UNIVERSITY OF BIRMINGHAM | BIRMINGHAM BUSINESS SCHOOL

Birmingham Business School

Discussion Paper Series

Innovative Events

Max Nathan

Anna Rosso

2019-01

***

***

INNOVATIVE EVENTS

Max Nathan[1] and Anna Rosso[2]

[1] University of Birmingham and CEP. m.nathan@bham.ac.uk
[2] University of Milan and Centro Studi Luca d'Agliano. anna.rosso@unimi.it

Abstract

We take a fresh look at firms' innovation-productivity linkages, using novel data capturing new aspects of innovative activity. We combine UK administrative microdata, media and website content to develop experimental metrics – new product/service launches – for a large panel of SMEs. Extensive validation and descriptive exercises show that launches complement patents, trademarks and innovation surveys. We also establish connections between launches and previous innovative activity. We then link IP, launches and productivity, controlling for media exposure and firm heterogeneity. Launch activity is associated with higher SME productivity, especially in the service sector. High-quality launches and medium-size firms help drive this result.

Keywords: innovation, productivity, ICT, data science
JEL: C55, L86, O81

Acknowledgements

Thanks to seminar/conference participants at ZEW Economics of ICT workshop 2017, 4th 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. Aslam Ghumra, Neeraj Baruah and Christoph Stich provided outstanding technical 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.

This version: March 2019

1. Introduction

This paper takes a fresh look at a long-standing unresolved question: the nature of links between firm innovation and productivity. We use novel data that captures new aspects of company-level innovative activity. We fit this in a large panel of UK small and medium-size enterprises (SMEs).

The role of innovation in explaining economic performance has been a focus of economic research for decades. Endogenous growth theory and related empirics help explain country-level innovation-productivity linkages (Lucas, 1988; Romer, 1990). Schumpetarian frameworks explain these via entrepreneurial entry and competition (Schumpeter, 1962; Aghion et al., 2009). But as Hall (2011) and Mohnen and Hall (2013) point out in recent reviews, an empirical link between same-firm innovation and performance is not clearly established.

The empirical literature linking innovation and productivity in firms dates back to Griliches (1979; 1986), whose work links R&D to productivity outcomes. Crepon, Duguit and Mairesse (1998) (hence CDM) develop an influential structural model linking innovation decisions, innovation outputs and firm productivity. Their analysis on French microdata shows positive links between productivity and innovation, measured by either patents or innovation-driven sales. A wave of subsequent empirical studies adopts CDM-type models in a range of countries and cross-country settings.[1] Reviewing this literature, Hall (2011) and Mohnen and Hall (2013) note some major limitations. First, most studies rely on small-*n* cross-sectional data. A few recent studies (Howell, 2015; Fernandes and Paunov, 2015; Baumann and Kritikos, 2016; Audretsch et al, 2018; Morris, 2018) use panel settings, but this is still rare. Second, studies suffer from measurement issues. In practice, most firms do not use formal IP protection, with one recent UK study finding that just 1.6% of firms file patents (Hall et al., 2013). Self-reported innovation surveys also have drawbacks (Hall and Harhoff, 2012; Mairesse and Mohnen, 2010). Growing digitisation makes this problem worse. Third, few studies look at the service sector (Audretsch et al, 2018) or at SMEs, despite the fact that SMEs typically comprise the majority of firms in an economy.

[1] For example Loof et al (2001), Klomp and Van Leeuwen (2001); Griffith et al (2006), Crespi and Zuniga (2012), Hashi and Stojčić (2013) among many others.

In theory, the innovation-productivity link is ambiguous for SMEs. It may further vary along age, size or sector lines. Entrants may also be more likely to engage in Schumpeterian radical/disruptive innovation. But innovation is resource-consuming for smaller, younger companies (Acs and Audretsch, 1987), who may also suffer credit constraints in funding R&D (Audretsch et al, 2018). Smaller, younger firms may also lack absorptive capacity relative to larger incumbents (Cohen and Levinthal, 1990). Regardless of age and size, minimum efficient scale for innovation may be lower in services than manufacturing, given lower capital requirements (Audretsch et al, 2018). The relative strength of these channels will determine a) the level of innovative activities in a given SME and b) the chances of productivity-enhancing discovery.

In a recent meta-analysis, Rosenbusch et al (2011) find generally positive innovation-productivity links for SMEs. These are stronger for younger firms. In some studies, innovation output seems to decrease with firm size (Hashi and Stojčić, 2013). In others, there is no difference between micro manufacturing firms and others (Baumann and Kritikos, 2016). Innovation has been found to reduce the risk of market failure of startups, by increasing firm profits and productivity (Howell, 2015). But this positive effect is smaller for single-product/non-exporting innovating firms as they rely on a single revenue source (Fernandes and Paunov, 2015). These studies all suffer from one or both of the limitations identified above.

This paper makes two linked contributions to these debates. First, we use a novel mix of UK administrative microdata, media and website content to develop novel measures of firm innovation. These complement existing metrics such as patents, trademarks and self-reported innovation surveys. Second, we match these data to a large cross-sector panel of UK SMEs, allowing us to explore innovation-productivity connections with more richness and robustness than previous studies.

To do this, we exploit a cutting-edge dataset developed by the data science firm Growth Intelligence (GI), which uses machine-learning routines on company website and media content to model firms' lifecycle 'events' (for example, new product/service launches, mergers and acquisitions, new hiring, or joint ventures). We clean and refine these variables, using structural topic modelling to better align the reported GI data with underlying real-world activity. We also develop measures of launch quality analogous to patent citations. We then build a short panel dataset matching these 'innovative events' to UK administrative firm level microdata, patents, trademarks and UK Innovation Survey data. This dataset of 4.9m observations is the first firm-level resource of its kind that we are aware of.[2]

In descriptive exercises we show that launches complement existing metrics such as patents, trademarks and surveys, providing high volume (for single-plant SMEs we find 10,349 UK launches in 2014/2015, versus 2,908 patent applications and 3,902 trademarks filed) and even cross-sector coverage. Using a knowledge production function, we go on to show positive links from past IP activity to current launches, with cross-industry variation along expected lines. We then test associations between launch activity and firm productivity, controlling for past IP and for firm heterogeneity. We pay careful attention to the fact that events exposure is not random, and that events are reported, not directly observed. We find that launch activity is associated with higher SME productivity, especially in the service sector. A subset of high-quality launches helps drive the main result.

The paper provides a fresh contribution to well-established literatures on innovation in firms – where patents and surveys remain the dominant metrics – by developing new ways to view firms' innovative activity based on web scraping and natural language processing.[3] This allows us broader coverage than the pioneering studies of media coverage and innovation. Katila and Ahuja (2002) and Fosfuri et al (2008), for example, are restricted to a few hundred firms in single sectors. We also advance on rich data papers such as Hall et al (2013) who combine conventional administrative data, patents, trademarks and innovation surveys for 8,600 UK firms.

Our paper further contributes to a growing empirical literature that uses 'big data' sources and data science techniques.[4] Unusually for this field, we combine commercial big data and large, high quality administrative data sets. The latter provides a clear sampling frame that helps understand implicit sampling issues in the former. This aids inference and interpretation (Einav

[2] Existing datasets of news events such as GDELT and Events Registry are designed for country-level analysis, especially politics / current affairs. Proprietary firm-level datasets such as Mattermark (US) and Beauhurst (UK) are restricted to small numbers of 'high-potential' businesses. Crunchbase is a global wiki-type dataset for the tech sector with good US coverage but limited coverage for other countries, as well as significant quality concerns due to its self-reported nature (Motoyama and Bell-Masterson, 2014).

[3] See Arora et al (2017), Hall and Harhoff (2012), and Mairesse and Mohnen (2010) for recent reviews of patents and surveys in innovation research. Trademarks are increasingly used alongside these metrics; see Block et al (2015). Gentzkow et al (2017) review economic applications of natural language processing.

[4] For reviews see Varian (2014) and Einav and Levin (2014). Guzman and Stern (2015) provide an example of 'nowcasting' and 'placecasting' entrepreneurship, using cross-validation on a very large sample of US data.

and Levin, 2014). The closest comparator is Kelly et al (2018), who use text analysis of historical patents data to develop measures of breakthrough innovations, linking these to industry and macro growth.

2. Data

We use modelled company 'events' to develop new measures of innovative activity. Each 'event' derives from content taken from 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 covering financial years 2014 and 2015. Figure 1 provides two examples for product / service launches, the event type we focus on, showing both raw inputs and modelled outputs. Specifically, GI match raw text to a UK-wide company register, Companies House, using firm names and contextual information. Using both event text and information from company websites, they then use supervised learning to classify the activity described as one of several event types. Nathan and Rosso (2015) give more detail.[5] We focus on 'product/service launch'; other types include 'alliance/joint venture', 'contract awarded', 'management change' and 'merger/acquisition'.

Figure 1 about here

The intuition behind using 'events' is that one can exploit how companies describe themselves or their activities – and how these are reported by others – to understand things that companies do or that happen to them. Ideally, each event observation represents a distinct thing that happened to some firm, or an action that the firm does. In practice, we need to substantially clean the data to get closer to this ideal. Cleaning details are provided in Appendix A1 and summarised here. Given that events derive from media reports, we also need to consider underlying factors driving media exposure. We return to this issue in subsequent sections.

[5] Text fragment for illustration. GI use the full page of content to assign text to a subject company and to classify the activity. Where text describes more than one subject company, as in mergers or joint ventures, GI assign to pair or n-groups. GI also filter to remove results from irrelevant domains (for example, mentions of companies in celebrity magazines, or results from sites that largely or wholly deal in markets outside the UK).

We first remove duplicates and control for 'farmed' content.[6] We next run a quality check on GI's syntax parsing and matching routines, for a sample of 5,000 'hard cases' where ascription errors are most likely to occur. Specifically, we sample observations ascribed to the largest ICT/tech companies by revenue (such as Google, Facebook and Microsoft), or to the largest media companies by market share (Reuters, PA, PR Newswire). These are cases where company names are likely to feature as context as well as subject, so that content might be especially error-prone. Analysis using title and text fragment fields suggests around 16% error rates. Our focus on single-plant SMEs (see below) removes these hard cases from the data, minimising the ascription error rate on the rest of the sample. However, to the extent that mis-ascription 'gives' events to large tech and media firms that actually belong to SMEs, we have a lower bound on the true level of event activity for our firms of interest.

We further improve the realism of the data using topic modelling. In its raw form, event observations may not reflect the importance of the underlying (real world) event. For example, a major product launch is likely to be reported hundreds of times; in the raw data each is reported as a distinct event. Even if such major events are rare, not controlling for them biases the distribution of launches. We use structural topic modelling (STM) to deal with this, clustering text fragments that talk about the same topic in different ways, using different text but similar content words (Roberts et al., 2016). We cluster raw events with similar content – that are likely to refer to the same real-world event – into single reported instances. STM substantially reduces the count of event observations, to 202,912.

We then exploit the number of raw observations per modelled event to make measures of event 'significance' and 'importance' analogous to patent citations. Not surprisingly for a panel of single plant SMEs, only 2% of launches with more than one underlying mention. For each firm-year cell we make counts of mentions, a dummy for whether or not a firm has one of the 2% of 'important' launches, and a count of such launches.[7]

---

[6] Recent structural changes to the media industry – notably, the rise of online platforms – may be reducing levels of quality and scrutiny, for example through 'content farming' and 'churnalism' (Viner, 2016; Gentzkow and Shapiro, 2010; Davies, 2009). The first leads to duplicate reported events; the second alter the distribution of event activity. Both may be particularly prevalent in the ICT sector (Lafrance, 2016). We identify duplicate observations events using all available variables except the source and time. Within each group we just keep one event, so that we are not selecting events on the basis of the quality of the source.

[7] Future analysis could also exploit datasets like Prodcom, as another way to give a sense of product/service quality.

We combine cleaned events data with other sources. Our 'base layer' is the Business Structure Database (BSD) (Office of National Statistics, 2017). This high quality administrative microdata covers 99% of UK enterprises, and gives a clearly-defined sampling frame. Our 'bridging layer' is the Companies House dataset, an open dataset of UK-registered companies that provides unique company identifiers. Finally, we use various matching routines[8] to link US, European and other patents data (from Orbis, application years 1900-2015), UK trademarks data (from the UK Intellectual Property Office, 2012 - 2015) and UK Innovation Survey data (2002-2014). Build and variables are described in detail in Appendix A2.

We restrict the sample to single plant SMEs, allowing us to cleanly ascribe events to single firms and locations. We also remove outliers: specifically in each year we remove SME observations with an event count higher than 1 standard deviation of the mean event count (this drops 84 observations). Overall, these steps reduce the number of event observations to 26,622 from the original 318,899. Figure 2, below, shows the distribution of all events across all firms in the raw data vs. the single plant SMEs in the estimation sample. Overall, our approach is conservative, removing much of the firm-level variation in event activity from the events data.

Figure 2 about here

3. Descriptive analysis

The panel contains 4,878,532 observations for 1.364m single-plant SMEs in the financial years 2014-2017. For 2014-2015, the years when events are observable, we have 2,643,043 observations for 1.36m firms. Firms are mobile across industries (given by 2-digit SIC2003 codes) and across locations (here, Travel to Work Areas). 7.3% of firms are SIC movers, 5.95% of firms are TTWA movers. Table B1 in the Appendix provides summary statistics.[9]

---

[8] Bureau Van Dijk identifiers; or firm name and full postcode. An alternative approach would be the automated method developed by Autor et al, (forthcoming) which exploits internet search results.

[9] We have relatively few firms with UKIS coverage. We run Hotelling tests to check whether this set of firms is systematically different from the rest of the sample. Results, which are all significant at 1%, show that the UKIS subsample differs on a large set of observable characteristics (Hotelling's T2 = 24866.03, F(29,2468834) = 857.440***). In particular, UKIS firms have a substantially higher probability of events exposure (2.91% vs 0.86%) and launch activity (1.03% vs 0.31%). Firms with events exposure who are in the UKIS subsample are also systematically different from other firms with events exposure (Hotelling's T2 = 487.983, F(21868) = 17.407***). We therefore treat the UKIS subsample as different from the rest of the panel.

Table 1 about here

Table 1 gives more detail on events coverage. Around 1% of firms have event coverage (panel A), and over a third of these have product / service launches (Panel B). Firms with event coverage have around 3.4 events, of which around 1.09 are product/service launches. This distribution is broadly stable across the two sample years (Panel C).

