

# INNOVATIVE EVENTS: PRODUCT LAUNCHES, INNOVATION AND FIRM PERFORMANCE

Max Nathan

*Affiliations:* UCL, CEP and IZA;

*email:* max.nathan@ucl.ac.uk

*Present Address:*

CASA, 90 Tottenham Court Road, London, W1T 4TJ UK

Anna Rosso (Corresponding author)

*Affiliations:* Università degli Studi di Milano, CEP and LdA;

*email:* anna.rosso@unimi.it.

*Present Address:*

Università Statale di Milano, Dipartimento di Economia, Management e Metodi Quantitativi  
Via Conservatorio 7, 20122, Milano Italy

## Abstract

In this paper, we shed new light on the links between firm-level innovation and growth. We introduce data that capture a difficult-to-observe aspect of firms' innovative activity – new product/service launches – at scale. We show that our novel measures complement existing innovation metrics. We build a simple framework covering firm-level innovation, launches and revenue productivity. Then, we show positive linkages between past patenting and launches and between launches and performance for a large panel of small and medium-sized enterprises (SMEs) in the UK. We go on to explore the roles of age, size, industry and product/service quality in these relationships. A subset of SMEs with high-quality launches explains our results.

**Keywords:** innovation, productivity, ICT, data science

**JEL:** C55, L86, O40

## Acknowledgements

Thanks to seminar/conference participants at ZEW Economics of ICT workshop 2017, 4<sup>th</sup> VPDE-BRICK workshop, SpazioDati, CTCS 2016, Uddevalla 2016, AAG 2016, University of Birmingham, CIRCLE and NIESR, and to Cathy Atkinson, Steve Dempsey, Rob Elliott, Bill Kerr, Francesco Rentocchini and Brandon Stewart for helpful comments. Many thanks to Tom Gatten, Prash Majmudar and Francois Bouet at Growth Intelligence for data. Camilla Andretta, Neeraj Baruah, Aslam Ghumra, and Christoph Stich provided outstanding technical and research assistance. This work includes analysis based on data from the Business Structure Database and the UK Innovation Survey, produced by the Office for National Statistics (ONS) and supplied by the Secure Data Service at the UK Data Archive. The data is Crown copyright and reproduced with the permission of the controller of HMSO and Queen's Printer for Scotland. The use of the ONS statistical data in this work does not imply the endorsement of the ONS or the Secure Data Service at the UK Data Archive in relation to the interpretation or analysis of the data. This work uses research datasets that may not exactly reproduce National Statistics aggregates. All the outputs have been granted final clearance by the staff of the SDS-UKDA. This research builds on projects funded by NESTA and by Google UK. It represents our views not theirs.

**ACCEPTED VERSION, FORTHCOMING IN RESEARCH POLICY**

## **1/ Introduction**

There is decades' worth of research exploring the role of innovation in explaining economic performance. Endogenous growth theory helps explain country-level innovation-productivity linkages (Lucas, 1988; Romer, 1990). Schumpeterian frameworks highlight the roles of entrepreneurial entry, competition and factor reallocation (Schumpeter, 1962; Aghion and Howitt, 1992). They also provide firm-level microfoundations linking R&D activity, innovation and new products and services (see Akcigit (2017) for a review). Evolutionary perspectives emphasise that firms' capabilities vary greatly. Firms' resources accumulate over time, and it is difficult to shift them (Nelson and Winter 1982, Dosi et al 2000). This variation in firms' capabilities makes predictors of average firm performance difficult to identify (Nightingale and Coad, 2013).

The empirical literature on firm-level R&D, innovation and productivity dates back to Griliches (1979; 1986) and Crepon, Duguet and Mairesse (1998). Most subsequent studies use R&D, patent or innovation survey data to identify cross-sectional links between these factors and firm performance (see Hall, 2011; Syverson, 2011; Mohnen and Hall, 2013 and Audretsch et al, 2014 for reviews). More recent contributions use panel data and more sophisticated estimators (for example Fernandes and Paunov, 2015; Howell, 2015; Coad et al, 2016; Coad et al, 2016; Baumann and Kritikos, 2016; Bianchini et al, 2018; Morris, 2018; Grillitsch et al, 2019; Spescha and Woerter, 2019; Audretsch and Belitski, 2020; Audretsch et al, 2020).

Consistent with evolutionary theory, these studies broadly confirm a positive linkage between innovation and firm performance – but with much heterogeneity across firm characteristics, behaviours and macro factors. The overall relationship is thus hard to pin down. For example, Mohnen (2019) shows that innovation has long-term effects on economic growth as measured by TFP, both at the firm level and the aggregate level, confirming Schumpeter's view of

innovation. Conversely, Guarascio and Tamagni (2019) suggest that innovation-sales links are largely random, consistent with Gibrat's Law (Coad, 2009).

Starting with Mendonca et al. (2004), a more recent stream of work finds positive links between trademarks and innovation (see Taques et al. (2021), Castaldi et al. (2020), and Schautschick and Greenhalgh (2016) for reviews). These studies concentrate on start-ups, SMEs, and service industry settings (for example, Crass, 2020; Flikkema et al, 2019; De Vries et al, 2017; Block et al, 2015; Flikkema et al, 2014; Gotsch and Hipp, 2012).

One limitation of both bodies of work is that much innovative activity is informal and unobserved. Most companies rely on tools such as lead time, design complexity, or less often, NDAs or other forms of secrecy (Hall et al 2014). Only 1.6% of UK businesses file patents (Hall et al, 2013), and some firms employ patents defensively (Noel and Schankerman, 2013). While companies use trademarks more widely, including for non-technological and service innovations, it is still unclear which trademarks relate to innovation (Castaldi et al, 2020; Flikkema et al, 2019).<sup>1</sup> Low response rates and small samples can limit the usefulness of innovation surveys, as do the widely varying answers given by respondents (Mairesse and Mohnen, 2010).<sup>2</sup>

One newer strand of empirical work seeks to close these gaps with novel innovation metrics derived from firms' website text (Axenbeck and Breithaupt, 2019; Kinne and Lenz, 2019; Lenz and Winker, 2020), patents (Arts et al, 2021; Kelly et al, 2018) or regulatory filings (Saunders and Tambe, 2015; Hoberg and Philips, 2016; Kogan et al, 2017). These studies typically involve larger and/or listed firms rather than the SMEs that make up the bulk of the economies of more

---

<sup>1</sup> Castaldi et al. (2020) show that firms use trademarks for multiple purposes, including securing market position (allowing markups or deterring entry) and attracting resources (from venture capitalists and other investors).

<sup>2</sup> For example, the response rate of the 2017 UK Innovation Survey was 43 percent. This rather modest response rate mainly affects the creation of a balanced panel of firms over consecutive CIS years (Mairesse and Mohnen, 2010).

developed countries. Another strand of research uses product-level data to examine innovation-growth links (Bottazzi et al, 2001; Stam and Wennberg, 2009; Corsino and Gabriele, 2010; Cucculelli and Ermini, 2012; Argente et al 2018; 2019, Bokhari et al, 2020) or, relatedly, the timing of new product introductions (Ortega and Garcia-Villaverde, 2011; Rodriguez-Pinto et al, 2012; Hsiao et al, 2017). Data constraints mean that this literature usually involves single or restricted industry cases: many studies use bespoke, small-*n* surveys.<sup>3</sup>

This paper makes three linked contributions to these debates. First, we use a novel mix of UK administrative microdata and media content to develop novel measures of innovative activity at the firm level across all industries and firm types. We have substantially broader coverage than both the abovementioned product data analyses and pioneering studies on media coverage and innovation.<sup>4</sup> To do this, we exploit a cutting-edge dataset developed by the data science firm Growth Intelligence (GI). This firm uses machine-learning routines on company website and media content to model firms' lifecycle 'events'. We focus on one of these variables – new products/services reported in the news. We also exploit overall event exposure to aid our research design. We clean and refine these data using structural topic modelling to better align reported and real-world activity. We also advance innovation-output studies, developing measures of launch quality analogous to patent citations.

Second, we show that these reported launches complement measures of formal IP. This includes frequency and industry/geographic coverage. For example, for single-plant SMEs, we find 24,720 UK launches in 2014/2015, versus 4,194 patent applications and 4,510 trademarks filed.

---

<sup>3</sup> For example, Bokhari et al, Bottazzi et al, Corsino and Gabriele, and Ortega and Garcia-Villaverde all examine single industries; Rodriguez-Pinto et al survey 136 manufacturing firms, Cucculelli and Ermini investigate 204 SMEs, and Stam and Wennberg analyse approximately 2,000 start-ups. Focusing on publicly listed firms, Argente et al (2018, 2019) use barcode data with much broader, fine-grained coverage closer in spirit to our data.

<sup>4</sup> Katila and Ahuja (2002) and Fosfuri et al (2008), for example, are restricted to a few hundred firms in single sectors. These studies use the 'counting innovation' method, to examine a selection of innovations introduced in a given year and reported in trade journals. See Kleinknecht et al (1993), Coombs et al (1996) and Santarelli and Piergiovanni (1996).

We then show positive links between past IP activity and current launches. We find cross-industry and trademark-type variation consistent with prior literature and our framework's predictions.

Third, we develop a simple framework linking firm-level IP, launches and performance and test its predictions on a panel of UK SMEs. In particular, we test whether a company grows by transforming new products/services into higher revenue per worker and explore the roles of age, size, industry and product/service quality in explaining our results. We pay careful attention to the fact that event exposure is not random and that we are working with reported rather than observed activity. We find that launch activity is associated with higher SME revenue per worker, especially in the service sector, among medium-sized firms and among firms with specialised trademarks. Consistent with a world of heterogeneous firm capabilities, a small subset of high-quality launches helps drive the main result. Robustness checks on all firms with event exposure find relatively weak links – consistent with larger, multi-plant firms having multiple sources of revenue growth, such as advertising existing products.

To our knowledge, this study is the first to use such data, at scale, to study innovation and firm performance. More broadly, it contributes to the literature on firm growth (Gilbert and Newbery, 1982; Coad, 2009; Audretsch et al, 2014; Castaldi et al, 2020), IP choices (Hall et al, 2014), determinants of firm productivity (Syverson, 2011), high-growth businesses (Coad et al, 2014), and economic applications of natural language processing (Gentzkow et al, 2019). Crucially, we combine text-based measures of innovative activity with high-quality administrative microdata. This gives us a clear sampling frame, aiding inference and interpretation (Einav and Levin, 2014). We also advance on rich data papers such as those of Hall et al. (2013) and Coad et al. (2016), who combine conventional administrative data, patents, trademarks and innovation surveys to examine smaller groups of firms. Overall, we

focus on demonstrating data use cases and estimating clean associations rather than causal effects. Our dataset is replicable and extendable for future research.<sup>5</sup>

## 2/ Data

We use modelled company ‘events’ to develop new measures of innovative activity. Our method extends the innovative outputs approach (Corsino and Gabriele, 2010; Cucculelli and Ermini, 2012; Bokhari et al, 2020). Each ‘event’ derives from article text taken from 3,740 online news sources (including major sources such as Reuters and Yahoo news and industry sources such as IT Briefing and PRWeb).<sup>6</sup> Our raw data consist of 318,899 observations corresponding to 30,205 companies during financial years 2014 and 2015. Growth Intel matches the text of each article to the UK company register (Companies House) using firm names and contextual information, then uses supervised learning to classify the text as one of several event types. Nathan and Rosso (2015) provide more details on the data sources and workflow.<sup>7</sup> We focus on one event type: ‘product/service launch’. Figure 1 provides two examples of product/service launches, showing both raw inputs and modelled outputs.

---

<sup>5</sup> A sample of the GI data, plus cleaning, topic modelling and matching code is available at <https://osf.io/bjykc/>. Our data is part of a growing body of similar resources. Existing datasets on news events such as GDELT and Events Registry are designed for country-level analyses, especially those on politics/current affairs. Other proprietary firm-level datasets such as Matternmark (US) and Beauhurst (UK) provide rich information from a range of sources but are restricted to small numbers of ‘high-potential’ businesses. Crunchbase (US) is a global wiki-type dataset regarding the tech sector with good US coverage but limited coverage for other countries, as well as significant quality concerns due to the self-reported nature of its data (Motoyama and Bell-Masterson, 2014). Glass AI (UK) draws data from firms’ websites with some observations now linked back to administrative data; see Siepel et al (2020) for one recent use case. SpazioDati (Italy) provides similar website-sourced data, as do the German datasets used in Axenbeck and Breithaupt (2019) and Kinne and Lenz (2019) to measure firm-level innovation.

<sup>6</sup> We cannot split the sample of product launches based on sources. In future research, it would be interesting to compare events reported via online news outlets (as we have here) with events directly reported on company websites (which we do not include). Similarly, we lack detailed information on local vs. national online news sources.

<sup>7</sup> Text fragment for illustration. GI uses a full page of content to assign text to a subject company and to classify the related activity. Where a text describes more than one subject company, as in mergers or joint ventures, GI assigns the event to a pair of companies or *n*-groups. GI also filters the data to remove results from irrelevant domains (for example, mentions of companies in celebrity magazines or results from sites that largely or wholly deal with markets outside the UK).

*Figure 1 about here*

In theory, each launch event represents a new product/service that a given firm releases into the world and that is then covered by at least one of our media sources. In practice, we need to deal with three ascription challenges. First, our coverage may be uneven, analogous to the well-discussed limitations of patents (Hall and Harhoff, 2012). Second, media exposure is determined by a combination of firm decisions, firm media-generating capacity and reporting norms, some of which are difficult to observe. Third, and relatedly, reported launch content might reflect firms' advertising or PR rather than 'true' innovation.

We deal with these issues as follows. First, we carefully clean the raw data, as detailed below. Additionally, we show that launch coverage is substantially more frequent and even across industries and regions than patents and trademarks.

Second, to address selection issues, we use firms' overall 'event exposure' or 'coverage'; that is, whether a firm has *any* reported events of *any* kind. We show that event exposure is not random and that it is correlated with a range of observables, while launch exposure is more balanced. We focus on SMEs with event exposure and then show that our main results hold for larger samples.

The third issue is harder to disentangle. Our input data are news articles, not advertising copy. However, we may be (a) reporting trivial innovations, (b) missing important innovations or both. To address (a), we measure the number of raw mentions of each launch and use this to construct a proxy measure of launch 'importance' analogous to patent citations. Existing literature provides some reassurance on (b). Firms' predominant use of informal IP protection means that *all* innovation measures undercount the true level of innovative activity (Hall et al,

2014). However, coverage of variables based on observations of products and services in the marketplace is then more affected while that of measures based on formal IP practices is less affected.

## 2.1/ Build

Our data cleaning is summarised here and detailed in Appendix A1. We first remove duplicates and control for ‘farmed’ content.<sup>8</sup> Next, we run two quality checks on GI’s ascription routines. We then improve the realism of the data using structural topic modelling (STM). A major product launch is likely to be reported hundreds of times; in the raw data, each launch is reported as a distinct event. STM is used to cluster text fragments that refer to the same topic into single observations representing the underlying real-world launch event (Roberts et al, 2016). Overall, our cleaning steps substantially reduce the number of event observations from 318,899 observations for 30,205 firms, to 257,056 observations for 30,187 firms. STM accounts for the bulk of this reduction.

We combine this cleaned event data with data from other sources. To do this, we link Companies House identifiers to the Business Structure Database (BSD) (Office of National Statistics, 2017). The high-quality administrative microdata in this database cover 99% of UK enterprises and provide a clearly defined sampling frame. We then use various matching routines, detailed in Appendix A2,<sup>9</sup> to link US, European and other patent data (from Orbis,

---

<sup>8</sup> Recent structural changes to the media industry – notably, the rise of online platforms – may be the reducing levels of data quality and scrutiny in this industry, for example through ‘content farming’ and ‘churnalism’ (Viner, 2016; Gentzkow and Shapiro, 2010; Davies, 2009). The first leads to duplicate reported events, while the second alters the distribution of event activity. Both may be particularly prevalent in the information and communications technology (ICT) sector (Lafrance, 2016). We identify duplicate observations of events using all available variables except the source and time. Within each group, we keep only one event; thus, we are not selecting events on the basis of source quality.

<sup>9</sup> Bureau Van Dijk identifiers or firm name and full postcode. An alternative approach is the automated method developed by Autor et al (2020), which exploits internet search results.



application years 1900-2015) and UK trademark data (from the UK Intellectual Property Office, 2012 - 2015).<sup>10</sup> The resulting dataset includes 1,399,146 firms and is an unbalanced panel of 5,039,811 firm-year observations from 2014-2017. Within this set, there are 22,497 firms with event exposure and 212,426 unique events.

In our descriptive analysis (Section 3), we use data for the subset of single-plant SMEs from 2014-2015 – the years in which events are observed – to compare the characteristics of GI product launches with those of traditional IP measures. Crucially, using single-plant SMEs allows us to cleanly ascribe a launch to a given firm and location. Removing the largest firms also reduces ascription error (see Appendix A1). Over 95% of firms in the BSD are single-plant SMEs.<sup>11</sup>

In our regression analysis (Sections 4-6), we further focus on the subset of single-plant SMEs with ‘event exposure’, which we define as firms that had an event (of any kind) during 2014, 2015 or both. This allows us to work with the variation in launch activity across firms and time, conditional on non-random event exposure. Given that these restrictions remove much of the firm-level variation in event activity – see Figure 2 below – for robustness, we rerun our analysis for a) all firms with event exposure, including large and multi-plant businesses, and b) all SMEs.

---

<sup>10</sup> We also link our data with data from the UK Innovation Survey, although this is less successful. Specifically, we link our data with Waves 4-9 of the UK Innovation Survey (UKIS), covering the years 2002-2014. We match 26,708 firms, of which 1,173 are single-plant SMEs. We run Hotelling tests to determine whether this set of firms is systematically different from the rest of the sample. The results, which are all significant at the 1% level, show that the UKIS subsample differs on a large set of observable characteristics (Hotelling's  $T^2 = 24866.03$ ,  $F(29,2468834) = 857.440^{***}$ ). In particular, the UKIS firms have a substantially higher probability of event exposure (2.91% vs 0.86%) and launch activity (1.03% vs 0.31%). The firms with event exposure that are in the UKIS subsample are also systematically different from the other firms with event exposure (Hotelling's  $T^2 = 487.983$ ,  $F(21868) = 17.407^{***}$ ). Given these substantial differences, we do not use UKIS data in the subsequent analysis.

