Industrial and Corporate Change, 2025, **00**, 1–29
DOI: https://doi.org/10.1093/icc/dtaf004
Original Article

Overcoming the innovation threshold through innovative public procurement: evidence from CERN

Andrea Bastianin[1,2],Paolo Castelnovo[2,3] and Lorenzo Zirulia[1,*]

[1]Department of Economics, Management and Quantitative Methods, University of Milan, via Conservatorio 7, Milan 20122, Italy. e-mail: andrea.bastianin@unimi.it, [2]Fondazione Eni Enrico Mattei, Corso Magenta 63, Milan 20123, Italy. e-mail: paolo.castelnovo@uninsubria.it and [3]Department of Economics, University of Insubria, via Monte Generoso 71, Varese 21100, Italy. e-mail: lorenzo.zirulia@unimi.it

*Main author for Correspondence.

This paper contributes to the literature on the impact of large-scale research infrastructure projects on innovation. We use public procurement data to investigate the impact of CERN—the European Organization for Nuclear Research—on the likelihood of firms becoming innovators. Specifically, we assess the effect of CERN on the probability of a firm filing a patent application for the first time, and the timing of this effect. Using survival models and Propensity Score Matching to construct a counterfactual sample of firms, we show that qualifying for an industrial procurement contract with CERN increases the probability of filing a first patent. This effect emerges with a lag of 3–7 years from the start of the collaboration, suggesting a relatively slow process of the absorption of new ideas. We find heterogeneity in this effect: in fact, it occurs mostly within small firms, whose relationship with CERN involves more sophisticated technological problems and a higher frequency of interaction.

JEL Classification: C14; C23; C25; H57; O31

1. Introduction

Decades of active innovation policies in both advanced and developing countries have produced various policy instruments, whose evaluation poses major challenges (Bloom *et al.*, 2019). One instrument is public procurement for innovation (PPI), which is often used for large-scale research infrastructure projects (LSRIs). Two factors play a major role when deciding which LSRIs need to be funded: the scientific evidence about their discovery potential and the societal benefits they yield. Innovation spillovers to collaborating firms are one of the most relevant societal benefits generated by LSRIs (Bastianin *et al.*, 2023). For this reason, both scholars and policymakers have thoroughly analyzed the LSRI–industry nexus (Salter and Martin, 2001; Edler and Georghiou, 2007; Edquist and Zabala-Iturriagagoitia, 2012; OECD, 2014; Scarrà and Piccaluga, 2022).

In this paper, we assess innovation spillovers for firms involved in the construction of the Large Hadron Collider (LHC) at CERN.[1] In particular, we investigate whether PPI can help firms to overcome the so-called "innovation threshold" (Geroski *et al.*, 1997; Cefis, 2003; Cefis and Marsili, 2015). Innovating is hard for all firms, but it is especially hard for firms that have

---

[1] CERN—the European Organization for Nuclear Research—operates the largest particle collider in the world, the Large Hadron Collider or LHC. This machine is used to advance knowledge in the field of fundamental physics. See https://home.cern/science/accelerators/large-hadron-collider.

not innovated before, as it requires investment (e.g., in facilities and personnel) that is sunk and uncertain. Consequently, achieving an initial level of innovation—such as launching the first product or securing the first patent—is more challenging for a firm than subsequent innovations of comparable impact. We show that the beginning of a procurement relationship with CERN has a positive impact on the probability that a firm starts patenting.[2] This occurs with a delay of some years, pointing to a relatively slow process for the absorption of new ideas. Finally, the positive effect is mainly due to the impact on small firms—for whom it is presumably more difficult to overcome the innovation threshold—and on firms whose relationship with CERN involves more sophisticated technological problems and a higher frequency of interaction, thus guaranteeing more opportunities for technological learning.

Our paper contributes to the literature on the impact of PPI on innovation, focusing on new innovators. While most analyses rely on qualitative methods such as surveys and case studies (e.g., Autio *et al.*, 2004), few works have estimated the increase in the number of patent applications filed by firms collaborating in relation to LSRIs (Castelnovo *et al.*, 2018; Bastianin *et al.*, 2021), as we do. We use two different datasets to investigate whether becoming a supplier for CERN's LHC is associated with an increase in the probability of first-time patenting. To establish a directional association from CERN procurement to patenting activity, we first extend the dataset collected by Bastianin *et al.* (2021) and Castelnovo *et al.* (2018), which contains information on firms that were awarded a contract for the construction of the LHC. Second, we employ Propensity Score Matching (PSM) to create ex novo a control group of firms that have not supplied LHC orders but possess comparable characteristics to the treated group.