Table 2 about here

As anticipated, events coverage is not random, with firms selected into events exposure on observables. Table 2 shows the mean characteristics of firms with and without reported events, and with/without launches, for the short panel 2014-2015. We can see there are large mean differences between firms with and without events exposure; rank-sum tests confirm significant mean differences on all observables. Specifically, firms with reported events are on average older, bigger in terms of employment and revenue, with higher employment, revenue and revenue productivity growth, and are more IP-active (with more patents and trademarks filed). Firms with events also report more product and process innovation than firms without events. However, they are notably less likely to have 'high-growth' episodes on OECD definitions of revenue/worker, revenue or employment growth.

By contrast, for firms with events, differences between those with and without launch activity are rather smaller.[10] Companies with launches are more likely to patent and have trademarks and report innovations than those without launches. They have significantly lower productivity, are less likely to have high revenue/worker growth episodes, are more likely to be foreign-owned and part of a larger business group, and more likely to be a listed company rather than a partnership. However, they balance on age, share of start-ups, number of employees, share in an urban TTWA, and on employment and revenue in levels, growth and high-growth status.

Table 3 about here

Table 3 compares the coverage of patents, trademarks, reported launches and reported innovation at aggregated industry level in our data. We use the short panel 2014-2015 in order to

[10] 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.

compare coverage only in years where we observe events and launches. Panel A shows coverage of events, launches, patents, trademarks across SIC1 bins. Columns show shares of firms with at least one event, launch, patent, TM, or reported product/process innovation. Panel B repeats the analysis for firms with events exposure. The results suggest that launches are a useful complement to conventional innovation metrics. Event and launch coverage is more evenly distributed across sectors than patenting, TMs and reported innovations. This holds whether or not we restrict to firms with events exposure. As expected, patenting is most concentrated in manufacturing, but is also present in parts of the services sector, notably business services (including software and other 'knowledge-intensive' activities (Castellacci, 2008). Given their broader functionality, trademarks are more evenly distributed, with most activity in manufacturing, wholesale /retail/repair and social/personal services. Launch activity is also higher-frequency. For the full sample, we observe 8,275 launches, 1,941 patents and 3,164 trademarks. When we restrict to firms with launch activity, we see order of magnitude differences. We report the same Table in the Appendix (Table B2) with the same information on the distribution of events across sectors for matched UKIS companies. Among the UKIS matched sample, on average 27% of companies report some innovation, with this share being higher for the manufacturing and real estate sector. While in the very small sample of companies with events, on average 40% of companies report some innovation.

Table 4 about here

Table 4 shows coverage of events, launches, patents, trademarks and reported innovations across urban travel to work areas (TTWAs) and London in the short panel. TTWAs approximate spatial economies; urban TTWAs are defined as those containing an urban core of 125,000 people or more. Over 78% of observations are in an urban TTWA; just under 23% are in London (Panel A). For firms with events, urban and London shares are even higher (Panel B). In both samples, launches and other innovation metrics for these single-plant SMEs are highly urbanised, in line with the wider firm literature (Audretsch and Feldman, 1996). Rank-sum tests suggest these differences are significant, except for urban / non-urban patenting; reported innovation samples are too small to give significant differences.

Figure 3 about here

Figures 3 and 4 extend the geographical analysis. Figure 3 is a simple scatterplot of launches, patents and trademarks across TTWAs. In raw counts, coverage across spatial economies appears even, although launch counts are substantially higher. London is a big outlier in counts, even for single plant SMEs.

Figure 4 about here

To correct for this, Figure 4 plots TTWA counts weighted by the number of firms in each TTWA. Very interestingly, we can see that when local economic conditions are taken into account, launches have a far more even distribution across space than either patents or trademarks.

4. Framework

Innovation in firms is a multi-stage process. The knowledge production function paradigm pioneered by Griliches (1979) links upstream inputs (internal R&D, external knowledge), intermediate outputs (inventions) and firm 'performance' (productivity, stock market value and so on). Performance is partly driven by inventions successfully deployed internally, and/or commercialised (innovations). In practice, knowledge may emerge from interactions with customers, suppliers and peers (Chesborough, 2003; Von Hippel, 2005) as well as a firm's asset base, and is shaped by firms' absorptive capacity and evolution paths (Cohen and Levinthal, 1990; Blundell et al., 1995; Teece et al., 1997).

Intermediates are product/process innovations protected either formally (via patents, trademarking or designs) or informally (via secrecy, confidentiality agreements or lead times) (Hall et al., 2014). Surveys suggest that firms typically use a range of IP protection tools, both formal and informal, if they do so at all (*ibid*). In particular, trademarks are an important formal complement to patents, both for IP protection (via legal protection for brands and marketing assets), but also to aid product differentiation and as a way to signal innovativeness to potential investors (Block et al., 2015; Helmers and Rogers, 2010; Fosfuri et al., 2008). Among informal tools, lead time seems the most widely used (Hall et al., 2014).

We argue that product/service launch events – like those in Figure 1 – are a measure of firms' 'downstream' innovative activity: specifically, they represent inventions that have been commercialised into new-to-the-firm products and services. Crucially, while launches do not capture innovations protected by secrecy, they can (in theory) pick up *any* other public innovation however protected. This suggests 1) some positive link between prior formal IP filings and subsequent launches, and that 2) in turn, differences in firm-level launch activity may feed through to subsequent firm-level productivity differences.

As discussed in Section 1, these linkages are *a priori* ambiguous for SMEs and require empirical testing. In the spirit of the CDM approach, we proceed in two stages. We first estimate a knowledge production function that links launches to past patenting, trademarking and self-reported product/process innovation. We run our estimations for the full sample, using placebo tests to capture the role of underlying events exposure once observables are controlled for. Next, we explore the link between launch activity, past IP and firm productivity.

Given the media-reported nature of launches, identifying the link between launch activity and productivity is challenging. In theory we could estimate, for firm *i* in year *t* in TTWA *a* and sector *s*:

$$Y_{itas} = F(L_{it-n}, \mathbf{IP}_{\mathrm{it-n}}, \mathbf{X}_{it-n}, T_t, A_a, S_s, u_{itas}) \qquad (1)$$

Where Y is productivity, L is some measure of launch activity, **IP** is a vector of past patenting and trademarking, **X** is a control vector, and T, A and S are fixed effects. As confirmed in Section 3, events exposure varies on observable characteristics. However, events exposure may condition both sides of (1). That is, events exposure might be driven by unobservables, and these might be time-varying. More mechanically, launches are only observed conditional on an event being observed. (Our short panel also makes it challenging to fit firm fixed effects I_i. Blundell et al (1995) propose firm-specific 'levels effects' based on historic patenting activity, an approach we follow here.)

We can think of a firm's selection into events / media exposure as a decision shaped by demand side and supply side factors. On the demand side, the strategic value of media exposure will vary across firms, given an individual firm's strategic choices; management capacity (Cohen and Levinthal, 1990); resources and other observable characteristics (such as age, size, legal and corporate structure) (Teece et al., 1997); position in a value chain (B2B or B2C); industry characteristics and trends; and larger forces -such as national / international policy regimes, trade frictions and changes in these (Cockburn et al., 2016). On the supply side, the availability of coverage for a given new product and service will vary across industry (some sectors are more likely to have newsworthy content than others), location (physical proximity to media producers), time-varying trends in the media (Davies, 2009; Viner, 2016), as well as the individual firm characteristics listed previously.

Many of these factors are directly observed in our data or can be handled via fixed effects. If selection into events varies only on observables, and controllable wider conditions, then we can cleanly identify the link between L and Y in (1). However, if event exposure is also driven by time-varying unobservables, then we cannot fully separate out the effect of launch activity from underlying events exposure, then in this case our best option is to estimate (1) for all firms *and* for those with events, then compare coefficients of L. Our preferred estimates will be for the events subset, where we effectively estimate the role of launches conditional on events exposure.

5. Launches as an innovation measure

We begin by exploring launches' place in knowledge production. If launches are in fact measures of 'downstream' innovative activity, we should expect a positive significant link from firms' past IP activity (reflecting 'upstream' patenting, trademarking and before that, R&D) to launch activity, but not vice versa. Given their respective functions, we might also expect larger/stronger links from patenting than from trademarking. Raw industry-level correlations seem to bear this out (Figure 5).

Figure 5 about here

More formally, we can represent these relationships as a modified knowledge production function, in which launch activity L for firm *i* in year *t* is a function of past IP in period *t-n*, firm characteristics **X**, and wider local (*a*) and sectoral (*s*) conditions:

$$\mathrm{L}_{itas} = a + b\mathrm{PATS}_{it-n} + c\mathrm{TM}_{it-n} + \boldsymbol{d}\mathbf{PAST\_P}_{i} + \boldsymbol{e}\mathbf{X}_{it} + \mathrm{T}_{t} + \mathrm{A}_{a} + \mathrm{S}_{s} + e_{itas} \qquad (2)$$

We define L as either a product launch dummy taking the value 0 or 1, or the count of launches l, where $l = 0, ... l$. PATS are patent stocks with a standard 15% depreciation rate (Hall and Harhoff, 2012), which we vary in sensitivity tests. TM stocks are constructed the same way. We define 'recent' patenting as occurring in any given five year period, so that n takes the value 0, 1 ... 5 for patents, for EPO/US/PCT filings in any given year back to 2009.[11] For trademarks, n takes the value 0, 1 or 2 given available data. Given the short panel, firm fixed effects are too demanding to run. Instead, following Blundell et al (1995) we use individual firms' historic patent stocks as proxies for firm-level experience, absorptive capacity and other unobservables.[12] We define 'historic' patenting as taking place pre-2009. Specifically, **PAST_P** includes a dummy taking the value 1 if the firm patents before this date, and an average of pre-2009 patenting activity which takes values $p = 0, ... p$. **X** includes predictors of firm growth suggested by the descriptive analysis, such as lagged log turnover, age, startup dummy, firm size group dummies, company legal status and structure dummies, plus an urban TTWA dummy. We include TTWA and 2-digit industry fixed effects as well as a time dummy. Standard errors are clustered on two-digit SICs. We estimate (1) for the full sample and for the UKIS subsample, where we additionally fit a dummy for lagged reported product/process innovations. We estimate in OLS because nonlinear estimates typically 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.

There are three main caveats to this exploratory exercise. First, the time decay function from IP to launches is unclear. Depreciation of patents and TM stocks may matter less for recent IP activity. Conversely, if historic IP indicates some generalised absorptive capacity, this link may be more important than recent IP. Second, measurement error on both sides of (1) will affect our estimates. The majority of UK innovations are not protected with formal IP, even for R&D-intensive companies (Hall et al., 2013). Many new products / services involve multiple patents (e.g. the iPhone reportedly has over 100).[13] Reported launches are likely to be a lower bound on true levels of launch activity. We also know (from Section 2) that GI's modelling has ascription

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

[12] KPF approaches normally include R&D, as in the CDM model. Our data makes this challenging. We have matched UKIS data to our panel, but the sample is small and highly selected. Commercial sources such as Orbis have limited direct coverage (7,600 'industrial companies' in the UK with R&D expenditure in annual accounts); UK SMEs file minimal returns to Companies House, so that standard proxies are hard to reconstruct. For this reason we rely on past patenting and trademarks to provide a (lower bound) approximation of underlying R&D.

[13] E.g: https://www.quora.com/How-many-patents-does-the-iPhone-use?, accessed 23 February 2017.

error that tends to allocate events away from single plant SMEs, the firms in our sample. We are testing aggregate links for each firm using many years of patents and TMs, but only two financial years' worth of reported launches. While the measurement error in patents and trademarks may downward bias the estimates, we can consider the error in the product launch as good as random, conditional on the set of variables we have in the regression. Third, at this stage we are not explicitly conditioning on events exposure. We fit (1) for the full sample and run robustness checks to understand the role of events exposure in driving our results.

Table 5 about here

In Table 5 we report results from the regression shown above on the probability to have a product launch (Panel A) and the count of product launches (Panel B). In the first column we only control for patent activity and trademarks, lagged one period and depreciated at 15%; in column two we add controls; column 3 adds historic IP and column 4 year, area and industry dummies.[14] For the linear probability model (Panel A) we can see that coefficients past IP are significant, with one extra patent raising launch probability by 0.5% points the following year; for trademarks, the link is slightly larger, at 0.7% points. Historic patenting activity is a significant predictor of current launch activity, with coefficients substantially larger than recent IP: firms with historic IP are 3.6 percentage points more likely to have a launch in any sample year. The number of historic patents is a negative predictor. As above, this suggests historic IP is a good proxy for absorptive capacity, with old patents suffering from significant depreciation. We see similar patterns for the launch counts model (Panel B): 10 additional patents in a given year is linked to just over 0.2 extra launch events the following year, with 10 extra trademarks yielding 0.17 extra launches.

In column 5, we also show results using only the UKIS sample. For the LPM, both patent and trademark links disappear, while we find a strong and significant result of lagged reported innovation on the probability of launching (1.4% points). As discussed in Section 3, UKIS SMEs are observably different from the rest of our sample, including on launch exposure. Their average probability to have a launch is 1.03% (compared to 0.31% in the full sample), so just being a UKIS active firm increases the baseline probability of launching by 135% compared to non-UKIS firms. While this tells us something about the extensive margin for launches, it does

[14] Sample size changes across results in different columns. In order to make sure the small differences in the results were not 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.

not seem to be informative on the intensive margin, with no significant effect on the count of launches.

These results survive an extensive set of sensitivity checks. In the Appendix, Table B3 varies the lags for patents and trademarks, for all firms and the UKIS subsample. In further checks we add controls for high-growth firms; add technology field fixed effects; re-specify patents using cumulative patent counts; use 40% depreciation rates, following Li and Hall (2016); and break out results for manufacturing and services subsamples. Our main results are confirmed for the full sample (Tables B4-B5) and the UKIS subsample (Tables B6-B7), with expected difference in the importance of patents vs trademarks between manufacturing and services firms.

In Table B8, we re-estimate (1) on launch mentions, the ‘important launch’ dummy and counts of important launches. Table B9 repeats the analysis for the UKIS subsample. We find no significant links from IP to launch significance or importance, except for past trademarks in the UKIS sample. We speculate that for this set of outcomes, some firm-level unobservables likely drive both events exposure and launch activity the results. We address this issue further in the next section.

Finally, we run two falsification tests. To check the direction of the IP-launch relationship, we first run a cross-sectional placebo test where we regress current (2015) launches on past (2014) patents, then vice versa. If products are the 'downstream' result of 'upstream' inventive activity, the coefficient of past patents on present launches will be positive significant, and the coefficient of present launches on past IP will be zero, insignificant or both. Table B10 gives the results for OLS, and for models with area and industry dummies. The launch-to-patent relationship is close to zero, and orders of magnitude smaller than the patent-to-launch relationship.

