." This indicates that the patent contains information about how this object is constructed. In this sense, the patent has both a product component since it describes features of a physical object that Micron will sell, but it also has a process component since it describes how to manufacture the product.

In order to break patents down into individual components that can be classified as product or process innovations, I turn to the publication claims of the patents. U.S. patents contain individual publication claims describing precisely what they protect. The claims enumerate each invention's individual innovations.[10] According to the USPTO it is improper for a single patent claim to be directed to both a product and a process.[11] This allows me to individually classify each claim as a product or process innovation. The current patent of consideration, U.S. patent number 7271654, has four independent claims:

1. A method of manufacturing a device comprising...
2. A device comprising...
3. A method of operating a set of differential pairs comprising...
4. An input buffer comprising...

The first claim refers to a process innovation since it discusses a process used to create an object that the firm will sell. According to the definition used in this paper, the next three claims pertain to product innovations that describe the CMOS differential amplifier, along with descriptions about how to use it. But according to the USPTO definition, the third claim would be a process claim since it refers to a "method to be performed," despite the fact that it is describing how to use the product of a firm. To capture the fact that this patent contains both product and process innovations, I assign this patent a process share of 0.25 where the process share is the proportion of a patent's claims that are process innovations. In the previous two examples, all the claims were either product innovations, as in the case of the memory cell in U.S. patent number 6952359, or process innovations, as in the case of the thermal conditioning apparatus in U.S. patent number 6051074. This gives the preceding patents a process share of 0.0 and 1.0 respectively. I apply this method of individually classifying patent claims as product or process innovations and then calculate a product share for each patent. This method captures the fact that patents can contain both product and process innovations.

3.3 Classifying Patents

The final hurdle is determining a method of systematically, accurately, and efficiently classifying the over one million patents in my sample. There are two approaches I could take. The first is specifying an algorithm for deciding whether a publication claim is a product or process innovation. In the current literature, the algorithm is often an indicator function that classifies the claim as a process innovation when certain keywords are in the claim text (Banholzer et al. 2019; Liu et al. 2020; Bena and Simintzi 2022; Ganglmair et al. 2022). I could follow the spirit of this approach and create a more complex classification rule that takes into account other features of the patent such as the CPC code, firm, industry, and year. Alternatively, I could hand classify a sample of claims and then use predictive methods, such as machine learning, to predict the status of claims that I have not hand classified. This approach has been used recently in the economics literature to classify various types of patents (Chen et al. 2019; Clemens and Rogers 2020; Lerner et al. 2021) and has been used by Liu et al. 2020 to expand their USPTO keyword based classification to international patents.

I chose the latter approach for several reasons. First, after reading several hundred patents, it was clear to me that the functional form which would best identify whether a claim is a product or process innovation is a very complicated function that would be extremely difficult for me to specify. For example, the phrase "a method" followed by words such as "of" or "comprising" and then followed by a verb often indicates that the claim should be classified as a process innovation. This is true in the case of U.S. patent number 8185230, assigned to Advanced Micro Devices Inc (AMD), where the first claim states: "a method comprising: ... performing a first fabrication process on the semiconductor device..." Since AMD sells semiconductors, the claim describes a process used to create an object they will sell. This makes the claim a process innovation. Although the presence of keywords such as "a method" often indicate a claim is a process innovation, this is not always the case. For example, the last claim of U.S. patent number 8185666, assigned to Texas Instruments (TI), states: "A method of executing a single instruction, comprising...comparing the array index value to the array size value..." This is a claim related to executing instructions on a microprocessor. Since TI is a company that sells semiconductors, the claim describes a product that they will sell and should be classified as a product innovation despite the presence of the process keyword "a method." This example illustrates the difficulty in creating a pre-determined rule that could distinguish between these situations.

To classify claims as product or process innovations, I start by hand classifying the claims of 100 patents for each of the 100 4-digit SIC industries. I classify industries separately since there is significant heterogeneity in how vocabulary is linked with the distinction between product and process innovations. For industries where prediction was less accurate, I classified more claims in order to improve the accuracy and precision of the prediction. My hand classification process involved evaluating each claim in the sample, one industry at a time. For each new firm where I was unfamiliar with the products the firm sold, I retrieved several of their 10-Ks and read their product description section in order to understand what their final products were and how they may have evolved over time. I would then read the text of each claim, examining whether it met the definition of a product or process innovation as outlined above and reviewing the 10-Ks if I needed more information to make a decision. In the end, I hand classified 40,682 claims (over 14,000 patents). I then cleaned the claims text of all patents by removing stopwords, punctuation, whitespace, and numbers. Further, I lemmatize the text which involves reducing each word to its lemma in order to analyze it as a single concept. For example, "forming" would be transformed to its lemma, "form."

For each industry, I then assess the performance of sixty different specifications.[12] Each specification has three components: a ML classifier, a text feature set (which form the independent variables), and dummies for whether to drop certain coefficients. I use three different ML classifiers: multinomial naive bayes, complement naive bayes, and a passive-aggressive classifier developed by Crammer et al. 2006. I experiment with ten different text feature sets and whether to keep all features in a given feature set or drop features that are below median "importance" in predicting the outcome.[13] In total this gives me $3 \times 10 \times 2 = 60$ specifications for each industry. In Appendix A.1, I include a comprehensive list of the elements that make up each specification. I now turn to discussing how I choose a model for each industry and how I validate the quality of my classifications.

3.4 Model Validation

3.4.1 Quantitative Validation

For each of these 60 specifications, I assess its quality using repeated k-fold cross validation where I choose k = 5 (Raschka 2020). This process works by taking 20% of my hand classified claims data for the industry as the evaluation sample. I then fit the model to the other 80% of the data. I use this fitted model to make predictions about whether each claim in the 20% evaluation sample is a product or process innovation. I repeat this procedure 5 times so that each claim is in the evaluation sample exactly once. I then repeat this entire procedure k = 5 times, randomly shuffling the data each time. In the end, each claim will be classified five times. It is important to note that each time a model is fitted, the training is "one-shot" in the sense that the model finds the optimal predictive coefficients only using the data it was most recently exposed to.[14] After running this procedure for all 60 specifications, I choose the model that has the highest correlation coefficient between the vector of binary hand classifications and the vector of binary predictions. In a final step, I estimate seven more specifications where I add various features to the final selected model which again are described in detail in Appendix A.1. If any of these extra seven specifications obtain a higher correlation coefficient between the truth and the prediction then the new model is chosen;

otherwise I retain the previously chosen model. I use the correlation coefficient to assess the performance of each model as it effectively trades off type-I and type-II error, and it performs well in the case of imbalance, where one category far outnumbers the other (Chicco and Jurman 2020). This is relevant in my case since most industries are imbalanced, with fewer process innovations. In practice, the model with the highest correlation coefficient often achieves the highest balanced accuracy, which is another popular evaluation metric that also performs well in the presence of imbalance and is used in Clemens and Rogers 2020.

Before displaying aggregate statistics on quality of my match across industries, Table 1 presents the confusion matrix from the model that achieved the highest correlation coefficient for the most innovative industry in my data, the semiconductor industry. As with all the models, it was evaluated using 5-fold cross validation meaning that whether each of my 2,332 hand classified claims was a product or process innovation was predicted five times. This is reflected in the 11,660 total classifications in the bottom right-hand corner of Table 1. The row and column labeled "Total" respectively reflects the number of true and predicted claims falling into each category, where the true status of the claim is determined by my hand classification and the prediction is determined by the selected ML model. For example, 9,784 of the claims in the data where predicted to be product claims while only 9,125 were truly product claims. Of the total predicted product claims, 9,018 (92.2%) of the claims that were predicted to be product claims were actually product claims while 766 (7.8%) were actually process claims. This captures "Product Precision" which is defined as the share of claims that are predicted to be product claims which are truly product claims. "Process Precision" is similarly defined and comes in at 94.3%, indicating that when the model predicts that a claim is a process claim, it is correct 94.3% of the time.[15] "Recall" refers to a model's ability to find objects of a specific category, with "Product Recall" being defined as the share of claims which are truly product claims that are predicted to be product claims. Product recall is high at 98.8%, indicating that nearly all the claims which are truly product claims have been identified by the model.[16] "Process Recall" is substantially lower at 69.8%, indicating that nearly one third of claims which were truly process were not identified as process claims by the model.[17] Balanced accuracy combines recall across both classes and is the unweighted average of product recall and process recall and comes in at 84.3%. Balanced accuracy intuitively captures the notion that a good model will be able to recall large shares of all classes. Overall, the results from the confusion matrix indicate that the selected model performs well in classifying product and process innovations but struggles by systematically classifying a portion of the process claims as product claims.

With this illustration of the various diagnostic statistics in mind, Table 2 displays the results across industries where, for each industry, I select the model with the highest correlation coefficient and then perform 5-fold repeated cross validation for each specification. The summary statistics are calculated using weights where the weights correspond to the number of patents in the industry.

I achieve a mean correlation coefficient of 0.74 with a standard deviation of 0.09 across industries and balanced accuracy of 85%. One thing to notice from Table 2 is that process recall is lower than process precision, meaning that with respect to process innovations the ML models are making more type-II (false negative) errors than type-I (false positive). This is consistent with what was seen in the Table 1, the selected model has a hard time recalling all the process claims. To contextualize these results, I compare them to two patent categorization projects in the economics literature. Chen et al. 2019 assign financial

Table 1: Semiconductor Industry Confustion Matrix (SIC 3674)

		Prediction		
		Product	Process	Total
	Product	9,018	107	9,125
Truth	Process	766	1,769	2,535
	Total	9,784	1,876	11,660

Notes: This table presents a confusion matrix for the semiconductor (SIC 3674) industry. Rows correspond to counts of publication claims that were hand classified as product or process publication claims. Columns correspond to counts of publication claims that were classified as product or process innovations by the machine learning classifier. I hand classified the claims of 2,332 semiconductor patents, leading to 2,332*5 = 11,660 predictions since I am using 5-fold repeated cross validation. Process recall is defined as the fraction of claims which are truly process that are predicted to be process. This is calculated as $\frac{1,769}{2,535} = 69.8\%$. Product recall is defined in an analogous way and receives a score of $\frac{9,018}{9,125} = 98.8\%$. Process precision is defined as the fraction of claims that are predicted to be process which are truly process. This is calculated as $\frac{1,769}{1,876} = 94.3\%$. Product precision is defined in an analogous way and receives a score of $\frac{9,018}{9,784} = 92.2\%$. Balanced accuracy is the unweighted average of product and process recall and is $\frac{69.8+98.8}{2} = 84.3\%$

Table 2: Prediction Diagnostics

	Mean	St. Dev.	25%	50%	75%
Correlation Coefficient	0.74	0.09	0.65	0.77	0.80
Process Recall	0.73	0.13	0.65	0.77	0.81
Product Recall	0.97	0.03	0.96	0.98	0.99
Process Precision	0.84	0.11	0.81	0.85	0.91
Product Precision	0.95	0.05	0.94	0.95	0.97
Balanced Accuracy	0.85	0.06	0.81	0.86	0.89
Observations	100				

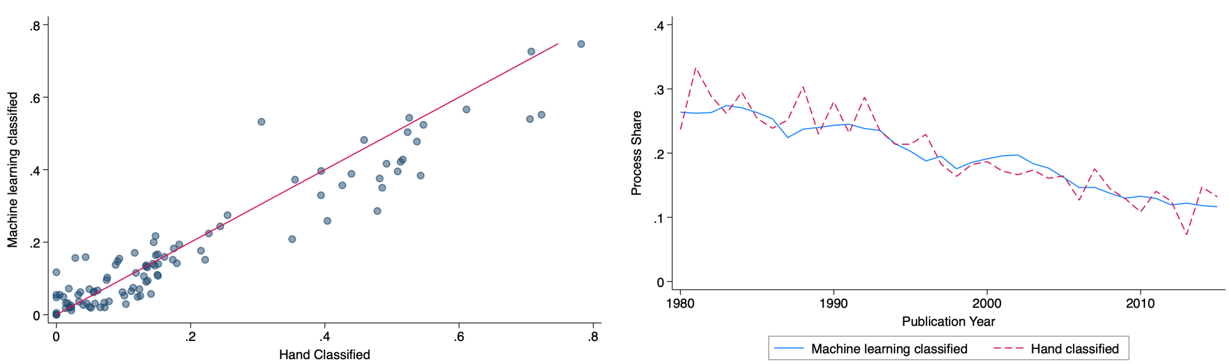
Notes: This table presents summary statistics across the 100 4-digit SIC industries weighted by the number of patents in the industry. The correlation coefficient is the correlation between the vector of publication claims predictions and true classifications. Process (product) recall refers to the fraction of true process (product) claims that were correctly classified as process (product). Process (product) precision refers the fraction of publication claims that were predicted as process (product) that are truly process. Balanced accuracy is defined as the unweighted mean of product and process recall.

patents to FinTech technologies and Clemens and Rogers 2020 categorize various features of prosthetic device patents. The diagnostic statistics in Table 2 exceed those of Chen et al. 2019 across all reported metrics, but Clemens and Rogers 2020 are able to achieve a balanced accuracy above 90%, exceeding my 85%. Therefore, the quality of my classification falls between these two patent classification projects.