Our analysis provides evidence of the impact of CERN procurement on supplier innovation. Using Latent Dirichlet Allocation (LDA) on 19 200 patent titles, we find a strong alignment between patent topics and CERN activity codes, suggesting an association between CERN orders and subsequent patents filed by suppliers. Cox proportional hazard models reveal a directional relationship, with CERN procurement increasing the likelihood of first-time patenting by LHC suppliers after a lag of at least 3 years. The PSM further supports these findings, showing that LHC suppliers are more likely to patent compared to similar firms without CERN contracts, indicating that CERN procurement influences suppliers' innovation activity.

The remainder of the paper is organized as follows. In Section 2, we briefly survey the relevant literature. Section 3 presents the data, while Section 4 illustrates the empirical strategy. Section 5 discusses the results and Section 6 concludes. An Appendix completes the paper.

2. Related literature

2.1. Public procurement for innovation

Public procurement plays a crucial role in promoting innovation and economic growth, by fostering partnerships between the public and private sectors. In recent years, there has been a growing interest in how PPI influences the innovation performance of firms (Florio *et al.*, 2018). PPI differs from regular procurement: the latter involves the purchase of standard products and services that do not require research and development activities, while PPI refers to products or services that are not yet available, thus requiring an innovative effort to deliver what is expected (Edquist *et al.*, 2000; Salter and Martin, 2001).

By offering contracts for innovative solutions, governments effectively incentivize firms to allocate more resources to R&D, leading to the creation of novel products and technologies (Raiteri, 2018). In that respect, PPI constitutes an important form of demand-side innovation policy (Edler and Georghiou, 2007; Uyarra and Flanagan, 2010; Edquist and Zabala-Iturriagagoitia, 2012; Edquist *et al.*, 2015) and is often more effective than other policies, such as R&D subsidies and tax credits (Lichtenberg, 1988; Aschhoff and Sofka, 2009; Guerzoni and Raiteri, 2015).

The literature has postulated, and then empirically looked for several benefits that can accrue to firms involved in public procurement contracts such as those initiated by LSRIs. These benefits can be grouped into two broad categories. The first of these is associated with a "demand pull" effect: the firm can improve its opportunities for innovation because of the market for its

---

[2] See Section 4 for a discussion of possible concerns about using patents as a proxy for innovation in our setting.

products generated by the research infrastructure and by additional customers acquired through signalling and reputation building. The second category, on the other hand, is associated with a "knowledge" effect arising from the technological learning that comes from the interaction between the firm and the research infrastructure.

As for the "demand pull" effect, participating in PPI provides firms with an opportunity to showcase their innovative products and services to a broader market. In addition, successful contracts can act as a valuable reference, attracting new clients and investment, and ultimately aiding in the growth of the firm. Finally, PPI can mitigate the risks associated with innovation investment by providing a stable and predictable market for innovative products and services: firms are more likely to take risks and invest in breakthrough technologies when they have a secured public buyer (Malerba *et al*., 2007; Mazzucato, 2016).

As for the "knowledge" effect, the literature has suggested that the technological learning occurring in LSRI–business collaborations materializes in a user–producer interaction (Lundvall, 1992), as the LSRI provides access not only to financial resources and reputational capital, but also to scientific and technological capabilities, which are inputs in the process of knowledge creation and technological capital accumulation by firms (Hameri, 1996). Rather than being mere market-based relations, public procurement contracts often entail a deeper form of cooperation: "Not only demand as such, but also the interaction between demand and supply has crucial implications for innovation dynamics" (Edler and Georghiou, 2007, 949).