Next, we check for the role of underlying events exposure in the launch-IP relationship. Ideally, past IP predicts launches better than it predicts other kinds of events, such as mergers or joint ventures, where the role of innovative activity is likely to be less relevant. Conversely, events exposure may affects both launch activity and IP even after conditioning on observables. Table B11 shows the results of a test where we regress events and event counts on past IP, for the set of firms without product launches (i.e. we test links from IP to all other event types). The effect of patenting on other events and event counts is smaller and weaker than for launches, especially for event counts. The effect of past trademarking are the same or larger than in our main results, likely reflecting their multi-purpose role (Block et al 2015). For the UKIS sample, only reported innovations robustly predict event activity, although past patenting is marginally significant. Overall, these results are reassuring, although we cannot rule out the role of media (events) exposure –related unobservables in (2). In the next section we address this identification challenge directly.

6. Linking IP, launches and firm performance

Having established connections between past inventive and IP activity and present launches, we now turn to innovation-productivity links. Specifically, we look at how far launch activity explains firm productivity on top of patenting, trademarks and self-reported innovation. We specify a growth model in which productivity is a function of firms' recent IP, recent launches, controls, and wider time, area and sector conditions:

$$\mathrm{Y}_{itas} = a + bL_{it-1} + c\mathrm{PATS}_{it-2} + d\mathrm{TM}_{it-2} + \boldsymbol{e}\mathbf{PAST\_P}_{i} + \boldsymbol{e}\mathbf{X}_{it-n} + \mathrm{T}_{t} + \mathrm{A}_{a} + \mathrm{S}_{s} + e_{itas} \quad (3)$$

As in existing studies, we specify productivity as revenue per worker (Mohnen and Hall, 2013), where Y is either log revenue/worker, % revenue worker growth / year, or a dummy for whether a firm has at least one 'high-growth' episode. High-growth episodes are specified as per OECD definitions. Lagged launches, patents and trademarks are specified as in (2): we lag the latter back two periods to allow 'upstream' IP to influence 'downstream' launches, as established previously. **X** is specified as in (2) except that instead of firm size dummies and lagged turnover, we fit five-period lags of revenue and employment to control for firms' pre-sample characteristics. As before, given the short panel we use a historic patenting dummy and mean pre-2009 patenting (**PAST_P**) as a proxy for firm-level heterogeneity. Time, area and SIC2 sector dummies are fitted as before. Standard errors are clustered at SIC2.

As discussed in Section 4, selection into events / media exposure is a potentially important omitted variable in (3). We go some way towards controlling for the firm-level and wider factors that shape media exposure. However, as Section 5 makes clear, unobservable characteristics driving events exposure are likely to shape the launch-IP relationship, and these factors plausibly influence productivity too. We therefore estimate (3) for the full sample, then for firms with events exposure, comparing coefficients of *b*, our parameter of interest.

Tables 6 and 7 give results for the full sample, for the launch dummy and launch counts respectively. For each outcome we fit the model with launches (columns 1, 3, and 5) and without (columns 2, 4 and 6).

Table 6 about here

Table 6 shows that product launches have a positive and very significant effect on log revenue productivity, at least in levels. With underlying events exposure uncontrolled for, firms with product launches have 45% higher productivity than those without, and this is significant at 1% (column 1). That is, an increase of 1 standard deviation in the probability to launch (0.055) increases productivity by 2.5%. However, we find no significant links to productivity growth, or to high productivity growth activity. Notably, for single plant SMEs, recent past patenting has no robust positive link to productivity, although it is marginally significant in predicting high-growth episodes. In contrast, trademark counts are positively correlated in all three specifications. When omitting launches, model fit is slightly lower, while other coefficients stay the same. These results suggest that product launches provide additional insight into firms' productivity drivers than using IP measures alone.

Table 7 about here

We find similar results for the count of launches, with a positive significant effect on log productivity. Specifically, each additional launch raises revenue productivity by 4.7%, albeit with underlying media exposure uncontrolled for. While there is no link to productivity growth, we find a small (0.1%), marginally significant link to high-growth episodes. Results for patent and trademark counts, as well as historic IP, are similar to the dummy model. A simple quantification exercise shows us that 1 standard deviation increase in the number of launches in this sample (0.34, as reported in Table 1), increases the productivity by 1.6%.

Tables 8 and 9 repeat the analysis for the sub-sample of firms with events. Here, we interpret coefficients of b as expressing the association between launches and productivity conditional on media exposure.

Table 8 about here

As expected, selection into the events sample drives much of the large associations found earlier. In Table 8, single plant SMEs with events and launch exposure now have 6.4% higher productivity than other firms with events (column 1). This lower results also reflects the fact that the variable product launch has a higher standard deviation with respect to the sample of all firms used in Table 6: having a probability of 1 standard deviation (0.42) higher is associated with a 2.7% higher productivity, therefore slightly higher with respect to the gap observed in all other firms (Table 6, column 1). While recent patenting has no link to productivity for these SMEs, as before, recent trademarking has a positive significant association, with each additional trademark associated to 8.1% higher productivity (column 1). Model fit drops slightly once launches are removed, suggesting they have some additional explanatory power. As before, we find no link between launch activity and productivity growth or high-growth episodes; patent and trademark links are also weak or non-significant here (columns 3 and 5).

Table 9 about here

Table 9 fits the count of launches. As expected, controlling for underlying media exposure substantially reduces the launch-productivity link. Each additional launch is now linked to a 1.7% increase in productivity (column 1), down from 4.7% for the full sample, but remains significant at 1%. Also in this case, the quantification allows us to compare results across samples: having a 1 standard deviation more launches (2.3) is linked to a 3.9% higher productivity, 2.4 times higher than results shown in Table 7. As in Table 8, we find no link from recent patenting, but a clear positive link from recent trademarks which is larger than that of launches. Removing launches from the model (column 2) reduces model fit, as before. As in the full sample, we find that launch counts do not predict productivity growth or high-growth episodes.

We run a battery of robustness checks on these findings. Tables B12-B14 in the Appendix give results for log productivity, productivity growth and high-growth episodes respectively. Each table shows coefficients of the launch dummy and launch counts for tests that: include pre-2014 controls for firms' high-growth episodes; change the lag of patents and trademarks; depreciate patents at 40%, not 15%; drop firms that change industry or move across areas; fit 4-digit industry dummies rather than 2-digit; fit industry-year fixed effects; fit IPC1 technology field-year fixed effects; fit two-way clustered standard errors on industry and area; and include a simple dummy for London location rather than area fixed effects. Results are robust to all of these alternative specifications.

6. Extensions

We briefly show three extensions to these main results. First, we split the events sub-sample to look at launch-productivity links in manufacturing and services industries. Tables B15 and B16 in the Appendix give results, for the LPM and counts models respectively. For both the LPM and the counts model, overall positive launch-productivity links are driven by firms in the service sector. Services firms also drive the trademarking result. For manufacturing firms, recent patenting is linked to lower productivity growth, but historic patenting correlates to higher productivity growth. Overall, the results are consistent with Audretsch et al (2018) who suggest that barriers to (reported) innovation are lower for services firms than manufacturers.

Second, we look at links from launch quality to productivity. We use the number of media reports per event as a proxy for quality, as set out in Section 2. We re-estimate (3) using four alternative quality measures: a simple count of the number of reports across launches, per firm per year; firm-year counts weighted by the number of launches; a dummy for whether a firm has an ‘important’ launch with a high number of mentions; and the count of important launches per firm per year. Table 10 gives 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 from report counts to productivity in levels, and very small negative links to productivity growth. We find zero links for weighted report counts. We find large, significant effects from having an important launch to productivity, and from the count of important launches. Specifically, SMEs with at least one important launch have about 17% higher productivity than other SMEs with media exposure; each additional important launch adds almost 22% to a firm’s revenue / worker. This suggests that our main results, linking launch activity to SME productivity, are significantly driven by a small set of high-profile, important product and services launches.

Table 10 about here

Finally, we look at interactions between firms' productivity, age, size, lunch activity. US and UK studies suggest that 'small, young' firms disproportionately account for employment growth (Haltiwanger et al 2013, NESTA 2010). Testing whether innovative small, young firms drive productivity is a policy-relevant next step. In the events subsample, we identify 'young' firms as the most recent quartile. These are firms up to nine years old, close to the group identified by Haltiwanger et al (2013). We group firms by size using OECD definitions of micro, small and medium businesses.

Interacting patenting and trademark activity with age and size groups yields non-significant results. Tables B17 and B18 re-run (3) for productivity in levels, interacting IP measures with a dummy for young firms; Tables B19 and B20 repeat the analysis for size group dummies, with medium-size firms the reference category. Table 11 shows this analysis for launch activity. In each case, we first fit the re-specification of (3) with age and size dummies (columns 1 and 4), then size group interactions (columns 2 and 5), then age interactions (columns 3 and 6). Overall, the results are mainly driven by medium-size firms while there is no effect of age. Interestingly, there is a negative effect of being a micro firm on the extensive margin (they are less likely to launch, column 2), yet on the intensive margin the coefficient of the effect of innovation on productivity is not statistically different from that of medium size firms (column 5). If it is a small firm launching the product the effect is positive but half of that found for medium firms. The result on the counts is similar: positive but smaller.

Table 11 about here

In Table B21 we estimate *only the interactions* of launch activity with all age and all size groups, without the underlying main effects. This approach allows us to directly read off *absolute effects* for the interactions of interest. For the LPM we find large productivity effects for young micro and young medium firms, but these are only marginally significant. For the counts model, we find significant productivity effects for young firms of all sizes, with slightly higher coefficients for young small and young micro firms. This suggests that launch-productivity links on the extensive margin are driven by a very small number of outliers; while for young firms with launch exposure, adding a further launch raises productivity similarly across size groups.

7. Conclusions

This paper introduces a new and experimental measure of firm-level innovation: product/service launches modelled from media reports, which we argue complement existing innovation metrics and provide a potentially important tool for decision-makers. We provide evidence that past patenting, trademarking and self-reported innovation predict launch activity at the firm level, for a panel of UK SMEs. We then look at links from launches and IP to SME productivity. Launch activity is associated with higher firm productivity, controlling for IP, drivers of underlying media exposure and firm heterogeneity. Firms in the service sector and those with high-profile launches seem to drive much of this result.

References

Acs ZJ and Audretsch DB. (1987) Innovation, Market Structure, and Firm Size. *The Review of Economics and Statistics* 69: 567-574.

Aghion P, Blundell R, Griffith R, et al. (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.

Arora A, Belenzon S and Sheer L. (2017) Back to Basics: Why do Firms Invest in Research? *NBER Working Paper 23187.* Cambridge, MA: NBER.

Audretsch D and Feldman M. (1996) R&D Spillovers and the Geography of Innovation and Production. *American Economic Review* 86: 630-640.

Audretsch D, Hafenstein M, Kritikos AS, et al. (2018) Firm Size and Innovation in the Service Sector. *IZA DP 12035.* Bonn: IZA.

Autor D, Dorn D, Hanson G, Pisano G, and Shu P (forthcoming) Foreign Competition and Domestic Innovation: Evidence from US Patents. *American Economic Review: Insights.*

Baumann J and Kritikos AS. (2016) The link between R&D, innovation and productivity: Are micro firms different? *Research Policy* 45: 1263-1274.

Block JH, Fisch CO, Hahn A, et al. (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.

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.

Chesborough H. (2003) *Open Innovation: The New Imperative for Creating and Profiting from Technology,* Cambridge, MA: Harvard Business School Press.

Cockburn IM, Lanjouw JO and Schankerman M. (2016) Patents and the Global Diffusion of New Drugs. *The American Economic Review* 106: 136-164.

Cohen WM and Levinthal DA. (1990) Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly* 35: 128-152.

Crepon B, Duguet E and Mairesse J. (1998) Research, Innovation And Productivity: An Econometric Analysis At The Firm Level. *Economics of Innovation and New Technology* 7: 115-158.

Crespi, G., Zuniga, P., 2012. Innovation and Productivity: Evidence from Six Latin American Countries. World Development 40, 273–290.

Davies N. (2009) *Flat Earth News,* London: Vintage.

Einav L and Levin JD. (2014) The Data Revolution and Economic Analysis. *Innovation Policy and the Economy* 14.

Fernandes AM and Paunov C. (2015) The Risks of Innovation: Are Innovating Firms Less Likely to Die? *The Review of Economics and Statistics* 97: 638-653.

Fosfuri A, Giarratana MS and Luzzi A. (2008) The Penguin Has Entered the Building: The Commercialization of Open Source Software Products. *Organization Science* 19: 292-305.

Gentzkow M and Shapiro JM. (2010) What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica* 78: 35-71.

Gentzkow M, Kelly BT and Taddy M. (2017) Text as Data. *NBER Working Paper 23276.* Cambridge, MA: NBER.

Griffith R, Huergo E, Mairesse J and Peters B (2006). Innovation and Productivity Across Four European Countries. *Oxford Review of Economic Policy* 22, 483–498.
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.
Guzman J and Stern S. (2015) Where is Silicon Valley? *Science* 347: 606-609.
Hall BH. (2011) Innovation and Productivity. *NBER Working Paper 17178.* Cambridge, MA: NBER.
Hall BH and Harhoff D. (2012) Recent Research on the Economics of Patents. *NBER Working Paper No. 17773.* Cambridge, MA: NBER
Hall BH, Helmers C, Rogers M, et al. (2013) The Importance (or not) of Patents to UK Firms. *National Bureau of Economic Research Working Paper 19089.* Cambridge, MA: NBER.
Hall BH, Helmers C, Rogers M, et al. (2014) The Choice Between Formal and Informal Intellectual Property: A review. *Journal of Economic Literature* 52: 375-423.
Haltiwanger J, Jarmin R, S. and Miranda J. (2013) Who Creates Jobs? Small versus Large versus Young. *The Review of Economics and Statistics* 95: 347-361.
Hashi I and Stojčić N. (2013) The impact of innovation activities on firm performance using a multi-stage model: Evidence from the Community Innovation Survey 4. *Research Policy* 42: 353-366.
Helmers C and Rogers M. (2010) Innovation and the Survival of New Firms in the UK. *Review of Industrial Organization* 36: 227-248.
Howell A. (2015) ‘Indigenous’ innovation with heterogeneous risk and new firm survival in a transitioning Chinese economy. *Research Policy* 44: 1866-1876.
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, et al. (2018) Measuring Technological Innovation over the Long Run. *NBER Working Paper 252266.* Cambridge, MA: NBER.
Klomp, L., Van Leeuwen, G., 2001. Linking Innovation and Firm Performance: A New Approach. *International Journal of the Economics of Business* 8, 343–364.
Lafrance A. (2016) ‘Access, Accountability Reporting and Silicon Valley. *Nieman Reports*, 17 August.
Lööf H, Heshmati A, Asplund R and Nåå, S-O. (2001) Innovation and Performance in Manufacturing Industries. *SSE/EFI Working Paper Series in Economics and Finance No. 457.*
Li W and Hall BH. (2016) Depreciation of Business R&D Capital. *NBER Working Paper No. 22473.* Cambridge, MA: NBER.
Lucas R. (1988) On the Mechanics of Economic Growth. *Journal of Monetary Economics* 22: 3-42.
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.
Mohnen P and Hall BH. (2013) Innovation and Productivity: An Update. *Eurasian Business Review* 3: 47-65.
Morris DM. (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.