To further examine the quality of the match, Panel (a) of Figure 1 examines how the process share in the hand classified and ML classified samples differs across industries where the hand classified sample comprises all the patents in an industry which I hand classify and the machine learning classified sample comprises the remaining patents in the industry which have not been hand classified. The process share for a group is defined as the mean process share across patents in the group. The points cluster around the 45 degree line with no industries being clear outliers, indicating that the mean process share within industries aligns well across the hand classified and machine learning samples. Panel (b) compares the process share in the hand classified and ML classified groups across time. Both series follow each other and display a steady downtrend over time with the hand classified sample exhibiting more variation from year to year which is consistent with the hand classified sample being much smaller than the ML classified sample. Figure A.2 confirms that the quality of the classification does not vary significantly over time as the aggregate correlation coefficient and balanced accuracy remain steady from 1980-2015.

Figure 1: Classification Robustness

(a) Process Share by Industry (b) Process Share over Time



Notes: The process share on a given patent is defined as the proportion of a patent's claims that are process innovations. Panel (a) displays the mean process share across industries with the 45 degree line corresponding to the same process share in the hand classified and ML classified data. Panel (b) displays the mean process share over time for both the ML and hand classified data.

As a validation of whether product patenting measures true product innovation, I use data provided by Hoberg and Phillips 2024 which conducts text analysis on a firm's 10-K documents to measure how many distinct product markets a firm sells products in. Using this data, I test whether patents with a high product share are associated with firms increasing the new of product markets they operate in. I estimate Equation (1) via OLS where $\#$ of Product Industries$_{fst}$ denotes the number of product industries that firm f, belonging to 4-digit SIC industry s, is operating in during year t:

$$\#\text{ of Product Industries}_{fst} = \beta \text{Product Share}_{pfst} + \phi(f, s, t) + \mathrm{X}_{ft} + \varepsilon_{pfst} \tag{1}$$

β is the coefficient of interest. If $\beta > 0$, this indicates that product patenting is associated with the firm operating in more product industries. $\phi(f, s, t)$ is a set of fixed effects which vary by specification and X_{ft} is a vector of controls which uses data from Hoberg and Maksimovic 2022 to measure where the firm is at in their product life cycle. Table 3 presents the results showing a positive and statistically significant association between product patenting and the number of product industries a firm operates in. The results indicate that product patenting is associated with the entry of firms into new product markets.

3.4.2 Qualitative Validation

In the three Micron patents shown in Figure A.1, we saw that my definition of product and process innovation yielded different hand-classifications relative to the USPTO based definitions that are used in other classifications of product and process innovations (Banholzer et al. 2019; Liu et al. 2020; Bena and Simintzi 2022; Ganglmair et al. 2022). To see if this distinction came through in claims that were classified by the machine learning classifiers, I took a random set of patents where the Banholzer et al. 2019, Bena and Simintzi 2022, and Ganglmair et al. 2022 product shares disagreed with my machine learning classified product share by at least 0.5. Consider the first of these, U.S. patent number 5086041 assigned to Monsanto Co. and falling under the NBER drugs & medical category. Its first claim is:

Table 3: Number of Product Industries and Product Patenting

	# of Product Industries		
	(1)	(2)	(3)
Product Share	0.162*	0.188**	0.111**
	(0.098)	(0.079)	(0.053)
Product Development	5.758*	4.152	3.424
	(3.487)	(3.226)	(3.320)
Process Optimization	3.039	-0.049	-0.909
	(4.467)	(2.370)	(2.009)
Mature Product	7.826**	2.790	1.263
	(3.973)	(2.523)	(2.265)
$\overline{\text{Y}}$	11.8	11.8	11.8
Year FE	✓		
Sector × Year FE		✓	
Industry × Year FE			✓
Observations	633,250	633,250	633,250

Notes: This table presents results from estimating Equation (1) via OLS. Standard errors are clustered at the firm level and shown in parentheses. * (p<0.1), ** (p<0.05), *** (p<0.01).

> A method for achieving prolonged release of a biologically active somatotropin into the circulatory system of an animal which comprises parenteral administration...

This claim is classified as a product claim by my ML classifier, while the presence of the word "method" causes it to be classified as a process innovation for BBG. The claim describes a drug Monsanto sells and does not mention method of producing the drug. The claim is therefore a product innovation in the economic sense. Misclassifying product innovations as process innovations is not isolated to drug and medical patents; it is also prevalent in the computer and communications category. Consider U.S. patent number 6469707, assigned to

NVIDIA Corp who operates in the semiconductor industry. The first claim of this patent is:

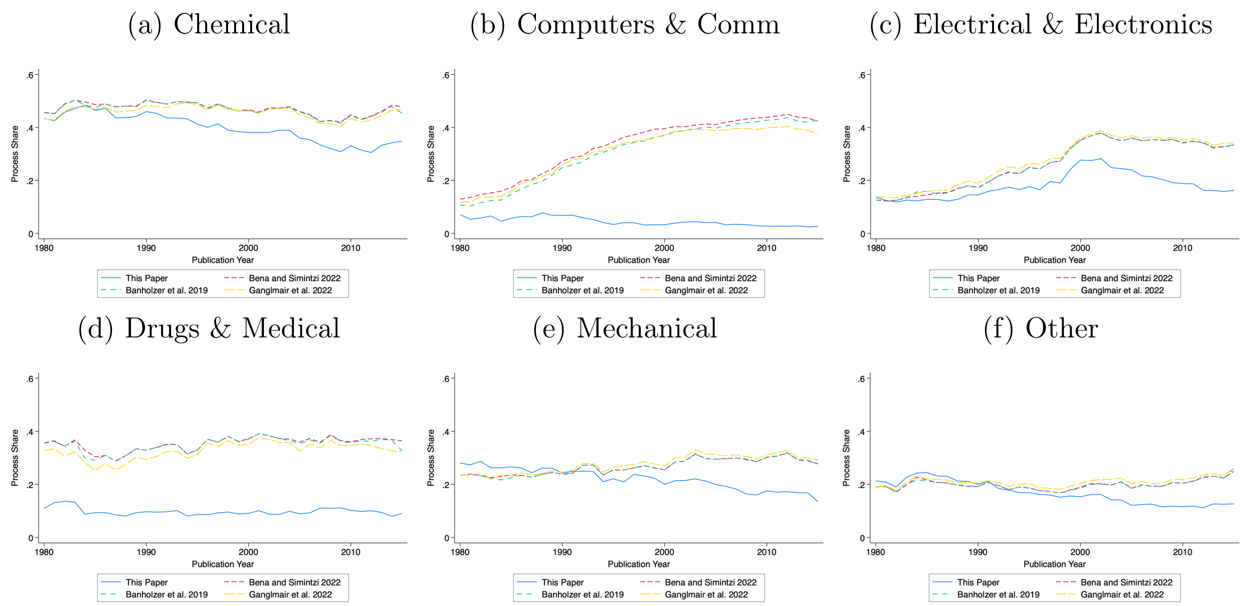
> A method for efficiently rendering and displaying color intensity information of pixels in a computer system, the pixels including a plurality of fragments, the method comprising the steps of...

Describing a computational product using "process" language is prevalent in computer and communication patents. Given that NVIDIA Corp makes graphical processing units, this claim is clearly a product innovation according to my definition, but it is identified as a process innovation using the USPTO definition since it relates to a "method to be performed." I inspected ten randomly selected patents in total and found that the patterns outlined in these two selected examples are common and that my ML classifiers are more accurately able to distinguish product and process innovations even for claims where keywords used by Banholzer et al. 2019, Bena and Simintzi 2022, and Ganglmair et al. 2022 would indicate that the claim is a process innovation. This suggests that my classification contains significantly fewer cases where product innovations are misclassified as process innovations according to the economic definition that I laid out.

How large is this systematic misclassification? Figure 2 plots the mean process share by year across broad six-digit NBER categories of patents using my classification and the classifications of BBG.

Note that the process shares in the BBG classifications (which use the USPTO definition of product and process innovation) are almost indistinguishable in every NBER category. However, they exhibit significant differences with my classification. The difference in the computer and communication category is particularly stark as the classifications using the

Figure 2: Comparison with Alternative Product/Process Classifications



Notes: Panels (a)-(f) of this figure plot the process share over time from my classification (solid navy line), the Banholzer et al. 2019 classification (dashed green line), the Bena and Simintzi 2022 classification (dashed red line), and the Ganglmair et al. 2022 classification (dashed yellow line) by NBER six-digit category.

USPTO definition of product and process innovations all exhibit a large and steady increase in the process share. In contrast, my classification has a small process share that remains steady over time. This finding is consistent with my qualitative examination of the NVIDIA patent where I found that my ML classification was able to identify the innovation as a product innovation even in the presence of a process keyword. Given that the use of process keywords to describe a firm's products is particularly common in the computer and communications sector, the divergence in classifications is not surprising. In the drugs and medical category, all four classifications exhibit flat trends, but there is a significant level difference in the share of process innovations between my measure and the USPTO based measures. Again, the lower process share in my data is consistent with my qualitative findings that many pharmaceutical products use process language to describe how they work. The other categories exhibit more similarity in levels, but in all cases my measure exhibits a steeper decline relative to the USPTO based classifications. The results presented in Figure 2 indicate that the misclassification of product innovations as process innovations is not an isolated occurrence but is large and systematic, especially in the computer & communications and drug & medical categories.

In order to help provide some color to the data, Table 4 lists the industries with the highest and lowest process shares.[18] The industry with the second-highest process share is petroleum refining which is consistent with Cohen and Klepper 1996 who use a set of hand classified patents in 1974 and comment that petroleum refining firms spend almost three-quarters of their total R&D on process innovations. Other industries with high shares of process innovation include industries engaged in metal, food, and chemical manufacturing. On the other hand, industries with low process shares are generally those producing machinery, highly specialized equipment[19], or computer and communication devices. Using the same data as Cohen and Klepper 1996, Scherer 1983 finds that process R&D was 24.6 percent of total 1974 company-financed R&D spending. Not only that, but the National Science Foundation estimated that in 1981 about 75% of industry R&D was directed to product innovations (Gilbert 2006). My estimates are similar, I find that the process share of innovations is 27.9% in 1980.