Unsurprisingly, given its relevance, the case of CERN has already been investigated to assess, such potential benefits empirically, as the instruments demanded by CERN and the technological problems it faces have a degree of complexity that usually require innovative solutions. Autio *et al*. (2003), through a survey of CERN suppliers, found that collaborations with CERN are associated with both a "demand pull" and a "knowledge" effect. While almost all CERN's suppliers appear to derive great value from the collaboration in terms of marketing reference, it has been found that an increase in product innovation, the initiation of new R&D projects and technological learning are more concentrated among such suppliers, and are positively associated with the frequency of interaction and the relational social capital that characterizes the relationship between the supplier and CERN. Florio *et al*. (2018) also used survey data and confirmed that cooperative relations between CERN and its suppliers are key for innovation performance; the importance of interaction is also confirmed by the interviews carried out by Aberg and Bengtson (2015). As for studies measuring innovation through patent data, the econometric evidence found by Castelnovo *et al*. (2018) and Bastianin *et al*. (2021) showed that, after becoming CERN industrial partners, firms generally experienced a rise in innovation-related variables such as R&D, patents, and productivity.[3]

2.2. Overcoming the "innovation threshold"

New innovators play a pivotal role in shaping the evolution of technologies and industries. The introduction of novel ideas, processes, and products by these pioneering firms can lead to disruptive innovations and paradigm shifts, thereby altering the competitive landscape and driving economic growth. If we draw on Joseph Schumpeter's concept of creative destruction, new innovators embody the spirit of entrepreneurship and disruption, as they can displace established players, leading to a reconfiguration of industry dynamics. This process of creative destruction allows for the replacement of obsolete technologies with more efficient and productive ones, fostering economic progress.[4]

---

[3] Beyond the CERN cases, Fernandes *et al*. (2014) found technological benefits for firms involved in the European Southern Observatory procurement and Castelnovo and Dal Molin (2021) found the same for 150 suppliers to the Italian Institute of Nuclear Physics.

[4] The relative role of new innovators (vis-à-vis established ones) constitutes one of the main themes in the literature on technological regimes and the pattern of innovation (Malerba and Orsenigo, 1995, 1997; Breschi *et al*., 2000). Malerba and Orsenigo (1999) consider European Patent Office data for the period 1979–1991 for six countries (Germany, France, the UK, Italy, the USA, and Japan). They show that firms innovating for the first time in a given year are a significant fraction of all inventors (on average, higher than 40%), although they account for a lower share in terms of patents. Differences across sectors are more relevant than those across countries.

However, what do we know about the innovation process in firms that have not yet been innovative? The literature has identified the existence of an "innovation threshold" (Geroski *et al.*, 1997; Cefis and Orsenigo, 2001; Cefis, 2003; Cefis and Marsili, 2015). The results on this have been obtained within the broader literature investigating the extent to which firms innovate persistently (that is, innovate continuously after their first innovation).[5] In a nutshell, moving away from the status of a non-innovative firm is difficult, and is more difficult than continuing to innovate after making an initial start. This may be because of the sunk cost nature of R&D investment, which thus represents a barrier to entry for non-innovating firms On the other hand, once the status of an innovative firm has been reached, past innovation activities can mitigate the financial constraints often associated with R&D investments (Flaig and Stadler, 1994; Czarnitzki and Hottenrott, 2010), particularly because previous innovation success allows the firm to reinvest profits into R&D (Nelson and Winter, 1982). In addition, learning by doing and dynamic economies of scale in R&D may kick in (Arrow, 1962), as the generation of new knowledge, on which innovation builds, is usually a cumulative process (Rosenberg, 1976; Nelson and Winter, 1982; Weitzman, 1998).

Using patent application data, Cefis and Orsenigo (2001) followed a transition probability matrix approach to show that non-innovators (and great innovators) have a high probability of remaining in the same state, so that, for these firms, innovation is persistent. Related to this, Cefis (2003) showed that the probability of moving from zero to one patent application is much lower than the probability of moving from n to $n+1$ applications. A similar result was found by Geroski *et al.* (1997), who used both patent and innovation data for the UK. More recently, Cefis and Marsili (2015) showed, using Community Innovation Survey data, that mergers and acquisitions increase the probability of the transition from being a non-innovator to being an active innovator.[6]

3. Data

In this section, we describe the two datasets used in the empirical analysis. The first consists of firms that received contracts to build the LHC, which represents our treated group. The second dataset was constructed using PSM to identify a control group of firms by matching LHC suppliers to untreated firms with comparable observable characteristics.