NESTA. (2010) The vital 6 per cent: How high-growth innovative businesses generate prosperity and jobs. London: NESTA.
Office of National Statistics. (2017) Business Structure Database, 1997-2017 SN: 6697. *Secure Data Service Access [computer file]*. Colchester: UK Data Archive.
Roberts ME, Stewart BM and Tingley D. (2016) stm: R Package for Structural Topic Models. http://www.structuraltopicmodel.com.
Romer P. (1990) Endogenous Technological Change. *Journal of Political Economy* 98: 71-102.
Rosenbusch N, Brinckmann J and Bausch A. (2011) Is innovation always beneficial? *Journal of Business Venturing* 26: 441-457.
Schumpeter J. (1962) *The Theory of Economic Development*, Berlin: Springer.
Teece DJ, Pisano G and Shuen A. (1997) Dynamic Capabilities and Strategic Management. *Strategic Management Journal* 18: 509-533.
Varian HR. (2014) Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives* 28: 3-28.
Viner K. (2016) How Technology Disrupted the Truth. *Guardian.*
Von Hippel E. (2005) *Democratising Innovation*, Cambridge, MA: MIT Press.

Figures and tables.

Table 1. Event coverage and type, 2014-2015.

A. All firms	**Obs**	**Mean**	**Std. Dev.**
Firm has event	2,643,043	0.009	0.092
Event count	2,643,043	0.030	0.663
Firm has product launch	2,643,043	0.003	0.056
Product launch count	2,643,043	0.009	0.347
B. Firms with events	**Obs**	**Mean**	**Std. Dev.**
Event count	22,622	3.452	6.293
Firm has product launch	22,622	0.366	0.482
Product launch count	22,622	1.093	3.591
C. Coverage by year	**Year**	**Events**	**%**
	2014	12,018	53.13
	2015	10,604	46.87
	Total	*22,622*	*100*

Source: GI. Panel A shows all firms in the sample. Panel B shows all observations where events or launches are observed, as applicable. Panel C gives the breakdown across years.

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

Variable	Mean for firms with			
	No events	**Events**	**Events, no launches**	**Events and launches**
Patent count	0.001	0.042	0.03	0.062 ***
Weighted patent count	0.001	0.041	0.029	0.062 ***
EPO/US/PCT patents	0.001	0.023	0.015	0.037 ***
Weighted EPO/US/PCT patents	0.001	0.023	0.014	0.037 ***
TM count	0.002	0.015	0.011	0.022 ***
Firm reports product or process innovation	0.276	0.41	0.354	0.512 **
Rev per worker two-year average	146.55	781.3	966.7	461.6 ***
Annual % rev per worker growth	-0.006	0.017	0.019	0.0132
High rev per worker growth firm	0.129	0.148	0.154	0.14 ***
Revenue two-year average	811	13752	14032	13264
Annual % revenue growth	0.011	0.049	0.05	0.048
High revenue growth firm	0.15	0.215	0.218	0.208
Employment two-year average	5.1	21.2	21.2	21.3
Annual % employment growth	0.017	0.032	0.029	0.036
High jobs growth firm	0.014	0.06	0.06	0
Age entered BSD / incorporated	12.4	17.9	17.8	17.9
Year incorporated	2004	1998	1998	1998
Startup	0.142	0.028	0.028	0.028
Firm has 1-9 staff	0.892	0.571	0.573	0.568
Firm has 10-49 staff	0.086	0.284	0.281	0.289
Firm has 50-249 staff	0.013	0.124	0.123	0.125
Immediate foreign ownership	0.165	0.328	0.263	0.44 ***
Firm is in a group of enterprises	0.003	0.055	0.06	0.047 ***
Number of companies in the group	0.008	0.187	0.226	0.119 ***
Firm is a company	0.942	0.903	0.887	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	0.898	0.858 ***
Urban TTWA	0.788	0.838	0.839	0.836
Greater London	0.228	0.303	0.31	0.291 ***
Observations	*2,620,421*	*22,622*	*14,347*	*8,275*

Source: BSD / CH / GI / Orbis / UKIPO / UKIS. Firms with events gives all observations where events are observed. Stars give the results of rank-sum tests for events sub-sample. *** 1% significance, ** 5% significance. All mean differences between firms with and without events are significant.

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

		A. All firms. % of firms with coverage.					**B. Firms with events exposure. % of firms with coverage**			
sic03	**sic03 section name**	**events**	**launch**	**patent**	**tm**	***N***	**launch**	**patent**	**tm**	***N***
A	Agriculture, hunting and forestry	0.29	0.1	.	0.06	*40,735*	33.33	.	.	*117*
B	Fishing	.	.	.	.	*2,964*	.	.	.	*9*
C	Mining and quarrying	6.1	1.27	.	.	*869*	20.75	.	.	*53*
D	Manufacturing	1.25	0.57	0.33	0.25	*196,577*	45.65	2.93	1.46	*2,458*
E	Electricity, Gas and Water Supply	1.81	.	.	.	*3,483*	.	.	.	*63*
F	Construction	0.34	0.08	0.01	0.02	*293,384*	23.42	.	.	*1,012*
G	Wholesale and retail trade, etc	0.73	0.38	0.06	0.23	*405,338*	52.71	0.81	0.78	*2,956*
H	Hotels and restaurants	0.28	0.09	.	0.05	*120,108*	31.44	.	.	*334*
I	Transport, storage and communications	0.88	0.32	0.03	0.07	*94,905*	36.56	.	.	*837*
J	Financial intermediation	2.18	0.41	.	0.09	*48,217*	18.93	.	1.05	*1,051*
K	Real estate, renting and business activities	0.94	0.33	0.08	0.1	*1,152,392*	34.67	1.19	0.88	*10,812*
L	Public administration and defence, etc	.	.	.	.	*15*	.	.	.	.
M	Education	1.16	0.27	.	0.13	*34,850*	23.21	.	.	*405*
N	Health and social work	0.73	0.18	0.03	0.07	*79,130*	25.3	.	.	*577*
O	Other community, social and personal services	1.14	0.41	0.03	0.14	*169,995*	36.22	.	1.14	*1,938*
P	Household domestic employment	.	.	.	.	*47*	.	.	.	.
Q	Extra-terrestrial organisations, bodies	.	.	.	.	*34*	.	.	.	.
	Average coverage	*0.86*	*0.31*	*0.07*	*0.12*	*2,643,043*	*36.58*	*1.09*	*0.88*	*22,622*
	Observations (with coverage)	*22,622*	*8,275*	*1,941*	*3,164*		*8,275*	*246*	*198*	

Source: BSD / CH / GI / Orbis / UKIPO / UKIS. Cells with under 10 observations suppressed to avoid disclosure. Each entry represents the share of companies with coverage (event, launch, patent, tm or ukis). Panel A reports all the companies in the sample (2.6m) and UKIS column entries refer to the sample of UKIS companies matched in the data.

Table 4. Area coverage for events, launches, patents, trademarks and UKIS, 2014-15.

A. All firms (%)	**Non-urban**	**Urban**	**Not London**	**London**
No events	21.24	78.76	77.2	22.8
Events	16.19	83.81	69.68	30.32
No launches	21.22	78.78	77.16	22.84
Launches	16.41	83.59	70.9	29.1
No patents	21.2	78.8	77.13	22.87
Patents	21.48	78.52	84.96	15.04
No TMs	21.2	78.8	77.14	22.86
TMs	19.66	80.34	74.37	25.63
No reported innovation	28.72	71.28	85.61	14.39
Reported innovation	29.16	70.84	87.21	12.79
All	*21.2*	78.8	*77.14*	*22.86*
	560,364	*2,082,679*	*2,038,730*	*604,313*
B. Firms with events (%)	**Non-urban**	**Urban**	**Not London**	**London**
No launches	16.06	83.94	68.98	31.02
Launches	16.41	83.59	70.9	29.1
No patents	16.13	83.87	69.53	30.47
Patents	21.14	78.86	82.93	17.07
No TMs	16.18	83.82	69.69	30.31
TMs	16.67	83.33	68.69	31.31
No reported innovation	19.44	80.56	77.78	22.22
Reported innovation	.	84	84	.
All	*16.19*	*83.81*	*69.68*	*30.32*
	3,662	*18,960*	*15,763*	*6,859*

Source: BSD / CH / GI / Orbis / UKIPO / UKIS. Cells with under 10 observations suppressed to avoid disclosure.

Table 5. Linking past IP activity to product launches. Stepwise regressions, all firms.

A. Pr(launch)	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
L1.15% depreciated PCT / EPO / US patent count	0.009***	0.007***	0.006***	0.005***	-0.003
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
L1.15% depreciated TM count	0.010***	0.008***	0.007***	0.007***	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Ave pre-2009 patenting			-0.007**	-0.007**	-0.010*
			(0.003)	(0.003)	(0.005)
Firm patents pre-2009			0.041***	0.036***	0.076*
			(0.010)	(0.009)	(0.041)
L1.firm reports product or process innovation					0.014**
					(0.006)
Observations	2643043	866076	866076	858096	3347
R^2	0.0016	0.0081	0.0086	0.0148	0.1521
B. Launch counts	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
L1.15% depreciated PCT / EPO / US patent count	0.037***	0.027***	0.027***	0.026***	-0.000
	(0.003)	(0.007)	(0.006)	(0.007)	(0.007)
L1.15% depreciated TM count	0.028***	0.019***	0.019***	0.017***	0.072***
	(0.007)	(0.006)	(0.006)	(0.005)	(0.022)
Ave pre-2009 patenting			-0.023***	-0.023***	-0.004
			(0.005)	(0.005)	(0.018)
Firm patents pre-2009			0.077***	0.061**	0.059
			(0.027)	(0.028)	(0.185)
L1.firm reports product or process innovation					0.100
					(0.076)
Observations	2643043	866076	866076	858096	3347
R^2	0.0007	0.0030	0.0031	0.0063	0.1328
Controls	N	Y	Y	Y	Y
Pre-sample patenting	N	N	Y	Y	Y
Year, area and industry dummies	N	N	N	Y	Y

Source: BSD / CH / GI / Orbis / UKIPO / UKIS. Controls fitted include log mean turnover and employment, startup dummy, firm size dummies, company legal status and structure dummies, plus an urban TTWA dummy Controls lagged one year except age. Standard errors clustered on 2-digit SIC. *** significant at 1%, ** significant at 5%, * significant at 10%. Constant not shown.

Table 6. Effect of innovation (launch dummy) on firm productivity, all firms.

	Log revenue/worker		Rev/worker growth		High-growth episodes	
	(1)	(2)	(3)	(4)	(5)	(6)
L.new product launch	0.452***		0.006		0.006	
	(0.018)		(0.006)		(0.004)	
L2.15% depreciated PCT / EPO / US patent count	-0.003	0.001	-0.002	-0.002	0.002*	0.002*
	(0.005)	(0.005)	(0.002)	(0.002)	(0.001)	(0.001)
L2.15% depreciated TM count	0.181***	0.184***	0.011**	0.011**	0.010***	0.010***
	(0.017)	(0.017)	(0.004)	(0.004)	(0.004)	(0.004)
Ave pre-2009 patenting	0.053	0.050	0.007	0.007	0.017***	0.016***
	(0.036)	(0.036)	(0.007)	(0.007)	(0.006)	(0.006)
Firm patents pre-2009	0.084	0.100*	-0.003	-0.003	-0.005	-0.005
	(0.060)	(0.060)	(0.013)	(0.013)	(0.011)	(0.011)
Observations	1580303	1580303	1596775	1596775	1596775	1596775
R^2	0.0849	0.0842	0.0015	0.0015	0.0097	0.0097

Source: BSD / CH / GI / Orbis / UKIPO / UKIS. All models fit controls, area, year and SIC2 dummies. Controls fitted include 5-year lag of log turnover and mean employment, startup dummy, firm size dummies, company legal status and structure dummies, plus an urban TTWA dummy. Controls lagged one year except age. Standard errors clustered on 2-digit SIC. *** significant at 1%, ** significant at 5%, * significant at 10%. Constant not shown.

Table 7. Effect of innovation (launch counts) on firm productivity, all firms.

	Log revenue/worker		Rev/worker growth		High-growth episodes	
	(1)	(2)	(3)	(4)	(5)	(6)
L.new product launch count	0.047***		0.001		0.001*	
	(0.005)		(0.001)		(0.001)	
L2.15% depreciated PCT / EPO / US patent count	-0.001	0.001	-0.002	-0.002	0.002*	0.002*
	(0.005)	(0.005)	(0.002)	(0.002)	(0.001)	(0.001)
L2.15% depreciated TM count	0.183***	0.184***	0.011**	0.011**	0.010***	0.010***
	(0.017)	(0.017)	(0.004)	(0.004)	(0.004)	(0.004)
Ave pre-2009 patenting	0.052	0.050	0.007	0.007	0.017***	0.016***
	(0.036)	(0.036)	(0.007)	(0.007)	(0.006)	(0.006)
Firm patents pre-2009	0.095	0.100*	-0.003	-0.003	-0.005	-0.005
	(0.060)	(0.060)	(0.013)	(0.013)	(0.011)	(0.011)
Observations	1580303	1580303	1596775	1596775	1596775	1596775
R^2	0.0845	0.0842	0.0015	0.0015	0.0097	0.0097

Source: BSD / CH / GI / Orbis / UKIPO / UKIS. All models fit controls, area, year and SIC2 dummies. Controls fitted include 5-year lag of log turnover and mean employment, startup dummy, firm size dummies, company legal status and structure dummies, plus an urban TTWA dummy. Controls lagged one year except age. Standard errors clustered on 2-digit SIC. *** significant at 1%, ** significant at 5%, * significant at 10%. Constant not shown.