Table 4: Industries with Highest and Lowest Process Shares

SIC	SIC Desc	Process Share	Patents
Panel A: Top 15 Process Share			
1311	Crude Petroleum and Natural Gs	0.747	7,541
2911	Petroleum Refining	0.726	25,777
3720	Aircraft and Parts	0.594	2,060
3312	Steel Works and Blast Furnaces	0.568	1,038
2052	Cookies and Crackers	0.555	2,235
2000	Food and Kindred Products	0.543	874
3350	Rolling and Draw Nonfer Metal	0.530	839
3221	Glass Containers	0.525	1,256
3290	Abrasive, Asbestos, Misc Minrl	0.504	2,274
2040	Grain Mill Products	0.481	1,162
2810	Indl Inorganic Chemicals	0.429	13,718
2821	Plastics,Resins,Elastomers	0.428	5,452
2860	Industrial Organic Chemicals	0.424	23,102
2631	Paperboard Mills	0.421	2,653
3411	Metal Cans	0.419	889
Panel B: Bottom 15 Process Share			
3559	Special Industry Machy, Nec	0.000	15,335
3578	Calculate, Acct Mach, Ex Comp	0.005	1,560
3579	Office Machines, Nec	0.012	3,504
3533	Oil and Gas Field Machy, Equip	0.020	1,891
3990	Misc Manufacturng Industries	0.020	3,258
3576	Computer Communications Equip	0.020	18,853
3540	Metalworking Machinery and Eq	0.021	5,193
3829	Meas and Controlling Dev, Nec	0.022	2,201
3523	Farm Machinery and Equipment	0.022	5,845
3669	Communications Equip, Nec	0.026	1,035
3651	Household Audio and Video Eq	0.027	2,413
3826	Lab Analytical Instruments	0.030	15,688
3571	Electronic Computers	0.031	13,983
3661	Tele and Telegraph Apparatus	0.031	5,256
3823	Industrial Measurement Instr	0.031	8,050

Notes: This table presents statistics on the 15 industries with the highest process shares and the lowest process shares. Industries with less than 500 total patents from 1980-2015 are excluded from the analysis.

4 Describing Product and Process Innovation

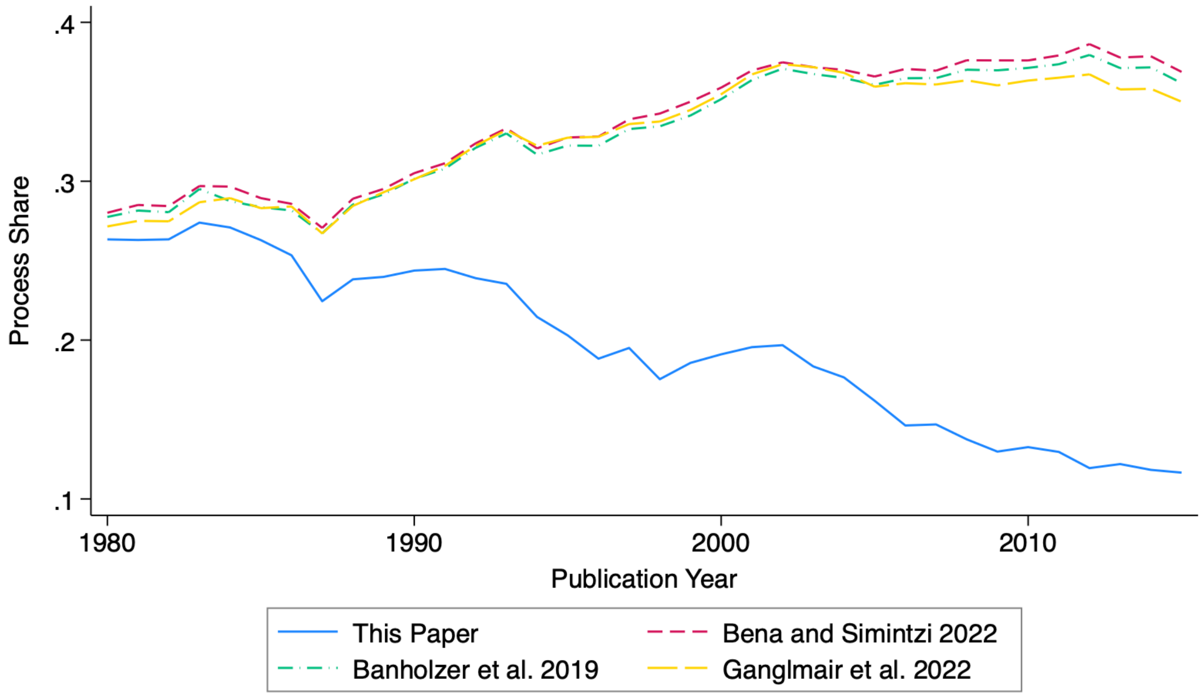
4.1 Product and Process Innovation Over Time

4.1.1 Decompositon of Time Trend

I now turn to understanding the time trends in the process share of innovation. Figure 3 reveals that my classification exhibits a large and secular decline in the process share of innovation which went from 26% in 1980 to 12% in 2015. This stands in contrast to the increase in process innovation when using the BBG data that relies on the USPTO definitions of product and process patenting. The results here are consistent with Figure 2 and the qualitative analysis of the NVIDIA and Monsanto patents which showed that the USPTO classification more often classifies a firm's products as process innovations due to the presence of process keywords that describe the functions of a firm's products.

To better understand the time trend in the process share of innovation using this paper's data, I perform a within-between decomposition which seeks to address whether the decline in the process share is due to a decline in the process share within different groups of innovative activity or a shift towards innovation that has a lower process share. There are reasons to suppose that the within or between components play a larger role. Figure 2 shows that the computer & communications category of innovation has a low process share across the entire time period. A larger share of patenting in these technological areas could cause the process share to decline through a composition effect. On the other hand, Figure 2 shows that most NBER categories of patenting saw a decline in the process share within their category over time. This suggests that within effects could explain the overall decline of the process share.

Figure 3: Process Share Over Time



Notes: This figure shows the overall process share over time for patents in my sample using the data from this paper and the data in BBG.

To formally test how much of the effect can be explained by within and between components, I follow Baily et al. 1992 and decompose the change in the process share into within, between, cross, entry, and exit components to Equation (2) where S, E, X respectively denote the set of surviving groups, entering groups, and exiting groups.

$$
\begin{aligned}
\Delta\text{Process Share} = &\sum_{g \in S} \Delta\text{Process Share}_g \times \text{Patent Share}_{g,t-1} + \\
&\sum_{g \in S} \Delta\text{Patent Share}_g \times \text{Process Share}_{g,t-1} + \\
&\sum_{g \in S} \Delta\text{Process Share}_g \times \Delta\text{Patent Share}_g + \\
&\sum_{g \in E} \text{Process Share}_{g,t} \times \text{Patent Share}_{g,t} + \\
&\sum_{g \in X} -1 \times \text{Process Share}_{g,t-1} \times \text{Patent Share}_{g,t-1}
\end{aligned}
\tag{2}
$$

The first term[20] captures how much the process share changes within a group holding its share of patents constant. The second term[21] captures how much of the change in the total process share is explained by the patent share shifting to groups with different process shares. The third term[22], referred to as the cross component, captures the fact that groups which have large changes in their process share may also have large changes in their share of patents. The fourth term[23] captures the contribution of entering groups to the change in the process share which can conceptually be thought of as a between change brought by new entrants. The last term[24] is the exit component which measures the contribution of exiting groups to the overall change in the process share.

Table 5 displays the results when the 4-digit SIC industry of the assignee firm is used as the group variable. In 1980, the overall process share was 26.3%, falling to 11.7% by the end of the 2010-2015 time period, a decline of approximately 14.7 percentage points. In the four

Table 5: Decomposition of Process Share Time Trends Over SIC Industry

Time Period	Process Share$_0$	Within	Between	Cross	Entry	Exit	Total
1980-1985	26.34	.03	.11	-.21	.05	-.02	-.04
1985-1990	26.3	.2	-2.32	.19	0	0	-1.92
1990-1995	24.38	-.78	-3.33	.03	.01	0	-4.08
1995-2000	20.3	-.36	-.99	.5	0	-.35	-1.2
2000-2005	19.1	-2.82	-.01	-.23	.18	-.04	-2.92
2005-2010	16.18	-3.22	-.1	.4	0	-.01	-2.92
2010-2015	13.26	-1.33	-.36	.1	.01	0	-1.6
Total		-8.28	-7	.78	.25	-.42	-14.68

Notes: This table presents the decomposition of the percentage point change in the mean process share of patents into within, between, cross, entry, and exit components with 4-digit SIC industries of the assignee firm being the group variable. Process Share$_0$ indicates the overall process share at the beginning of the time period.

five-year intervals going from 1980-2000, the within industry decline in the process share was relatively small, only amounting to a 0.92 percentage point decline while the between component accounted for 6.52 percentage points of the 7.24 percentage point decline. From 1980-2000 the cross term is negligible. The entry and exit terms are included as some industries do not have patenting activity in a given year, but these terms are negligible as well. The results clearly show that a reallocation of patenting activity towards industries with lower process shares was the primary driver of the decline in the process share from 1980-2000. Table A.3 uses the NBER category of patents, which is a more broad measure of a patent's location in technological space, as the group measure and confirms that the between component is responsible for the majority of the decline in the process share from 1980-2000. Table A.4 uses the firm as the group which naturally places more weight on entry and exit as firms are more likely to enter and exit the sample than industries or NBER categories. The results in Table A.4 highlight that a large portion of the between component from 1980-2000 is explained by firms with large process shares exiting.

Table 5 shows that the results flip during the 2000-2015 time period when the within component explains nearly all the decline in the process share, accounting for 7.38 percentage points of the 7.44 percentage point decline in the overall process share, indicating that industries became less process intensive from 2000-2015. Table A.3 and Table A.4 show that the within component explains less of the 2000-2015 decline in the process share, indicating that reallocation of innovative activity within industries but between NBER categories and firms plays a role in explaining the large within component from 2000-2015. Figure A.3 presents counterfactual situations where only the within or between components are allowed to contribute to the change in the process share, confirming that between industry reallocation was the main driver of the decline from 1980-2000 while within industry declines drove the 2000-2015 decline in the process share. Overall, the results indicate that the share of U.S. innovation being devoted to process innovation has been falling since 1980 and that within industry changes have been driving the recent decline in the process share.

4.1.2 China Shock

Given the large within industry decline in the process share from 2000-2015, it is natural to ask what factors contribute to this aggregate shift in the orientation of patenting? One of the largest shocks to the US economy during this time period was the rise of Chinese manufacturing. Autor et al. 2020 show that firms operating in industries that experienced more Chinese import penetration reduced their overall patenting output. But if the China shock disproportionately reduced process patenting, then this could explain the large within industry decline in the process share from 2000-2015. To examine whether Chinese import penetration had differential effects on product and process patenting, I follow the empirical strategy of Autor et al. 2020 and estimate the following stacked first differences specification over the 1991-1999 and 1999-2007 time periods

$$\Delta Y_{fsj\tau} = \beta \Delta \mathrm{IP}_{j\tau}^{US} + \gamma_s + \delta_\tau + \mathrm{X}_{fsj0} + \varepsilon_{fsj\tau}$$

$\Delta Y_{fsj\tau}$ is a differenced outcome variable of interest over time period, τ, for firm f, belonging to manufacturing sector s, operating in 4-digit SIC j. As in Autor et al. 2020, the dependent variable across specifications is the DHS growth rate[25] of patenting for a firm which allows me to control for both intensive and extensive margin changes in the dependent variable (Davis et al. 1998). Manufacturing sector dummies (γ_s) effectively limit comparisons to be made within sectors. Autor et al. 2020 show the importance of comparing within sector as sectors were on different growth paths[26] over the 1975-2007 time period. Time period dummies control for average growth rates of patenting during the time period. X_{fsj0} is a vector of firm level controls measured before the rise of China. I follow Autor et al. 2020 and weight all regressions by the average total number of patents over the start and end periods.