<sup>11</sup> We also remove outliers: for each year, we remove observations with an event count higher than 1 standard deviation above the mean event count. This eliminates 84 observations.

*Figure 2 about here*

Table 1 provides a comparison of our full dataset and subsamples. For each, the first panel gives the number of firm-year observations for each year, the years during which we observe events, and the number of unique firms. The next panel gives the number of firms for which we observe events of any kind and the total number of events for each year. We repeat this for launches and then for patents and trademarks. Overall, 1.6% of all the firms and 1.3% of the SMEs have event exposure; in both cases, over 37% of the firms also have launch exposure. Events and launches are well balanced across all the years. There are rather smaller shares of firms with patents and trademarks, with more uneven coverage over time.

*Table 1 about here*

### **3/ Descriptive analysis**

We now explore how modelled launches compare with other innovation measures. We make a launch dummy that takes the value of one when a firm has at least one launch during a given year. We also count each firm's launches in that year. We then use the number of raw observations per modelled event to create measures of launch 'importance'. For each firm-year cell, we make a count of mentions, a dummy for whether a firm has an 'important' launch, and a count of such launches. Further details are given in Appendix A1.

Single-plant SMEs with event coverage have approximately 2.2 events, of which approximately 0.7 are product/service launches. The firms have fewer patents and trademarks. While the average launch has over 250 raw media reports, this is driven by a small number of high-profile

events; only 2% of the SMEs with events have ‘important’ launches with more than one underlying media report. Appendix Table B1 provides details.

*Table 2 about here*

As anticipated, event exposure is not random, with firms selected on a range of observables. Table 2 shows the mean characteristics of SMEs with and without events and launches for 2014-2015, the years during which we observe event activity. We can see that the mean differences between firms with and without event exposure are large; rank-sum tests confirm significant mean differences for all observables. In contrast, for firms with events, differences between those with and without launch activity are rather small and often nonsignificant.<sup>12</sup> Companies with launches are more likely to obtain patents and trademarks. They are older, have significantly lower revenue productivity, are less likely to have high revenue and growth episodes, are more likely to be foreign-owned, and are more likely to be listed companies than partnerships. However, the balance on shares of start-ups, sole proprietors and small firms, business group structure, number of employees, urban location, revenue, revenue per worker growth, employment level, growth and high-growth episodes. Therefore, we focus the second part of our analysis on SMEs with events.

*Table 3 about here*

Table 3 compares coverage of patents, trademarks and reported launches at the industry level for 2014-2015, the years during which we observe events. We show launch, patent, and trademark coverage across SIC1 bins for firms with event exposure (Table B3 repeats this analysis for all SMEs, showing similar results). Overall, launches have a wider industry spread

---

<sup>12</sup> Rank-sum tests are preferred, as we do not know the underlying distribution of events. T-tests give virtually identical results and are available on request.

than patents or trademarks; for a minority of firms, we also find correlations between the three measures.<sup>13</sup> Patenting is concentrated in manufacturing, but it is also present in services, notably business services (including software and other ‘knowledge-intensive’ activities (Castellacci, 2008)). Given their broader functionality, trademarks are more evenly distributed, with most related activity in manufacturing, wholesale/retail/repair and social/personal services.

Finally we turn to spatial variation. Appendix Figure B1 shows geographical coverage of events, launches, patents and trademarks across urban travel to work areas (TTWAs), which approximate local spatial economies. Panel A provides a simple scatterplot of launches, patents and trademarks across TTWAs. Based on raw counts, coverage across spatial economies appears even, although launch counts are substantially higher than patent or trademark counts. London is a major outlier in terms of counts, even for single-plant SMEs.<sup>14</sup> To correct for this, Panel B plots TTWA counts weighted by the number of firms in each TTWA. We can see that when local economic conditions are taken into account, launches have a far more even geographical distribution than either patents or trademarks.

## **4/ Research design**

### 4.1/ Theoretical framework

We develop a simple framework to formally explore links between launches, other innovation measures and firm performance, which we take to data in the rest of this paper. We start with

---

<sup>13</sup> Specifically, Table B2 further checks for the dispersion of launch activity. It reports the number of launches for firms with one patent or one trademark in either 2014 or 2015 for all SMEs, and for those with event exposure. While the majority have zero launches to show for the IP, a small number have two or more reported launches per patent or trademark. Table B3 reports the industry coverage for all single plant SMEs.

<sup>14</sup> Table B4 shows the urban/non-urban and London/non-London shares of launches, patents and trademarks for single-plant SMEs. Launches and other innovation metrics are highly urbanised.

the quality ladder model developed by Aghion and Howitt (1992), in which product innovation stems from a combination of entrants seeking market share (and thus temporary revenue markups) and incumbents seeking to protect it. Innovation at the firm level leads to increased average product quality, knowledge spillovers across firms, and the reallocation of capital, workers and products across and within firms as products and businesses enter and exit the market (Klette and Kortum, 2004; Lentz and Mortensen, 2008).<sup>15</sup> In turn, these three forces drive aggregate revenue and productivity growth. In practice, any firm's growth path is further shaped by its individual endowments, including capabilities built up over time (Nelson and Winter, 1982; Cohen and Levinthal, 1990; Dosi et al, 2000). Firms and their managers also operate with bounded rationality and may individually maximise profits, growth or some combination of the two (Marris, 1963). These factors skew growth distributions so that a small number of high-growth/superstar firms typically has a disproportionate impact on aggregate outcomes (Nightingale and Coad, 2013).

Given our short timeframe, we say a firm has fixed capability endowments and grows through developing products and services – either new to the firm or the market. Specifically, innovation is reflected in IP and is a function of costly past and current R&D and managerial and organisational capabilities. This can come from either inside a firm or external sources via spillovers (Akcigit and Kerr, 2018). Incumbents with valuable IP may be less likely to source innovations externally (Bei, 2019). A firm's probability of conducting a successful launch then depends on existing capabilities and knowledge stocks accumulated through past product innovations. Higher-quality products/services, produced by a subset of more capable and/or experienced firms, allow for higher markups but are also costlier to develop.

---

<sup>15</sup> Most of the earlier literature on reallocation has focused on input markets, combining labour market and establishment data (see, for example, Foster et al, 2016). In contrast, Argente et al (2018) focus on output markets.

Firms also need to decide how to protect innovations and to what ends. As noted above, the majority of firms avoid using formal IP tools or combine formal and informal tools (Hall et al, 2014). In principle, patents indicate ‘upstream’ inventions and trademarks denote ‘downstream’ commercialisation (Castaldi et al, 2020; Flikkema et al 2019). In practice, they are used as complements or substitutes (Llerena and Millot, 2020; De Vries et al, 2017) at varying points in the innovation process (Seip et al, 2018), and varying across industry and market contexts (Jensen and Webster, 2009). In particular, brand creation trademarks and trademarks with a narrow scope may be important indicators of innovation in start-ups and young firms (Flikkema et al, 2019).

Formally, Argente et al. (2019) distinguish between ‘productive’ IP, which creates new products and protects revenue markups, and ‘protective’ IP, which creates future revenue sources. Large firms worry about cannibalizing their existing products; thus, returns to new products decrease with firm size while returns to patenting increase. Large firms are also better able than small firms to bear the costs of formal IP protection. This implies that small firms below a certain size cut-off may not patent or trademark at all, while large businesses have multiple filings per new product; the cost differentials between patents and trademarks also influence this (Castaldi et al, 2020). Additionally, firms can choose between engaging in product/service innovation and advertising/marketing their existing offerings. Advertising may complement innovation by increasing markups on new products or substitute for it by increasing revenues on existing products (Bokhari et al, 2020; Cavenaile and Roldan 2020). Larger (and older) firms tend to prefer advertising their existing products over product innovation due to spillovers from umbrella branding.

#### 4.1/Empirical specification

This framework generates a number of predictions that we can explore in our data – both on links between IP activity and launches and from launches to firm growth (proxied by revenue per worker, as is usual in this literature). First, we should expect a positive link between past IP (patents and trademarks) and launches. The timing of IP activity is ambiguous. Knowledge stock decay implies that past activities have weaker links, but to the extent that past IP proxies for individual firm capabilities, it positively predicts launch activity. Per Flikkema et al. (2019), we should expect stronger links for more narrowly focused trademarks and for trademarks related to services. Per Argente et al. (2018), larger firms should produce more IP per launch than smaller firms; the exact size cut-off is an empirical question. Relatedly, younger firms should have higher probabilities of launching and lower returns to formal IP. Sectors where the cost of R&D is lower should exhibit stronger IP-launch links – for example, services should have stronger links than manufacturing (Audretsch et al, 2020).

Second, we should expect a positive relationship between firm launch activity and levels of revenue productivity, as innovations generate temporary monopolies for their producers. To the extent that they reflect higher-quality products/services, more ‘important’ launches should generate stronger revenue/worker effects. In contrast, the link to revenue productivity *growth* is ambiguous, as we do not observe different levels of overall competition in our data.

In theory, to estimate the link between IP and launches, we can estimate for firm  $i$  in year  $t$ , TTWA  $a$  and sector  $s$ :

$$L_{itas} = F(\mathbf{IP}_{it-n}, \mathbf{X}_{it-n}, T_t, A_a, S_s, e_{itas}) \quad (1)$$

where  $L$  is a measure of launch activity, including proxies for launch importance/quality;  $\mathbf{IP}$  is a vector of past patenting, trademarking and self-reported innovation;  $\mathbf{X}$  is a vector of time-

varying controls; and T, A and S are year, area and industry fixed effects, respectively.

Similarly, to estimate the launch-growth link, we can estimate:

$$Y_{itas} = F(L_{it-n}, \mathbf{IP}_{it-n}, \mathbf{X}_{it-n}, T_t, A_a, S_s, u_{itas}) \quad (2)$$

where Y is a measure of revenue productivity, and other terms are defined as above.

This design leaves us with three main challenges. First, in our framework, firms have fixed, individual capability endowments; additionally, the decision to innovate varies at the firm level. Only some relevant determinants are observable, and our short panel makes it challenging to fit firm fixed effects. Blundell et al. (1995) propose using firm-specific ‘level effects’ based on historic patenting activity, and we follow this alternative to capture firm-level heterogeneity. Thus we estimate the following:

$$L_{itas} = F(\mathbf{IP}_{it-n}, \mathbf{X}_{it-n}, \mathbf{HP}_i, T_t, A_a, S_s, e_{itas}) \quad (3)$$

$$Y_{itas} = F(L_{it-n}, \mathbf{IP}_{it-n}, \mathbf{X}'_{it-n}, \mathbf{HP}_i, T_t, A_a, S_s, u_{itas}) \quad (4)$$

Second, launches (and events, more broadly) are media-reported rather than directly observed. For a given firm, event exposure is determined by a) a firm’s decision to seek coverage, b) its capacity to do so, and c) media interest in reporting the firm’s activity. The value of media coverage varies across firms and is a function of management strategy. The capacity to achieve is a function of management quality (Cohen and Levinthal, 1990), resources and other characteristics (such as age, size, legal and corporate structure) (Teece et al, 1997). Both firm choices and capacity are also shaped by industry characteristics, trends and macro forces, such as national/international policy regimes, trade frictions and changes in these factors (Cockburn



et al, 2016). Media interest may vary across industries (for instance, on levels of newsworthy content) and locations (physical proximity to media producers), and is affected by media industry trends related to reporting capacity and coverage (Davies, 2009; Viner, 2016).

Much of this can be addressed with controls and fixed effects, while our single-country setting eliminates cross-country differences.<sup>16</sup> Nevertheless, unobservables that affect event exposure may also condition both sides of equations (3) and (4). Because launches are observed only conditional on event exposure, we cannot directly control for the latter. While we could in principle use a Heckman or IV estimator to handle selection, in this case there is no obvious instrument. Thus, our preferred approach is to estimate equations (3) and (4) for the sample of SMEs with events so that we can estimate the examined linkages conditional on all drivers of event exposure. We also run diagnostic/falsification tests on this sample, showing that past IP is linked to launches and not other event counts, while other events are, as expected, correlated with our level effect (past patenting). This provides further support that we are both estimating a true IP-launch relationship and controlling for unobservables that drive both selection and outcomes. In robustness checks, we quantify these sources of bias by re-estimating (3) and (4) for all SMEs and for all firms with events.

Third, per Cavenaile and Roldan (2020), launches have measurement error to the extent that media coverage functions as a form of advertising for firms. It is extremely challenging to distinguish the direct revenue effect of a new product/service from that of the launch process. Nevertheless, an observation of the predicted significant positive relationship between past IP and launches would support the idea that reported launches are linked to innovation rather than being purely a form of advertising. Relatedly, we lack data on advertising spending at the firm

---

<sup>16</sup> Additionally, we assume that media interest in any given firm is equal, conditional on sector, year and individual level effects. While we might worry that individual firms could influence media interest through their market position or by buying advertising, this is less plausible in our main sample of single-plant SMEs.

level.<sup>17</sup> However, if advertising spending is equal within firm size and industry bins, our specification eliminates spending as an unobservable.

## 5/ Results

First, we examine the link between patenting/trademarks and the extensive and intensive margins of product launches. We then move to the gains of innovation, estimating the link between launches and revenue productivity.

### 5.1/Linking past IP with launch activity

Table 4 gives the results of equation (3): for SMEs with events, we regress launch activity on past IP stocks, controlling for a range of firm characteristics, local and sectoral conditions. In IP, patents and trademark stocks are depreciated with the standard 15% depreciation rate (Hall and Harhoff, 2012).<sup>18</sup> Trademark stocks are constructed in the same way. We define 'recent' patenting as that occurring in a given five-year period such that  $n$  takes the value 0, 1... 5 for patents and for EPO/US/PCT filings in any given year since 2009.<sup>19</sup> For trademarks,  $n$  takes the value 0, 1 or 2 based on the available data. As discussed above, following Blundell et al. (1995), we use individual firms' historic patent stocks as proxies for firm-level experience, absorptive capacity and other unobservables.<sup>20</sup> We define 'historic' patenting as that taking place before

---

<sup>17</sup> While the UK Innovation Survey asks many questions about firm spending, it does not cover advertising.

<sup>18</sup> This 15% rate is varied in sensitivity tests.

<sup>19</sup> We use filings to these offices as a proxy for invention quality: inventions filed in international domains rather than to a single country are 'worth' more to applicants (Helmers and Rogers, 2010). Alternatives are triadic patent family constructs as an ex-ante measure of quality or patent citations as an ex-post measure.

<sup>20</sup> Many of the cited approaches normally include R&D and advertising expenditures. Our data makes this challenging. We do not observe firm-level advertising spending. The UKIS data contain R&D spend information, and we match this to our panel, but the sample is small and highly selective. Commercial sources such as Orbis have limited direct coverage (7,600 'industrial companies' in the UK with R&D expenditures in their annual accounts);

2009. Specifically, in **HP**, we include a dummy taking the value 1 if a firm has patented before this date and an average of pre-2009 patenting activity taking the values  $p = 0 \dots p$ .<sup>21</sup>

*Table 4 about here*

Table 4 gives results for both the product launch dummy (Panel A) and the count of product launches (Panel B), fitting progressively more demanding specifications.<sup>22</sup> Overall, we find the positive link between patents and launches predicted in our framework. For our preferred linear probability model (Panel A, column 4), past patenting increases launch probabilities by 0.7% points the following year. For trademarks, conversely, the overall relationship is close to zero and non-significant: this varies when we decompose trademarks by type and scope, as seen below. Consistent with our framework, historical patenting predicts current launch activity: firms with some historical IP are 9.7 percentage points more likely to have a launch in any sample year. The *number* of historical patents is a significant negative predictor, however, consistent with depreciation from bigger stocks of older patents. We see similar patterns for the launch count model (Panel B): 10 additional patents in a given year are linked to over 0.5 extra launch events the following year, while trademarks have no effect. Here, historical patenting has no significant link with the intensive margin of launches.<sup>23</sup>

---

UK SMEs file minimal returns with Companies House, so it is difficult to reconstruct standard proxies. As an alternative, we follow Audretsch et al. (2020) and infer the role of R&D activity by subsetting it across industry bins.

<sup>21</sup> We estimate using OLS because nonlinear estimates converge to OLS results once converted to marginal effects (Angrist and Pischke, 2009). OLS is also more efficient given the very large number of fixed effects in our data. The functionally ‘correct’ estimation methods are the Zero-inflated Poisson or Zero-inflated Negative Binomial methods. Angrist and Pischke (2009) convincingly show that once the raw coefficients produced by these estimators are converted to marginal effects, the results are essentially identical to those of OLS.

<sup>22</sup> Sample size changes drive results in different columns. To make sure the small differences in the results are driven by sample selection, we run the same regressions, keeping the sample size constant. Results are qualitatively the same, with very minor changes to the coefficients.

<sup>23</sup> These results survive an extensive set of sensitivity checks and re-running on different samples. In Table B5 we vary the lags for patents and trademarks; in Tables B6-B7 we add controls for past high-growth episodes; add technology field fixed effects; re-specify patents using cumulative patent counts; and use 40% depreciation rates, following Li and Hall (2020).

Then, we focus on other predictions. We re-estimate (3) separately for product, service and specialised trademarks – defined using NICE codes (see Appendix 2 for details). Table B8 shows the results. We find larger IP-launch dummy coefficients for service trademarks than for product trademarks, although we have significantly fewer of the former.<sup>24</sup> In line with Flikkema et al. (2019), for specialised trademarks, the links are larger still and are marginally significant.