3.1. Treated group (LHC suppliers)

A database maintained by the CERN Procurement and Industrial Services Group was used to identify the orders placed with firms involved in the construction of the LHC over the period 1995–2008. Starting from this database, we identified 1296 suppliers that received at least one LHC-related order for above 10,000 CHF and retrieved information on the date marking the beginning of the procurement relationship, the "activity codes" used by CERN to classify purchases from its suppliers, and the number and the total value of orders supplied by the firm. Next, we used the Orbis database, maintained by Bureau van Dijk, to obtain the firms' balance sheet data, including their geographical location, size, incorporation date, sector of activity (based on the NACE 2-digit codes), and intangible fixed assets. Lastly, we retrieved the firms' patent data from the Orbis Intellectual Property database. The process of merging data from these three sources left us with a panel dataset of 896 firms. Since our goal was to assess whether the procurement collaboration with CERN encouraged firms to file a patent for the first time, we excluded firms that had filed a patent before 1995. Our final sample included 741 firms. Before 1995, none of these companies had received an LHC-related order. Among them, 158 (21.4%) filed their first patent after the beginning of their collaboration with CERN. The first part of

---

[5] Examples of contributions in this stream of literature are the works by Arroyabe and Schumann (2022), Bartoloni (2012), Duguet and Monjon (2004) Fontana and Vezzulli (2016), Manez *et al.* (2015), Peters (2009), Raymond *et al.* (2010), and Roper and Hewitt-Dundas (2008).

[6] Hall *et al.* (2021) considered barriers to entry into patenting and showed, in particular, how these can be high in the presence of patent thickets, i.e., when technologies are protected by hundreds or even thousands of patents.

our analysis exploits the time variation of the firms' status—from "not-yet-supplier" to "supplier"—to establish how CERN procurement affected the firms' likelihood of filing a patent for the first time.

3.2. Control group

We constructed a control group of firms using PSM to perform a counterfactual analysis that allows us to provide a causal interpretation of our results. To estimate the probability of being treated, given a set of observable characteristics (i.e., the propensity score), we first selected a large sample of untreated firms to match the treated ones. To this end, we identified all active firms in the Orbis database with the same 4-digit NACE Rev. 2 codes as the CERN suppliers, but restricted them to those located in the USA or Japan, which are not CERN member states. There are two reasons for this choice. First, the USA and Japan are among the most innovative countries in the world (see e.g., the 2023 Global Innovation Index ranking), and are therefore likely to host highly innovative companies. Second, CERN allocates procurement contracts to entities within its member states, meaning that all contracts in our database were allocated to European countries. Thus, geographical location should provide an exogenous treatment assignment between treated (suppliers) and untreated (non-suppliers). Because of computational constraints, we randomly selected a subsample of 100,000 units from the reference population of approximately 6.882 million US and Japanese firms.

For the matching, we used a k-nearest neighbour (NN) matching algorithm with $k = 3$. Given the categorical nature of the variable capturing the industrial sector (identified by the NACE code), we imposed exact matching on sectors to make the matching procedure more reliable. Since the treated firms became CERN suppliers in different years, we replicated the matching procedure year by year. That is, for each treatment year (i.e., the year in which a contract was awarded), we selected from the treated group only those firms that became CERN suppliers in that specific year, and matched these with firms from the untreated group. As we are interested in understanding whether becoming a supplier to CERN encourages a firm that had never filed a patent prior to the procurement relationship to start patenting, we dropped from the control sample firms with patents filed before the year of the first order received by the corresponding treated counterpart. Control firms were assigned the treatment year of their corresponding treated counterparts. This approach allowed us to construct a dummy variable that also identifies the pre- and post-treatment periods for the control group firms.

By implementing this matching procedure, we obtained a final dataset consisting of 640 treated firms and 2968 untreated counterparts (see Section 5.3 for details on the matching procedure). As a robustness check of the validity of our matching strategy, the matching algorithm was also run with $k = 5$. In this case, the number of untreated firms increased to 6263.

4. Empirical models

The empirical analysis is divided into three main steps. First, we use a comprehensive database of textual information to show that the content of the LHC orders and the titles of the patents filed after collaboration with CERN largely overlap. Details of the methods used to implement the textual analysis are provided in Appendix B. Second, we exploit the database with detailed information on the LHC suppliers and orders to assess the association between the status of the CERN contractor and the probability of filing a first patent. Finally, using our original database obtained by matching LHC suppliers with control firms, we perform a counterfactual analysis that allows us to move from results that show the directionality of the association between CERN and first patenting to results that speak to causality. This section discusses the methods underlying the second and third stages of the analysis and looks at the use of patents as a proxy for measuring innovation.