The independent variable of interest is $\Delta \mathrm{IP}_{j\tau}^{US}$ which is the change in Chinese import penetration[27] over the time period. As changes in import penetration to the U.S. are endogenous, I follow Autor et al. 2020 and instrument for U.S. import penetration with the change in Chinese import penetration experienced by eight other high income (OTH) countries. This instrument is valid when demand shocks across the U.S. and the OTH countries are uncorrelated. While this condition is unlikely to be met entirely, I perform pre-trend analyses to show that future Chinese import competition is not predictive of past changes in outcomes. I standardize import penetration to have mean zero and standard deviation of one for ease of interpretation.

Table 6: China Shock

	1991-1999 & 1999-2007				1983-1991	
	(1) Product	(2) Product	(3) Process	(4) Process	(5) Product	(6) Process
ΔIP	-10.01***	-10.60***	-10.87**	-11.70**	-1.55	-2.10
	(3.52)	(3.06)	(5.53)	(4.56)	(8.38)	(8.93)
$\frac{\text{Computer Invest}_{i,90}}{\text{Total Invest}_{i,90}}$		-3.19***		-4.00***	-1.06	-1.97*
		(1.19)		(1.27)	(0.95)	(1.10)
$\frac{\text{High Tech Invest}_{i,90}}{\text{Total Investment}_{i,90}}$		-133.39		-158.22	111.58	419.15
		(219.38)		(224.04)	(222.42)	(305.94)
$\frac{\text{Production Emp}_{i,91}}{\text{Total Emp}_{i,91}}$		-171.41***		-189.15***	-5.69	0.48
		(38.85)		(47.29)	(70.18)	(71.09)
$\frac{\text{Capital}_{i,91}}{\text{Value Added}_{i,91}}$		-11.58		-10.43	-12.96	-9.60
		(11.51)		(12.44)	(10.70)	(10.87)
$\ln(\overline{\text{Wage}_{i,91}})$		12.20		25.38	23.48	67.00
		(26.89)		(32.79)	(59.46)	(61.60)
F-Stat	43.67	69.56	43.67	69.56	92.00	92.00
Sector FE	✓	✓	✓	✓	✓	✓
Time Period FE	✓	✓	✓	✓	✓	✓
Observations	1,871	1,871	1,871	1,871	574	574

Notes: Columns (1)-(4) present results from estimating regressions where stacked DHS growth rates (multiplied by 100) between the firm's 1991-1999 and 1999-2007 product or process patent stock is the dependent variable. The independent variable of interest is the change in 4-digit SIC industry import penetration from China to the US which is instrumented for by the change in import penetration experienced by other developed countries (see Autor et al. 2020 for more details). Manufacturing sector dummies are included in all specifications. Columns (5)-(6) utilize the same specification but for the single time period 1983-1991. The sample is comprised of firms covered by the EPP (Davison 2023). Kleibergen-Paap F-statistics are used and standard errors are clustered at the industry level and shown in parentheses. * (p<0.1), ** (p<0.05), *** (p<0.01).

Table 6 presents the results. In all specifications, the Kleinbergen-Paap F-statistic exceeds 40, eliminating the concern for weak instruments. In column (1) which includes manufacturing sector and time period fixed effects, but no other controls, the effect of a one standard deviation increase in Chinese import penetration is to lower product patenting by 10%. In column (3), I change the dependent variable to be the DHS growth rate of process patenting, but I get a very similar magnitude decline in process patenting in response to Chinese import penteration. In both cases, columns (2) and (4) reveal that adding controls does little to alter the result. Finally, columns (5) and (6) indicate that the results are not a continuation of pre-existing trends as the change in Chinese import penetration from the 1991-2007 time period is not correlated with the growth of product or process patenting during the 1983-1991 time period.

Overall, the results indicate that the China shock, writ large, is not responsible for the decline in the process share of innovation as it negatively affected product and process patenting equally. This result must be put in context of Bena and Simintzi 2022 who find that offshoring to China reduced process patenting for US firms. The results in Table 6 are consistent with the large decline in product patenting that Bena and Simintzi 2022 find, but I also observe that the rise of China led to an equally large decrease in US product patenting, leading to no change in the process share of innovation.

4.2 Product and Process Innovation Over the Firm's Life Cycle

While Section 4.1 addressed how the overall process share evolved over time, I now turn to empirically examining how the process share of innovation varies over the product life cycle of the firm. To measure where a firm is located in their product life cycle, I use the measure from Hoberg and Maksimovic 2022 which places firm $\times$ year observations into the four stages of the Utterback and Abernathy 1978 model: product development, process optimization, mature product, and product decline. Hoberg and Maksimovic 2022 use the 10-K filings of firms along with text analysis techniques in order to determine the intensity with which firms refer to words and phrases which are associated with the four life cycle stages. Recognizing that firms are not exclusively located in one of the four stages, Hoberg and Maksimovic 2022 instead measure each firm's location in the product life cycle using a four-element vector, $\{\text{Product Development}, \text{Process Optimization}, \text{Mature Product}, \text{Product Decline}\}$, where each element in the vector corresponds to the number of paragraphs in the firm's 10-K which are associated with the life cycle stage, scaled by four times the summed paragraph count. The vector sums to unity and is populated by non-negative elements which are bounded between zero and one inclusively. The measure is only available for firms from 1997-2017, and I match it to patents based on the patent's year of application.

To test the whether the process share of innovation varies over the life cycle as predicted by Klepper 1996 and Utterback and Abernathy 1975, I estimate regressions of the following form where $\text{Process Share}_{pfst}$ denotes the share of publication claims which are process innovations for patent p, firm f, 4-digit SIC industry s, and publication year t:

$$\begin{aligned}\text{Process Share}_{pfst} =& \beta_1 \text{Product Development} + \beta_2 \text{Mature Product} + \\ & \beta_3 \text{Product Decline} + \phi(f, s, t) + \text{X}_{ft} + \varepsilon_{pfst}\end{aligned} \tag{3}$$

Since the life cycle vector sums to unity, I choose the process optimization stage to serve as the omitted category as the theories predict that firms will engage in the most process innovation when in the process optimization life cycle stage. $\phi(f, s, t)$ is a set of fixed effects which vary by specification and X_{ft} is a vector of controls which includes a firm's age.[28] Standard errors are clustered at the firm level.

Table 7: Process Share Over the Life Cycle

	Process Share		
	(1)	(2)	(3)
	OLS	OLS	OLS
Product Development	-0.061**	-0.099**	-0.075*
	(0.029)	(0.044)	(0.040)
Mature Product	-0.020	-0.040*	-0.026
	(0.021)	(0.023)	(0.020)
Product Decline	-0.019	-0.043	-0.032
	(0.037)	(0.036)	(0.034)
Firm Age			0.005***
			(0.002)
$\overline{Y}$	0.15	0.15	0.15
Firm FE	✓	✓	✓
Year FE	✓		
4-digit SIC × Year FE		✓	✓
Observations	630,690	630,690	630,690

Notes: This table presents results from estimating Equation (3) using OLS. Standard errors are clustered at the firm level and shown in parentheses. *(p<0.1), **(p<0.05), ***(p<0.01).

Column (1) of Table 7 includes firm and year fixed effects, removing time-invariant firm heterogeneity and common annual effects. The point estimates of β_1, β_2, and β_3 are negative but only statistically distinguishable from zero in the case of the coefficient on product development. The results indicate that a firm who is entirely in the product development stage will have patents that have six percentage points lower process share. Off a mean process share of 0.15, this would be a 40% reduction in the process share. This aligns with Klepper 1996 and Utterback and Abernathy 1975 who hypothesize that the process share of innovation will be lowest in the earliest stages of the life cycle when the firm has a small scale and is focused most on product development. Next, the firm enters a stage where they focus on process improvement and increase their scale which allows them to take advantage of the fact that the return to process innovations increase with the size of the firm (Cohen and Klepper 1996). The results in column (1) indicate that this omitted process improvement stage is the stage where the process share of innovation is the highest. Next the firm enters the mature product stage where they have large scale, which incentivizes process innovation, but they no longer are in the stage of rapid process improvement as the firm's processes become more integrated and developed. Finally, the firm enters the stage of product decline where they have even less incentive to do process innovation since their scale begins to decline. In column (2) I tighten the specification by replacing year fixed effects with industry × year fixed effects. In this tighter specification, the estimates become larger in absolute value and the coefficient on the mature product stage becomes marginally significant. The results continue to align with the hypothesis that the product development stage will have the lowest process intensity, the process improvement stage will have the highest process intensity, and the final two stages will have intermediate levels of process intensity. In column (3) I control for firm age, which is related but distinct from the notion of where the firm is in the product life cycle. The coefficient on firm age is positive and highly significant, consistent with the hypothesis of Klepper 1996 that firms increase their process intensity as they age. Although the life cycle estimates become less precise in column (3), the same general pattern holds.

To examine whether the results change when using the alternate process classifications of BBG, I estimate Equation (3) but with the process share of each patent taken from the three alternate classifications. Table A.5, Table A.6, and Table A.7 provide the results with the coefficients being largely statistically insignificant and positive, indicating no significant difference in the process share of innovation as firms move through their life cycle.[29] The results highlight the importance of using data based on an economically grounded definition of product/process innovation for understanding how firms change their process innovation throughout their life cycle.

4.3 Product and Process Innovation and Firm Size

One of the theorized mechanisms for why firm increase their process innovation intensity as they move through their life cycle is the issue of scale (Klepper 1996). The returns to process innovation are naturally increasing in the output of the firm as a larger firm can apply their process innovation to a larger amount of output (Cohen and Klepper 1996). To directly test whether this mechanism is at play, I alter Equation (3) by replacing the measures of a firm's position in their life cycle (Product Development, Mature Product, Product Decline) with a measure of firm size. I expect to find a positive relationship between the size of the firm and their process intensity. Table 8 reports the results when I use the natural logarithm of deflated total revenue of the firm in a given year as a measure of firm size. In column (1), the estimate on log revenue is positive and highly significant, indicating a positive relationship between the revenue of the firm and their process intensity. Column (2) accounts for time-varying industry heterogeneity in the process share by including industry $\times$ year fixed effects, leaving the coefficient on log revenue materially unchanged. As firms move through their life cycle they grow, creating a correlation between firm size, age, and product life cycle (Haltiwanger et al. 2013). In column (3), I control for a firm's stage in the life cycle and I continue to find a highly significant and positive relationship between log revenue and the process share. This suggests that the positive relationship between firm size and the process share of innovation is not entirely accounted for by the firm's position in their life cycle. The coefficients on the life cycle stages continue to be negative and significant except in the case of the product decline stage, indicating that life cycle plays an important role in determining process intensity even after controlling for firm size. In column (4), I directly control for a firm's age which approximately halves the coefficient on log revenue and leaves it statistically insignificant, suggesting that while firm size plays a role in determining the process share of a firm, the life cycle and age of the firm play an even more important role. Table A.8 shows that the results are robust to using log number of employees as the measure of firm size. When using the BBG process classifications to examine the relationship between firm size and the process share, I find statistically insignificant and largely negatively signed coefficients on log revenue in Table A.9, Table A.10, and Table A.11. Again, the results indicate the distinction between an economically grounded definition of product and process innovation and the USPTO definition is important for understanding the relationship between firm size and the process share. Overall, the results indicate that both firm size and a firm's position in the life cycle and age play an important role in determining how a firm chooses between product and process innovation.