We also divide the panel into manufacturing and services subsamples (based on SIC1 classifications) and subgroups based on firm size (following OECD definitions of sole traders, micro, small and medium-sized firms) and age (those under 10 years old, the youngest 25% of the sample, and the remainder). The results are given in Tables B9 and B10. Although we do not observe R&D costs directly, we can infer them from sectoral information (Audretsch et al 2020): as expected, we find stronger links between patents and launches in the services sector, where R&D capital costs are lower than they are in the manufacturing sector. Consistent with Argente et al. (2019), since larger firms file more patents per launch, the coefficients of patents on launch probabilities and launch counts are overall increasing with firm size. There is one exception: micro firms (with 1-9 staff) are more likely than medium-sized firms (with 25-249 staff) to launch with a prior patent, and they generate more launches per patent. Consistent with our framework, we also find that the youngest 25% of firms have higher probabilities of launching (per their past IP) and lower returns to patenting than more established businesses (nonsignificant coefficients of IP on launch count, versus a link that is significant at the 1% level for the oldest 75%).<sup>25</sup>

---

<sup>24</sup> See Table B1. We have 8,493 trademarks across 5,189 firms, of which 4,744 have only product NICE codes and 1,969 have only service NICE codes. When we re-run this test for all SMEs, we find the same pattern of results, but effect sizes are larger and all significant. Launch activity is unaltered, so this is driven by the larger sample plus more variation in trademarking activity. Results are available on request.

<sup>25</sup> In Table B11, we examine links between past IP activity and launch quality/importance measures, but we find no significant linkages, and these factors are instead linked to revenue per worker (Section 6). In Table B12, we rerun our main analysis on all single-plant SMEs (Panel A) and on all firms with event exposure (Panel B). For all SMEs, we find slightly smaller coefficients, with a significant link to both patenting and trademarking; for all firms with events, we find similar results on the extensive margin, but nonsignificant links to launch counts. This implies that in explaining the IP-launch relationships across the examined population of firms, event exposure may be less salient than the role of firm size, although as explained in Section 2, in our all-firm sample, we risk error in ascribing launches to specific plants and locations.

There are three main caveats to these exercises. First, although our results are robust to varying the lag, the true time decay function between IP and launches is unclear. Second, measurement error on both sides of equation (3) affects our estimates. The majority of UK innovations are not protected with formal IP (Hall et al, 2013). Many new products/services involve multiple patents; for instance, the iPhone reportedly has over 100 (Mazzucato, 2013). We also test aggregate links for each firm using many years of patents and trademarks, but only two financial years' worth of reported launches. While measurement error related to patents and trademarks may downward bias the estimates, we can consider the error in product launches to be as good as random conditional on observables.

Third, even conditioning on event exposure may not fully control for unobservables. Per our framework, past IP should have a stronger link with launch activity than with other types of events, such as mergers or staff changes. However, if *any* kind of event exposure is a proxy for underlying knowledge capabilities (Klette and Kortum, 2004), it may affect both IP and launch activity differently across individual firms. We test this in Table B13, restricting to the set of SMEs with event exposure but no launches, then regressing this alternative event count on **IP**. Reassuringly, we find that recent patenting and trademarks are not associated with non-launch event counts. Additionally, we find positive, significant links between firms' historic patenting and other event exposure. This provides further evidence that we are estimating a link between past IP and launches specifically, which is not affected by individual variation in other types of event exposure, and that our individual-level effects capture relevant firm-level heterogeneity.

## 5.2/ Linking IP, launches and firm performance

To explore links between IP/launch activity and firm performance, for SMEs with events, we estimate:

$$Y_{itas} = a + bL_{it-1} + cPATS_{it-2} + dTM_{it-2} + \mathbf{HP}e_i + \mathbf{X}f_{it-n} + T_t + A_a + S_s + u_{itas} \quad (5)$$

This function allows us to study the link between launch activity and subsequent performance changes at the firm level, conditional on previous patenting and trademarking. As before, we then examine subsamples to explore heterogeneity in the innovation-launch-performance relationship. As in existing studies, we specify Y as revenue per worker (Mohnen and Hall, 2013; Klette and Kortum, 2004). We fit Y in both levels (log revenue/worker) and changes (% revenue worker growth/year). Alternately, we specify Y as a dummy indicating whether a firm has at least one ‘high-growth’ revenue growth episode – per the OECD definition – during the sample period. This last specification focuses on the most dynamic firms in the sample. Given the short panel, levels is likely to be more informative than changes or growth episodes.

Controls, lagged launches,<sup>26</sup> patents and trademarks are specified as in the previous subsection: we lag the latter two periods to allow ‘upstream’ IP to influence ‘downstream’ launches. Given the short panel, we use each firm’s pre-2009 patenting activity as a proxy for firm-level heterogeneity.

Tables 5 and 6 give results for the subsample of firms with event exposure. For each outcome, we fit the model with launches (columns 1, 3, and 5) and without launches (columns 2, 4 and 6). We interpret coefficients of *b* as expressing the association between launches and revenue productivity, conditional on media exposure.

*Table 5 about here*

---

<sup>26</sup> We can only lag launches by one year. We are aware that any estimated correlations may be industry-specific. We do not run each regression separately by sector, although we control for industry fixed effects (and in some specifications, for industry-by-year fixed effects) to account for average sectoral heterogeneity.

As suggested by our framework, Table 5 shows that product launches have a positive, significant relationship with log revenue productivity. Specifically, SMEs with launches have 6.4% more revenue per worker than firms with other types of event exposure (column 1): a 1 standard deviation (0.42) increase in the average launch probability is associated with a 2.7% increase in revenue productivity. While recent patenting has no relationship with revenue per worker, recent trademarking has a positive, significant association, as each additional trademark is linked to an 8.1% increase in revenue productivity (column 1). This result is in opposition to our earlier IP-launch results, suggesting both that patents affect revenue per worker through launches and that patents and trademarks are complements, consistent with De Vries et al (2017). We speculate that the two-year lagged trademark results may partly reflect revenue markups from launches prior to 2014, which we do not observe in our data. We find no link between launch activity and revenue productivity growth or high-growth episodes; patent and trademark links are also weak or non-significant here (columns 3 and 5). Given our short panel, this is, perhaps, not surprising.

*Table 6 about here*

Table 6 gives results for launch counts. Each additional launch is linked to a 1.7% increase in revenue productivity (column 1), and this significant at the 1% level. As in Table 5, we find no link with recent patenting, but we see a clear, positive link with recent trademarks, which is larger than that with launches. Removing launches from the model (column 2) reduces the model fit, as before. As before, we find that launch counts do not predict revenue per worker growth or high-growth episodes.

These results are robust to a battery of robustness checks. Tables B14-B16 in the Appendix give results for our three dependent variables. Table B17 reruns the levels result for alternate

samples. For all SMEs (Panel A), selection into the events sample drives many of the major associations in our main results. Firms with product launches have 45% higher revenue productivity than those without, and this is significant at the 1% level (column 1). We find similar results regarding the launch count, which has a positive, significant relationship with log revenue productivity. As expected, not controlling for underlying media exposure substantially strengthens the launch count-performance link. Specifically, each additional launch increases revenue productivity by 4.7%, though underlying media exposure is uncontrolled. In contrast, adding in larger, multiplant firms with events (Panel B) only increases the launch-performance link to 8.5%, from 6.4% in our main results, and there is now a nonsignificant link between revenue productivity and the number of launches. Per our framework and Cavenaile and Roldan (2020), this is consistent with larger firms being more likely to have other sources of revenue such as advertising and with the presence of spillover effects from existing products and services.

## **6/ Extensions**

We extend our main results in four ways. First, we decompose trademarks into product, service and specialised categories, as before (Table B18). We find that coefficients of launch activity are essentially unchanged; however, consistent with Castaldi et al. (2020), counts of specialised trademarks are more strongly linked to revenue per worker than the simple trademark count is.<sup>27</sup> Second, we split the events subsample to examine the launch-revenue productivity links in manufacturing and services industries. Tables B19 and B20 give results for the linear probability and count models, respectively. In both cases, overall, positive links are driven by firms in the services sector. Service firms also drive the trademark results, consistent with

---

<sup>27</sup> As before, when we rerun this test for all SMEs, we find the same pattern of results but with larger, more robust effect sizes. These results are available on request.



previous studies (Flikkema et al, 2019; Castaldi et al, 2020). For manufacturing firms, recent patenting is linked to lower revenue productivity growth, but historical patenting is correlated with higher revenue per worker growth. Overall, these results are consistent with Audretsch et al. (2020), who suggest that barriers to (reported) innovation are lower for service firms than manufacturing firms.

Third, we investigate the role of firm size and age in explaining our results (Table B21). As before, we group firms into size bins using OECD definitions, and define young firms as the youngest 25% of firms in the events sample. For launch dummies, we first fit our main regression with age and size dummies (column 1), then add size and age group interactions (columns 2 and 3, respectively). In columns 4-6, this is repeated for launch counts. Overall, the results are driven by medium-sized firms, while there is no effect of age. Specifically, for the extensive margin, we find a positive revenue/worker link for small firms, but this is half as strong as that found for medium-sized firms.

Finally, we examine the link between launch quality and revenue productivity. We use the number of media reports per event as a proxy for quality, as detailed in Section 2. We re-estimate equation (5) using four *alternative* quality measures in separate regressions: 1) a simple count of the number of reports across each firm's launches per year; 2) firm-year counts weighted by the number of launches; 3) a dummy for whether a firm has an 'important' launch with many mentions; and 4) the number of important launches per firm per year. Table 7 gives the results when we look at counts for the main event topic (using counts across all topics and counts weighted by topics give identical findings). We find very small positive links between the report counts and levels of revenue productivity and very small negative links to revenue productivity growth. We do not find links for weighted report counts. We find large, significant associations between having an important launch and revenue per worker and between having an important launch and the count of important launches. Specifically, SMEs with at least one

important launch have approximately 17% higher revenue productivity than other SMEs with media exposure; each additional important launch increases revenue per worker by nearly 22%. This suggests that our main results, which link launch activity to SME revenue productivity, are significantly driven by a small set of high-profile, important product and service launches.

*Table 7 about here*

## **7/ Conclusions**

A vast field of literature explores the links between innovation and economic performance (Romer, 1990; Aghion and Howitt, 1992; Coad, 2009; Akcigit, 2017). Four streams of empirical work unpack connections at the firm level. An established set of studies uses R&D, patents and innovation surveys; newer analyses use trademarks, text-based measures or product-level data. However, this body of work has two constraints: the informality of much innovative activity (Hall et al, 2014) and practical limits in processing richer text or product-level information. Our paper makes three practical and empirical advances on this literature. First, we develop novel product-level innovation metrics that extend existing studies. Second, we show that our new measures complement existing, formal IP metrics. Third, we find positive links between SME launch activity and revenue per worker. Importantly, we also find that industry, size, IP strategy and launch quality differences moderate our main results.

We highlight three main lessons for existing research, and for practice. First, our results further confirm the overall positive links between firm-level innovation and growth found in many previous studies (Audretsch et al, 2014). Our findings are also consistent with more recent work on trademarks and innovation (Castaldi et al, 2020). Second, however, and consistent with extreme heterogeneity, we show that a subset of high-growth firms can drive overall innovation

and growth outcomes (Nightingale and Coad, 2013). Developing policy tools to identify and support such firms is both important and highly challenging. Third, we show the value of monitoring innovations not captured by formal IP or surveys – and the rich potential of text-based sources to achieve this (Gentzkow et al, 2019).

Four limitations of our work may inform future research. First, we explore heterogeneity mainly via subsamples. One could instead use data-driven approaches to identify high-growth businesses, as in Coad et al. (2016), or richer firm-level information that covers management strategy, as in Grillitsch et al. (2019). Second, it would be valuable to link our data to information on firm advertising, capital intensity and R&D, as in Hall et al. (2013) and Cavenaile and Roldan (forthcoming). Third, we utilize a short timeframe; a longer time series would allow an analysis of macro conditions, as in Spescha and Woerter (2019). Finally, our analysis is not causal. Future work could improve on this issue by exploiting policy evaluation settings or finding viable instruments.

## References

- Akcigit, U. (2017) "Economic Growth: The Past, the Present, and the Future." - Commemorative essay in "The Past, Present, and Future of Economics: A Celebration of the 125 Year Anniversary of the JPE and of Chicago Economics," edited by John List and Harald Uhlig. *Journal of Political Economy* 125(6): 1736-1747.
- Akcigit, U. and Kerr, W. (2018) Growth through Heterogenous Innovations. *Journal of Political Economy* 126(4): 1347-1443.
- Aghion, P. and Howitt, P. (1992), A Model of Growth through Creative Destruction, *Econometrica*, 60(2): 323-51.
- Aghion, P., Blundell, R., Griffith, R., Howitt, P. and Prantl, S. (2009) The Effects of Entry on Incumbent Innovation and Productivity. *Review of Economics and Statistics* 91: 20-32.
- Angrist, J and Pischke, J-S. (2009) *Mostly Harmless Econometrics*, Princeton: Princeton University Press.
- Argente, D., Lee, M. and Moreira, S., (2018), Innovation and product reallocation in the great recession, *Journal of Monetary Economics*, 93:. 1-20
- Argente, D, Baslandze, S, Hanley, D, Moreira, S. (2019) Patents to Products: Innovation, Product Creation and Firm Growth. *Mimeo*.
- Arts, S., Hou, J., Gomez, J.C., (2021) Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures. *Research Policy* 50, 104144.
- Audretsch, D. B., and Belitski, M. (2020) The role of R&D and knowledge spillovers in innovation and productivity. *European Economic Review*, 123: 103391
- Audretsch, D.B., Coad, A., Segarra, A. (2014) Firm growth and innovation. *Small Business Economics* 43, 743-749.
- Audretsch, D.B, Hafenstein, M., Kritikos, A.S. (2020) Microfirms and innovation in the service sector. *Small Business Economics*, 55: 997–1018.
- Autor, D., Dorn, D., Hanson, G., Pisano, G., and Shu, P. (2020) Foreign Competition and Domestic Innovation: Evidence from US Patents. *American Economic Review: Insights* 2(3):357-74.
- Axenbeck, J. and Breithaupt, P. (2019), Web-Based Innovation Indicators - Which Firm Website Characteristics Relate to Firm-Level Innovation Activity?, ZEW Discussion Paper No. 19-063
- Baumann, J. and Kritikos, A.S. (2016) The link between R&D, innovation and productivity: Are micro firms different? *Research Policy* 45: 1263-1274.
- Bei, X., (2019) Trademarks, specialized complementary assets, and the external sourcing of innovation. *Research Policy* 48, 103709.
- Bianchini, S., Pellegrino, G., Tamagni, F., (2018) Innovation complementarities and firm growth. *Industrial and Corporate Change* 27, 657-676.
- Block, JH., Fisch, CO., Hahn, A., Sandner, PG. (2015) Why do SMEs file trademarks? Insights from firms in innovative industries. *Research Policy* 44: 1915-1930.
- Blundell, R., Griffith, R. and Van Reenen, J. (1995) Dynamic Count Data Models of Technological Innovation. *The Economic Journal* 105: 333-344.
- Bokhari, F.A.S., Mariuzzo, F., Bennato, A.R., (2020) Innovation and growth in the UK pharmaceuticals: the case of product and marketing introductions. *Small Business Economics*.
- Bottazzi, G., Dosi, G., Lippi, M., Pammolli, F., Riccaboni, M. (2001) Innovation and corporate growth in the evolution of the drug industry. *International Journal of Industrial Organization* 19, 1161-1187.
- Castaldi, C., Block, J., Flikkema, M.J., (2020) Editorial: why and when do firms trademark? Bridging perspectives from industrial organisation, innovation and entrepreneurship. *Industry and Innovation* 27, 1-10.

- Castellacci, F. (2008) Technological paradigms, regimes and trajectories: Manufacturing and service industries in a new taxonomy of sectoral patterns of innovation. *Research Policy* 37: 978-994.
- Cavenaile, L. and Roldan, P., (2020) Advertising, Innovation and Economic Growth. *AEJ: Macroeconomics, forthcoming*.
- Coad, A., (2009) *The Growth of Firms: A Survey of Theories and Empirical Evidence*. Edward Elgar, Cheltenham.
- Coad, A., Pellegrino, G., Savona, M., (2016) Barriers to innovation and firm productivity. *Economics of Innovation and New Technology* 25, 321-334.
- Coad, A., Segarra, A., Teruel, M., (2016) Innovation and firm growth: Does firm age play a role? *Research Policy* 45, 387-400.
- Cockburn, I.M., Lanjouw, J.O. and Schankerman, M. (2016) Patents and the Global Diffusion of New Drugs. *The American Economic Review* 106: 136-164.
- Cohen, W.M. and Levinthal, D.A. (1990) Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly* 35: 128-152.
- Coombs, R., Narandren, P. and Richards, A. (1996), A literature-based innovation output indicator. *Research Policy* 25: 403-413.
- Corsino, M., Gabriele, R., (2010) Product innovation and firm growth: evidence from the integrated circuit industry. *Industrial and Corporate Change* 20, 29-56.
- Crass, D., (2020) Which firms use trademarks? Firm-level evidence from Germany on the role of distance, product quality and innovation. *Industry and Innovation* 27, 730-755.
- Crepon, B., Duguet, E., & Mairesse, J. (1998). Research, Innovation And Productivity: An Econometric Analysis At The Firm Level. *Economics of Innovation and New Technology*, 7(2), 115-158.
- Cucculelli, M., Ermini, B., (2012) New product introduction and product tenure: What effects on firm growth? *Research Policy* 41, 808-821.
- Davies, N. (2009) *Flat Earth News*, London: Vintage.
- De Vries, G., Pennings, E., Block, J.H., Fisch, C., (2017) Trademark or patent? The effects of market concentration, customer type and venture capital financing on start-ups' initial IP applications. *Industry and Innovation* 24, 325-345.
- Dosi, G., Nelson, R., Winter, S., (2000) *The nature and dynamics of organizational capabilities*. OUP, Oxford.
- Einav, L. and Levin, J.D. (2014) The Data Revolution and Economic Analysis. *Innovation Policy and the Economy* 14.
- Fernandes, A.M. and Paunov, C. (2015) The Risks of Innovation: Are Innovating Firms Less Likely to Die? *The Review of Economics and Statistics* 97: 638-653.
- Flikkema, M., Castaldi, C., de Man, A.-P., Seip, M., (2019) Trademarks' relatedness to product and service innovation: A branding strategy approach. *Research Policy* 48, 1340-1353.
- Flikkema, M., De Man, A.-P., Castaldi, C., (2014) Are Trademark Counts a Valid Indicator of Innovation? Results of an In-Depth Study of New Benelux Trademarks Filed by SMEs. *Industry and Innovation* 21, 310-331.
- Fosfuri, A., Giarratana, M.S. and Luzzi, A. (2008) The Penguin Has Entered the Building: The Commercialization of Open Source Software Products. *Organization Science* 19: 292-305.
- Foster, L., Grim, C. and Haltiwanger, J. (2016), Reallocation in the great recession: cleansing or not?, *Journal of Labor Economics*, 34 (S1), S293–S331.
- Gentzkow, M. and Shapiro, J.M. (2010) What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica* 78: 35-71.
- Gentzkow, M., Kelly, B.T. and Taddy, M. (2019) Text as Data. *Journal of Economic Literature* 57(3):535-574
- Gilbert, R.J. and Newbery, D.M.G. (1982) Preemptive Patenting and the Persistence of Monopoly. *The American Economic Review* 72: 514-526.