4.1. CERN procurement and the hazard of filing a patent

We study the impact and timing of CERN procurement on suppliers' hazard of filing a patent for the first time using a Cox Proportional Hazard (Cox-PH) model (see Cox, 1972). The Cox-PH model specifies the hazard of patenting as:

$$\lambda\left(t|\mathbf{X}_{isc,t}\right) = \lambda_0\left(t\right) exp\left(\gamma_k CERN^k_{i,t} + \mathbf{X}_{i,t}\beta + \sum_s \alpha_s D_s + \sum_c \alpha_c D_c\right) \tag{1}$$

The hazard rate, λ(.), is the instantaneous probability that firm *i* starts patenting in year *t* for the first time, provided that it has not yet filed a patent by that date. The baseline hazard, $\lambda_0(t)$, is a function of time alone. The exponential function ensures that the hazard rate is non-negative and acts as a scale factor that makes $\lambda_0(t)$ proportional to the vector of the covariates $\mathbf{X}_{isc,\,t}$.

The dummy variable $CERN^k_{i,t}$ is set to one if the procurement relationship with CERN started *k* years ago. In practice, we consider 10 different models, for $k = 0, 1, \ldots, 8$ and $k \geq 9$, to identify the timing of the effect. Unobservable country- and sector-specific heterogeneity is captured by the dummy variables D_c and D_s.

The vector $\mathbf{X}_{i,\,t}$ contains firm-level control variables. Firm size is captured by dichotomous variables constructed starting from the "size" variable provided in the Orbis database, which classifies companies as very large, large, medium, and small, based on their number of employees, total assets, and operating revenue. A dummy variable ($Hi\text{-}tech_i$) exploits the technological intensity classification of orders developed by experts at CERN to define firms as high- or low-tech. High-tech firms are those that have supplied orders with customization and co-design involving CERN staff. Lastly, the (logarithm of the) total order value received by firm *i* ($Order_i$) is built as the sum of the monetary value of all the LHC-related orders received by the supplier during its collaboration with CERN. This proxies the involvement and continuity of the procurement relationship with CERN. See Appendix A for further details.

Counterfactual impact assessment. In the third part of the empirical analysis, we rely on a sample made of LHC suppliers and a control group of firms with comparable characteristics constructed with PSM. To implement a difference-in-differences analysis, the empirical specification in model (1) needs to be modified as follows[7]:

$$\begin{aligned}\lambda\left(t|\mathbf{X}_{isrc,t}\right) = \lambda_0\left(t\right) exp\left(\gamma\ supplier_i + \delta\ post_{it} + \rho\ supplier_i \times post_{it}\right.\\ \left. + \theta pat_r + \mathbf{Z}_{i,t}\beta + \sum_s \alpha_s D_s + \sum_c \alpha_c D_c\right)\end{aligned} \tag{2}$$

where "$supplier_i$" takes the value one for LHC suppliers and zero for firms in the control group, while "$post_{it}$" is a dummy variable that takes the value one in the years following the start of the procurement relationship and is zero otherwise. As explained in Section 5.3, the matching algorithm uses data at the regional level, and we therefore also include the number of regional patents per inhabitant, pat_r, where the subscript *r* indicates the region where firm *i* is located. The vector $\mathbf{Z}_{i,\,t}$ contains firm-level control variables, including "$size_i$", the dummy variable $EUHT_{is}$, which takes the value one for firms in sectors classified by Eurostat as either "high-tech knowledge-intensive services" or "high-tech manufacturing",[8] and the firm's Age_i, built as the difference between the year in which collaboration with CERN began and the firm's incorporation year. Note that it is no longer possible to rely on the variable $Hi\text{-}tech_i$ to define firms as high- or low-tech, as we did in model (1), since this variable is not available for firms in the control group. For the same reason, it is not possible to include the variable $Order_i$ in model (2). The estimate of the coefficient *ρ*—associated with the interaction between "$supplier_i$" and "$post_{it}$"—is the main object of interest in this specification; in fact, it captures the effect of procurement on the hazard that a firm starts filing patents after its collaboration with CERN begins.

4.2. On the use of patents as a proxy for innovation

Both equations (1) and (2) rely on patents as a proxy for innovation. Patent data have advantages and disadvantages compared to other proxies for innovation. The advantages are well known. Patents provide objective data that avoid the biases associated with innovation surveys, such

---

[7] See e.g., Clotfelter *et al.* (2008) for an example of the use of the difference-in-differences framework within hazard models.