Table 8: Process Share and Log Revenue

	Process Share			
	(1)	(2)	(3)	(4)
	OLS	OLS	OLS	OLS
ln(Revenue)	0.010**	0.009***	0.010***	0.004
	(0.004)	(0.003)	(0.003)	(0.003)
Product Development			-0.101**	-0.077*
			(0.045)	(0.040)
Mature Product			-0.044*	-0.029
			(0.023)	(0.020)
Product Decline			-0.039	-0.032
			(0.035)	(0.035)
Firm Age				0.005***
				(0.002)
$\overline{\text{Y}}$	0.15	0.15	0.15	0.15
Firm FE	✓	✓	✓	✓
Year FE	✓			
4-digit SIC × Year FE		✓	✓	✓
Observations	629,733	629,733	629,733	629,733

Notes: This table presents results from estimating Equation (3) using OLS where a measure of firm size is added. Standard errors are clustered at the firm level and shown in parentheses. *(p<0.1), **(p<0.05), ***(p<0.01).

5 Conclusion

This paper introduces a novel measurement of a firm's product and process innovation using the text of patents and ML classifiers. Using this data, I document a large secular decline in the process share of innovation from 1980-2015. I find that reallocation of patenting activity between industries explains the decline in the process share from 1980-2000 while within industry declines in the process share explain the decline from 2000-2015. Other classifications, which use the USPTO definition of product and process innovation find that the process share of innovation has been increasing over time. I provide evidence that this discrepancy is due to innovations which are economically product innovations being marked as process innovations due to the presence of process language[30]. Next, I examine how the process share of innovation varies over the firm's product life cycle. I find that firms have low process shares early in the product life cycle, followed by high process shares and then plateauing at an intermediate process share. These findings are consistent with the life cycle theory of Utterback and Abernathy 1975. Replicating the same empirical methodology with the USPTO based classifications of BBG produce statistically insignificant results.

In regard to the economic effects of this declining process share, other work has provided evidence that process innovations generate fewer knowledge spillovers relative to product innovations (Kotabe and Murray 1990; Kraft 1990; Ornaghi 2006; Davison 2022). Given the importance of knowledge spillovers in driving a wedge between the social and private values of innovation, this decline in the process share has implications for the socially optimal amount of innovation (Nelson 1959). Further, for Aghion et al. 2023, the difference in knowledge spillovers plays a key role in determining the social planner's preference for product innovation over process innovation. Exploring the economic implications and the welfare effects of this decline in the process share is a rich area for future research.

How the firm innovates over its product life cycle has been highlighted as an important factor in determining market structure and productivity growth (Dasgupta and Stiglitz 1980; Klepper 1996; Huergo and Jaumandreu 2004; Coad et al. 2016). Given the empirical support that this paper provides for the product life cycle theory of Utterback and Abernathy 1975 and the connection between firm size and process innovation (Cohen and Klepper 1996), more work examining the implications of these findings would be valuable. For example, does the path of process innovation over the product life cycle provide explanations for the resulting market structure in industries? In addition, does the connection between firm size and process innovation imply anything about optimal market structure?

While this paper documents the large decline in the process share of innovation for U.S. manufacturing firms from 1980−2015, I do not advance an explanation for this decline. In different settings, Branstetter et al. 2021 and Bena and Simintzi 2022 use alternate classifications of process innovation and find negative effects of offshoring on the process share of innovation. These results suggest that the rise in U.S. offshoring may play a key role in explaining the declining process share. A further examination of this and other factors which contribute to the declining process share would be welcome. In addition, if the U.S. continues to focus on product innovation while other countries focus on process innovation, it may be the case that global process innovation is increasing despite the decline in the U.S. process innovation share. Better measurement of global process innovation would help us understand how the specialization of R&D activities has affected aggregate innovation.

Future research could examine the complementarity of product and process innovation. Do firms who outsource their process innovation see the quality of their product innovation suffer? While outsourcing manufacturing and process innovation provides short-term financial gain, it is possible that it atrophies a firm's product innovation capabilities in the long-run. To help address these questions and further our understanding of product and process innovation, this paper provides a new publicly available classification of product and process innovations, referred to as the Economically Based Product Process Patent Dataset (EPP), which will allow researchers to make progress on these questions (Davison 2023).

References

Aghion, Philippe et al. (2023). "Good Rents versus Bad Rents: R&D Misallocation and Growth". en.

Antonucci, Tommaso and Mario Pianta (July 2002). "Employment Effects of Product and Process Innovation in Europe". In: *International Review of Applied Economics* 16.3. Publisher: Routledge, pp. 295–307. ISSN: 0269-2171. DOI: 10.1080/02692170210136127.

Arora, Ashish, Sharon Belenzon, and Andrea Patacconi (2018). "The decline of science in corporate R&D". en. In: *Strategic Management Journal* 39.1, pp. 3–32. ISSN: 1097-0266. DOI: 10.1002/smj.2693.

Arora, Ashish, Sharon Belenzon, and Lia Sheer (Mar. 2021). "Knowledge Spillovers and Corporate Investment in Scientific Research". en. In: *American Economic Review* 111.3, pp. 871–898. ISSN: 0002-8282. DOI: 10.1257/aer.20171742.

Arundel, Anthony et al. (Dec. 2007). "How Europe's economies learn: a comparison of work organization and innovation mode for the EU-15". In: *Industrial and Corporate Change* 16.6, pp. 1175–1210. ISSN: 0960-6491. DOI: 10.1093/icc/dtm035.

Autor, David et al. (Sept. 2020). "Foreign Competition and Domestic Innovation: Evidence from US Patents". en. In: *American Economic Review: Insights* 2.3, pp. 357–374. DOI: 10.1257/aeri.20180481.

Baily, Martin Neil et al. (1992). "Productivity Dynamics in Manufacturing Plants". en. In: *Brookings Papers on Economic Activity. Microeconomics* 1992, p. 187. ISSN: 10578641. DOI: 10.2307/2534764.

Banholzer, Nicolas et al. (2019). *Knowledge Spillovers from Product and Process Inventions in Patents and their Impact on Firm Performance.*

Bena, Jan and Elena Simintzi (Dec. 2022). *Machines Could Not Compete with Chinese Labor: Evidence from U.S. Firms' Innovation*. en. SSRN Scholarly Paper. Rochester, NY. DOI: 10.2139/ssrn.2613248.

Bertschek, Irene (1995). "Product and Process Innovation as a Response to Increasing Imports and Foreign Direct Investment". In: *The Journal of Industrial Economics* 43.4. Publisher: Wiley, pp. 341–357. ISSN: 0022-1821. DOI: 10.2307/2950548.

Bloom, Nicholas et al. (Apr. 2020). "Are Ideas Getting Harder to Find?" en. In: *American Economic Review* 110.4, pp. 1104–1144. ISSN: 0002-8282. DOI: 10.1257/aer.20180338.

Branstetter, Lee G. et al. (Aug. 2021). *Does Offshoring Production Reduce Innovation: Firm-Level Evidence from Taiwan*. Working Paper. DOI: 10.3386/w29117.

Chen, Mark A, Qinxi Wu, and Baozhong Yang (May 2019). "How Valuable Is FinTech Innovation?" In: *The Review of Financial Studies* 32.5, pp. 2062–2106. ISSN: 0893-9454. DOI: 10.1093/rfs/hhy130.

Chicco, Davide and Giuseppe Jurman (Jan. 2020). "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation". In: *BMC Genomics* 21.1, p. 6. ISSN: 1471-2164. DOI: 10.1186/s12864-019-6413-7.

Classen, Nicolas et al. (Mar. 2014). "Innovation in family and non-family SMEs: an exploratory analysis". en. In: *Small Business Economics* 42.3, pp. 595–609. ISSN: 1573-0913. DOI: 10.1007/s11187-013-9490-z.

Clemens, Jeffrey and Parker Rogers (Jan. 2020). *Demand Shocks, Procurement Policies, and the Nature of Medical Innovation: Evidence from Wartime Prosthetic Device Patents.* Working Paper. DOI: 10.3386/w26679.

Coad, Alex, Agustí Segarra, and Mercedes Teruel (Mar. 2016). "Innovation and firm growth: Does firm age play a role?" In: *Research Policy* 45.2, pp. 387–400. ISSN: 0048-7333. DOI: 10.1016/j.respol.2015.10.015.

Cohen, Wesley M. and Steven Klepper (1996). "Firm Size and the Nature of Innovation within Industries: The Case of Process and Product R&D". In: *The Review of Economics and Statistics* 78.2. Publisher: The MIT Press, pp. 232–243. ISSN: 0034-6535. DOI: 10.2307/2109925.

Crammer, Koby et al. (2006). "Online Passive-Aggressive Algorithms". In: *Journal of Machine Learning Research* 7.19, pp. 551–585. ISSN: 1533-7928.

Damanpour, Fariborz (2010). "An Integration of Research Findings of Effects of Firm Size and Market Competition on Product and Process Innovations". en. In: *British Journal of Management* 21.4, pp. 996–1010. ISSN: 1467-8551. DOI: 10.1111/j.1467-8551.2009.00628.x.

Dasgupta, Partha and Joseph Stiglitz (1980). "Uncertainty, Industrial Structure, and the Speed of R&D". In: *The Bell Journal of Economics* 11.1. Publisher: [RAND Corporation, Wiley], pp. 1–28. ISSN: 0361-915X. DOI: 10.2307/3003398.

Davis, Steven J., John Haltiwanger, and Scott Schuh (1998). *Job Creation and Destruction.*

Davison, Colin (2022). "Knowledge Spillovers of Products and Processes". en.

— (Sept. 2023). *Economically Based Product Process Patent Dataset (EPP)*. DOI: 10.5281/zenodo.8364951.

Demmel, M. Constanza et al. (2017). "Product and process innovation and total factor productivity: Evidence for manufacturing in four Latin American countries". en. In: *Review of Development Economics* 21.4, pp. 1341–1363. ISSN: 1467-9361. DOI: 10.1111/rode.12323.

Dickinson, Victoria (Nov. 2011). "Cash Flow Patterns as a Proxy for Firm Life Cycle". In: *The Accounting Review* 86.6, pp. 1969–1994. ISSN: 0001-4826. DOI: 10.2308/accr-10130.

Diéguez-Soto, Julio, Aurora Garrido-Moreno, and Montserrat Manzaneque (June 2018). "Unravelling the link between process innovation inputs and outputs: The moderating role of family management". In: *Journal of Family Business Strategy* 9.2, pp. 114–127. ISSN: 1877-8585. DOI: 10.1016/j.jfbs.2017.11.007.

Fritsch, Michael and Monika Meschede (Nov. 2001). "Product Innovation, Process Innovation, and Size". en. In: *Review of Industrial Organization* 19.3, pp. 335–350. ISSN: 1573-7160. DOI: 10.1023/A:1011856020135.

Ganglmair, Bernhard, W. Keith Robinson, and Michael Seeligson (Apr. 2022). *The Rise of Process Claims: Evidence from a Century of U.S. Patents*. en. SSRN Scholarly Paper. Rochester, NY. DOI: 10.2139/ssrn.4069994.