Table 8. Effect of innovation (launch dummy) on firm productivity, firms with events.

	Log revenue/worker		Rev/worker growth		High-growth episodes	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
L.new product launch	0.064***		0.000		-0.006	
	(0.019)		(0.007)		(0.005)	
L2.15% depreciated PCT / EPO/US patent count	0.004	0.005	-0.006*	-0.006*	0.002	0.002
	(0.007)	(0.007)	(0.003)	(0.003)	(0.002)	(0.002)
L2.15% depreciated TM count	0.081***	0.081***	0.002	0.002	0.003	0.003
	(0.024)	(0.024)	(0.005)	(0.005)	(0.005)	(0.005)
Ave pre-2009 patenting	0.072	0.070	0.009	0.009	0.029*	0.029*
	(0.057)	(0.058)	(0.022)	(0.022)	(0.016)	(0.016)
Firm patents pre-2009	-0.223*	-0.217*	0.014	0.014	-0.019	-0.020
	(0.116)	(0.116)	(0.043)	(0.043)	(0.029)	(0.029)
Observations	27019	27019	27794	27794	27794	27794
R^2	0.1663	0.1659	0.0108	0.0108	0.0232	0.0232

Source: BSD / CH / GI / Orbis / UKIPO / UKIS. All models fit controls, area, year and SIC2 dummies. Controls fitted include 5-year lag of log turnover and mean employment, startup dummy, firm size dummies, company legal status and structure dummies, plus an urban TTWA dummy. Controls lagged one year except age. Standard errors clustered on 2-digit SIC. *** significant at 1%, ** significant at 5%, * significant at 10%. Constant not shown.

Table 9. Effect of innovation (launch counts) on firm productivity, firms with events.

	Log revenue/worker		Rev/worker growth		High-growth episodes	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
L.new product launch count	0.017***		0.000		0.001	
	(0.005)		(0.001)		(0.001)	
L2.15% depreciated PCT / EPO / US patent count	0.004	0.005	-0.006*	-0.006*	0.002	0.002
	(0.007)	(0.007)	(0.003)	(0.003)	(0.002)	(0.002)
L2.15% depreciated TM count	0.081***	0.081***	0.002	0.002	0.003	0.003
	(0.024)	(0.024)	(0.005)	(0.005)	(0.005)	(0.005)
Ave pre-2009 patenting	0.073	0.070	0.009	0.009	0.029*	0.029*
	(0.058)	(0.058)	(0.022)	(0.022)	(0.016)	(0.016)
Firm patents pre-2009	-0.220*	-0.217*	0.014	0.014	-0.020	-0.020
	(0.116)	(0.116)	(0.043)	(0.043)	(0.029)	(0.029)
Observations	27019	27019	27794	27794	27794	27794
R^2	0.1672	0.1659	0.0108	0.0108	0.0232	0.0232

Source: BSD / CH / GI / Orbis / UKIPO / UKIS. All models fit controls, area, year and SIC2 dummies. Controls fitted include 5-year lag of log turnover and mean employment, startup dummy, firm size dummies, company legal status and structure dummies, plus an urban TTWA dummy. Controls lagged one year except age. Standard errors clustered on 2-digit SIC. *** significant at 1%, ** significant at 5%, * significant at 10%. Constant not shown.

Table 10. Launch quality, launch importance and firm productivity, events subsample.

	(1) Log rev/worker	(2) Rev/worker growth	(3) High growth episodes
L.total launch reports, main topic	0.000***	-0.000**	-0.000
	(0.000)	(0.000)	(0.000)
	0.1664	*0.0109*	*0.0232*
L.weighted launch reports, main topic	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
	0.1659	*0.0108*	*0.0232*
L.firm has important launch, main topic	0.168***	-0.027	-0.007
	(0.052)	(0.018)	(0.013)
	0.1662	*0.0108*	*0.0232*
L.count of important launches, main topic	0.218***	-0.014	-0.002
	(0.069)	(0.016)	(0.014)
	0.1665	*0.0108*	*0.0232*
Observations	27019	27794	27794

Source: BSD / CH / GI / Orbis / UKIPO / UKIS. Each specification shows coefficient of *b* in equation (2), standard error in parenthesis and R^2 in italics. All models fit controls, area, year and SIC2 dummies. Controls fitted include 5-year lag of log turnover and mean employment, startup dummy, firm size dummies, company legal status and structure dummies, plus an urban TTWA dummy. Controls lagged one year except age. Standard errors clustered on 2-digit SIC. *** significant at 1%, ** significant at 5%, * significant at 10%.

Table 11. Productivity, launch dummies, age and size. Events subsample.

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

Source: BSD / CH / GI / UKIPO / UKIS. All models fit controls and fixed effects. Age and size variables and interactions are lagged one period. 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. Reference categories are older and medium-sized firms. Standard errors clustered on 2-digit SIC. *** significant at 1%, ** significant at 5%, * significant at 10%.

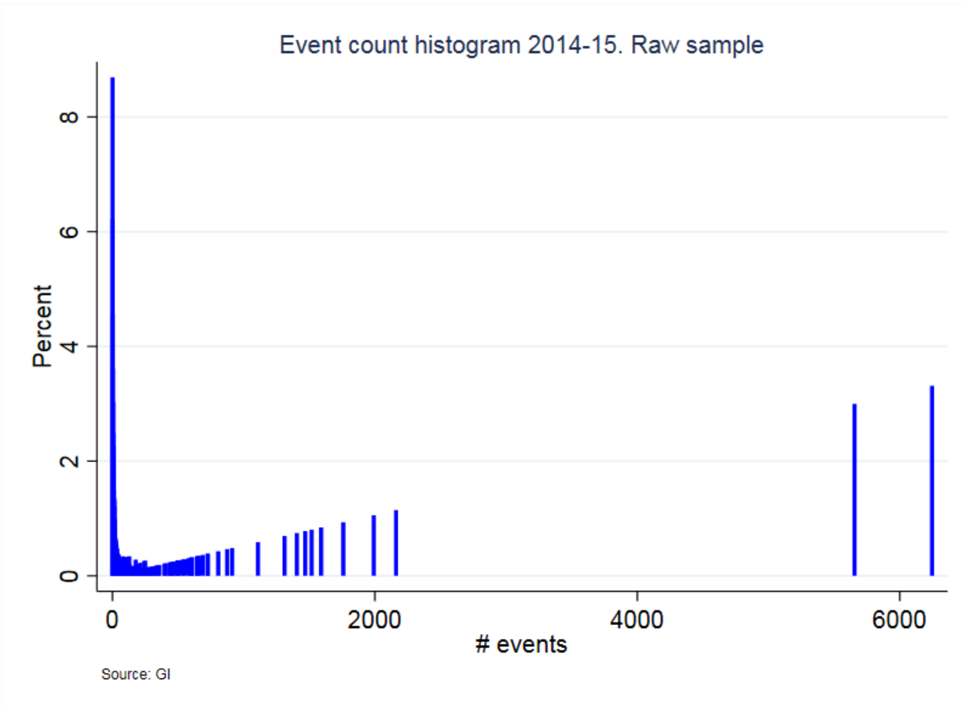
Figure 1. Example 'events', showing raw text and classification.

Sample fragment	**Masterwork** 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	http://www.XXX/NewsStory.aspx?i=2296
event_date	2014-03
source_name	xxx
company_id	13724
event_type_id	product_launch

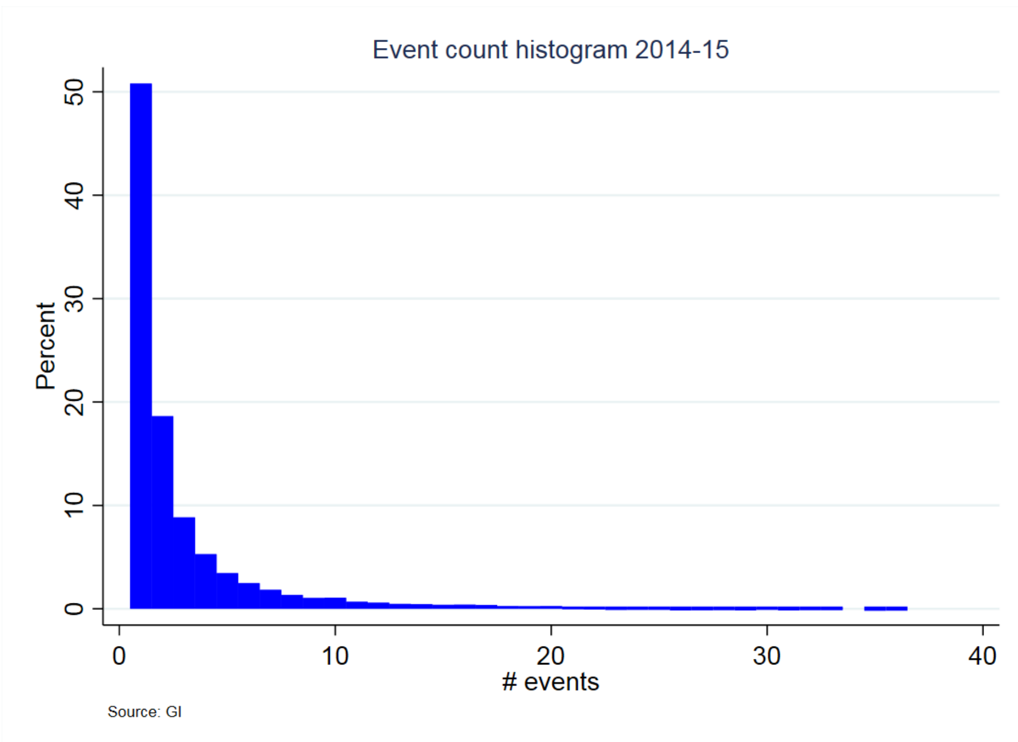
Sample fragment	**Hammond Electronics** has launched a range of designspecific moulded enclosures to support the new types of credit card sized, low cost bare board computers, which, typically running Linux, provide basic functionality across a wide range of applications…
doc_title	Enclosures for credit-card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

Source: GI.
Note: 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 info. 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), compared with estimation sample of single-plant SMEs (bottom).



A. Raw sample, all firms.



B. Estimation sample of single-plant SMEs. Disclosive cell counts suppressed.

Figure 3. Counts of launches, patents and trademarks across TTWAs, 2014-2015.

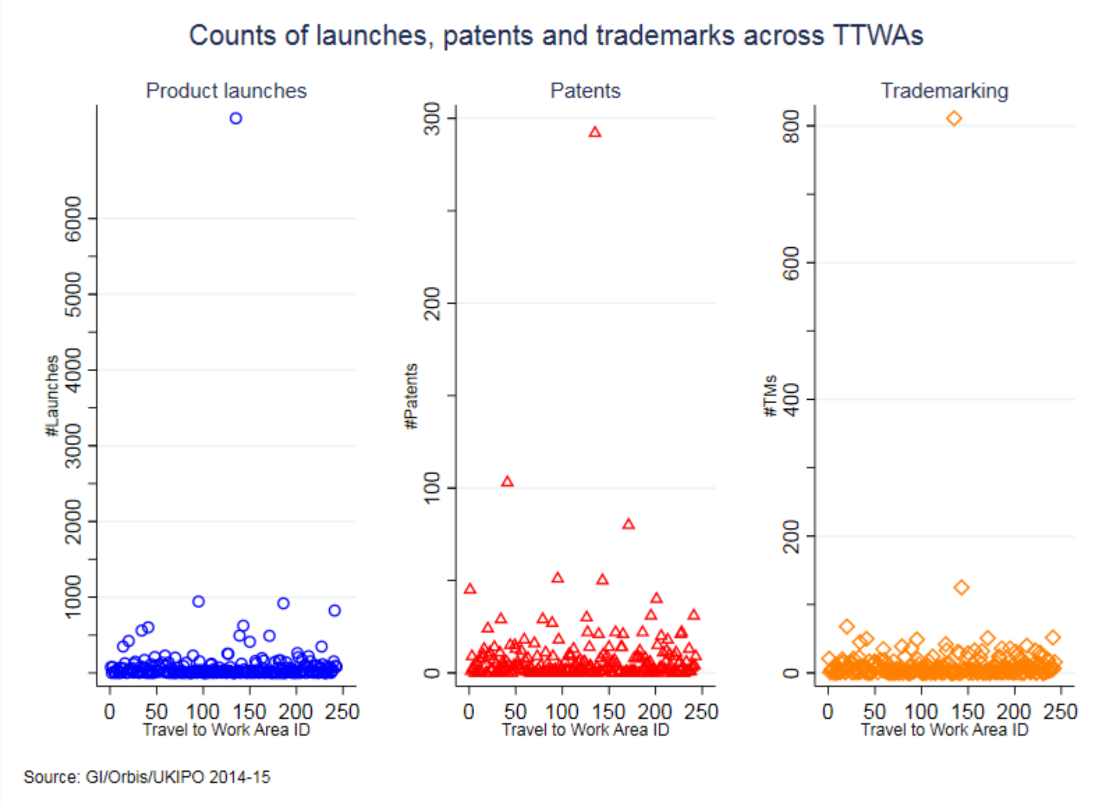


Figure 4. Weighted counts of launches, patents and trademarks, TTWAs, 2014-2015.