- Gotsch, M., Hipp, C., (2012) Measurement of innovation activities in the knowledge-intensive services industry: a trademark approach. *The Service Industries Journal* 32, 2167-2184.
- Griliches, Z. (1979) Issues in assessing the contribution of R&D to productivity growth. *Bell Journal of Economics* 10.
- Griliches, Z. (1986) Productivity, R&D, and basic research at the firm level in the 1970s. *American Economic Review* 76.
- Grillitsch, M., Schubert, T., Srholec, M., (2019) Knowledge base combinations and firm growth. *Research Policy* 48, 234-247.
- Guarascio, D., Tamagni, F., (2019) Persistence of innovation and patterns of firm growth. *Research Policy* 48, 1493-1512.
- Hall, B.H. (2011) Innovation and Productivity. *NBER Working Paper 17178*. Cambridge, MA: NBER.
- Hall, B.H.; Harhoff, D. (2012). Recent Research on the Economics of Patents, *Annual Review of Economics*, 4 (1), 541-565.
- Hall B.H., Helmers, C., Rogers, M., Sena, V. (2013) The Importance (or not) of Patents to UK Firms. *National Bureau of Economic Research Working Paper 19089*. Cambridge, MA: NBER.
- Hall, B.H., Helmers, C., Rogers, M., Sena, V. (2014) The Choice Between Formal and Informal Intellectual Property: A review. *Journal of Economic Literature* 52: 375-423.
- Helmers, C. and Rogers, M. (2010) Innovation and the Survival of New Firms in the UK. *Review of Industrial Organization* 36: 227-248.
- Hoberg, G. and Phillips, G. (2016) Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy* 124: 1423-1465.
- Howell, A. (2015) 'Indigenous' innovation with heterogeneous risk and new firm survival in a transitioning Chinese economy. *Research Policy* 44: 1866-1876.
- Hsiao, Y.-C., Chen, C.-J., Guo, R.-S., Hu, K.-K., (2017) First-mover strategy, resource capacity alignment, and new product performance: a framework for mediation and moderation effects. *R&D Management* 47, 75-87.
- Jensen, P.H., Webster, E., (2009) Another Look At The Relationship Between Innovation Proxies\*. *Australian Economic Papers* 48, 252-269.
- Katila, R. and Ahuja, G. (2002) Something Old, Something New: A Longitudinal Study of Search Behavior and New Product Introduction. *The Academy of Management Journal* 45: 1183-1194.
- Kelly, B., Papanikolaou, D., Seru, A., Taddy, M., 2018. Measuring Technological Innovation over the Long Run, NBER Working Paper 252266. NBER, Cambridge, MA.
- Kinne, J. and Lenz, D. (2019), Predicting Innovative Firms Using Web Mining and Deep Learning, ZEW Discussion Paper No. 19-001, Mannheim.
- Kleinknecht, A., Reijnen, J.O.N. and Smits, W. (1993), Collecting literature-based innovation output indicators. The experience in the Netherlands, in: A. Kleinknecht and D. Bain (Editors), *New Concepts in Innovation Output Measurement* (Macmillan, London) pp. 42-84.
- Klette, T.J. and Kortum, S. (2004). Innovating Firms and Aggregate Innovation. *Journal of Political Economy* 112: 986-1018.
- Kogan, L., Papanikolaou, D., Seru, A., Stoffman, N. (2017) Technological Innovation, Resource Allocation, and Growth. *Quarterly Journal of Economics* 132: 665-712.
- Lafrance, A. (2016) 'Access, Accountability Reporting and Silicon Valley. *Nieman Reports*, 17 August.
- Lenz, D. and Winker, P. (2020), Measuring the Diffusion of Innovations with Paragraph Vector Topic Models. *PLOS ONE* 15(1): 1-18.
- Lentz, R., and Mortensen, D. (2008). An Empirical Model of Growth through Product Innovation. *Econometrica*, 76(6): 1317-1373.
- Li, W. and Hall, B.H. (2020) Depreciation of Business R&D Capital. *Review of Income and Wealth* 66(1): 161-180.

- Llerena, P., Millot, V., (2020) Are two better than one? Modelling the complementarity between patents and trademarks across industries. *Industry and Innovation* 27, 52-79.
- Lucas, R. (1988) On the Mechanics of Economic Growth. *Journal of Monetary Economics* 22: 3-42.
- Mazzucato, M. (2013) *The Entrepreneurial State: debunking public vs. private sector myths*, Anthem Press.
- Mairesse, J. and Mohnen, P. (2010) Using Innovation Surveys for Econometric Analysis. In: Bronwyn HH and Nathan R (eds) *Handbook of the Economics of Innovation*. North-Holland, 1129-1155.
- Marris, R., (1963) A Model of the “Managerial” Enterprise. *The Quarterly Journal of Economics* 77, 185-209.
- Mendonca, S., Pereira, T.S., Godinho, M.M., (2004) Trademarks as an indicator of innovation and industrial change. *Research Policy* 33: 1385-1404.
- Mohnen, P. and Hall, B.H. (2013) Innovation and Productivity: An Update. *Eurasian Business Review* 3: 47-65.
- Mohnen, P. (2019) R&D, Innovation and Productivity. In: ten Raa T., Greene W. (eds) *The Palgrave Handbook of Economic Performance Analysis*.
- Morris, D.M. (2018) Innovation and productivity among heterogeneous firms. *Research Policy* 47: 1918-1932.
- Motoyama, Y. and Bell-Masterson, J. (2014) Beyond Metropolitan Start-Up Rates. Kansas City: Kauffman Foundation.
- Nathan, M. and Rosso, A. (2015) Mapping digital businesses with Big Data: some early findings from the UK *Research Policy* 44: 1714-1733.
- Nelson, R., Winter, S., (1982) *An Evolutionary Theory of Economic Change*. Harvard University Press, Cambridge, MA.
- Nightingale, P., Coad, A., (2013) Muppets and gazelles: political and methodological biases in entrepreneurship research. *Industrial and Corporate Change* 23, 113-143.
- Noel, M., Schankerman, M., (2013) Strategic Patenting and Software Innovation. *The Journal of Industrial Economics* 61: 481-520.
- Office of National Statistics. (2017) Business Structure Database, 1997-2017 SN: 6697. *Secure Data Service Access [computer file]*. Colchester: UK Data Archive.
- Ortega, M.J.R., García-Villaverde, P.M., (2011) Pioneer orientation and new product performance of the firm: Internal contingency factors. *Journal of Management & Organization* 17, 474-497.
- Roberts, M.E., Stewart, B.M. and Tingley, D. (2016) stm: R Package for Structural Topic Models. <http://www.structuraltopicmodel.com>.
- Rodríguez-Pinto, J., Rodríguez-Escudero, A.I., Gutiérrez-Cillán, J., (2012) How market entry order mediates the influence of firm resources on new product performance. *Journal of Engineering and Technology Management* 29, 241-264.
- Romer, P. (1990) Endogenous Technological Change. *Journal of Political Economy* 98: 71-102.
- Santarelli, E. and Piergiovanni, R. (1996), Analysing literature-based innovation output indicators: The Italian experience *Research Policy* 25: 689-71.
- Saunders, A. and Tambe, P. (2015) Data Assets and Industry Competition: Evidence from 10-K Filings. Available at SSRN: <https://ssrn.com/abstract=2537089>
- Schautschick, P., Greenhalgh, C., (2016) Empirical studies of trade marks – The existing economic literature. *Economics of Innovation and New Technology* 25, 358-390.
- Schumpeter, J. (1962) *The Theory of Economic Development*, Berlin: Springer.
- Siepel, J., Camerani, R., Masucci, M., Velez Ospina, J., Casadei, P. and Bloom M. (2020) Creative Industries Radar: Mapping the UK’s creative clusters and microclusters. Available at <https://www.pec.ac.uk/research-reports/creative-radar>
- Seip, M., Castaldi, C., Flikkema, M., De Man, A.-P., 2018. The timing of trademark application in innovation processes. *Technovation* 72-73, 34-45.

- Spescha, A., Woerter, M., (2019) Innovation and firm growth over the business cycle. *Industry and Innovation* 26, 321-347.
- Stam, E., Wennberg, K., (2009) The roles of R&D in new firm growth. *Small Business Economics* 33, 77-89.
- Syverson, C. (2011) What Determines Productivity? *Journal of Economic Literature* 49: 326–365.
- Taques, F.H., López, M.G., Basso, L.F., Areal, N., (2021) Indicators used to measure service innovation and manufacturing innovation. *Journal of Innovation & Knowledge* 6, 11-26.
- Teece, D.J, Pisano, G. and Shuen, A. (1997) Dynamic Capabilities and Strategic Management. *Strategic Management Journal* 18: 509-533.
- Viner, K. (2016) How Technology Disrupted the Truth. *Guardian*.



## Figures and tables

**Figure 1. Example ‘events’, showing raw text and classification.**

Sample fragment	<b>Masterwork</b> goes large with new die cutter. Postpress equipment manufacturer Masterwork Graphic Equipment has expanded its range of products with the addition of the MK1450ER large-format die cutter with stripping and blanking facilities....
doc_title	Masterwork goes large with new die cutter
url	<a href="http://www.XXX/NewsStory.aspx?i=2296">http://www.XXX/NewsStory.aspx?i=2296</a>
event_date	2014-03
source_name	xxx
company_id	13724
event_type_id	product_launch

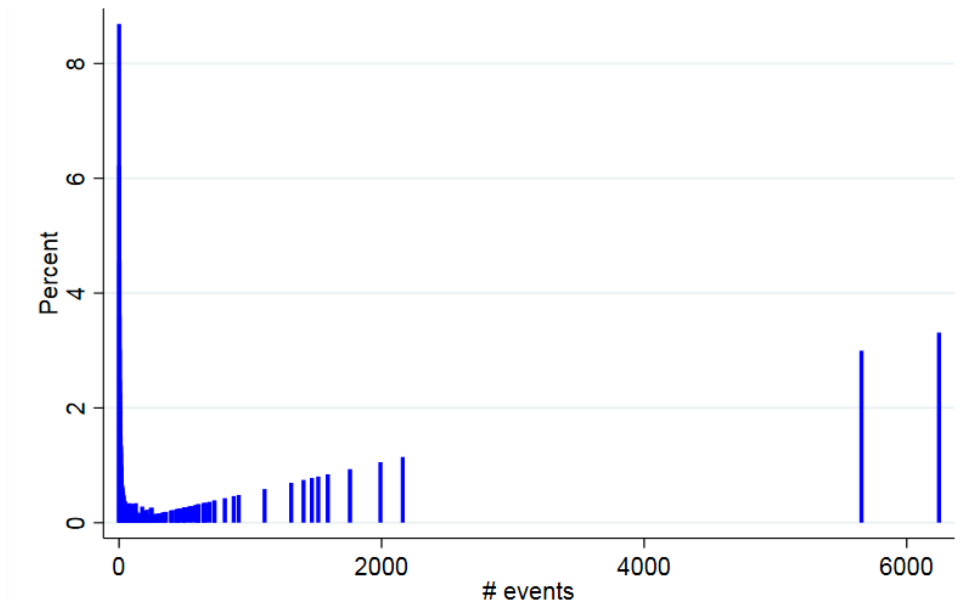
Sample fragment	<b>Hammond Electronics</b> has launched a range of design specific moulded enclosures to support the new types of credit card sized, low cost bare board computers, which, typically running Linux, provide basic functionality across a wide range of applications...
doc_title	Enclosures for credit-card sized computers
url	<a href="http://www.XXXX/content/enclosures-credit-card-sized-computers">http://www.XXXX/content/enclosures-credit-card-sized-computers</a>
source_name	xxx
event_date	2013-12
company_id	1542955
event_type_id	product_launch

Sample fragment	New social housing energy switching service has been launched this week with Wigan and Leigh Housing, <b>Salix Homes</b> in Salford and Blackburn-based Twin Valley Homes, who together are responsible for a total of more...
doc_title	New social housing energy switching service launches
url	<a href="http://www.XXXX/energy-news/17541-new-social-housing-energy-switching-service-launches">http://www.XXXX/energy-news/17541-new-social-housing-energy-switching-service-launches</a>
source_name	xxx
event_date	2013-12
company_id	625998
event_type_id	product_launch

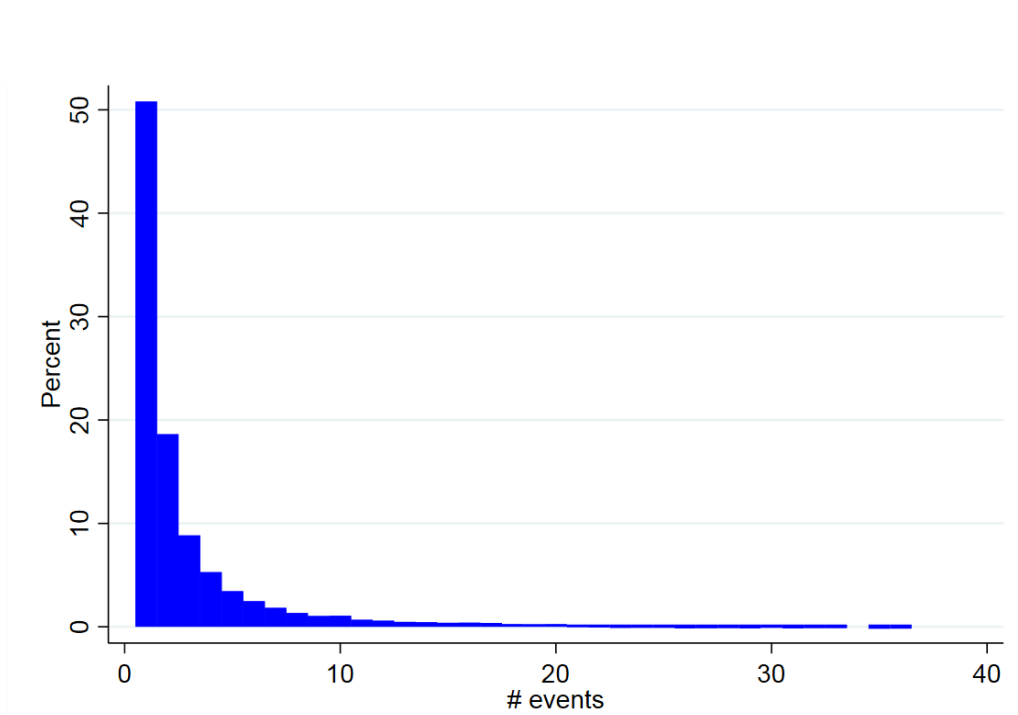
Source: GI. Each example shows the workflow from raw data to modelled variable. GI start with the raw text. We show a sample text fragment here with the company subject in **bold**. Title, URL, date and source name provide further information. As agreed with the data provider we cannot report the source name or the full text. The company ID field shows the match to Companies House data. Event type ID is the eventual classification into an event type: in both cases, these are new product launches.

**Figure 2. Histogram of events activity, 2014-2015. Raw sample (top), single-plant SMEs (bottom).**

A. Raw sample, all firms with events exposure.



B. Single-plant SMEs with events exposure. Disclosive cell counts suppressed.



Source: GI data, years 2014-2015.

**Table 1. Panel characteristics, 2014-2017.**

	<b>All firms</b>	<b>Single plant SMEs</b>	<b>SMEs with events</b>	<b>All firms with events</b>
Observations, all years	5,039,811	4,878,532	67,739	87,390
<i>Observations, 2014-15</i>	<i>2,723,875</i>	<i>2,643,043</i>	<i>35,289</i>	<i>44,763</i>
Unique firms	1,399,146	1,364,624	17,905	22,497
Firms with events	22,497	17,905	17,905	22,497
#events	212,426	78,090	78,090	212,426
<i>of which 2014</i>	<i>113,423</i>	<i>42,225</i>	<i>42,225</i>	<i>113,423</i>
<i>of which 2015</i>	<i>99,003</i>	<i>35,865</i>	<i>35,865</i>	<i>99,003</i>
Firms with product launches	8,435	6,640	6,640	8,435
#launches	89,027	24,720	24,720	89,027
<i>of which 2014</i>	<i>47,236</i>	<i>12,527</i>	<i>12,527</i>	<i>47,236</i>
<i>of which 2015</i>	<i>41,791</i>	<i>12,193</i>	<i>12,193</i>	<i>41,791</i>
Firms with patents	2,355	1,795	295	645
#patents	9,064	4,194	1,141	4,474
<i>of which 2014</i>	<i>8,235</i>	<i>3,892</i>	<i>1,055</i>	<i>3,393</i>
<i>of which 2015</i>	<i>829</i>	<i>302</i>	<i>86</i>	<i>481</i>
Firms with trademarks	3,961	3,164	280	589
#TMs	6,407	4,510	491	1,322
<i>of which 2014</i>	<i>6,407</i>	<i>4,510</i>	<i>491</i>	<i>1,322</i>
<i>of which 2015</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>0</i>

Source: BSD / GI / Companies House / Orbis / UKIPO.