[8] See https://ec.europa.eu/eurostat/Knowledge-intensive_services_(KIS) and https://ec.europa.eu/eurostat/High-tech_classification_of_manufacturing_industries

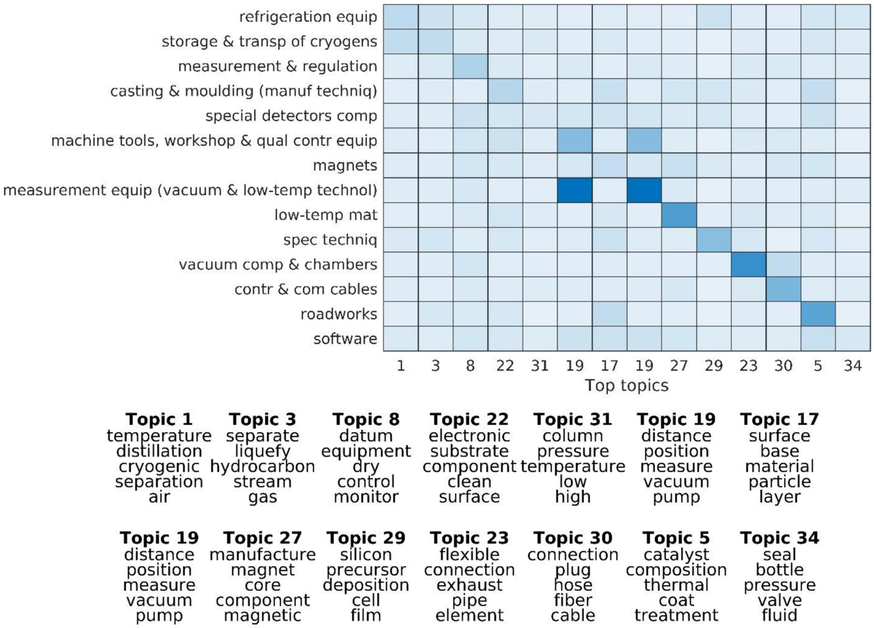


Figure 1. CERN activity codes and the content of patents.

as the low compliance rate and the fact that responding firms are often the most efficient. More importantly, and unlike most surveys, patent data are available for a long period of time, allowing the construction of a panel dataset with a significant time dimension (Griliches, 1990). In our setting, this allows us to study the dynamics of the procurement effect on firms' outcomes.

Nevertheless, as the existing literature acknowledges, patents are an imperfect proxy for firms' innovation. In particular, patents may under-represent innovation if firms choose not to patent their inventions but instead to protect them by other means. This may be an issue for our exercise if a firm has an (unpatented) innovation prior to the start of its relationship with CERN (which may even be the reason why the firm and CERN start collaborating), and this innovation is then converted into a patent after the start of the collaboration. We believe that this possibility is unlikely for two reasons. First, CERN's policy is that any intellectual property generated in the course of a procurement contract is to be owned exclusively by CERN. For a company to enter into a relationship with CERN with an invention that is ready for use but unpatented, the details of which may be disclosed in the interaction, seems a risky strategy. Second, contracts between CERN and suppliers often require technological solutions that are specific to CERN's problems, especially (by definition) for those collaborations involving companies that we classify as "high-tech". It follows that it is unlikely that inventions ready for patenting but not yet patented will exist from the outset, especially for collaborations with "high-tech" firms.

5. Results

5.1. Are patents related to CERN orders?

In this subsection, we provide evidence of the association between patents filed by LHC suppliers and CERN orders, relying on textual data analysis. With this aim, the codes assigned by CERN to its LHC contracts are exploited to assess how similar the orders are to the content of the patents subsequently filed by the firms.[9] Topic modeling—also known as Latent Dirichlet Allocation

---

[9] This analysis builds upon and extends the findings of Bastianin *et al.* (2021), who utilized the same dataset but employed a simpler approach based on measures of association for the categorical data to validate the coherence between WIPO codes and CERN activity codes.

(LDA)—is employed to identify the topics that recur most in the titles of 19,200 patents (Blei *et al.*, 2003). Figure 1 displays the *predicted topic probability*—defined as the probability that a topic appears in the title of a patent—for each of the 14 most common CERN activity codes. Darker shades indicate a higher probability. The lower panel displays the five most probable words in each topic, presented as word clouds. It can be observed that in some cases a single topic is highly likely for more than one CERN activity code. This is exemplified by Topic 19, which is associated with two CERN activity codes: "measurement" and "quality control equipment".