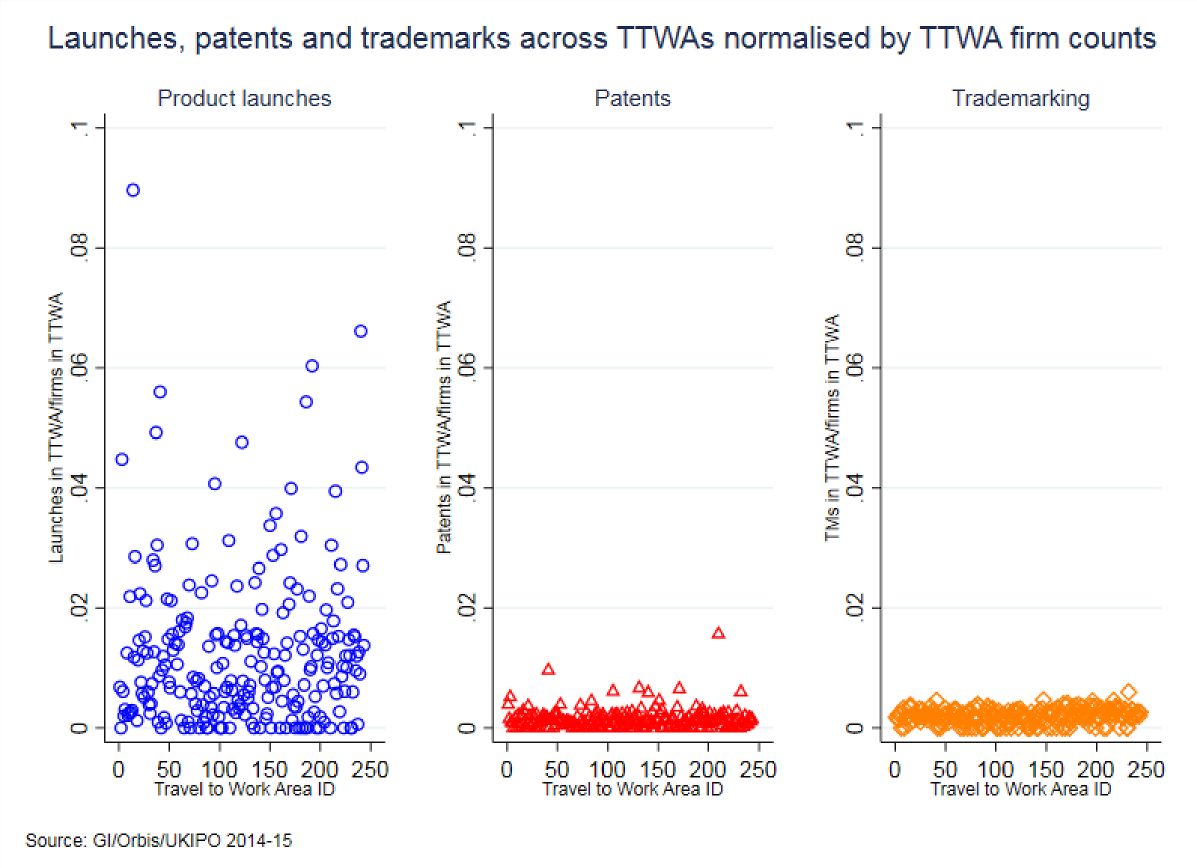
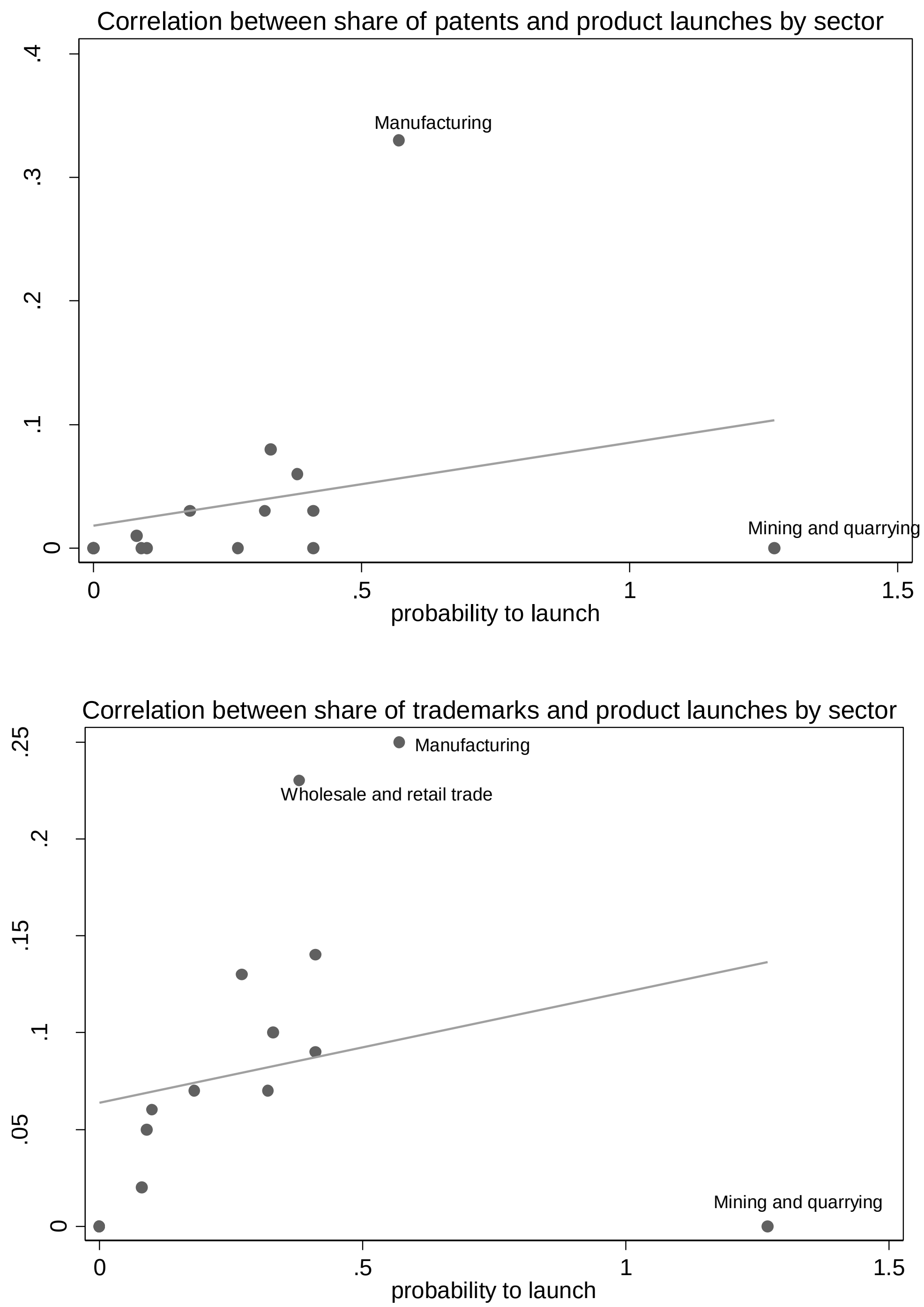


Figure 5. Correlation between share of companies patenting/trademarking and launching a product. 1-digit industry level.



Source: GI / CH / Orbis / UKIPO.

INNOVATIVE EVENTS: 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 company websites or from 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 covering 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.

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. To give a sense of this upper bound, below we set out some analysis of the hard cases subset. Note that in our main analysis, we restrict the sample to single plant SMEs, removing these hard cases from the data and further guarding against error.

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 (hence 'big digital', 12.5% of the 40,000 observations ascribed to such firms). Analysis using title and text fragment fields suggests around 16% content farming error in the 'big digital' sample, 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). Note that GI's 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 also conduct further, experimental tests on a sub-sample of the 'big digital' companies. We use the URL field to re-extract the original text, then to reverse-engineer GI's feature extraction and syntax parsing routines. We are only able to perform this exercise on websites that are a) scrapable b) active (return a 200 to standard HTTP requests). This reduces our sample size to

1,746. 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 abd 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 therefore drop 413 cases where no company is mentioned in the article title.

This leaves us with three scenarios. 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 $C_{candidates.}$ For the 'digital' sample, we find 95.1% incorrect ascription on the 977 remaining observations. Note that this is a lower bound on the true error rate: if we assume all 'probably correct' cases are incorrect, the error rate rises to 99% of cases.[1] Note that our focus on single plant SMEs removes hard cases from the data, minimising the ascription error rate on the rest of the sample. However, to the extent that mis-ascription 'gives' events to large tech and media firms which actually belong to SMEs, we have a lower bound on the true level of event activity for our firms of interest.

Event observations vs. real-world events

[1] We check for out-of-sample error rates by running these routines for the full set of 40,000 observations, with very similar results.

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.[2]

Before modelling the data, we stem the fragments (reducing the words to their roots) and remove stopwords (definite and indefinite articles, pronouns, etc.).[3] We then group individual event observations according to three variables – type of event, company and event date[4] – and run the model within each group. If an event is reported by several sources in different formats on the

[2] This threshold can be modified.
[3] For more precise information on the model and on the implementation in R see Roberts et al (2016).
[4] 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.

same day, the STM algorithm identifies the repetitions and keeps one of them. STM processing substantially reduces the number of events observations, to 257,056.

Event importance

We 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 follow 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. 98% of launches only have one mention: given that our firms are single plant SMEs this is not surprising. We dub the remaining 2% of launches '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.

A2 / Panel build

To build the panel, we match Companies House companies to enterprises in the Business Structure Database (BSD). Growth Intelligence (GI) data is pre-matched to Companies House identifiers. We then match in patents, trademarks and UK Innovation Survey (UKIS) data.

BSD-Companies House-GI matching

Company-level data (Companies House and GI) is based on companies active as of August 2012 (financial year 2013). The UK Data Service team therefore matches companies to enterprises using the 2013 BSD cross-section, which comprises 1,818,263 unique entrefs (which denote individual enterprises in the BSD). The initial matching rate 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.[5] 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.

[5] 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).

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 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.[6]

At the end of these further cleaning steps we have 1,364,624 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. This reduces the panel to 5,039,811 observations. In merging, we drop large firms and multi-plant firms, which reduces the panel to 5,013,702 and then 4,878,646 observations respectively. Finally, we remove outliers: specifically in each year we observations with an event count higher than 1 standard deviation of the mean event count. This drops 84 observations, giving us a final 'estimation panel' of 4,878,532 firm*year observations for 1,364,624 single-plant SMEs.

The raw 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

[6] 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's 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.

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.

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. 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.[7] The overall match rate for 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.

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

[7] 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.

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.

UK Innovation Survey

Finally, we match in information on firms' self-reported innovative activity from the UK Innovation Survey (UKIS). Specifically, we use UKIS4-UKIS9, covering the years 2002-2014 inclusive, where 2014 is the most recent information available. Since each survey covers three years, we ascribe data to the middle year. For example, UKIS9, which covers 2012-2014 inclusive, is ascribed to panel year 2013. This gives us 85,834 UKIS observations for 56,473 firms in 2003, 2005, 2007, 2009, 2011 and 2013.

There are three main constraints in merging this data to our panel. First, we only have data to 2014. Second, we use single plant SMEs, not all firms. Response bias with innovation surveys may mean such smaller firms are under-represented in raw UKIS data. Third, we merge onto using Enterprise Group numbers rather than entrefs, as these are the only available IDs provided. This allows for an enterprise-level match in many cases, but matches will fail if the group consists of more than one enterprise. Given these constraints, we successfully merge UKIS information for 43,211 observations, around 50% of the initial UKIS sample, with 26,708 unique firms. After restricting the panel to single plant SMEs 2014-2017, we have 4,195 UKIS observations for the same number of firms. Given the restrictiveness of the match, we run Hotelling Tests and other diagnostics, finding that the subset of UKIS firms differs on observables from the rest of the data.

Panel and variables

The final panel contains 4,878,532 observations for 1.364m enterprises in the financial years 2014-2017. For 2014-2015, the years when events are observable, we have 2,643,043 observations for 1.36m firms.

Besides events and launch measures, 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 ±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.
- Reported innovation – for firms in the UKIS sample (4,195 observations) a dummy taking the value 1 if the company reports a product innovation, a process innovation or both in a given wave.
- 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

Hall BH and Harhoff D. (2012) Recent Research on the Economics of Patents. *NBER Working Paper No. 17773*. Cambridge, MA: NBER.

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

Table B1. Summary statistics for single plant SMEs, 2014-2017.

VARIABLES	N	mean	sd
Firm has event	2,643,043	0.00856	0.0921
Total events	2,643,043	0.0295	0.663
New product launch	2,643,043	0.00313	0.0559
Total product launches	2,643,043	0.00935	0.347
Total launch reports main topic	2,643,043	3.398	211.0
Mean launch reports / topics	2,643,043	2.776	174.7
Firm has important launch (main topic)	2,643,043	0.000313	0.0177
Count of important launches (main topic)	2,643,043	0.000267	0.0170
Patent count	4878562	0.000860	0.0911
Weighted patent count	4878562	0.000855	0.0908
TM count	4878562	0.000924	0.0507
Firm reports product or process innovation in UKIS	4,195	0.280	0.449
Rev per worker two-year average	4829893	161.9	6,497
Annual % rev per worker growth	4878562	0.00501	0.477
High rev per worker growth firm	4730583	0.133	0.340
Revenue two-year average	4878562	1,037	74,487
Annual % revenue growth	4878562	0.0101	0.465
High revenue growth firm	4730583	0.148	0.355
Employment two-year average	4878562	5.613	14.49
Annual % employment growth	4878562	0.00434	0.302
High jobs growth firm	4730583	0.0135	0.116
Services sector	4878562	0.907	0.290
Firm has 1-9 staff	4829893	0.884	0.321
Firm has 10-49 staff	4829893	0.0929	0.290
Firm has 50-249 staff	4829893	0.0147	0.121
Number of companies per entref	4878562	0.0107	0.654
Enterprise has >1 associated company	4878562	0.00388	0.0622
Public company	3382694	0.940	0.237
Non-profit making body	3382694	0.0238	0.153
Partnership	3382694	0.0142	0.118
Public corporation	3382694	0.000212	0.0145
Sole proprietor	3382694	0.0215	0.145
Age since BSD entry OR incorporation	4878562	13.67	12.25
Firm is 3 years old or less	4878562	0.0767	0.266
Urban TTWA	4878562	0.785	0.411
Greater London	4878562	0.224	0.417

Source: BSD / CH / GI / Orbis / UKIPO / UKIS. Minima and maxima suppressed by UK Data Service. Patents are weighted by number of applicants. High growth firms (jobs / revenue / revenue per worker) are defined using the OECD definition of high growth firms. See Appendix A2 for further variable details.

Table B2. Coverage by SIC1 sectors for product launch, patents and trademarks, 2014-15.

		A. All firms. % of firms with coverage.	B. Firms with events exposure. % of firms with coverage
sic03	**sic03 section name**	**ukis**	**ukis**
A	Agriculture, hunting and forestry	27.03	.
B	Fishing	.	.
C	Mining and quarrying	.	.
D	Manufacturing	38.71	.
E	Electricity, Gas and Water Supply	.	.
F	Construction	21.03	.
G	Wholesale and retail trade, etc	23.22	44
H	Hotels and restaurants	.	.
I	Transport, storage and communications	.	.
J	Financial intermediation	.	.
K	Real estate, renting and business activities	30.65	35.48
L	Public administration and defence, etc	.	.
M	Education	.	.
N	Health and social work	.	.
O	Other community, social and personal services	28.79	.
P	Household domestic employment	.	
Q	Extra-terrestrial organisations, bodies	.	
	Average coverage	*27.96*	*40.98*
	Observations (with coverage)	*1,173*	*50*

Source: BSD / CH / GI / Orbis / UKIPO / UKIS. In panel A the total number of UKIS matched observations is 4,195, while in panel B the total number of observations of companies with events matched to the UKIS is 122.