**Table 2. Comparing observable characteristics across samples, 2014-2015.**

Variables	A. Single plant SMEs		B. SMEs with events		
	No events	Events	No launches	Launches	(3)-(4)
	(1)	(2)	(3)	(4)	
Patent count	0.001	0.042	0.023	0.062	***
Weighted patent count	0.001	0.041	0.023	0.062	***
EPO/US/PCT patents	0.001	0.023	0.013	0.037	***
Weighted EPO/US/PCT patents	0.001	0.023	0.013	0.037	***
TM count	0.002	0.015	0.011	0.022	***
Rev per worker two-year average	146.55	781.3	900.341	461.6825	***
Annual % rev per worker growth	-0.006	0.017	0.019	0.0132	
High rev per worker growth firm	0.129	0.148	0.151	0.139	**
Revenue two-year average	811	13752	12932.89	13264.97	***
Annual % revenue growth	0.011	0.049	0.05	0.048	
High revenue growth firm	0.15	0.215	0.22	0.208	**
Employment two-year average	5.1	21.2	21.254	21.352	
Annual % employment growth	0.017	0.032	0.03	0.036	
High jobs growth firm	0.014	0.06	0.059	0.059	
Age entered BSD / incorporated	12.4	17.9	17.421	17.943	***
Startup	0.142	0.028	0.028	0.028	
Firm has 1-9 staff	0.892	0.571	0.592	0.568	***
Firm has 10-49 staff	0.086	0.284	0.278	0.289	*
Firm has 50-249 staff	0.013	0.124	0.11	0.125	***
Immediate foreign ownership	0.165	0.328	0.269	0.44	***
Firm is in a group of enterprises	0.003	0.055	0.048	0.047	
Number of companies in the group	0.008	0.187	0.168	0.119	
Firm is a company	0.942	0.903	0.899	0.931	***
Firm is a sole proprietor	0.021	0.004	0.004	0.003	
Firm is a partnership	0.014	0.004	0.005	0.002	***
Firm is a public company	0	0.001	0.001	0.001	
Firm is a non-profit / social enterprise	0.023	0.088	0.103	0.063	***
Services sector	0.909	0.883	0.891	0.858	***
Urban TTWA	0.788	0.838	0.832	0.836	
Greater London	0.228	0.303	0.292	0.291	
<i>Observations</i>	2,643,043		35,289		
<i>Unique firms</i>	1,346,719	17,905	11,265	6,640	

Source: BSD / CH / GI / Orbis / UKIPO. The table shows mean differences between all single-plant SMEs with and without events exposure (Panel A) and for SMEs with and without launches (Panel B). For Panel B, the stars in the last column give the results of rank-sum tests for each variable between columns 3 and 4. \*\*\* denotes 1% significance, \*\* denotes 5% significance.

**Table 3. Coverage by SIC1 sectors for product launch, patents and trademarks, 2014-15.**

SIC1	Section Name	SMEs with events exposure.			
		% of firms with coverage, within sectors			
		Launch	Patent	TM	N
A	Agriculture, hunting and forestry	19.5	.	.	200
B	Fishing	.	.	.	12
C	Mining and quarrying	15.71	.	.	70
D	Manufacturing	29.24	2.66	1.33	3,837
E	Electricity, Gas and Water Supply	.	.	.	92
F	Construction	14.11	.	.	1,680
G	Wholesale and retail trade, etc	33.92	0.74	0.85	4,593
H	Hotels and restaurants	19.34	.	.	543
I	Transport, storage and communications	23.2	.	.	1,319
J	Financial intermediation	12.43	.	0.87	1,601
K	Real estate, renting and business activities	22.3	1.05	0.73	16,814
L	Public administration and defence, etc	.	.	.	.
M	Education	14.62	.	.	643
N	Health and social work	15.35	.	.	951
O	Other community, social and personal services	23.93	.	1.02	2,934
P	Household domestic employment	.	.	.	.
Q	Extra-terrestrial organisations, bodies	.	.	.	.
<i>Average coverage, %</i>		<i>23.45</i>	<i>0.96</i>	<i>0.79</i>	
<i>Observations</i>		<i>8,275</i>	<i>339</i>	<i>280</i>	<i>35,289</i>
<i>Unique firms</i>		<i>6,640</i>	<i>295</i>	<i>280</i>	

Source: BSD / CH / GI / Orbis / UKIPO. For each sector, the table shows the share of firms with coverage reported in the column (event, launch, patent, tm). N is the total number of firms in each sector. Panel A reports all single plant SMEs, Panel B reports single plant SMEs with at least one event in some year. Observations are instances of a firm having at least one event, launch, patent or TM in that year. Trademark data is only available to 2014, so that observations are the same as unique firms. Cells with under 10 observations are suppressed to avoid disclosure.

A. Probability to Launch

L1.15% depreciated PCT / EPO / US patent count	0.011*** (0.002)	0.010*** (0.002)	0.008*** (0.002)	0.007*** (0.001)
L1.15% depreciated TM count	0.009 (0.010)	0.006 (0.010)	0.006 (0.010)	-0.001 (0.008)
Ave pre-2009 patenting			-0.026*** (0.007)	-0.022*** (0.007)
Firm patents pre-2009			0.138*** (0.036)	0.097*** (0.035)
Observations	29528	29528	29528	29189
R <sup>2</sup>	0.0012	0.0059	0.0070	0.0614

**B. Launch counts**

L1.15% depreciated PCT / EPO / US patent count	0.063*** (0.009)	0.056*** (0.010)	0.059*** (0.010)	0.055*** (0.011)
L1.15% depreciated TM count	0.010 (0.030)	-0.009 (0.030)	-0.011 (0.030)	-0.032 (0.029)
Ave pre-2009 patenting			-0.138** (0.053)	-0.120* (0.067)
Firm patents pre-2009			0.371 (0.250)	0.196 (0.268)
Observations	29528	29528	29528	29189
R <sup>2</sup>	0.0008	0.004	0.004	0.031
Controls	N	Y	Y	Y
Pre-sample patenting	N	N	Y	Y
Year, area and industry dummies	N	N	N	Y

Source: BSD / CH / GI / Orbis / UKIPO. The dependent variable is a dummy for whether the firm has a product launch in a given year (Panel A) and the count of a firm's product launches in that year (Panel B). We control for log mean turnover and employment, age, firm size dummies, company legal status and structure dummies, and an urban TTWA dummy. Controls are lagged one year except age. Pre-sample patenting levels effects are detailed in the main text. Standard errors are clustered on 2-digit SIC. \*\*\* denotes a result significant at 1%, \*\* significant at 5%, \* significant at 10%. Constant not shown.

**Table 5. Linking launch dummies and firm revenue productivity, SMEs with events, 2014-2017.**

	Log revenue/worker		Rev/worker growth		High-growth episodes	
	(1)	(2)	(3)	(4)	(5)	(6)
L.new product launch	0.064*** (0.019)		0.000 (0.007)		-0.006 (0.005)	
L2.15% depreciated PCT / EPO/US patent count	0.004 (0.007)	0.005 (0.007)	-0.006* (0.003)	-0.006* (0.003)	0.002 (0.002)	0.002 (0.002)
L2.15% depreciated TM count	0.081*** (0.024)	0.081*** (0.024)	0.002 (0.005)	0.002 (0.005)	0.003 (0.005)	0.003 (0.005)
Ave pre-2009 patenting	0.072 (0.057)	0.070 (0.058)	0.009 (0.022)	0.009 (0.022)	0.029* (0.016)	0.029* (0.016)
Firm patents pre-2009	-0.223* (0.116)	-0.217* (0.116)	0.014 (0.043)	0.014 (0.043)	-0.019 (0.029)	-0.020 (0.029)
Observations	27019	27019	27019	27019	27019	27019
R <sup>2</sup>	0.166	0.165	0.010	0.010	0.023	0.023

Source: BSD / CH / GI / Orbis / UKIPO. The dependent variables are log revenue per worker, annual growth in revenue per worker, and a dummy for whether a firm has a high-growth episode, per the OECD definition. L2 is the stock of patents or trademarks two years before. All models fit controls for log turnover and employment, age, firm size dummies, company legal status and structure dummies and an urban TTWA dummy. Controls are lagged one year except age. All models also fit TTWA, 2-digit industry and year dummies. Pre-sample patenting levels effects detailed in the main text. Standard errors are clustered on firms. \*\*\* denotes a result significant at 1%, \*\* significant at 5%, \* significant at 10%. Constant not shown.

**Table 6. Linking launch counts and firm revenue productivity, SMEs with events, 2014-2017.**

	Log revenue/worker		Rev/worker growth		High-growth episodes	
	(1)	(2)	(3)	(4)	(5)	(6)
L.new product launch count	0.017*** (0.005)		0.000 (0.001)		0.001 (0.001)	
L2.15% depreciated PCT / EPO / US patent count	0.004 (0.007)	0.005 (0.007)	-0.006* (0.003)	-0.006* (0.003)	0.002 (0.002)	0.002 (0.002)
L2.15% depreciated TM count	0.081*** (0.024)	0.081*** (0.024)	0.002 (0.005)	0.002 (0.005)	0.003 (0.005)	0.003 (0.005)
Ave pre-2009 patenting	0.073 (0.058)	0.070 (0.058)	0.009 (0.022)	0.009 (0.022)	0.029* (0.016)	0.029* (0.016)
Firm patents pre-2009	-0.220* (0.116)	-0.217* (0.116)	0.014 (0.043)	0.014 (0.043)	-0.020 (0.029)	-0.020 (0.029)
Observations	27019	27019	27019	27019	27019	27019
R <sup>2</sup>	0.167	0.165	0.010	0.010	0.023	0.023

Source: BSD / CH / GI / Orbis / UKIPO. Notes as in Table 5.



**Table 7. Launch quality, launch importance and firm revenue productivity, SMEs with events, 2014-2017.**

	<b>Log rev/worker</b>	<b>Rev/worker growth</b>	<b>High growth episodes</b>
	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>
A. L.total launch reports, main topic	0.000*** (0.000)	-0.000** (0.000)	-0.000 (0.000)
<i>R</i> <sup>2</sup>	<i>0.166</i>	<i>0.010</i>	<i>0.023</i>
B. L.weighted launch reports, main topic	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
<i>R</i> <sup>2</sup>	<i>0.165</i>	<i>0.010</i>	<i>0.023</i>
C. L.firm has important launch, main topic	0.168*** (0.052)	-0.027 (0.018)	-0.007 (0.013)
<i>R</i> <sup>2</sup>	<i>0.166</i>	<i>0.010</i>	<i>0.023</i>
D. L.count of important launches, main topic	0.218*** (0.069)	-0.014 (0.016)	-0.002 (0.014)
<i>R</i> <sup>2</sup>	<i>0.166</i>	<i>0.010</i>	<i>0.023</i>
Observations	27019	27019	27019

Source: BSD / CH / GI / Orbis / UKIPO. Each panel of the table represents a different regression for equation (2), with dependent variables specified A-D. For each panel, each cell is a different specification showing the coefficient of  $b$  in equation (2), with standard errors in parentheses and  $R^2$  in italics. All models fit controls for log turnover and employment, age, firm size dummies, company legal status and structure dummies and an urban TTWA dummy. Controls are lagged one year except age. All models also fit TTWA, 2-digit industry and year dummies, plus pre-sample patenting levels effects detailed in the main text. Standard errors are clustered on firms. \*\*\* denotes a result significant at 1%, \*\* significant at 5%, \* significant at 10%. Constant not shown.

**INNOVATIVE EVENTS:  
PRODUCT LAUNCHES, INNOVATION AND FIRM PERFORMANCE  
ONLINE APPENDIX**

## Appendix A: variables and build

### A1 / Events data

This paper uses variables that model events in a company's lifecycle (hence 'events'), developed by the data science firm Growth Intelligence (GI). Each 'event' is based on content taken from one of the 3,740 online news sources, including major sources such as Reuters or Yahoo news, as well as industry sources such as IT Briefing and PRWeb. Our raw data consists of 318,899 observations for 30,205 companies in financial years 2014 and 2015 (August 2013 to November 2014 inclusive). The fundamental challenge in using the events dataset for inference is dealing with its unstructured nature. We develop a number of substantive checks and improvements on the raw data.

#### GI data quality checks

We first clean the data to remove all-fields duplicates and the small number of events projected for dates in the future. Next, we remove 'farmed' content by not allowing identical text fragments to appear more than once a day anywhere in the data. Third, we conduct checks for the quality of GI's feature extraction and syntax parsing. Finally, we remodel the raw data for greater realism.

After cleaning, we begin with a simple manual check for 'negative events' – that is, reports describing something that has *not* occurred. On a 1% sample of product/service launch events, we find a negative event error rate of 0.6% (5/823). Next, we conduct more systematic checks on a sample of 'hard cases'. We define 'hard cases' as observations where there are *a priori* reasons to believe GI's ascription of news article text to a given company may be incorrect: specifically, because the text includes either a large tech company (e.g. Google, Facebook) or a large press agency (e.g. Reuters, Bloomberg). These company names often appear in everyday contexts outside activities by that company. For example, 'Google' is now commonly used as a noun or verb; many company websites and online news articles will include social media-related text along the lines of 'follow us on Facebook'; many news reports about other

companies are filed by large press agencies. In this way, the set of hard cases provides a natural upper bound on the error rate in GI's analysis. In the GI data, ascription error could arise from failure to extract text from credible online sources ('content farming error'), or, once text has been extracted, from failures of name entity recognition or selection ('algorithm error'). We define large tech and media companies through Wikipedia reports of global market cap or market share. We draw 5,000 event observations (news articles) ascribed to one of these companies. Analysis using title and text fragment fields suggests an error rate of around 16%, especially what we term 'copyright clutter' (where ascription has been done on article source/copyright text) and what we term 'social media clutter' (ascription based on 'follow us on Facebook' type text). Overall, we judge this to be an acceptable error rate, especially as 1) we are positively selecting cases where error is likely to be present and 2) GI's original ascription is based on the full text from each event text, not just the fields provided to us, so true error rates due to clutter may be lower than this. We can therefore treat the 16% error rate on hard cases as an upper bound on the true error rate across the dataset as a whole. By working with SMEs in our main analysis, we also reduce the likelihood of such ascription errors substantially.<sup>1</sup>

---

<sup>1</sup> We also conduct further, less precise tests on a sub-sample of the big tech firms. Specifically we use the URL field to re-extract the original text, then reverse-engineer GI's feature extraction and syntax parsing routines on pages scraped in 2013 and 2014. We can only perform this exercise on websites that are a) scrapable b) active (return a 200 to GET requests). This reduces our sample size to 1,746, which will be selected on currently active organisations (specifically, well-maintained websites). We then build a web crawler to retrieve the original webpage text, and train a Name Entity Recognition (NER) model to identify company names from the re-extracted data. The model is built from Stanford NER Conditional Random Field Classifiers, which is the current gold standard (with over 80% accuracy) (Jiang et al 2016). We use the CoNLL, MU6, MU7 and ACE 2002 training datasets, which are substantively based on news corpora. For each observation, we proceed as follows. We extract all company names  $C_{ner}$  (we already know the GI company name  $C_{gi}$ ). Let  $C_{candidates}$  be a subset of  $C_{ner}$  occurring in the title / headline of each article, and identified as potential subjects in the text. We assume that the correct subject(s) of the event described will be a) identified as subject at least once in the text (and probably multiple times), and b) be mentioned in the article title. For precision, we drop 413 cases where no company is mentioned in the article title. If GI's ascription is correct,  $C_{gi}$  is in  $C_{candidates}$ , and  $C_{candidates} = 1$ . If GI's ascription is probably correct,  $C_{gi}$  is in  $C_{candidates}$ , and  $C_{candidates} \geq 1$ . If GI's ascription is incorrect,  $C_{gi}$  is not in  $C_{candidates}$ . We find 95.1% incorrect ascription on the 977 remaining observations. We check for out-of-sample error rates by running these routines for the full set of 40,000 observations, with very similar results.

## Event observations vs. real-world events

A further substantive issue is that in its raw form, an event observation may not perfectly correspond to some underlying (real world) event. For example, a major merger is likely to be reported hundreds of times; each of these is currently reported as a distinct event occurrence. We use structural topic modelling (STM) to cluster raw events data in a more realistic fashion; we then exploit the raw event-level counts to make measures of modelled event ‘quality’ or ‘importance’ (see below).

Topic modelling algorithms cluster text fragments that talk about the same topic in different ways, using different text but similar content words (Roberts et al., 2016). In STM, each text fragment is modelled as a document. A topic is defined as a mixture over words where each word is associated to a probability of belonging to a topic. A document is a mixture over topics; therefore each document can be associated to multiple topics. For each text fragment we have a *topical prevalence* and a *topical content*. The prevalence refers to how much a document is associated with a topic, and it is computed using the shared words in the document, while the content refers to the words used within the topic. We use the topical prevalence to group event fragments within the same topic. We use the 90% threshold, so we assume that events belong to the same cluster if they share at least 90% of the content.<sup>2</sup>

Before modelling the data, we stem the fragments (reducing the words to their roots) and remove stopwords (definite and indefinite articles, pronouns, etc.).<sup>3</sup> We then group individual event observations according to three variables – type of event, company and event date<sup>4</sup> – and run the model within each group, using 200 number of topics. If an event is reported by several sources in different formats on the same day, the STM algorithm identifies the repetitions and keeps one of them. Overall, our cleaning steps

---

<sup>2</sup> This threshold can be modified.

<sup>3</sup> For more precise information on the model and on the implementation in R see Roberts et al (2016).

<sup>4</sup> We use the day, but future analysis will be extended using a longer time frame (days or weeks) as the same event may be reported for more than a day. Variations on this might include allowing a weekly bound. However, a bound is hard to identify as we do not know when the actual event took place. Also bounds may differ across event types. Is it better to use the first day that event appeared or is it better to use the day with the highest frequency, or is it better to use the last day the event is reported as we can be more confident that on that day the event has already happened.

substantially reduce the count of GI event observations, from 318,899 for 30,205 firms to 257,056 observations for 30,187 firms. STM accounts for the bulk of this.

### Variable build

We make a launch dummy, taking the value one when a firm has at least one launch in a given year. We also count each firm's launches in that year.

We then use the number of raw reports / mentions for each launch as a proxy measure for that launch's 'significance' or 'quality'. The intuition is similar to patent citations - as more cites indicate a more significant patent, so more mentions suggest a more significant new product or service. We make the following measures for each cleaned launch event: 1) # mentions across all topics; # mentions in main topic; 3) # mentions / # topics. Of these, 2 and 3 are preferred measures – the former looks at mentions in the most relevant topic, and latter penalises poorly identified real-world events.

For each firm\*year cell, we sum these measures 1) - 3). We also build weighted measures, where weights are #launches in a firm/year cell. We can think of this as analogous to weighting patent counts by inventors. We dub all launches with more than one underlying mention, 'important' launches. This gives us two further measures: a dummy for whether or not a firm has an important launch in a given year; and the count of important launches in a given year. In our main panel of SMEs – see below – 98% of launches only have one mention.