All in all, the topics identified with LDA contain words related to orders with the relevant CERN activity code, providing evidence in favor of a sample association between CERN orders and patent content. Appendix B provides further details on the implementation of the LDA analysis, as well as additional textual analysis (based on word clouds) to support and strengthen the validity of our findings.

5.2. CERN procurement and the hazard of patenting

We estimate Cox-PH models to assess the hazard for LHC contractors of beginning to patent, and the timing of such an event. Working with a sample consisting entirely of CERN suppliers has some advantages, particularly in terms of the depth of the empirical analyses, but it also limits the conclusions that can be drawn from them. The complexity and lack of standardization in many LHC contracts present challenges in establishing a control group, particularly due to the often-unique characteristics of CERN suppliers. Furthermore, our database includes comprehensive details regarding orders, such as their value and their classification by CERN experts as either high- or low-tech. These factors serve as valuable control variables in empirical analyses and cannot be replicated for the control group. In addition, our dataset provides the date of the start of the procurement relationship, which is crucial for assessing the timing of the CERN effect on the likelihood of patenting. Such information is not available for the firms in the control group, thus preventing us from learning anything about the time lag between the start of the procurement relationship and the patent filing. For these reasons, in this subsection, we exploit the information in the database of LHC suppliers and focus on establishing a directional association running from CERN procurement to patenting activity.

Table 1 shows 12 different specifications of the Cox-PH model. We start the analysis by focusing on columns (3)–(12) where we report estimates of the coefficients of "$CERN_{i,t}^{k}$" for $k = 0$, 1, …, 8 and $k \geq 9$. The coefficient of "$CERN_{i,t}^{0}$" measures the change in the hazard rate in the year of the first LHC-related order. Similarly, "$CERN_{i,t}^{1}$" captures the variation of the hazard rate one year after the first order, and so on for each k. The estimated coefficients of "$CERN_{i,t}^{k}$" are positive and monotonically increasing up to $k = 6$, and they start decreasing thereafter. We can also see that for $k = 3, \ldots, 7$ the coefficients are statistically distinguishable from zero.

Since we want to be sure that the change in the status of a firm from non-patenting to patenting post-dates the start of its collaboration with CERN, the specifications in columns (1) and (2) include leads of the dummy variable marking the start of the procurement relationship. More precisely, "$CERN_{i,t}^{-1}$" and "$CERN_{i,t}^{-2}$" indicate that firm i will receive its first order from CERN in the next one or two years, respectively. Columns (2) and (1) of Table 1 show that the coefficients for such lead variables are statistically indistinguishable from zero. The inclusion of lead variables can be interpreted as a "placebo" test confirming the expectation that any effect of LHC procurement on a firm's probability of starting patenting should occur after the beginning of the relationship with CERN.[10]

The estimates of the hazard ratios (HR) and their 90% confidence intervals associated with the models in Table 1 are presented in Figure 2. An estimated HR equal to one indicates a lack of association, while an estimate greater than one suggests that CERN industrial partners face a higher "risk" of filing their first patent compared to "not-yet-suppliers". As shown in Figure 2, the 90% confidence bands for "$CERN_{i,t}^{k}$" lie above one only for $3 \leq k \leq 7$.

---

[10] The effect of the size variable is positive and statistically significant. The Wald tests of the null hypothesis that country and sector fixed effects are jointly equal to zero suggest rejecting the null for both sets of controls. All the remaining control variables are never statistically distinguishable from zero.