Gilbert, Richard (Jan. 2006). "Looking for Mr. Schumpeter: Where Are We in the Competition–Innovation Debate?" In: *Innovation Policy and the Economy* 6. Publisher: The University of Chicago Press, pp. 159–215. ISSN: 1531-3468. DOI: 10.1086/ipe.6.25056183.

Griffith, Rachel et al. (Dec. 2006). "Innovation and Productivity Across Four European Countries". In: *Oxford Review of Economic Policy* 22.4, pp. 483–498. ISSN: 0266-903X. DOI: 10.1093/oxrep/grj028.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg (Oct. 2001). *The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools*. Working Paper. DOI: 10.3386/w8498.

Haltiwanger, John, Ron S. Jarmin, and Javier Miranda (May 2013). "Who Creates Jobs? Small versus Large versus Young". In: *The Review of Economics and Statistics* 95.2, pp. 347–361. ISSN: 0034-6535. DOI: 10.1162/REST_a_00288.

Haneda, Shoko and Keiko Ito (Feb. 2018). "Organizational and human resource management and innovation: Which management practices are linked to product and/or process innovation?" In: *Research Policy* 47.1, pp. 194–208. ISSN: 0048-7333. DOI: 10.1016/j.respol.2017.10.008.

Hoberg, Gerard and Vojislav Maksimovic (Sept. 2022). "Product Life Cycles in Corporate Finance". In: *The Review of Financial Studies* 35.9, pp. 4249–4299. ISSN: 0893-9454. DOI: 10.1093/rfs/hhab134.

Hoberg, Gerard and Gordon M. Phillips (2024). "Scope, Scale, and Concentration: The 21st-Century Firm". en. In: *The Journal of Finance* n/a.n/a. _eprint: https://onlinelibrary.wiley.com/doi/pd ISSN: 1540-6261. DOI: 10.1111/jofi.13400.

Huergo, Elena and Jordi Jaumandreu (Apr. 2004). "Firms' age, process innovation and productivity growth". In: *International Journal of Industrial Organization* 22.4, pp. 541–559. ISSN: 0167-7187. DOI: 10.1016/j.ijindorg.2003.12.002.

Jones, Benjamin F. (Jan. 2009). "The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder?" In: *The Review of Economic Studies* 76.1, pp. 283–317. ISSN: 0034-6527. DOI: 10.1111/j.1467-937X.2008.00531.x.

Klepper, Steven (1996). “Entry, Exit, Growth, and Innovation over the Product Life Cycle”. In: *The American Economic Review* 86.3. Publisher: American Economic Association, pp. 562–583. ISSN: 0002-8282.

Kotabe, Masaaki and Janet Y. Murray (1990). “Linking Product and Process Innovations and Modes of International Sourcing in Global Competition: A Case of Foreign Multinational Firms”. In: *Journal of International Business Studies* 21.3. Publisher: Palgrave Macmillan Journals, pp. 383–408. ISSN: 0047-2506.

Kraft, Kornelius (Aug. 1990). “Are product and Process Innovations Independent of Each Other?” In: *Applied Economics* 22.8. Publisher: Routledge, pp. 1029–1038. ISSN: 0003-6846. DOI: 10.1080/00036849000000132.

Lerner, Josh et al. (July 2021). *Financial Innovation in the 21st Century: Evidence from U.S. Patents*. Working Paper. DOI: 10.3386/w28980.

Liu, Leo, Elvira Sojli, and Wing Wah Tham (Nov. 2020). *Advanced Manufacturing, Product Innovations, Productivity and Growth*. en. SSRN Scholarly Paper. Rochester, NY. DOI: 10.2139/ssrn.4104903.

Mairesse, Jacques et al. (2005). “The Importance of R&D and Innovation for Productivity: A Reexamination in Light of the French Innovation Survey”. In: *Annales d’Économie et de Statistique* 79/80. Publisher: [GENES, ADRES], pp. 487–527. ISSN: 0769-489X. DOI: 10.2307/20777586.

Marco, Alan C., Joshua D. Sarnoff, and Charles A. W. deGrazia (Nov. 2019). “Patent claims and patent scope”. In: *Research Policy* 48.9, p. 103790. ISSN: 0048-7333. DOI: 10.1016/j.respol.2019.04.014.

Nelson, Richard R. (June 1959). "The Simple Economics of Basic Scientific Research". In: *Journal of Political Economy* 67.3. Publisher: The University of Chicago Press, pp. 297–306. ISSN: 0022-3808. DOI: 10.1086/258177.

Ornaghi, Carmine (Mar. 2006). "Spillovers in product and process innovation: Evidence from manufacturing firms". In: *International Journal of Industrial Organization* 24.2, pp. 349–380. ISSN: 0167-7187. DOI: 10.1016/j.ijindorg.2005.07.002.

Pavitt, K., M. Robson, and J. Townsend (1987). "The Size Distribution of Innovating Firms in the UK: 1945-1983". In: *The Journal of Industrial Economics* 35.3. Publisher: Wiley, pp. 297–316. ISSN: 0022-1821. DOI: 10.2307/2098636.

Rammer, Christian (June 2023). "Measuring process innovation output in firms: Cost reduction versus quality improvement". In: *Technovation* 124, p. 102753. ISSN: 0166-4972. DOI: 10.1016/j.technovation.2023.102753.

Raschka, Sebastian (Nov. 2020). *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*. arXiv:1811.12808 [cs, stat]. DOI: 10.48550/arXiv.1811.12808.

Sastre, Juan Fernández (Jan. 2015). "The impact of R&D teams' gender diversity on innovation outputs". In: *International Journal of Entrepreneurship and Small Business* 24.1. Publisher: Inderscience Publishers, pp. 142–162. ISSN: 1476-1297. DOI: 10.1504/IJESB.2015.066154.

Scherer, F. M. (1982). "Inter-Industry Technology Flows and Productivity Growth". In: *The Review of Economics and Statistics* 64.4. Publisher: The MIT Press, pp. 627–634. ISSN: 0034-6535. DOI: 10.2307/1923947.

Scherer, F. M. (1983). "Concentration, R&D, and Productivity Change". In: *Southern Economic Journal* 50.1. Publisher: Southern Economic Association, pp. 221–225. ISSN: 0038-4038. DOI: 10.2307/1058053.

Scherer, Frederic (1984). "Using Linked Patent and R&D Data to Measure Interindustry Technology Flows". In: *R&D, Patents, and Productivity*. University of Chicago Press, pp. 417–464.

— (1991). *Innovation and Technological Change: An International Comparison*. en-US.

Shahzad, Farrukh et al. (Mar. 2022). "Innovation decisions through firm life cycle: A new evidence from emerging markets". In: *International Review of Economics & Finance* 78, pp. 51–67. ISSN: 1059-0560. DOI: 10.1016/j.iref.2021.11.009.

Tavassoli, Sam (Feb. 2015). "Innovation determinants over industry life cycle". In: *Technological Forecasting and Social Change* 91, pp. 18–32. ISSN: 0040-1625. DOI: 10.1016/j.techfore.2013.12.027.

Utterback, James M and William J Abernathy (Dec. 1975). "A dynamic model of process and product innovation". In: *Omega* 3.6, pp. 639–656. ISSN: 0305-0483. DOI: 10.1016/0305-0483(75)90068-7.

Utterback, William and James Abernathy (1978). "Patterns of Industrial Innovation". In: *Technology Review* 80.7, pp. 41–47.

Van Beveren, Ilke and Hylke Vandenbussche (Mar. 2010). "Product and process innovation and firms' decision to export". In: *Journal of Economic Policy Reform* 13.1. Publisher: Routledge, pp. 3–24. ISSN: 1748-7870. DOI: 10.1080/17487870903546267.

Vernon, Raymond (1966). "International Investment and International Trade in the Product Cycle". In: *The Quarterly Journal of Economics* 80.2. Publisher: Oxford University Press, pp. 190–207. ISSN: 0033-5533. DOI: 10.2307/1880689.

Notes

1. Available at: https://doi.org/10.5281/zenodo.8364951.

2. For example, Bena and Simintzi 2022 use "method" and "process" as process keywords while Liu et al. 2020 use the hypernym "activities" to generate the process innovation keyword list and the hypernym "physical entities" to generate the product innovation keyword list.

3. This is consistent with Liu et al. 2020 who use the hypernym "physical entities" to generate a list of product innovation keywords.

4. At the time of writing, Liu et al. 2020 have not made their data publicly available, but they report an 87% correlation between the number of process claims on a patent in their data and the data from Bena and Simintzi 2022.

5. One firm that I designate as not a manufacturing firm is IBM, due to their shift from IT manufacturing toward software services.

6. Some examples of firms who I classify as manufacturing, but would not be classified as manufacturing under the standard SIC definition are: General Electric, Honeywell, and Monsanto.

7. This is consistent with Liu et al. 2020 who use the hypernym "physical entities" to generate a list of product innovation keywords.

8. In the case of Liu et al. 2020, they rely on both the USPTO definition and definitions from international patent offices. Yet the various patent offices have general agreement about what constitutes a product or process innovation from the persepctive of the patent office.

9. 35 U.S.C. §101 states that a patent can be obtained for a process, machine, manufacture, or composition of matter. Machines, manufactures, and compositions of matter can be categorized as products.

10. Claims can be independent or dependent, with each independent claim standing on its own. Because dependent publication claims rely on independent claims, this paper restricts its attention to independent claims.

11. Ex parte Lyell, 1990 Pat. App. LEXIS 14, *12 (Bd. Pat. App. & Interferences August 16, 1990).

12. For some industries, I pool together several similar industries in order to increase accuracy and precision. Whenever industries are pooled together, the diagnostic statistics only reflect how well the model is able to predict for the focal industry.

13. Appendix A.1 lists all the features used. Importance is determined using a meta-transformer called "SelectFromModel" that is available in Python's scikit-learn package.

14. This is in contrast to neural networks which loop through the training data multiple times and iteratively tweak the coefficients, building upon what was seen prior in order to minimize the loss function.

15. This is calculated as $\frac{1{,}769}{1{,}876} = 94.3\%$

16. This is calculated as $\frac{9{,}018}{9{,}125} = 98.8\%$

17. This is calculated as $\frac{1{,}769}{2{,}535} = 69.8\%$

18. Industries with less than 500 patents from 1980-2015 are excluded from the analysis.

19. 4-digit SIC 3829 and 3823

20. $\sum_{g \in S} \Delta\text{Process Share}_g \times \text{Patent Share}_{g,t-1}$

21. $\sum_{g \in S} \Delta\text{Patent Share}_g \times \text{Process Share}_{g,t-1}$

22. $\sum_{g \in S} \Delta\text{Process Share}_g \times \Delta\text{Patent Share}_g$

23. $\sum_{g \in E}$ Process $\text{Share}_{g,t} \times$ Patent $\text{Share}_{g,t}$

24. $\sum_{g \in X} -1 *$ Process $\text{Share}_{g,t-1} \times$ Patent $\text{Share}_{g,t-1}$

25. $\Delta Y_{i\tau} = \frac{(Y_{i1} - Y_{i0})}{0.5(Y_{i1} + Y_{i0})}$

26. For example, patenting in the "computers" sector grew significantly both over the 1975-1991 and 1991-2007 time periods while patenting in the "chemicals" sector had slow growth over both time periods.

27. The change in import penetration is defined as the change in Chinese imports to the U.S. in 4-digit SIC, j, scaled by the industry's initial absorption (production plus imports minus exports) in 1991. Formally, this is $\Delta \mathrm{IP}_{j\tau}^{US} = \frac{\Delta Imports_{j\tau}^{US,China}}{Y_{j,1991} + Imports_{j,1991} - Exports_{j,1991}}$.