Table B3. Linking past IP activity to product launches. Variable lags.

A. Pr(launch)	**L1**	**L2**	**L5**	**L1**	**L2**	**L5**
Firm reports product or process innovation				0.014**	0.014**	0.014**
				(0.006)	(0.006)	(0.006)
PCT / EPO / US patent count	0.005***	0.008***	0.007	-0.003	-0.004	-0.029*
	(0.001)	(0.001)	(0.004)	(0.002)	(0.003)	(0.015)
TM count	0.007***	0.012**	0.013**	-0.002	0.019***	0.019***
	(0.002)	(0.005)	(0.005)	(0.003)	(0.006)	(0.006)
Ave pre-2009 patenting	-0.007**	-0.009***	-0.005***	-0.010*	-0.007	0.002
	(0.003)	(0.002)	(0.002)	(0.005)	(0.007)	(0.008)
Firm patents pre-2009	0.036***	0.037***	0.039***	0.076*	0.074*	0.081*
	(0.009)	(0.009)	(0.009)	(0.041)	(0.041)	(0.045)
Observations	858096	858096	858096	3347	3347	3347
R^2	0.0148	0.0148	0.0142	0.1521	0.1527	0.1547
B. Launch counts	**L1**	**L2**	**L5**	**L1**	**L2**	**L5**
Firm reports product or process innovation				0.100	0.106	0.106
				(0.076)	(0.081)	(0.080)
PCT / EPO / US patent count	0.026***	0.035***	0.018	-0.000	-0.003	-0.063
	(0.007)	(0.013)	(0.016)	(0.007)	(0.009)	(0.046)
TM count	0.017***	0.061	0.063	0.072***	0.808***	0.808***
	(0.005)	(0.045)	(0.044)	(0.022)	(0.155)	(0.154)
Ave pre-2009 patenting	-0.023***	-0.033***	-0.007	-0.004	-0.003	0.029*
	(0.005)	(0.005)	(0.005)	(0.018)	(0.016)	(0.017)
Firm patents pre-2009	0.061**	0.065**	0.073**	0.059	0.078	0.096
	(0.028)	(0.028)	(0.031)	(0.185)	(0.178)	(0.198)
Observations	858096	858096	858096	3347	3347	3347
R^2	0.0063	0.0063	0.0058	0.1328	0.1422	0.1424

Source: BSD / CH /Orbis / IPO / GI / UKIS. All models fit controls, area (TTWA), time (year) and 2-digit SIC industry dummies. Patent stocks are lagged back up to 5 periods, TM stocks up to 2 periods. Controls fitted as per Table 5 in the text. Standard errors clustered on 2-digit SIC. *** significant at 1%, ** significant at 5%, * significant at 10%.

Table B4. Sensitivity tests. Launch dummy.

All firms	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
L1.15% depreciated PCT / EPO / US patent count	0.005***	0.005***	0.005***	0.005***	0.006***	0.006***	0.006***	0.005***			0.009**	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)			(0.003)	(0.001)
L1.Cumulative PCT / EPO / US patent count									0.004***			
									(0.001)			
L1.40% depreciated PCT / EPO / US patent count										0.007***		
										(0.001)		
L1.15% depreciated TM count	0.007***	0.007***	0.007***	0.007***	0.007***	0.007***	0.007***	0.007***	0.007***	0.007***	0.003	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Ave pre-2009 patenting	-0.007**	-0.007**	-0.007**	-0.007**	-0.007**	-0.007**	-0.007**	-0.006**	-0.007**	-0.005*	-0.014*	-0.007**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.007)	(0.003)
Firm patents pre-2009	0.036***	0.036***	0.036***	0.036***	0.034***	0.034***	0.034***	0.035***	0.039***	0.036***	0.029*	0.050***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.016)	(0.013)
Observations	858096	858096	858096	858096	703154	703154	703154	858096	858096	858096	192096	665973
R^2	0.0148	0.0148	0.0149	0.0148	0.0165	0.0165	0.0165	0.0153	0.0148	0.0148	0.0217	0.0139

Source: BSD / CH /Orbis / IPO / GI / UKIS. All models fit controls as per Table 5, main paper, area, year and 2-digit SIC dummies, unless otherwise stated. Standard errors clustered on SIC2. Column 1 fits the 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. Columns 11 and 12 break out the sample into manufacturing and service sector firms respectively.

Table B5. Sensitivity tests. Launch counts.

All firms	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**
L1.15% depreciated PCT	0.026***	0.026***	0.026***	0.026***	0.029***	0.029***	0.029***	0.023***			0.018***	0.027***
/ EPO / US patent count	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.007)			(0.004)	(0.008)
L1.Cumulative PCT									0.016***			
/ EPO / US patent count									(0.006)			
L1.40% depreciated PCT										0.035***		
/ EPO / US patent count										(0.007)		
L1.15% depreciated TM	0.017***	0.017***	0.017***	0.017***	0.019***	0.019***	0.019***	0.018***	0.017***	0.017***	0.018	0.016***
count	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.015)	(0.004)
Ave pre-2009 patenting	-0.023***	-0.023***	-0.023***	-0.023***	-0.024***	-0.024***	-0.024***	-0.015**	-0.018***	-0.015**	-0.020	-0.028***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.004)	(0.007)	(0.015)	(0.007)
Firm patents pre-2009	0.061**	0.061**	0.061**	0.061**	0.054*	0.054*	0.054*	0.053*	0.075**	0.059**	0.024	0.117***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.027)	(0.029)	(0.029)	(0.035)	(0.041)
Observations	858096	858096	858096	858096	703154	703154	703154	858096	858096	858096	192096	665973
R^2	0.0063	0.0063	0.0063	0.0063	0.0072	0.0072	0.0072	0.0068	0.0061	0.0063	0.0112	0.0057

Source: BSD / CH /Orbis / IPO / GI / UKIS. All models fit controls as per Table 5, main paper, area, year and 2-digit SIC dummies, unless otherwise stated. Standard errors clustered on SIC2. Column 1 fits the 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. Columns 11 and 12 break out the sample into manufacturing and service sector firms respectively.

Table B6. Sensitivity tests. Launch dummy. UKIS subsample.

UKIS subsample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
L1.firm reports product	0.014**	0.014**	0.015**	0.014**	0.014**	0.014*	0.015*	0.014**	0.014**	0.014**	0.022**	0.014*
or process innovation	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.010)	(0.008)
L1.15% depreciated PCT /	-0.003	-0.003	-0.003	-0.003	-0.002	-0.003	-0.003	-0.000			-0.006	-0.003
EPO / US patent count	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)			(0.007)	(0.003)
L1.Cumulative PCT /									-0.001			
EPO / US patent count									(0.002)			
L1.40% depreciated PCT /										-0.003		
EPO / US patent count										(0.003)		
L1.15% depreciated TM count	-0.002	-0.002	-0.002	-0.002	-0.003	-0.002	-0.002	-0.002	-0.002	-0.002	-0.028	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.036)	(0.002)
Ave pre-2009 patenting	-0.010*	-0.010*	-0.010*	-0.010*	-0.009*	-0.009*	-0.009*	-0.014**	-0.011**	-0.012**	-0.121	-0.006***
	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.118)	(0.002)
Firm patents pre-2009	0.076*	0.076*	0.076*	0.076*	0.090*	0.089*	0.088*	0.081*	0.072*	0.074*	0.260	0.051
	(0.041)	(0.041)	(0.042)	(0.042)	(0.048)	(0.048)	(0.048)	(0.042)	(0.040)	(0.041)	(0.256)	(0.044)
Observations	3347	3347	3347	3347	3044	3044	3044	3347	3347	3347	845	2502
R^2	0.1521	0.1521	0.1529	0.1526	0.1462	0.1440	0.1443	0.1525	0.1517	0.1518	0.3613	0.1547

Notes as in Table B5.

Table B7. Sensitivity tests. Launch counts. UKIS subsample.

UKIS subsample	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**
L1.firm reports product or process innovation	0.100 (0.076)	0.102 (0.077)	0.102 (0.077)	0.101 (0.076)	0.103 (0.081)	0.106 (0.083)	0.108 (0.085)	0.100 (0.076)	0.101 (0.076)	0.100 (0.076)	0.034* (0.018)	0.120 (0.097)
L1.15% depreciated PCT / EPO / US patent count	-0.000 (0.007)	-0.000 (0.007)	-0.000 (0.007)	-0.000 (0.007)	0.003 (0.009)	0.002 (0.008)	0.001 (0.008)	0.006 (0.005)			-0.012 (0.015)	0.002 (0.012)
L1.Cumulative PCT / EPO / US patent count									0.004 (0.007)			
L1.40% depreciated PCT / EPO / US patent count										0.005 (0.009)		
L1.15% depreciated TM count	0.072*** (0.022)	0.072*** (0.022)	0.072*** (0.022)	0.072*** (0.022)	0.089*** (0.029)	0.091*** (0.030)	0.090*** (0.030)	0.072*** (0.022)	0.072*** (0.022)	0.072*** (0.022)	-0.090 (0.102)	0.075*** (0.019)
Ave pre-2009 patenting	-0.004 (0.018)	-0.005 (0.018)	-0.005 (0.018)	-0.005 (0.018)	-0.006 (0.014)	-0.004 (0.016)	-0.003 (0.017)	-0.013 (0.022)	-0.013 (0.016)	-0.009 (0.020)	-0.237 (0.230)	0.017 (0.027)
Firm patents pre-2009	0.059 (0.185)	0.064 (0.183)	0.062 (0.184)	0.060 (0.185)	0.087 (0.205)	0.078 (0.211)	0.075 (0.211)	0.069 (0.193)	0.058 (0.176)	0.055 (0.182)	0.523 (0.506)	-0.012 (0.263)
Observations	3347	3347	3347	3347	3044	3044	3044	3347	3347	3347	845	2502
R^2	0.1328	0.1330	0.1330	0.1328	0.1429	0.1406	0.1411	0.1328	0.1328	0.1328	0.3308	0.1495

Notes as in Table B5.

Table B8. Linking past IP to launch mentions and importance. All firms.

A. All firms	**raw counts**			**weighted counts**		
	1	**2**	**3**	**1**	**2**	**3**
L1.15% depreciated PCT / EPO / US patent count	7.321	3.734	3.605	1.422	0.648	0.559
	(6.076)	(3.182)	(3.121)	(1.024)	(0.490)	(0.453)
L1.15% depreciated TM count	12.954	6.220	5.584	5.690	2.595	2.250
	(8.517)	(4.283)	(4.033)	(3.580)	(1.733)	(1.564)
Ave pre-2009 patenting	-1.527	1.410	2.767	-1.146	-0.162	0.165
	(11.103)	(3.885)	(3.469)	(1.923)	(0.686)	(0.597)
Firm patents pre-2009	49.648	7.449	-2.859	5.219	0.343	-0.972
	(36.452)	(10.157)	(9.180)	(6.383)	(2.310)	(2.054)
Observations	858096	858096	858096	858096	858096	858096
R^2	0.0012	0.0013	0.0013	0.0005	0.0006	0.0006
B. All firms	**important launch dummy**			**# important launches**		
	1	**2**	**3**	**1**	**2**	**3**
L1.15% depreciated PCT / EPO / US patent count	0.000	0.000	0.000	0.000	0.000	0.000
	(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
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ave pre-2009 patenting	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Firm patents pre-2009	0.001	0.001	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	858096	858096	858096	858096	858096	858096
R^2	0.0015	0.0015	0.0015	0.0011	0.0011	0.0011

Source: BSD / CH /Orbis / IPO / GI / UKIS. Panel A gives results for raw mentions and mentions weighted by launches. Specifications 1-3 cover mentions by all topics; main topic; and main topic/#topics respectively. Other notes as in Table B4.

Table B9. Linking past IP to launch quality. UKIS subsample.

B. UKIS subsample	**raw counts**			**weighted counts**		
	1	**2**	**3**	**1**	**2**	**3**
L1.firm reports product or process innovation	31.257	15.003	13.030	0.809	0.128	0.055
	(29.387)	(13.695)	(11.863)	(2.319)	(0.830)	(0.692)
L1.15% depreciated PCT / EPO / US patent count	-0.549	-0.302	-0.278	-0.026	-0.037	-0.038
	(1.015)	(0.447)	(0.391)	(0.122)	(0.072)	(0.069)
L1.15% depreciated TM count	-9.211**	-2.744	-3.101*	-1.051***	-0.532***	-0.510***
	(4.440)	(1.802)	(1.542)	(0.371)	(0.155)	(0.138)
Ave pre-2009 patenting	5.402	2.875	2.524	0.590	0.408	0.389
	(5.429)	(2.829)	(2.488)	(0.674)	(0.495)	(0.479)
Firm patents pre-2009	15.544	2.716	1.890	2.229	-0.187	-0.391
	(19.399)	(10.323)	(9.184)	(3.719)	(2.662)	(2.579)
Observations	3347	3347	3347	3347	3347	3347
R^2	0.0579	0.0583	0.0580	0.0581	0.0537	0.0524
B. UKIS subsample	**important launch dummy**			**# important launches**		
	1	**2**	**3**	**1**	**2**	**3**
L1.firm reports product or process innovation	0.003	0.003	0.003	0.002	0.002	0.002
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
L1.15% depreciated PCT / EPO / US patent count	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
L1.15% depreciated TM count	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ave pre-2009 patenting	0.001	0.001	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Firm patents pre-2009	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	3347	3347	3347	3347	3347	3347
R^2	0.0620	0.0620	0.0620	0.0592	0.0592	0.0592

Source: BSD / CH /Orbis / IPO / GI / UKIS. Panel A gives results for raw mentions and mentions weighted by launches. Specifications 1-3 cover mentions by all topics; main topic; and main topic/#topics respectively. Other notes as in Table B4.

Table B10. Linking past IP to launch quality. Placebo test.