### A2 / Panel build

To build panels for analysis, we link GI data to enterprises in the Business Structure Database (BSD). Growth Intelligence (GI) data is pre-matched to Companies House company identifiers, which we can then match to firm identifiers in the BSD. We then match in patents and trademarks data.

## BSD-Companies House-GI matching

GI's data is based on companies active as of August 2012 (financial year 2013). The UK Data Service team matches companies to enterprises using the 2013 BSD cross-section, which comprises 1,818,263 unique entrefs (which denote individual real-world enterprises in the BSD, rather than the legal entities in Companies House). The initial matching rate of companies to enterprises is 61.1% (1,877,600 / 3,074,845 observations matched). Note that due to data protection legislation, we are unable to do this matching ourselves.

We then conduct a number of cleaning and matching sub-routines to optimise the match. Specifically, we drop all observations with no entref, neither in the 2013 BSD nor in the BSD-CH match; drop firms who left the BSD before 2012; drop public sector observations except public sector corporations (e.g. nationalised banks). At the end of these preliminary cleaning steps we have 1,423,558 observations, for 1,416,218 unique enterprises. This is 75.8% of the original matched sample. Some of the remaining enterprises are still matched to more than one legal entity (specifically, 78,379 observations, 1.6% of entrefs, 5.8% of observations). These firms are older, larger and richer than sample as a whole.<sup>5</sup> Because we do not have access to identifying information on the BSD side of the data, we are unable to observe the true corporate structures that match to each BSD enterprise. We therefore develop heuristics to give us a panel with 1:1 enterprise:company matching. The majority of corporate legal structures should reduce to this form, especially the single plant SMEs we focus on. We:

- 1) Keep companies in an enterprise:company group with non-missing year incorporated. Duplicates drop to 2.63% of observations from 5.8% of observations.
- 2) Keep companies in an enterprise:company group with non-missing CH revenue information and this reduces duplicates to 1.59% of observations. We prefer to have observations with revenues

---

<sup>5</sup> Specifically, the firms in these 1-to-many matches are older than average (mean incorporation 1990 vs. 2002); enter the BSD earlier (1984 vs. 2001); have more plants (94 vs. 6); have higher employment (3096 vs. 187) and employees (3095 vs. 187); have higher annual turnover (£1,200,313 vs. £70,983); are more likely to file revenue to Companies House; and report higher 2010-2013 revenue to Companies House (average £12.4bn vs £2.53bn).

rather than none. Given the observable characteristics of these firms, they are more likely to have revenues to report.

- 3) Keep companies in an enterprise:company group with highest-reported CH revenue. This step reduces duplicates to 0.08% of observations, as these are likely to be reporting the revenues of the other companies in the group.
- 4) Shuffle the data and drop any remaining duplicates.<sup>6</sup>

At the end of these steps we have 1,399,146 observations, for the same number of unique enterprises. This is 72.7% of the original matched sample.

We then match this cross-section to BSD panel data. We start with a panel of 16,274,552 BSD firm\*year observations for the years 1997-2017. Having built various lagged variables, we shorten the panel to 2014-2017, since events are only observed in 2014 and 2015. The final BSD-CH-GI dataset is an unbalanced panel of 5,039,811 firm\*year observations for 1,399,146 firms in 2014-2017. In 2014-15 we have 212,426 events and 89,027 launches for 22,497 and 8,435 firms respectively.

This panel is unbalanced because firms enter the BSD at different times, and because firms drop in and out when they no longer fulfil the BSD criteria (their turnover drops below VAT threshold; they have no employees on PAYE or both of these criteria). In some other cases, especially in earlier years, they file zero against employment or turnover. We fill in gaps in years, while preserving firms' different entry points to the panel. We use a simple interpolation rule to fill in time-varying variables for 4.9% of observations. We use linear interpolation and round interpolated values to the nearest whole number. This procedure assumes that (say) plant employment is a linear function of time in years.

---

<sup>6</sup> As a sensitivity check we compare characteristics of the retained observations against the modal values of group of linked companies. We find there's a 0.67\*\*\* in incorporation years; a 0.70\*\*\* correlation in modal founding years; a 0.86\*\*\* correlation in modal GI sector ; a 0.86\* correlation in group-modal GI products; there is a 0.82\*\*\* correlation with the retained and group-modal SIC5 codes. Overall, we conclude that these cleaning rules do not systematically misrepresent underlying corporate structure.



## Linking patents and trademarks data

We use fuzzy matching routines to match patents data and trademarks data to the panel. Raw patents data is taken from Orbis, which sources from the world's major patent offices, and covers 169,417 patents filed by 17,131 firms between 1900-2015. Patents are filed to UK, European (EPO), US, PCT and other offices. Patents are dated by priority year, that is, the first year an application enters any patent office in the world. Using application years places patenting activity as close as possible to the underlying invention. In our raw patents data, 10,360 patents are filed in 2014-2015 by 2973 firms, of which 6440 go to EPO/PCT/US. Orbis has pre-matched patent applicants to UK companies and provides Bureau van Dijk identifiers, which in the majority of cases are identical to, or slightly modified versions of, UK Companies House identifiers. In other cases we match patents to firms using fuzzy matching on company/applicant names and full UK postcodes.<sup>7</sup> The overall match rate for fuzzy-matched patents to BSD/CH/GI data is 80.5% for 2014-2015 (2683/3332 observations). We match for 82.5% of companies (2452/2973 firms) in 2014/15.

Our raw trademarks data covers calendar years 2012-2014, and comprises 8,493 UK trademarks filed by 5189 firms. 7129 trademarks are filed in 2014-2015 by 4395 firms. We use fuzzy matching based on company name and postcode to link trademark applicants and Companies House companies. The overall match rate is 89.1% for 2014-2015 (3918 / 4395 obs). We match for 89.1% of firms (3918 / 4395) in 2014/15.

---

<sup>7</sup> We also match a further 2,404 observations using variations on company name. We do not use these as we cannot be sure that applicants are based in the UK.

## Panel structure and variables

Our main data ‘Stack’ is an unbalanced panel of 5,039,811 firm\*year observations for 1,399,146 firms in 2014-2017. In 2014-15 we have 212,426 events and 89,027 launches for 22,497 and 8,435 firms respectively.

From this we extract two datasets for our main analysis. We build a panel of single plant SMEs which allows us to compare the characteristics of GI product launches with traditional IP measures. Crucially for this exercise, using single plant SMEs allows us to cleanly ascribe a launch to a given firm and location. Removing the largest firms also reduces likely ascription error. Over 95% of firms in the BSD are single plant SMEs: dropping large and multi-plant firms only reduces our data to 5,013,702 and then 4,878,646 observations respectively. We also remove outliers: in each year we remove observations with an event count higher than 1 standard deviation of the mean event count. This drops 84 observations. Our resulting SME panel has 4,878,532 unbalanced firm\*year observations for 1,364,624 single-plant SMEs 2014-17, including 78,090 events for 17,905 SMEs.

In our regression analysis we focus on the subset of single-plant SMEs with ‘events exposure’, defined as having an event (of any kind) in 2014, 2015 or both. This allows us to work with the variation in launch activity across firms and across time, conditional on events exposure. The panel of SMEs with events consists of 67,739 unbalanced firm\*year observations for 17,905 firms 2014-2017, again including 78,090 events for those SMEs.

In all panels, GI variables are defined as previously discussed. Control variables are defined as follows:

- Age – firms enter the BSD when they start paying UK sales tax (levied on companies with an annual revenue of £75,000 or more), have an employee, or both. Firms enter Companies House when they are incorporated – they may be pre-revenue and pre-employees. We set company age to be incorporation date. Where this is missing we use date of BSD entry.

- Employment and employment growth – following Haltiwanger et al (2013) we use a two-year moving average of employment to correct to regression to the mean. We then define employment growth as the change in  $E_t$  and  $E_{t-1}$ , weighted by the average of  $E_t$  and  $E_{t-1}$ . This bounds employment growth to  $\pm 200\%$ , removing outliers.
- Revenue and revenue growth – defined in the same way as employment, above.
- Productivity / revenue per worker – the BSD does not provide information on conventional labour productivity or TFP measures, but does allow us to directly observe revenue productivity. We define revenue productivity and its growth in the same way as revenue and employment.
- Patents – patents data is coded by application year, that is, the year in which a given patent submission was first submitted to any office in the world. We distinguish between patents filed at major patent offices (USPTO, EPO, PCT framework) and the entire pool of patents, which includes the above plus patents filed only with the UK Intellectual Property Office and with other single-country offices. We make unweighted counts and applicant-weighted counts, where raw patents are divided by the number of applicants. Our preferred measure is major office patent stock with a 15% annual rate depreciation (Hall and Harhoff 2012). In robustness checks, variants use a 40% depreciation rate and a simple cumulative measure.
- Trademarks – trademarks data are coded by application year to the UK IPO. We make simple counts and a TM stock measure specified with a 15% depreciation rate. Following Flikkema et al (2019), we also build counts of product, service and specialised trademarks, all depreciated on the same basis. Product trademarks are those in NICE classes 1-34; service trademarks are those in NICE classes 34-45; specialised trademarks are those where NICE classes do not cross the product / service boundary.
- High growth firms and gazelles – we follow the OECD definition of high-growth firms as those with a minimum of ten staff in a given period, where employment or revenue grows by at least 20% in the following three years inclusive. Gazelle firms are high-growth firms less than five years old. We also define high-growth and gazelle firms on the basis of revenue productivity.

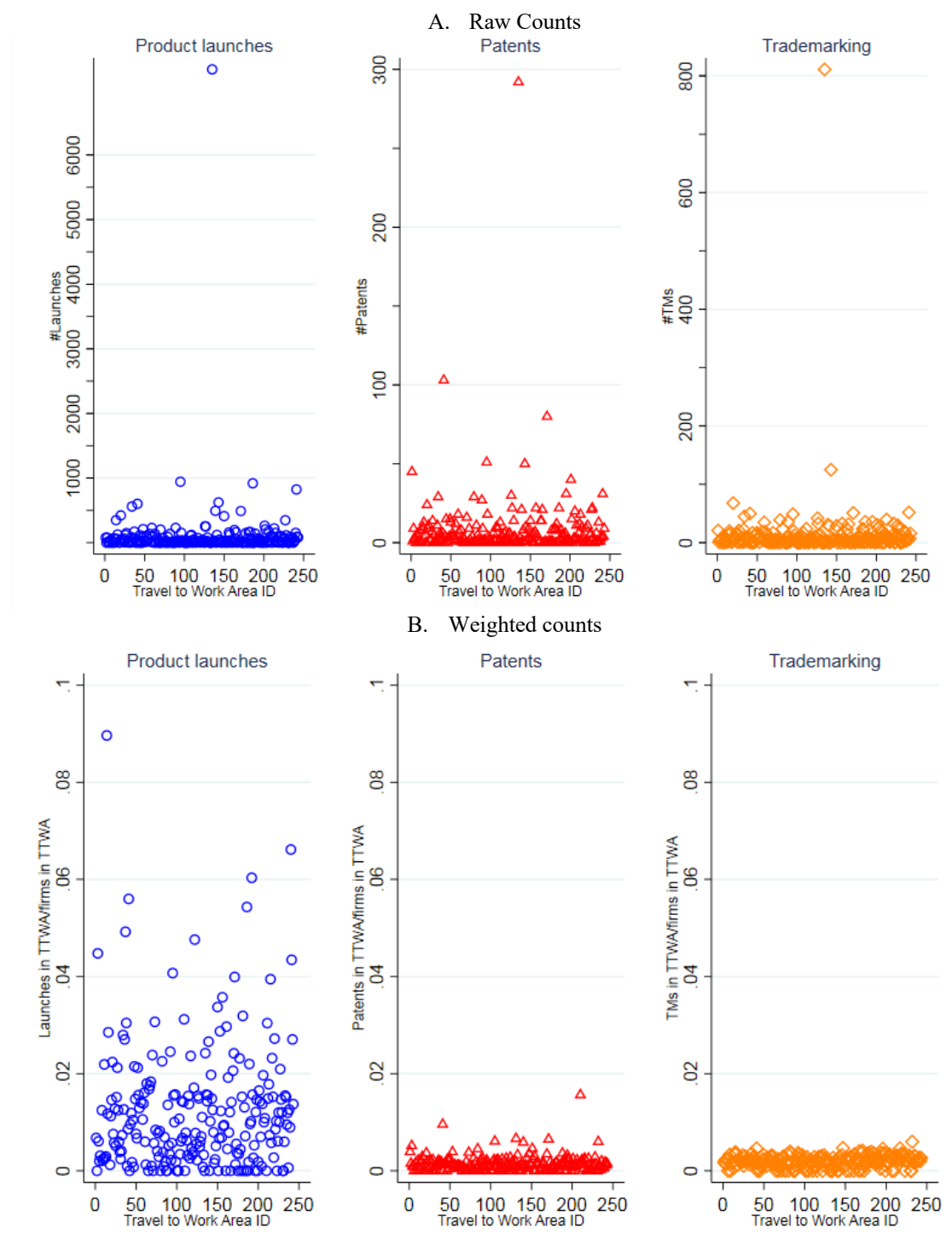
- Number of plants – the BSD allows enterprises to exist with zero plants (for example, when all staff are laid off for a period). For ease of interpretation, we set the minimum plant size to be one.
- Legal status – dummies taking the value 1 if the company is a PLC, sole proprietor or partnership / other.
- Enterprise group – a dummy variable taking the value 1 if firms are part of a larger group of companies.
- Companies per enterprise group – for firms in enterprise groups, a count of the number of companies in the group.
- Industry – we use 2-digit SIC 2003 codes as our basic industry unit (plus 4-digit SICs in robustness checks). 7.3% of companies in the BSD change SIC in our sample. In some cases this is due to change in company activity mix; in other cases ONS reclassifies to correct error, so that reported changes are an upper bound on actual changes.
- Area – we place enterprises in Travel to Work Areas (TTWAs), which are based on commuting patterns and are the best available proxy for local economies; there are 243 of these across the UK. We also use an urban/rural classification of TTWAs taken from Gibbons et al (2011), where ‘urban’ TTWAs contain at least one city of at least 125,000 people. 5.95% of enterprises change TTWA during the panel period.

## **References for Appendix A**

- Flikkema, M., Castaldi, C., de Man, A.-P., Seip, M., (2019) Trademarks' relatedness to product and service innovation: A branding strategy approach. *Research Policy* 48, 1340-1353.
- Gibbons, S.; Overman, H.G and G. Resende. 2011. Real Earnings Disparities in Britain. *SERC Discussion Paper DP0065*. London: SERC.
- Hall, B.H. and Harhoff, D. (2012) Recent Research on the Economics of Patents, *Annual Review of Economics* 4(1): 541-565, 07
- Haltiwanger J, Jarmin RS and Miranda J. (2013) Who Creates Jobs? Small versus Large versus Young. *The Review of Economics and Statistics* 95: 347-361.
- Jiang R, Banchs RE & Li H. (2016). Evaluating and Combining Named Entity Recognition Systems. *ACL* 2016, 21.
- Roberts ME, Stewart BM and Tingley D. (2016) stm: R Package for Structural Topic Models. <http://www.structuraltopicmodel.com>.

**Appendix B. Additional results**

**Figure B1. Counts of launches, patents and trademarks across Travel to Work Areas (TTWAs), 2014-2015.**



Source: GI data, years 2014-2015. In Panel A, we report the total number of product launches, patents and trademarks by Travel to Work Areas, while in Panel B we report the total number of product launches, patents and trademarks in each Travel to Work Area, normalized by the total number of firms in the area, for years 2014-2015, for all the firms included in the main sample.

**Table B1. Summary statistics for SMEs with events, 2014-2017.**

<b>Variable</b>	<b>N</b>	<b>Mean</b>	<b>SD</b>
Firm has event	35,289	0.641	0.480
Total events	35,289	2.213	5.304
New product launch	35,289	0.234	0.424
Total product launches	35,289	0.701	2.922
Total launch reports main topic	35,289	254.5	1,808
Mean launch reports / topics	35,289	207.9	1,497
Firm has important launch type 1	35,289	0.023	0.151
Count of important launches type 1	35,289	0.020	0.146
Firm has important launch (main topic)	35,289	0.023	0.151
Count of important launch (main topic)	35,289	0.020	0.146
Patent count	67,739	0.016	0.524
TM count	67,739	0.007	0.192
TM count, product NICE codes	67,739	0.004	0.170
TM count, service NICE codes	67,739	0.002	0.063
TM count, specialized NICE codes	67,739	0.002	0.047
Manufacturing sector	67,739	0.117	0.322
Services sector	67,739	0.883	0.322
Firm has 1-9 staff	66,457	0.581	0.493
Firm has 10-49 staff	66,457	0.283	0.450
Firm has 50-249 staff	66,457	0.117	0.322
Number of companies per entref	67,739	0.157	1.631
Enterprise has >1 associated company	67,739	0.047	0.213
Foreign ownership	27,723	0.302	0.459
Public company	58,003	0.906	0.292
Non-profit making body	58,003	0.085	0.279
Partnership	58,003	0.003	0.0623
Public corporation	58,003	0.001	0.0316
Sole proprietor	58,003	0.004	0.0652
Age since BSD entry OR incorporation	67,739	18.63	15.18
Firm is 3 years old or less	67,739	0.014	0.120
Urban TTWA	67,739	0.830	0.375
Greater London	67,739	0.290	0.454

Source: BSD / CH / GI / Orbis / UKIPO. The sample is single plant SMEs with events exposure. We observe events and launches in 2014-2015 only. Launch variables are defined as in Appendix A1. Patents are weighted by number of applicants. High growth firms (jobs/revenue/revenue per worker) are built using the OECD definition of high growth firms. See Appendix A2 for details of other variables.

**Table B2. Dispersion of launches for SMEs with a patent or trademark, 2014-2015.**

<b>A. All SMEs.</b>	<b>N (firms)</b>	<b>Number of launches</b>				
		<b>0</b>	<b>1</b>	<b>2</b>	<b>3-4</b>	<b>5 or more</b>
Firms with a patent	1,277	1,205	33	13	15	11
Firms with a trademark	2,502	2,439	30	19	.	.
<b>B. SMEs with events.</b>	<b>N (firms)</b>	<b>Number of launches</b>				
		<b>0</b>	<b>1</b>	<b>2</b>	<b>3-4</b>	<b>5 or more</b>
Firms with a patent	176	104	33	13	15	11
Firms with a trademark	194	131	30	19	.	.