28. Firm age is approximated using the COMPUSTAT listing vintage.

29. One marginally significant coefficient emerges on the product development stage when using the Ganglmair et al. 2022 classification, indicating that patenting tends to have higher process shares during the product development stage relative to the process improvement stage.

30. Process language refers to words or phrases such as: "a method for" which are associated with the USPTO notion of a process innovation.

A Appendix

A.1 Product Process Classification

The below list itemizes the elements that make up each of the 60 machine learning models used to predict the product/process status of patent claims.

- Machine Learning Models

  1. Multinomial Naive Bayes
  2. Complement Naive Bayes
  3. Passive Aggressive Classifier

- Text Features

  1. First two words and 4-digit CPC code
  2. First two words, first two words interacted with firm identifier, and 4-digit CPC code
  3. First two words, first two words interacted with 4-digit CPC code, and 4-digit CPC code
  4. First two words, first three words, ..., first ten words, and the 4-digit CPC code
  5. First two words, first three words, ..., first ten words, first two words interacted with firm identifier, first three words interacted with firm identifier, ..., first ten words interacted with firm identifier, and the 4-digit CPC code

6. First two words, first three words, ..., first ten words, first two words interacted with the 4-digit CPC code, first three words interacted with the 4-digit CPC code, ..., first ten words interacted with the 4-digit CPC code, and the 4-digit CPC code
7. Entire claims text and 4-digit CPC code
8. The abbreviated claims text (everything that comes before the first colon) and the 4-digit CPC code
9. The abbreviated claims text (everything that comes before the first colon), the abbreviated claims text interacted with a firm identifier, and the 4-digit CPC code
10. The abbreviated claims text (everything that comes before the first colon), the abbreviated claims text interacted with the 4-digit CPC code, and the 4-digit CPC code

- Trimming

1. Keep all features in a feature set.
2. Drop features that are below median "importance" in predicting the outcome variable.

- Extended Text Features

1. Number of words
2. Fraction of words that are stopwords
3. Average word length

4. Fraction of words that are nouns, verbs, and adjectives

5. Number of words, fraction of words that are stopwords, average word length, fraction of words that are nouns, verbs, and adjectives

6. Number of words, fraction of words that are stopwords, average word length, fraction of words that are nouns, verbs, and adjectives, number of words interacted with the 4-digit CPC code, fraction of words that are stopwords interacted with the 4-digit CPC code, average word length interacted with the 4-digit CPC code, fraction of words that are nouns, verbs, and adjectives interacted with the 4-digit CPC code

7. Number of words, fraction of words that are stopwords, average word length, fraction of words that are nouns, verbs, and adjectives, number of words interacted with a firm identifier, fraction of words that are stopwords interacted with a firm identifier, average word length interacted with a firm identifier, fraction of words that are nouns, verbs, and adjectives interacted with a firm identifier

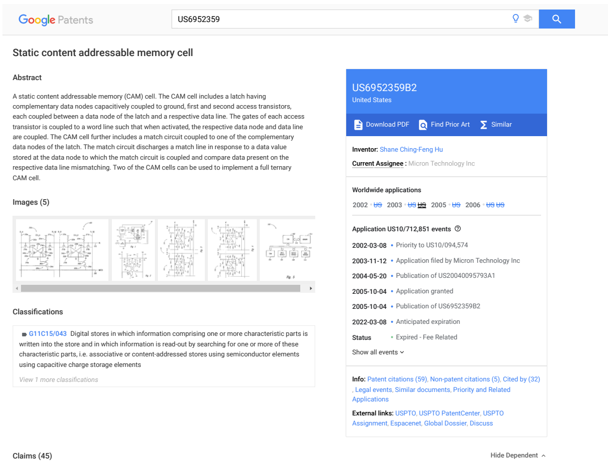
Table A.1: Defining Manufacturing Firms

		SIC Definition	
		Not Mfn	Mfn
Business Segment Sales Definition	Not Mfn	1,087	127
	Mfn	215	3,583

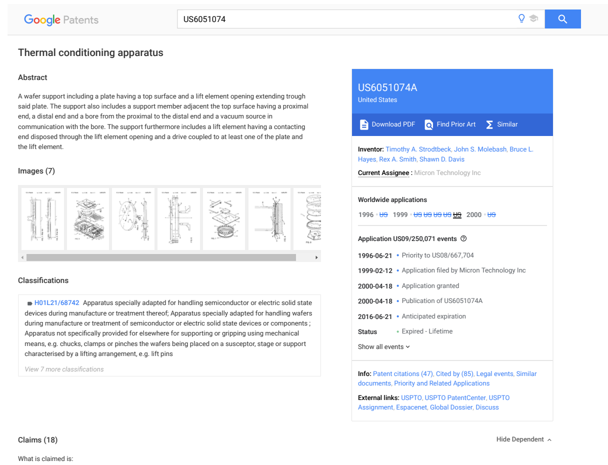
Notes: This table presents a cross-tabulation on the number of firms defined as manufacturing firms according to a SIC definition which defines all firms with primary SIC industry 2000-3999 as manufacturing firms and using a business segment sales definition that is outlined in the body of the paper.

Figure A.1: Three "Micron Technology" Patents

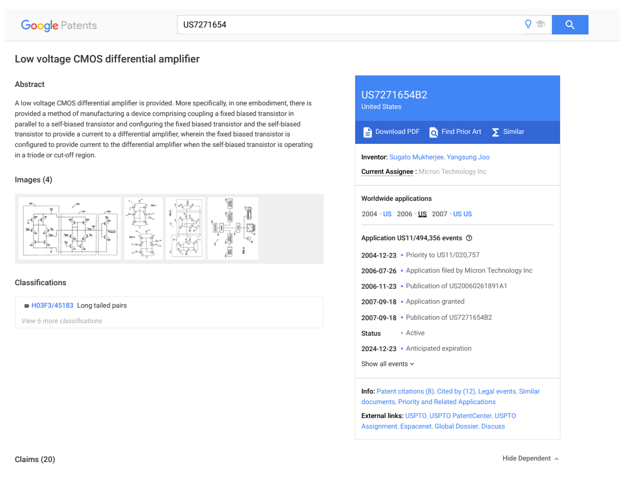
(a) Product



(b) Process

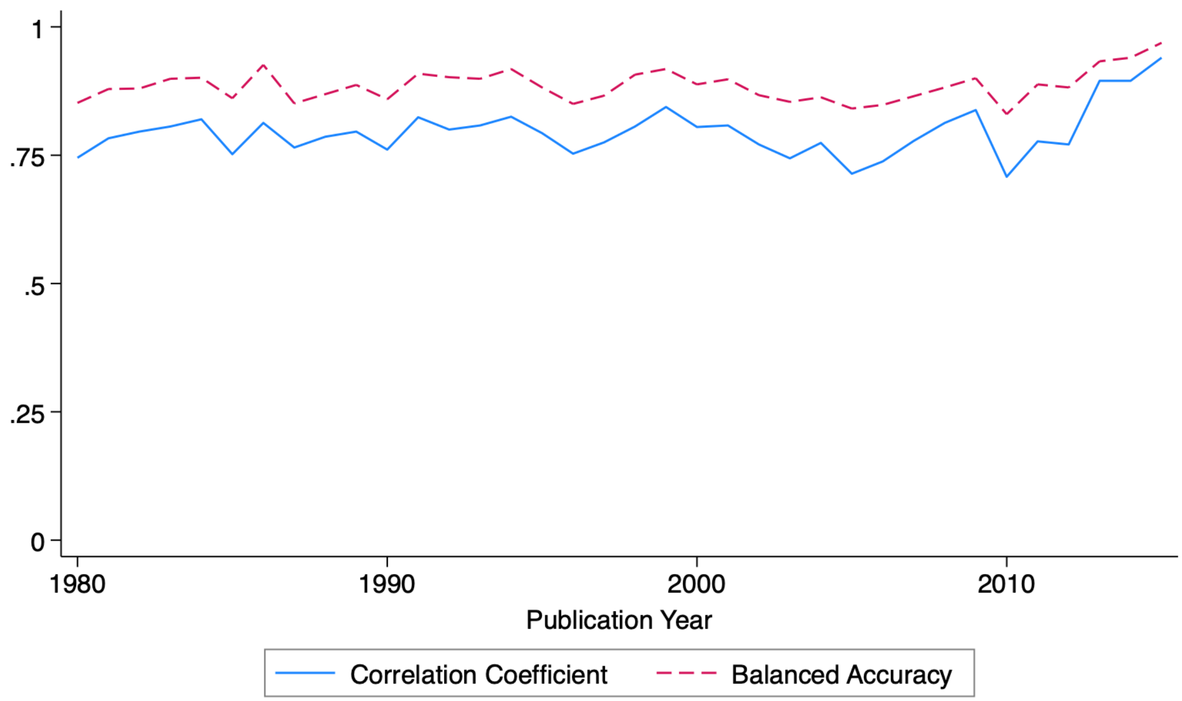


(c) Product & Process



Notes: This figure depicts the Google Patents webpages for three Micron Technology patents. Panel (a) depicts a product patent, US6952359. Panel (b) depicts a process patent, US6051074. Panel (c) depicts a patent that contains both product and process innovations, US7271654.

Figure A.2: Diagnostic Statistics Over Time



Notes: This figure shows the weighted average of the correlation coefficient and balanced accuracy over time where the weights correspond to the number of patents in the industry from 1980-2015.

A.2 Product and Process Innovation Over Time

Table A.2 presents results from estimating Equation (A.1) via OLS where the dependent variable is the process share of patent p, assigned to firm f, in 4-digit SIC industry s, CPC subclass c, and published in year t and standard errors are robust:

$$\begin{aligned}\text{Process Share}_{pfsct} = &\beta_1 \text{Time Trend}_t + \beta_2 \text{Time Trend}_t \times \mathbb{1}\{\text{Year} \geq 2000\} + \\ &\beta_3 \mathbb{1}\{\text{Year} \geq 2000\} + \phi(s, f, c) + \varepsilon_{pfsct}\end{aligned} \tag{A.1}$$

$\phi(s, f, c)$ are fixed effects which vary by specification and depend on industry, firm, and CPC subclass. β_1 captures the average annual change in the process share from 1980-2015 while β_2 captures how the time trend differs during the 2000-2015 time period. Table A.2 presents the results with column (1) showing a strong negative time trend. On average, the overall process share falls by approximately 0.5 percentage points a year. In column (2) when industry fixed effects are included the coefficient halves. Column (3) reveals that within industry there is a significant decline in the process share after 2000 but not before. While less pronounced, the results in columns (4) and (5) tell a similar story when firm and CPC subclass fixed effects are used.

Table A.2: Process Share Time Trend

	Process Share × 100				
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS
Time Trend	-0.484***	-0.249***	-0.018*	-0.066***	-0.106***
	(0.004)	(0.004)	(0.010)	(0.011)	(0.009)
Time Trend $\times \mathbb{1}\{\text{Year} \geq 2000\}$			-0.523***	-0.379***	-0.136***
			(0.013)	(0.014)	(0.012)
$\overline{\text{Y}}$	17.77	17.77	17.77	17.77	17.77
4-digit SIC FE		✓	✓		
Firm FE				✓	
CPC Subclass FE					✓
Observations	1,023,048	1,023,048	1,023,048	1,023,048	1,023,048

Notes: This table presents results from estimating Equation (A.1) using OLS. Standard errors are robust and shown in parentheses. *(p<0.1), **(p<0.05), ***(p<0.01).