	OLS	FE	OLS	FE
	2015 launch counts		2014 patent stocks	
2014 patent stocks	0.0204***	0.0194***		
	(0.003)	(0.003)		
2015 product launch counts			0.0009***	0.0008***
			(0.000)	(0.000)
Constant	0.0120***		0.0047***	
	(0.002)		(0.001)	
Observations	2370158	2347297	2370158	2347297
R^2	0.0000	0.0006	0.0000	0.0110

Source: BSD / CH /Orbis / IPO / GI / UKIS. OLS models fit bivariate model. FE models add Travel to Work Area and industry dummies.

Table B11. Linking past IP to launch quality. Falsification test.

	Events		Event counts	
	(1)	**(2)**	**(1)**	**(2)**
L1.15% depreciated PCT / EPO / US patent count	0.003**	0.002	0.008***	0.014*
	(0.001)	(0.002)	(0.002)	(0.008)
L1.15% depreciated TM count	0.009***	0.008	0.021***	0.005
	(0.002)	(0.015)	(0.008)	(0.013)
Ave pre-2009 patenting	0.010***	-0.018***	0.028***	-0.042***
	(0.003)	(0.004)	(0.007)	(0.005)
Firm patents pre-2009	0.004	0.031	-0.013	-0.114
	(0.006)	(0.020)	(0.018)	(0.121)
L1.firm reports product or process innovation		0.012**		0.018**
		(0.005)		(0.008)
Observations	854823	3308	854823	3308
R^2	0.0237	0.1302	0.0173	0.0999

Source: BSD / CH / GI / Orbis / UKIPO / UKIS. Controls fitted include log mean turnover and employment, startup dummy, firm size dummies, company legal status and structure dummies, plus an urban TTWA dummy Controls lagged one year except age. Standard errors clustered on 2-digit SIC. *** significant at 1%, ** significant at 5%, * significant at 10%. Constant not shown.

Table B12. Robustness checks: log productivity.

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

Source: BSD / CH / GI / Orbis / UKIPO / UKIS. Each specification shows coefficient of *b* in equation (2), standard error in parenthesis and R^2 in italics. All models fit controls, area, year and SIC2 dummies. Controls fitted include 5-year lag of log turnover and mean employment, startup dummy, firm size dummies, company legal status and structure dummies, plus an urban TTWA dummy. Controls lagged one year except age. Standard errors clustered on 2-digit SIC. *** significant at 1%, ** significant at 5%, * significant at 10%. Constant not shown.

Table B13. Robustness checks: productivity growth.

Check	Launch dummy	Launch counts	N
Main	0.000	0.000	27794
	(0.007)	(0.001)	
	0.0108	*0.0108*	
Pre-sample high-growth episodes	-0.002	0.000	26047
	(0.007)	(0.001)	
	0.0119	*0.0119*	
Patents 1-period lag	0.000	0.000	27794
	(0.007)	(0.001)	
	0.0107	*0.0107*	
40% depreciated patents	0.000	0.000	27794
	(0.007)	(0.001)	
	0.0107	*0.0107*	
Drop SIC switchers	-0.002	0.000	25419
	(0.007)	(0.001)	
	0.0111	*0.0111*	
Drop TTWA switchers	0.001	0.001	25707
	(0.007)	(0.001)	
	0.0114	*0.0115*	
SIC4 dummies, not SIC2	0.001	0.000	27794
	(0.007)	(0.001)	
	0.0230	*0.0230*	
Industry-year fixed effects	0.001	0.000	27794
	(0.007)	(0.001)	
	0.0291	*0.0291*	
IPC1-year fixed effects	0.001	0.000	27794
	(0.007)	(0.001)	
	0.0112	*0.0112*	
Industry-area clustering	0.001	0.000	27794
	(0.007)	(0.001)	
	0.0230	*0.0230*	
London dummy, not area FE	0.001	0.000	28173
	(0.007)	(0.001)	
	0.0164	*0.0164*	

Source: BSD / CH / GI / Orbis / UKIPO / UKIS. Each specification shows coefficient of *b* in equation (2), standard error in parenthesis and R^2 in italics. All models fit controls, area, year and SIC2 dummies. Controls fitted include 5-year lag of log turnover and mean employment, startup dummy, firm size dummies, company legal status and structure dummies, plus an urban TTWA dummy. Controls lagged one year except age. Standard errors clustered on 2-digit SIC. *** significant at 1%, ** significant at 5%, * significant at 10%. Constant not shown.

Table B14. Robustness checks: high productivity growth episodes.

Check	Launch dummy	Launch counts	N
Main	-0.006	0.001	27794
	(0.005)	(0.001)	
	0.0232	*0.0232*	
Pre-sample high-growth episodes	-0.006	0.001	26047
	(0.005)	(0.001)	
	0.0240	*0.0240*	
Patents 1-period lag	-0.006	0.001	27794
	(0.005)	(0.001)	
	0.0233	*0.0232*	
40% depreciated patents	-0.006	0.001	27794
	(0.005)	(0.001)	
	0.0232	*0.0232*	
Drop SIC switchers	-0.003	0.001	25419
	(0.005)	(0.001)	
	0.0264	*0.0264*	
Drop TTWA switchers	-0.004	0.001	25707
	(0.005)	(0.001)	
	0.0255	*0.0255*	
SIC4 dummies, not SIC2	-0.005	0.001	27794
	(0.005)	(0.001)	
	0.0428	*0.0428*	
Industry-year fixed effects	-0.005	0.001	27794
	(0.005)	(0.001)	
	0.0479	*0.0479*	
IPC1-year fixed effects	-0.006	0.001	27794
	(0.005)	(0.001)	
	0.0237	*0.0237*	
Industry-area clustering	-0.005	0.001	27794
	(0.005)	(0.001)	
	0.0428	*0.0428*	
London dummy, not area FE	-0.005	0.001	28173
	(0.005)	(0.001)	
	0.0341	*0.0341*	

Source: BSD / CH / GI / Orbis / UKIPO / UKIS. Each specification shows coefficient of *b* in equation (2), standard error in parenthesis and R^2 in italics. All models fit controls, area, year and SIC2 dummies. Controls fitted include 5-year lag of log turnover and mean employment, startup dummy, firm size dummies, company legal status and structure dummies, plus an urban TTWA dummy. Controls lagged one year except age. Standard errors clustered on 2-digit SIC. *** significant at 1%, ** significant at 5%, * significant at 10%. Constant not shown.

Table B15. Effect of innovation (launch dummy) on firm productivity, firms with events. Manufacturing versus services breakdowns.

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

Source: BSD / CH / GI / UKIPO / UKIS. Notes as in main paper.

Table B16. Effect of innovation (launch counts) on firm productivity, firms with events. Manufacturing versus services breakdowns.

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

Source: BSD / CH / GI / UKIPO / UKIS. Notes as in main paper.

Table B17. Young firms, innovative activity and productivity. Launch dummy. Firms with events.

	(1)	**(2)**	**(3)**	**(4)**
L.new product launch	0.064***	0.063***	0.064***	0.064***
	(0.019)	(0.019)	(0.019)	(0.019)
L2.15% depreciated PCT / EPO / US patent count	0.004	0.004	0.009	0.009
	(0.007)	(0.008)	(0.006)	(0.006)
L2.15% depreciated TM count	0.081***	0.079***	0.080***	0.080***
	(0.024)	(0.021)	(0.022)	(0.022)
Patent*young			-0.108*	
			(0.058)	
TM*young				-0.108*
				(0.058)
Micro firm		-0.084**	-0.085**	-0.085**
		(0.033)	(0.033)	(0.033)
Small firm		0.101***	0.102***	0.102***
		(0.031)	(0.031)	(0.031)
Young firm		-0.123***	-0.115***	-0.115***
		(0.027)	(0.027)	(0.027)
Ave pre-2009 patenting	0.072	0.063	0.053	0.053
	(0.057)	(0.058)	(0.057)	(0.057)
Firm patents pre-2009	-0.223*	-0.196*	-0.181	-0.181
	(0.116)	(0.118)	(0.118)	(0.118)
Observations	27019	26442	26442	26442
R^2	0.1663	0.1664	0.1668	0.1668

Source: BSD / CH / GI / UKIPO / UKIS. Column 1 fits the main specification. Column 2 refits with age and size group dummies. Columns 3 and 4 add interactions. IP interactions are lagged two periods. 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. Reference categories are older and medium-sized firms. All models fit controls and fixed effects. Other notes as in main paper.

Table B18. Young firms, innovative activity and productivity. Launch counts. Firms with events.

	(1)	(2)	(3)	(4)
L.total product launches	0.017***	0.017***	0.017***	0.017***
	(0.005)	(0.005)	(0.005)	(0.005)
L2.15% depreciated PCT / EPO / US patent count	0.004	0.004	0.008	0.008
	(0.007)	(0.008)	(0.006)	(0.006)
L2.15% depreciated TM count	0.081***	0.080***	0.080***	0.080***
	(0.024)	(0.022)	(0.022)	(0.022)
Patent*young			-0.109*	
			(0.058)	
TM*young				-0.109*
				(0.058)
Micro firm		-0.079**	-0.081**	-0.081**
		(0.033)	(0.033)	(0.033)
Small firm		0.103***	0.104***	0.104***
		(0.031)	(0.031)	(0.031)
Young firm		-0.121***	-0.113***	-0.113***
		(0.027)	(0.027)	(0.027)
Ave pre-2009 patenting	0.073	0.065	0.054	0.054
	(0.058)	(0.058)	(0.058)	(0.058)
Firm patents pre-2009	-0.220*	-0.193	-0.178	-0.178
	(0.116)	(0.118)	(0.118)	(0.118)
Observations	27019	26442	26442	26442
R^2	0.1672	0.1673	0.1678	0.1678

Source: BSD / CH / GI / UKIPO / UKIS. Column 1 fits the main specification. Column 2 refits with age and size group dummies. Columns 3 and 4 add interactions. IP interactions are lagged two periods. 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. Reference categories are older and medium-sized firms. All models fit controls and fixed effects. Other notes as in main paper.

Table B19. Micro and small firms, innovative activity and productivity. Launch dummy. Firms with events.

	(1)	**(2)**	**(3)**	**(4)**
L.new product launch	0.064***	0.063***	0.063***	0.063***
	(0.019)	(0.019)	(0.019)	(0.019)
L2.15% depreciated PCT / EPO / US patent count	0.004	0.004	0.012	0.012
	(0.007)	(0.008)	(0.008)	(0.008)
L2.15% depreciated TM count	0.081***	0.079***	0.080***	0.080***
	(0.024)	(0.021)	(0.022)	(0.022)
Patent*micro			0.002	
			(0.012)	
Patent*small			-0.030*	
			(0.017)	
TM*micro				0.002
				(0.012)
TM*small				-0.030*
				(0.017)
Micro firm		-0.084**	-0.081**	-0.081**
		(0.033)	(0.033)	(0.033)
Small firm		0.101***	0.107***	0.107***
		(0.031)	(0.031)	(0.031)
Young firm		-0.123***	-0.123***	-0.123***
		(0.027)	(0.027)	(0.027)
Ave pre-2009 patenting	0.072	0.063	0.081	0.081
	(0.057)	(0.058)	(0.060)	(0.060)
Firm patents pre-2009	-0.223*	-0.196*	-0.207*	-0.207*
	(0.116)	(0.118)	(0.118)	(0.118)
Observations	27019	26442	26442	26442
R^2	0.1663	0.1664	0.1666	0.1666

Source: BSD / CH / GI / UKIPO / UKIS. Column 1 fits the main specification. Column 2 refits with age and size group dummies. Columns 3 and 4 add interactions. IP interactions are lagged two periods. 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. Reference categories are older and medium-sized firms. All models fit controls and fixed effects. Other notes as in main paper.

Table B20. Micro and small firms, innovative activity and productivity. Launch counts. Firms with events.

	(1)	(2)	(3)	(4)
L.total product launches	0.017***	0.017***	0.017***	0.017***
	(0.005)	(0.005)	(0.005)	(0.005)
L2.15% depreciated PCT / EPO / US patent count	0.004	0.004	0.012	0.012
	(0.007)	(0.008)	(0.008)	(0.008)
L2.15% depreciated TM count	0.081***	0.080***	0.081***	0.081***
	(0.024)	(0.022)	(0.022)	(0.022)
Patent*micro			0.001	
			(0.013)	
Patent*small			-0.030*	
			(0.017)	
TM*micro				0.001
				(0.013)
TM*small				-0.030*
				(0.017)
Micro firm		-0.079**	-0.077**	-0.077**
		(0.033)	(0.033)	(0.033)
Small firm		0.103***	0.109***	0.109***
		(0.031)	(0.031)	(0.031)
Young firm		-0.121***	-0.121***	-0.121***
		(0.027)	(0.027)	(0.027)
Ave pre-2009 patenting	0.073	0.065	0.082	0.082
	(0.058)	(0.058)	(0.060)	(0.060)
Firm patents pre-2009	-0.220*	-0.193	-0.205*	-0.205*
	(0.116)	(0.118)	(0.119)	(0.119)
Observations	27019	26442	26442	26442
R^2	0.1672	0.1673	0.1675	0.1675

Source: BSD / CH / GI / UKIPO / UKIS. Column 1 fits the main specification. Column 2 refits with age and size group dummies. Columns 3 and 4 add interactions. IP interactions are lagged two periods. 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. Reference categories are older and medium-sized firms. All models fit controls and fixed effects. Other notes as in main paper.

Table B21. Three-way interactions, alternative specification. Events subsample.

	Dummy	**Counts**
Launch*young	-0.920	-0.483***
	(0.846)	(0.106)
Launch*old	0.286	0.157
	(0.482)	(0.147)
Launch*micro	-0.344	-0.153
	(0.483)	(0.147)
Launch*small	-0.126	-0.136
	(0.483)	(0.147)
Launch*medium	-0.007	-0.119
	(0.484)	(0.147)
Launch*micro*young	1.280*	0.649***
	(0.677)	(0.137)
Launch*small*young	1.066	0.665***
	(0.680)	(0.137)
Launch*medium*young	1.322*	0.622***
	(0.710)	(0.143)
Young*micro	0.849***	0.836***
	(0.252)	(0.236)
Young*small	1.004***	0.972***
	(0.254)	(0.238)
Young*medium	0.696**	0.754***
	(0.270)	(0.253)
Old*micro	0.968***	0.942***
	(0.251)	(0.235)
Old*small	1.109***	1.120***
	(0.251)	(0.235)
Old*medium	1.012***	1.037***
	(0.252)	(0.236)
Observations	26442	26442
R^2	0.1703	0.1706

Source: BSD / CH / GI / UKIPO / UKIS. All models fit lagged IP, levels effects, controls and fixed effects. 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. Other notes as in main paper.