Source: BSD / CH / GI / Orbis / UKIPO. The sample is single-plant SMEs with a single patent or single trademark. Each cell reports the number of firms with 0, ... 5 or more launches. Panel A reports results for all single plant SMEs, Panel B for single plant SMEs with events exposure. Cells with under 10 observations are suppressed to avoid disclosure.



**Table B3. Coverage by SIC1 sectors for product launch, patents and trademarks, 2014-15.**

SIC1	Section Name	Single plant SMEs. % of firms with coverage within sectors				
		Event	Launch	Patent	TM	<i>N</i>
A	Agriculture, hunting and forestry	0.29	0.1	.	0.06	40,735
B	Fishing	.	.	.	.	2,964
C	Mining and quarrying	6.1	1.27	.	.	869
D	Manufacturing	1.25	0.57	0.33	0.25	196,577
E	Electricity, Gas and Water Supply	1.81	.	.	.	3,483
F	Construction	0.34	0.08	0.01	0.02	293,384
G	Wholesale and retail trade, etc	0.73	0.38	0.06	0.23	405,338
H	Hotels and restaurants	0.28	0.09	.	0.05	120,108
I	Transport, storage and communications	0.88	0.32	0.03	0.07	94,905
J	Financial intermediation	2.18	0.41	.	0.09	48,217
K	Real estate, renting and business activities	0.94	0.33	0.08	0.1	1,152,392
L	Public administration and defence, etc	.	.	.	.	15
M	Education	1.16	0.27	.	0.13	34,850
N	Health and social work	0.73	0.18	0.03	0.07	79,130
O	Other community, social and personal services	1.14	0.41	0.03	0.14	169,995
P	Household domestic employment	.	.	.	.	47
Q	Extra-terrestrial organisations, bodies	.	.	.	.	34
<i>Average coverage, %</i>		<i>0.86</i>	<i>0.31</i>	<i>0.07</i>	<i>0.12</i>	
<i>Observations</i>		<i>35,289</i>	<i>8,275</i>	<i>1,941</i>	<i>3,164</i>	<i>2,643,043</i>
<i>Unique firms</i>		<i>17,905</i>	<i>6,640</i>	<i>1,795</i>	<i>3,164</i>	

Source: BSD / CH / GI / Orbis / UKIPO. For each sector, the table shows the share of firms with coverage reported in the column (event, launch, patent, trademark). *N* is the total number of firms in each sector. Observations are instances of a firm having at least one event, launch, patent or TM in that year. Note that trademark data is only available to 2014, so that observations are the same as unique firms. Cells with under 10 observations are suppressed to avoid disclosure.

**Table B4. Area coverage for events, launches, patents and trademarks, 2014-15.**

	<b>Non-urban</b>	<b>Urban</b>	<b>Not London</b>	<b>London</b>
	<b>SMEs with events (%)</b>			
No launches	16.82	83.18	70.8	29.2
Launches	16.41	83.59	70.9	29.1
No patents	16.68	83.32	70.71	29.29
Patents	20.65	79.35	82.6	17.4
No TMs	16.7	83.3	70.8	29.2
TMs	18.93	81.07	73.21	26.79
<i>All</i>	<i>16.72</i>	<i>83.28</i>	<i>70.82</i>	<i>29.18</i>
<i>Observations</i>	<i>5,901</i>	<i>29,388</i>	<i>24,993</i>	<i>10,296</i>

Source: BSD / CH / GI / Orbis / UKIPO. Cells are Travel to Work Areas (TTWAs). Urban TTWAs are defined as those containing a core city of 125,000 people or more, as defined in Appendix A2. Cells with under 10 observations are suppressed to avoid disclosure.

**Table B5. Linking past IP activity to product launches, SMEs with events, 2014-17. Variable lags.**

	L1	L2	L5	L1	L2	L5
	A. Pr(launch)			B. Launch counts		
	(1)	(2)	(3)	(4)	(5)	(6)
15% depreciated PCT / EPO / US patent count	0.007*** (0.001)	0.008*** (0.002)	-0.007 (0.010)	0.007*** (0.001)	0.008*** (0.002)	-0.007 (0.010)
15% depreciated TM count	-0.001 (0.008)	0.027 (0.043)	0.028 (0.043)	-0.001 (0.008)	0.027 (0.043)	0.028 (0.043)
Observations	29189	29189	28256	29189	29189	28256
R <sup>2</sup>	0.0614	0.0614	0.0620	0.0614	0.0614	0.0620

Source: BSD / CH / GI / Orbis / UKIPO. The dependent variable is a dummy for whether the firm has a product launch in a given year (Panel A) and the count of a firm's product launches in that year (Panel B). In each panel Patent stocks are lagged back up to 5 periods, TM stocks up to 2 periods. All models control for log mean turnover and employment, age, firm size dummies, company legal status and structure dummies, and an urban TTWA dummy. Controls are lagged one year except age. All models fit area, industry and time dummies, plus pre-sample patenting levels effects as in Table 4, main paper. Standard errors are clustered on 2-digit SIC. \*\*\* denotes a result significant at 1%, \*\* significant at 5%, \* significant at 10%. Constant not shown.

**Table B6. Sensitivity tests. Launch dummy, 2014-17.**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
L1.15% depreciated PCT / EPO / US patent count	0.007*** (0.001)	0.007*** (0.001)	0.007*** (0.001)	0.007*** (0.001)	0.007*** (0.001)	0.007*** (0.001)	0.007*** (0.001)	0.004* (0.002)		
L1.Cumulative PCT / EPO / US patent count									0.005*** (0.001)	
L1.40% depreciated PCT / EPO / US patent count										0.009*** (0.001)
L1.15% depreciated TM count	-0.001 (0.008)	-0.001 (0.008)	-0.000 (0.008)	-0.001 (0.008)	-0.000 (0.008)	0.000 (0.008)	0.000 (0.008)	-0.000 (0.008)	-0.000 (0.008)	-0.001 (0.008)
Ave pre-2009 patenting	-0.022*** (0.007)	-0.022*** (0.007)	-0.022*** (0.007)	-0.022*** (0.006)	-0.020*** (0.007)	-0.020*** (0.007)	-0.020*** (0.007)	-0.016** (0.006)	-0.022*** (0.006)	-0.019*** (0.007)
Firm patents pre-2009	0.097*** (0.035)	0.097*** (0.035)	0.098*** (0.035)	0.097*** (0.035)	0.092** (0.038)	0.093** (0.037)	0.092** (0.037)	0.090** (0.037)	0.101*** (0.036)	0.095** (0.036)
Observations	29189	29189	29189	29189	24587	24587	24587	29189	29189	29189
R <sup>2</sup>	0.0614	0.0615	0.0615	0.0614	0.0675	0.0672	0.0673	0.0622	0.0614	0.0615

Source: BSD / CH / GI / Orbis / UKIPO. The dependent variable is a dummy for whether the firm has a product launch in a given year. All models fitted as in Table 4, main paper, except where stated below. Column 1 fits our main specification. Columns 2-4 add 1-period lagged dummies for high-growth status in employment, revenue and revenue / per worker. Columns 5-7 repeat this for 5-period (pre-sample) lags. Column 8 fits technology field fixed effects for IPC1 classes. Column 9 fits cumulative patent counts. Column 10 first 40% depreciated patent stocks. \*\*\* denotes a result significant at 1%, \*\* significant at 5%, \* significant at 10%. Constant not shown.

**Table B7. Sensitivity tests. Launch counts, 2014-17.**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
L1.15% depreciated PCT / EPO / US patent count	0.055*** (0.011)	0.055*** (0.011)	0.055*** (0.011)	0.054*** (0.011)	0.053*** (0.011)	0.053*** (0.011)	0.053*** (0.011)	0.046*** (0.017)		
L1.Cumulative PCT / EPO / US patent count									0.037*** (0.011)	
L1.40% depreciated PCT / EPO / US patent count										0.067*** (0.012)
L1.15% depreciated TM count	-0.032 (0.029)	-0.032 (0.029)	-0.032 (0.029)	-0.033 (0.029)	-0.035 (0.028)	-0.035 (0.028)	-0.035 (0.028)	-0.034 (0.028)	-0.030 (0.028)	-0.032 (0.029)
Ave pre-2009 patenting	-0.120* (0.067)	-0.122* (0.068)	-0.120* (0.066)	-0.119* (0.066)	-0.117* (0.064)	-0.116* (0.063)	-0.116* (0.063)	-0.107* (0.056)	-0.118* (0.061)	-0.089 (0.067)
Firm patents pre-2009	0.196 (0.268)	0.198 (0.267)	0.196 (0.267)	0.193 (0.266)	0.174 (0.283)	0.173 (0.282)	0.173 (0.283)	0.213 (0.265)	0.224 (0.264)	0.179 (0.269)
Observations	29189	29189	29189	29189	24587	24587	24587	29189	29189	29189
R <sup>2</sup>	0.0314	0.0315	0.0315	0.0315	0.0348	0.0348	0.0348	0.0323	0.0313	0.0314

Source: BSD / CH / GI / Orbis / UKIPO. The dependent variable is the count of a firm's launches in a given year. All models fitted as in Table 4, main paper, except where stated below. Column 1 fits our main specification. Columns 2-4 add 1-period lagged dummies for high-growth status in employment, revenue and revenue / per worker. Columns 5-7 repeat this for 5-period (pre-sample) lags. Column 8 fits technology field fixed effects for IPC1 classes. Column 9 fits cumulative patent counts. Column 10 first 40% depreciated patent stocks. \*\*\* denotes a result significant at 1%, \*\* significant at 5%, \* significant at 10%. Constant not shown.

**Table B8. Past IP and product launches, trademark breakdowns. SMEs with events.**

<b>A. Launch dummies.</b>	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>	<b>(4)</b>
L1.15% depreciated PCT / EPO / US patent count	0.007*** (0.001)	0.007*** (0.001)	0.007*** (0.001)	0.007*** (0.001)
L1.15% depreciated TM count	-0.001 (0.008)			
L1.TM product NICE 15% depreciated		-0.003 (0.005)		
L1.TM service NICE 15% depreciated			0.061 (0.037)	
L1.TM specialised NICE 15% depreciated				0.077* (0.041)
Ave pre-2009 patenting	-0.022*** (0.007)	-0.022*** (0.006)	-0.022*** (0.006)	-0.021*** (0.007)
Firm patents pre-2009	0.097*** (0.035)	0.097*** (0.036)	0.097*** (0.036)	0.094*** (0.035)
Observations	29189	29189	29189	29189
R <sup>2</sup>	0.0614	0.0614	0.0616	0.0617
<b>B. Launch counts.</b>	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>	<b>(4)</b>
L1.15% depreciated PCT / EPO / US patent count	0.055*** (0.011)	0.054*** (0.011)	0.054*** (0.011)	0.054*** (0.011)
L1.15% depreciated TM count	-0.032 (0.029)			
L1.TM product NICE 15% depreciated		-0.030 (0.031)		
L1.TM service NICE 15% depreciated			-0.046 (0.107)	
L1.TM specialised NICE 15% depreciated				0.091 (0.272)
Ave pre-2009 patenting	-0.120* (0.067)	-0.120* (0.067)	-0.120* (0.067)	-0.119* (0.068)
Firm patents pre-2009	0.196 (0.268)	0.195 (0.268)	0.194 (0.268)	0.191 (0.274)
Observations	29189	29189	29189	29189
R <sup>2</sup>	0.0314	0.0314	0.0314	0.0314

Source: BSD / CH / GI / Orbis / UKIPO. The dependent variable is a dummy for whether the firm has a launch in a given year (Panel A) or the count of launches (Panel B). Controls, levels effects and clustered standard errors are the same as Table 4 in the main paper. \*\*\* denotes a result significant at 1%, \*\* significant at 5%, \* significant at 10%. Constant not shown.

**Table B9. Subsample tests by industry-size-age groups, SMEs with events, 2014-17. Launch dummy.**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
L1.15% depreciated PCT / EPO / US patent count	0.007*** (0.001)	0.006 (0.009)	0.007*** (0.001)	-0.013** (0.005)	0.010*** (0.001)	0.002 (0.003)	0.009*** (0.003)	0.026** (0.012)	0.006*** (0.001)
L1.15% depreciated TM count	-0.001 (0.008)	-0.009 (0.006)	0.004 (0.014)	0.067* (0.034)	0.006 (0.022)	-0.008 (0.009)	-0.006 (0.008)	0.012 (0.039)	-0.002 (0.008)
Ave pre-2009 patenting	-0.022*** (0.007)	-0.010 (0.039)	-0.030*** (0.006)	-0.100** (0.040)	-0.042** (0.018)	-0.033* (0.020)	0.011 (0.025)	-0.115*** (0.027)	-0.018** (0.007)
Firm patents pre-2009	0.097*** (0.035)	0.026 (0.068)	0.162*** (0.038)	0.388 (0.254)	0.136*** (0.036)	0.140*** (0.052)	0.038 (0.086)	0.229*** (0.054)	0.088** (0.040)
Observations	29189	5133	24056	583	16037	8593	3695	6754	22435
R <sup>2</sup>	0.0614	0.1133	0.0613	0.2582	0.0559	0.0990	0.1631	0.0737	0.0686

Source: BSD / CH / GI / Orbis / UKIPO. The dependent variable is a dummy for whether the firm has a product launch in a given year. All models fitted as in Table 4, main paper, except where stated below. Column 1 fits our main specification. Columns 2 and 3 re-run for manufacturing and services subsamples respectively. Columns 4-7 re-run for firms with no-employees (owner-managers / sole traders), micro firms (1-9 staff), small firms (10-24 staff) and medium-size firms (25-249 staff). Columns 8 and 9 re-runs for firms under 10 years old (the youngest quartile) and at least 10 years old respectively. \*\*\* denotes a result significant at 1%, \*\* significant at 5%, \* significant at 10%. Constant not shown.

**Table B10. Subsample tests by industry-size-age groups, SMEs with events, 2014-17. Launch counts.**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
L1.15% depreciated PCT / EPO / US patent count	0.055*** (0.011)	0.005 (0.034)	0.059*** (0.011)	-0.043** (0.019)	0.095*** (0.014)	0.001 (0.021)	0.063 (0.062)	0.147 (0.089)	0.051*** (0.012)
L1.15% depreciated TM count	-0.032 (0.029)	0.006 (0.043)	-0.066 (0.048)	-0.009 (0.064)	-0.100*** (0.029)	0.004 (0.053)	-0.048 (0.055)	0.099 (0.231)	-0.041 (0.030)
Ave pre-2009 patenting	-0.120* (0.067)	0.087 (0.139)	-0.174* (0.088)	-0.346** (0.154)	-0.132 (0.099)	-0.109 (0.106)	-0.010 (0.098)	-0.155 (0.228)	-0.119* (0.061)
Firm patents pre-2009	0.196 (0.268)	-0.347 (0.292)	0.591 (0.429)	1.191 (0.815)	0.027 (0.212)	0.508 (0.535)	-0.194 (0.429)	-0.572 (0.611)	0.209 (0.294)
Observations	29189	5133	24056	583	16037	8593	3695	6754	22435
R <sup>2</sup>	0.0314	0.0924	0.0315	0.2072	0.0371	0.0529	0.1191	0.0798	0.0363

Source: BSD / CH / GI / Orbis / UKIPO. The dependent variable is the count of a firm's launches in a given year. All models fitted as in Table 4, main paper, except where stated below. Column 1 fits our main specification. Columns 2 and 3 re-run for manufacturing and services subsamples respectively. Columns 4-7 re-run for firms with no-employees (owner-managers / sole traders), micro firms (1-9 staff), small firms (10-24 staff) and medium-size firms (25-249 staff). Columns 8 and 9 re-runs for firms under 10 years old (the youngest quartile) and at least 10 years old respectively. \*\*\* denotes a result significant at 1%, \*\* significant at 5%, \* significant at 10%. Constant not shown.



**Table B11. Linking past IP to launch mentions and importance. SMEs with events, 2014-17.**

A. Mentions	raw counts			weighted counts		
	1	2	3	1	2	3
L1.15% depreciated PCT / EPO / US patent count	4.997 (18.917)	1.040 (7.143)	-0.921 (4.663)	-4.387 (3.229)	-2.002 (1.475)	-1.670 (1.254)
L1.15% depreciated TM count	-26.279 (28.478)	-5.499 (13.182)	-1.895 (12.781)	-12.525 (14.685)	-4.921 (6.636)	-3.290 (5.728)
Ave pre-2009 patenting	-67.067 (84.914)	-8.360 (38.489)	5.170 (33.831)	-11.246 (20.046)	-4.205 (11.539)	-3.311 (10.691)
Firm patents pre-2009	382.185 (249.646)	113.728 (96.275)	67.201 (76.561)	-5.971 (77.253)	-0.485 (35.385)	7.225 (31.527)
Observations	29189	29189	29189	29189	29189	29189
R <sup>2</sup>	0.0127	0.0134	0.0134	0.0134	0.0136	0.0135
B. Importance	important launch dummy			# important launches		
	1	2	3	1	2	3
L1.15% depreciated PCT / EPO / US patent count	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
L1.15% depreciated TM count	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.002)	-0.001 (0.002)	-0.001 (0.002)
Ave pre-2009 patenting	-0.002 (0.003)	-0.002 (0.003)	-0.002 (0.003)	-0.000 (0.003)	-0.000 (0.003)	-0.000 (0.003)
Firm patents pre-2009	0.008 (0.008)	0.008 (0.008)	0.008 (0.008)	0.001 (0.005)	0.001 (0.005)	0.001 (0.005)
Observations	29189	29189	29189	29189	29189	29189
R <sup>2</sup>	0.0151	0.0151	0.0151	0.0151	0.0151	0.0151

Source: BSD / CH / GI / Orbis / UKIPO. Panel A gives results for raw mentions and mentions weighted by launches. Panel B gives results for a dummy for important launches (those mentioned more than once), and the count of important launches. For each panel, specifications 1-3 cover mentions by all topics; main topic; and main topic/#topics respectively. All models fitted as in Table 4, main paper. \*\*\* denotes a result significant at 1%, \*\* significant at 5%, \* significant at 10%. Constant not shown.