Table A.3: Decomposition of Process Share Time Trends Over NBER Category

Time Period	Process Share$_0$	Within	Between	Cross	Entry	Exit	Total
1980-1985	26.34	.88	-.84	-.08	0	0	-.04
1985-1990	26.3	-.67	-1.33	.08	0	0	-1.92
1990-1995	24.38	-1.98	-2.24	.14	0	0	-4.08
1995-2000	20.3	1.01	-2.54	.34	0	0	-1.2
2000-2005	19.11	-1.34	-1.66	.08	0	0	-2.92
2005-2010	16.18	-2.36	-.78	.22	0	0	-2.92
2010-2015	13.26	-.9	-.6	-.09	0	0	-1.6
Total		-5.36	-9.99	.69	0	0	-14.68

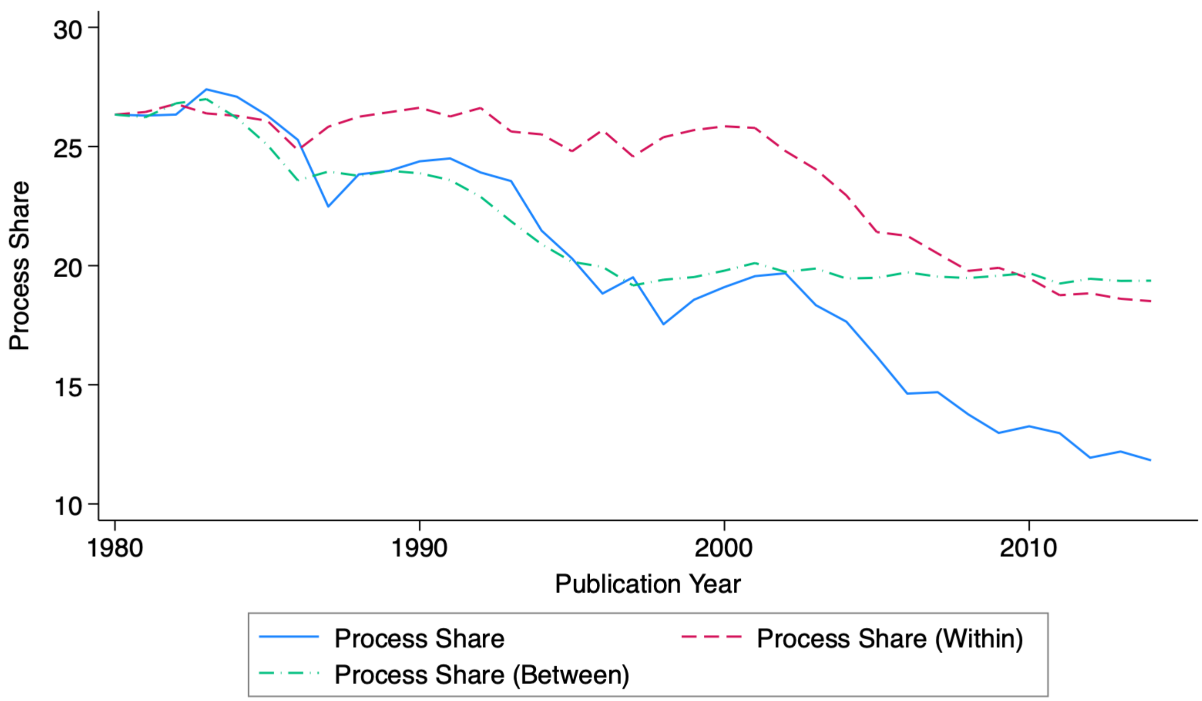
Notes: This table presents the decomposition of the percentage point change in the mean process share of patents into within, between, cross, entry, and exit components with the assignee firm being the group variable. Process Share$_0$ indicates the overall process share at the beginning of the time period.

Table A.4: Decomposition of Process Share Time Trends Over Firm

Time Period	Process Share_0	Within	Between	Cross	Entry	Exit	Total
1980-1985	26.34	-.13	1.2	-.41	.69	-1.39	-.04
1985-1990	26.3	.22	.83	-.09	.69	-3.58	-1.92
1990-1995	24.38	-.54	-4.45	.44	1.35	-.88	-4.08
1995-2000	20.3	-.79	1.5	.04	1.07	-3.02	-1.2
2000-2005	19.11	-1.66	-1.53	.35	1.01	-1.08	-2.92
2005-2010	16.18	-2.03	-.19	.29	.23	-1.23	-2.92
2010-2015	13.26	-1.82	.87	-.02	.6	-1.24	-1.6
Total		-6.75	-1.77	.6	5.64	-12.42	-14.68

Notes: This table presents the decomposition of the percentage point change in the mean process share of patents into within, between, cross, entry, and exit components with the assignee firm being the group variable. Process Share_0 indicates the overall process share at the beginning of the time period.

Figure A.3: Within and Between Components of the Process Share Over Time



Notes: This figure shows the overall process share over time along with how the overall process share would have evolved if only the "within" (red dashed line) or "between" (green dash-dot line) components were active.

A.3 Product and Process Innovation Over the Firm's Life Cycle

Table A.5: Process Share Over the Life Cycle (Banholzer et al. 2019)

	Process Share (Banholzer et al. 2019)		
	(1)	(2)	(3)
	OLS	OLS	OLS
Product Development	0.035	0.027	0.008
	(0.026)	(0.029)	(0.026)
Mature Product	0.010	0.026	0.016
	(0.025)	(0.022)	(0.021)
Product Decline	-0.006	-0.015	-0.024
	(0.049)	(0.039)	(0.035)
Firm Age			-0.004**
			(0.002)
$\overline{\text{Y}}$	0.37	0.37	0.37
Firm FE	✓	✓	✓
Year FE	✓		
4-digit SIC × Year FE		✓	✓
Observations	623,580	623,580	623,580

Notes: This table presents results from estimating Equation (3) using OLS. Standard errors are clustered at the firm level and shown in parentheses. *(p<0.1), **(p<0.05), ***(p<0.01).

Table A.6: Process Share Over the Life Cycle (Bena and Simintzi 2022)

	Process Share (Bena and Simintzi 2022)		
	(1)	(2)	(3)
	OLS	OLS	OLS
Product Development	0.041	0.036	0.016
	(0.028)	(0.031)	(0.028)
Mature Product	0.019	0.030	0.019
	(0.022)	(0.022)	(0.021)
Product Decline	0.007	-0.003	-0.011
	(0.050)	(0.041)	(0.037)
Firm Age			-0.004**
			(0.002)
$\overline{\text{Y}}$	0.37	0.37	0.37
Firm FE	✓	✓	✓
Year FE	✓		
4-digit SIC × Year FE		✓	✓
Observations	630,690	630,690	630,690

Notes: This table presents results from estimating Equation (3) using OLS. Standard errors are clustered at the firm level and shown in parentheses. *(p<0.1), **(p<0.05), ***(p<0.01).

Table A.7: Process Share Over the Life Cycle (Ganglmair et al. 2022)

	Process Share (Ganglmair et al. 2022)		
	(1)	(2)	(3)
	OLS	OLS	OLS
Product Development	0.042*	0.026	0.018
	(0.025)	(0.027)	(0.025)
Mature Product	0.014	0.022	0.017
	(0.024)	(0.020)	(0.019)
Product Decline	0.024	-0.013	-0.016
	(0.044)	(0.035)	(0.033)
Firm Age			-0.002
			(0.002)
$\overline{\text{Y}}$	0.36	0.36	0.36
Firm FE	✓	✓	✓
Year FE	✓		
4-digit SIC × Year FE		✓	✓
Observations	630,690	630,690	630,690

Notes: This table presents results from estimating Equation (3) using OLS. Standard errors are clustered at the firm level and shown in parentheses. *(p<0.1), **(p<0.05), ***(p<0.01).

Table A.8: Process Share and Log Number of Employees

	Process Share			
	(1) OLS	(2) OLS	(3) OLS	(4) OLS
ln(Employees)	0.009	0.010*	0.012**	0.005
	(0.006)	(0.006)	(0.005)	(0.005)
Product Development			-0.102**	-0.078*
			(0.046)	(0.041)
Mature Product			-0.044*	-0.028
			(0.024)	(0.021)
Product Decline			-0.040	-0.032
			(0.034)	(0.034)
Firm Age				0.005***
				(0.002)
$\overline{Y}$	0.15	0.15	0.15	0.15
Firm FE	✓	✓	✓	✓
Year FE	✓			
4-digit SIC × Year FE		✓	✓	✓
Observations	624,294	624,294	624,294	624,294

Notes: This table presents results from estimating Equation (3) using OLS where a measure of firm size is added. Standard errors are clustered at the firm level and shown in parentheses. *(p<0.1), **(p<0.05), ***(p<0.01).

Table A.9: Process Share and Log Revenue (Banholzer et al. 2019)

	Process Share (Banholzer et al. 2019)			
	(1) OLS	(2) OLS	(3) OLS	(4) OLS
ln(Revenue)	-0.003	-0.002	-0.003	0.002
	(0.005)	(0.005)	(0.005)	(0.004)
Product Development			0.028	0.008
			(0.029)	(0.027)
Mature Product			0.028	0.015
			(0.022)	(0.020)
Product Decline			-0.017	-0.024
			(0.039)	(0.035)
Firm Age				-0.004**
				(0.002)
$\overline{\text{Y}}$	0.36	0.36	0.36	0.36
Firm FE	✓	✓	✓	✓
Year FE	✓			
4-digit SIC × Year FE		✓	✓	✓
Observations	622,632	622,632	622,632	622,632

Notes: This table presents results from estimating Equation (3) using OLS where a measure of firm size is added. Standard errors are clustered at the firm level and shown in parentheses. *(p<0.1), **(p<0.05), ***(p<0.01).

Table A.10: Process Share and Log Revenue (Bena and Simintzi 2022)

	Process Share (Bena and Simintzi 2022)			
	(1) OLS	(2) OLS	(3) OLS	(4) OLS
ln(Revenue)	-0.005	-0.003	-0.004	0.001
	(0.005)	(0.005)	(0.005)	(0.005)
Product Development			0.037	0.016
			(0.031)	(0.028)
Mature Product			0.032	0.018
			(0.022)	(0.021)
Product Decline			-0.005	-0.012
			(0.041)	(0.037)
Firm Age				-0.005**
				(0.002)
$\overline{\text{Y}}$	0.37	0.37	0.37	0.37
Firm FE	✓	✓	✓	✓
Year FE	✓			
4-digit SIC × Year FE		✓	✓	✓
Observations	629,733	629,733	629,733	629,733

Notes: This table presents results from estimating Equation (3) using OLS where a measure of firm size is added. Standard errors are clustered at the firm level and shown in parentheses. *(p<0.1), **(p<0.05), ***(p<0.01).

Table A.11: Process Share and Log Revenue (Ganglmair et al. 2022)

	Process Share (Ganglmair et al. 2022)			
	(1) OLS	(2) OLS	(3) OLS	(4) OLS
ln(Revenue)	-0.002	-0.001	-0.002	0.000
	(0.005)	(0.004)	(0.004)	(0.004)
Product Development			0.027	0.019
			(0.027)	(0.025)
Mature Product			0.023	0.017
			(0.020)	(0.019)
Product Decline			-0.014	-0.017
			(0.035)	(0.033)
Firm Age				-0.002
				(0.002)
$\overline{\text{Y}}$	0.36	0.36	0.36	0.36
Firm FE	✓	✓	✓	✓
Year FE	✓			
4-digit SIC × Year FE		✓	✓	✓
Observations	629,733	629,733	629,733	629,733

Notes: This table presents results from estimating Equation (3) using OLS where a measure of firm size is added. Standard errors are clustered at the firm level and shown in parentheses. *(p<0.1), **(p<0.05), ***(p<0.01).