**Table B12. Linking past IP activity to product launches. Other samples, 2014-17.**

<b>A. All single-plant SMEs</b>		
	<b>Pr(launch)</b>	<b>Launch counts</b>
L1.15% depreciated PCT / EPO / US patent count	0.006*** (0.001)	0.031*** (0.004)
L1.15% depreciated TM count	0.008*** (0.002)	0.020*** (0.007)
Ave pre-2009 patenting	-0.004* (0.002)	-0.022** (0.009)
Firm patents pre-2009	0.033*** (0.006)	0.088** (0.033)
Observations	1762675	1762675
R <sup>2</sup>	0.0112	0.0038
<b>B. All firms with events exposure</b>		
L1.15% depreciated PCT / EPO / US patent count	0.004*** (0.001)	0.328 (0.209)
L1.15% depreciated TM count	0.007* (0.004)	0.374 (0.278)
Ave pre-2009 patenting	-0.006 (0.007)	-0.410 (0.505)
Firm patents pre-2009	0.081*** (0.026)	-1.623 (1.337)
Observations	38165	38165
R <sup>2</sup>	0.066	0.008
Controls	Y	Y
Pre-sample patenting	Y	Y
Year, area and industry dummies	Y	Y

Source: BSD / CH / GI / Orbis / UKIPO. Panel A gives results for all SMEs, Panel B for all firms with events exposure. All models are fitted as in Table 4, main paper. \*\*\* denotes a result significant at 1%, \*\* significant at 5%, \* significant at 10%. Constant not shown.

**Table B13. Falsification test: linking past IP to non-launch events.**

	Other event counts	
	No lag	1-year lag
15% depreciated PCT / EPO / US patent count	-0.010 (0.006)	-0.016* (0.008)
15% depreciated TM count	0.011 (0.053)	0.022 (0.041)
Ave pre-2009 patenting	0.125*** (0.040)	0.137*** (0.037)
Firm patents pre-2009	-0.249** (0.112)	-0.253** (0.112)
Observations	22278	22278
R <sup>2</sup>	0.0390	0.0390

Source: BSD / CH / GI / Orbis / UKIPO. The sample is the set of single-plant SMEs with non-launch events exposure. All models are fitted as in Table 4, main paper. \*\*\* denotes a result significant at 1%, \*\* significant at 5%, \* significant at 10%. Constant not shown.

**Table B14. Robustness checks: log revenue productivity, 2014-17.**

Check	Launch dummy	Launch counts	N
Main specification	0.064*** (0.019) <i>0.1663</i>	0.017*** (0.005) <i>0.1672</i>	27019
Pre-sample high-growth episodes	0.072*** (0.020) <i>0.1840</i>	0.016*** (0.005) <i>0.1848</i>	25313
Patents 1-period lag	0.063*** (0.019) <i>0.1665</i>	0.017*** (0.005) <i>0.1674</i>	27019
40% depreciated patents	0.064*** (0.019) <i>0.1663</i>	0.017*** (0.005) <i>0.1672</i>	27019
Drop SIC switchers	0.052*** (0.020) <i>0.1754</i>	0.016*** (0.005) <i>0.1764</i>	24712
Drop TTWA switchers	0.068*** (0.020) <i>0.1724</i>	0.017*** (0.005) <i>0.1733</i>	24976
SIC4 dummies, not SIC2	0.061*** (0.019) <i>0.2136</i>	0.016*** (0.005) <i>0.2145</i>	27019
Industry-year fixed effects	0.061*** (0.019) <i>0.2147</i>	0.016*** (0.005) <i>0.2155</i>	27019
IPC1-year fixed effects	0.063*** (0.019) <i>0.1664</i>	0.017*** (0.005) <i>0.1673</i>	27019
Industry-area clustering	0.061*** (0.020) <i>0.2136</i>	0.016*** (0.005) <i>0.2145</i>	27019
London dummy, not area FE	0.064*** (0.019) <i>0.1962</i>	0.017*** (0.005) <i>0.1972</i>	27387

Source: BSD / CH / GI / Orbis / UKIPO. The dependent variable is log revenue productivity. Each cell of the table is a different specification showing the coefficient of  $b$  in equation (5), with standard errors in parentheses and  $R^2$  in italics. Controls, fixed effects and standard errors are fitted as in Table 5, main paper, except where varied in the table. \*\*\* denotes a result significant at 1%, \*\* significant at 5%, \* significant at 10%. Constant not shown.

**Table B15. Robustness checks: revenue productivity growth, 2014-17.**

Check	Launch dummy	Launch counts	N
Main specification	0.000 (0.007) <i>0.0108</i>	0.000 (0.001) <i>0.0108</i>	27019
Pre-sample high-growth episodes	-0.002 (0.007) <i>0.0119</i>	0.000 (0.001) <i>0.0119</i>	25313
Patents 1-period lag	0.000 (0.007) <i>0.0107</i>	0.000 (0.001) <i>0.0107</i>	27019
40% depreciated patents	0.000 (0.007) <i>0.0107</i>	0.000 (0.001) <i>0.0107</i>	27019
Drop SIC switchers	-0.002 (0.007) <i>0.0111</i>	0.000 (0.001) <i>0.0111</i>	24712
Drop TTWA switchers	0.001 (0.007) <i>0.0114</i>	0.001 (0.001) <i>0.0115</i>	24976
SIC4 dummies, not SIC2	0.001 (0.007) <i>0.0230</i>	0.000 (0.001) <i>0.0230</i>	27019
Industry-year fixed effects	0.001 (0.007) <i>0.0291</i>	0.000 (0.001) <i>0.0291</i>	27019
IPC1-year fixed effects	0.001 (0.007) <i>0.0112</i>	0.000 (0.001) <i>0.0112</i>	27019
Industry-area clustering	0.001 (0.007) <i>0.0230</i>	0.000 (0.001) <i>0.0230</i>	27019
London dummy, not area FE	0.001 (0.007) <i>0.0164</i>	0.000 (0.001) <i>0.0164</i>	27387

Source: BSD / CH / GI / Orbis / UKIPO. The dependent variable is log revenue productivity growth. Each cell of the table is a different specification showing the coefficient of  $b$  in equation (5), with standard errors in parentheses and  $R^2$  in italics. Controls, fixed effects and standard errors are fitted as in Table 5, main paper, except where varied in the table. \*\*\* denotes a result significant at 1%, \*\* significant at 5%, \* significant at 10%. Constant not shown.

**Table B16. Robustness checks: high revenue productivity growth episodes, 2014-17.**

Check	Launch dummy	Launch counts	N
Main specification	-0.006 (0.005) <i>0.0232</i>	0.001 (0.001) <i>0.0232</i>	27019
Pre-sample high-growth episodes	-0.006 (0.005) <i>0.0240</i>	0.001 (0.001) <i>0.0240</i>	25313
Patents 1-period lag	-0.006 (0.005) <i>0.0233</i>	0.001 (0.001) <i>0.0232</i>	27019
40% depreciated patents	-0.006 (0.005) <i>0.0232</i>	0.001 (0.001) <i>0.0232</i>	27019
Drop SIC switchers	-0.003 (0.005) <i>0.0264</i>	0.001 (0.001) <i>0.0264</i>	24712
Drop TTWA switchers	-0.004 (0.005) <i>0.0255</i>	0.001 (0.001) <i>0.0255</i>	24976
SIC4 dummies, not SIC2	-0.005 (0.005) <i>0.0428</i>	0.001 (0.001) <i>0.0428</i>	27019
Industry-year fixed effects	-0.005 (0.005) <i>0.0479</i>	0.001 (0.001) <i>0.0479</i>	27019
IPC1-year fixed effects	-0.006 (0.005) <i>0.0237</i>	0.001 (0.001) <i>0.0237</i>	27019
Industry-area clustering	-0.005 (0.005) <i>0.0428</i>	0.001 (0.001) <i>0.0428</i>	27019
London dummy, not area FE	-0.005 (0.005) <i>0.0341</i>	0.001 (0.001) <i>0.0341</i>	27387

Source: BSD / CH / GI / Orbis / UKIPO. The dependent variable is a dummy taking the value one if a firm has a high-growth episode, as defined by the OECD. Each cell of the table is a different specification showing the coefficient of  $b$  in equation (5), with standard errors in parentheses and  $R^2$  in italics. Controls, fixed effects and standard errors are fitted as in Table 5, main paper, except where varied in the table. \*\*\* denotes a result significant at 1%, \*\* significant at 5%, \* significant at 10%. Constant not shown.

**Table B17. Linking innovation and log revenue per worker: alternate samples.**

<b>A. All SMEs</b>		
	<b>(1)</b>	<b>(2)</b>
L. new product launch	0.452*** (0.018)	
L. total product launches		0.047*** (0.005)
L2.15% depreciated PCT / EPO / US patent count	-0.003 (0.005)	-0.001 (0.005)
L2.15% depreciated TM count	0.181*** (0.017)	0.183*** (0.017)
Ave pre-2009 patenting	0.053 (0.036)	0.052 (0.036)
Firm patents pre-2009	0.084 (0.060)	0.095 (0.060)
Observations	1580303	1580303
R <sup>2</sup>	0.0849	0.0845
<b>B. All firms with events</b>		
L. new product launch	0.085*** (0.016)	
L. total product launches		0.000 (0.000)
L2.15% depreciated PCT / EPO / US patent count	0.003 (0.003)	0.003 (0.003)
L2.15% depreciated TM count	0.007 (0.013)	0.007 (0.014)
Ave pre-2009 patenting	0.058*** (0.019)	0.057*** (0.019)
Firm patents pre-2009	-0.083 (0.058)	-0.077 (0.058)
Observations	36291	36291
R <sup>2</sup>	0.198	0.197

Source: BSD / CH / GI / Orbis / UKIPO. Panel A gives results for launch dummies and counts across all SMEs; Panel B gives results for all firms with events exposure. Controls, fixed effects and standard errors are fitted as in Table 5, main paper. \*\*\* denotes a result significant at 1%, \*\* significant at 5%, \* significant at 10%. Constant not shown.

**Table B18. Linking launch activity to firm performance. Trademark breakdowns. SMEs with events.**

	A. launch dummies				B. launch counts			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
L.new product launch	0.064*** (0.019)	0.064*** (0.019)	0.063*** (0.019)	0.063*** (0.019)				
L.total product launches					0.017*** (0.005)	0.017*** (0.005)	0.017*** (0.005)	0.017*** (0.005)
L1.15% depreciated PCT / EPO / US patent count	0.004 (0.007)	0.005 (0.007)	0.005 (0.007)	0.005 (0.007)	0.004 (0.007)	0.005 (0.007)	0.005 (0.007)	0.004 (0.007)
L2.15% depreciated TM count	0.081*** (0.024)				0.081*** (0.024)			
L2.TM product NICE 15% depreciated		0.071*** (0.022)				0.072*** (0.022)		
L2.TM service NICE 15% depreciated			0.153* (0.080)				0.156* (0.080)	
L2.TM specialised NICE 15% depreciated				0.262*** (0.101)				0.263*** (0.101)
Ave pre-2009 patenting	0.072 (0.057)	0.070 (0.057)	0.069 (0.057)	0.072 (0.057)	0.073 (0.058)	0.072 (0.057)	0.071 (0.057)	0.073 (0.058)
Firm patents pre-2009	-0.223* (0.116)	-0.221* (0.116)	-0.217* (0.116)	-0.229** (0.116)	-0.220* (0.116)	-0.219* (0.116)	-0.214* (0.116)	-0.226* (0.116)
Observations	27019	27019	27019	27019	27019	27019	27019	27019
R <sup>2</sup>	0.1663	0.1661	0.1660	0.1662	0.1672	0.1671	0.1669	0.1671

Source: BSD / CH / GI / Orbis / UKIPO. The dependent variable is log revenue per works. Panel A gives results for launch dummies, Panel B for launch counts. Controls, levels effects and clustered standard errors are the same as Table 5, main paper. \*\*\* denotes a result significant at 1%, \*\* significant at 5%, \* significant at 1%. Constant not shown.



**Table B19. Effect of innovation (launch dummy) on firm revenue productivity, SMEs with events, 2014-17. Manufacturing versus services breakdowns.**

	Log revenue/worker			Rev/worker growth			High-growth episodes		
	All	MF	Services	All	MF	Services	All	MF	Services
L.new product launch	0.064*** (0.019)	0.018 (0.037)	0.070*** (0.022)	0.000 (0.007)	0.013 (0.018)	-0.003 (0.008)	-0.006 (0.005)	-0.001 (0.012)	-0.007 (0.006)
L2.15% depreciated PCT / EPO / US patent count	0.004 (0.007)	-0.028 (0.046)	0.006 (0.006)	-0.006* (0.003)	-0.030** (0.015)	-0.004 (0.003)	0.002 (0.002)	-0.005 (0.006)	0.003 (0.002)
L2.15% depreciated TM count	0.081*** (0.024)	0.032* (0.017)	0.119*** (0.026)	0.002 (0.005)	0.003 (0.005)	0.002 (0.008)	0.003 (0.005)	-0.007* (0.004)	0.011 (0.010)
Ave pre-2009 patenting	0.072 (0.057)	0.110 (0.155)	0.087 (0.064)	0.009 (0.022)	0.079* (0.045)	0.008 (0.025)	0.029* (0.016)	0.047 (0.031)	0.027 (0.017)
Firm patents pre-2009	-0.223* (0.116)	-0.116 (0.185)	-0.312* (0.161)	0.014 (0.043)	-0.040 (0.057)	-0.001 (0.057)	-0.019 (0.029)	-0.058 (0.040)	0.007 (0.038)
Observations	27019	3517	23502	27794	3526	24268	27794	3526	24268
R <sup>2</sup>	0.1663	0.2415	0.1741	0.0108	0.0543	0.0106	0.0232	0.1009	0.0224

Source: BSD / CH / GI / UKIPO. Panels cover our main dependent variables. Each panel shows results for our full sample (all), manufacturing (MF) and services subsamples. Controls, fixed effects and standard errors are fitted as in Table 5, main paper. \*\*\* denotes a result significant at 1%, \*\* significant at 5%, \* significant at 10%. Constant not shown.

**Table B20. Effect of innovation (launch counts) on firm revenue productivity, SMEs with events, 2014-17. Manufacturing versus services breakdowns.**

	Log revenue/worker			Rev/worker growth			High-growth episodes		
	All	MF	Services	All	MF	Services	All	MF	services
L.total product launches	0.017*** (0.005)	0.010 (0.007)	0.018*** (0.005)	0.000 (0.001)	-0.001 (0.002)	0.000 (0.001)	0.001 (0.001)	0.001 (0.002)	0.001 (0.001)
L2.15% depreciated PCT / EPO / US patent count	0.004 (0.007)	-0.028 (0.046)	0.006 (0.006)	-0.006* (0.003)	-0.030** (0.015)	-0.004 (0.003)	0.002 (0.002)	-0.005 (0.006)	0.003 (0.003)
L2.15% depreciated TM count	0.081*** (0.024)	0.032* (0.017)	0.120*** (0.026)	0.002 (0.005)	0.003 (0.005)	0.002 (0.008)	0.003 (0.005)	-0.007 (0.004)	0.011 (0.010)
Ave pre-2009 patenting	0.073 (0.058)	0.109 (0.155)	0.089 (0.064)	0.009 (0.022)	0.079* (0.045)	0.008 (0.025)	0.029* (0.016)	0.047 (0.031)	0.027 (0.017)
Firm patents pre-2009	-0.220* (0.116)	-0.110 (0.185)	-0.312* (0.161)	0.014 (0.043)	-0.040 (0.058)	-0.002 (0.057)	-0.020 (0.029)	-0.057 (0.040)	0.006 (0.038)
Observations	27019	3517	23502	27794	3526	24268	27794	3526	24268
R <sup>2</sup>	0.1672	0.2419	0.1751	0.0108	0.0542	0.0106	0.0232	0.1010	0.0224

Source: BSD / CH / GI / UKIPO. Panels cover our main dependent variables. Each panel shows results for our full sample (all), manufacturing (MF) and services subsamples. Controls, fixed effects and standard errors are fitted as in Table 5, main paper. \*\*\* denotes a result significant at 1%, \*\* significant at 5%, \* significant at 10%. Constant not shown.

**Table B21. Launches and revenue / worker. Size and age group interactions, 2014-17.**

	A. Launch Dummy			B. Launch Counts		
	(1)	(2)	(3)	(4)	(5)	(6)
Launch*medium	0.063*** (0.019)	0.290*** (0.047)	0.291*** (0.048)	0.017*** (0.005)	0.038*** (0.007)	0.038*** (0.007)
Launch*young			-0.021 (0.045)			0.005 (0.012)
Launch*micro		-0.329*** (0.055)	-0.325*** (0.056)		-0.033*** (0.012)	-0.034*** (0.013)
Launch*small		-0.151*** (0.055)	-0.149*** (0.055)		-0.016* (0.009)	-0.016* (0.009)
Micro firms	-0.084** (0.033)	-0.002 (0.036)	-0.002 (0.036)	-0.079** (0.033)	-0.052 (0.034)	-0.051 (0.034)
Small firms	0.101*** (0.031)	0.141*** (0.034)	0.141*** (0.034)	0.103*** (0.031)	0.119*** (0.032)	0.119*** (0.032)
Young firms	-0.123*** (0.027)	-0.123*** (0.027)	-0.119*** (0.029)	-0.121*** (0.027)	-0.122*** (0.027)	-0.124*** (0.027)
Observations	26442	26442	26442	26442	26442	26442
R <sup>2</sup>	0.1664	0.1678	0.1678	0.1673	0.1681	0.1681

Source: BSD / CH / GI / UKIPO. The dependent variable is log revenue per worker. Panel A gives results for the launch dummy, Panel B for launch counts. For each panel, estimates are the marginal effects of each category for launch and no launch activity. Young firms defined as those in the bottom 25% of the age distribution for the sample. Micro firms are those with 0-9 staff. Small firms are those with 10-24 staff. Controls, fixed effects and standard errors are fitted as in Table 5, main paper. \*\*\* denotes a result significant at 1%, \*\* significant at 5%, \* significant at 10%. Constant not shown.