Table 1. Cox proportional hazard models

	(1) $k=-2$	(2) $k=-1$	(3) $k=0$	(4) $k=1$	(5) $k=2$	(6) $k=3$	(7) $k=4$	(8) $k=5$	(9) $k=6$	(10) $k=7$	(11) $k=8$	(12) $k=9$
$CERN_{i,t}^{k}$	0.020	0.024	0.033	0.043	0.047	0.054^{**}	0.066^{***}	0.080^{***}	0.087^{***}	0.069^{*}	−0.014	0.026
	(0.037)	(0.043)	(0.047)	(0.037)	(0.029)	(0.022)	(0.023)	(0.021)	(0.022)	(0.040)	(0.059)	(0.072)
Hi-Tech	0.132	0.131	0.130	0.128	0.130	0.128	0.128	0.132	0.139	0.143	0.137	0.139
	(0.176)	(0.174)	(0.175)	(0.174)	(0.173)	(0.174)	(0.173)	(0.173)	(0.173)	(0.174)	(0.174)	(0.174)
Order	−0.001	−0.001	−0.006	−0.011	−0.011	−0.014	−0.017	−0.021	−0.018	−0.009	0.006	0.003
	(0.067)	(0.069)	(0.071)	(0.070)	(0.068)	(0.065)	(0.066)	(0.065)	(0.065)	(0.061)	(0.062)	(0.062)
Medium	0.517^{*}	0.516^{*}	0.515^{*}	0.516^{*}	0.517^{*}	0.522^{*}	0.529^{*}	0.532^{*}	0.529^{*}	0.523^{*}	0.519^{*}	0.520^{*}
	(0.275)	(0.275)	(0.273)	(0.273)	(0.273)	(0.272)	(0.272)	(0.275)	(0.275)	(0.275)	(0.280)	(0.279)
Large	1.165^{***}	1.166^{***}	1.160^{***}	1.158^{***}	1.158^{***}	1.163^{***}	1.169^{***}	1.173^{***}	1.165^{***}	1.160^{***}	1.169^{***}	1.165^{***}
	(0.402)	(0.404)	(0.396)	(0.393)	(0.394)	(0.394)	(0.395)	(0.398)	(0.402)	(0.407)	(0.409)	(0.408)
V. Large	2.067^{***}	2.065^{***}	2.069^{***}	2.072^{***}	2.078^{***}	2.088^{***}	2.106^{***}	2.120^{***}	2.114^{***}	2.089^{***}	2.064^{***}	2.067^{***}
	(0.371)	(0.370)	(0.368)	(0.362)	(0.361)	(0.358)	(0.359)	(0.357)	(0.360)	(0.363)	(0.378)	(0.376)

Sample: 741 firms Country (D_c) and sector (D_s) fixed effects are included in all the specifications.
$^{*}P < 0.10$.
$^{**}P < 0.05$.
$^{***}P < 0.01$.
Cluster robust standard errors (i.e., cluster is the 2-digit NACE code) in parentheses. The table shows estimates of coefficients, not hazard ratios. The dummy variable “$Small_i$” is excluded, as small firms are used as a reference category.

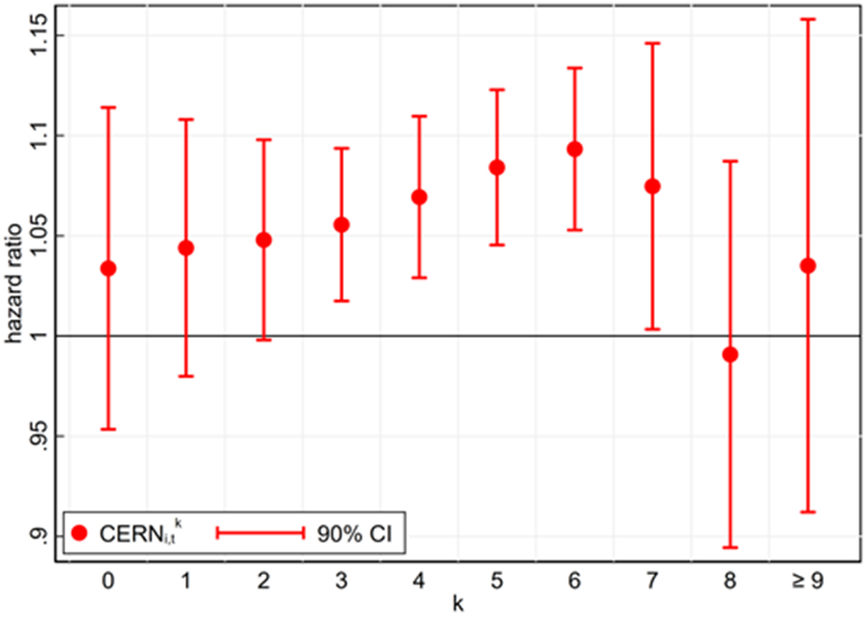


Figure 2. Estimates of the hazard ratio for LHC suppliers. Notes: Estimates of the hazard ratio (HR) associated with models in Table 1 and 90% confidence interval.

Table 2. Sectoral and size distribution of treated and control firms after matching

Sectoral distribution

NACE code (2 digits)	Treated (%)	Controls (%)
Wholesale trade, except of motor vehicles and motorcycles	24.77	21.75
Manuf. of computer, electronic and optical products	12.93	13.62
Manuf. of fabricated metal products, except machinery and equipment	9.35	10.86
Manuf. of machinery and equipment	8.57	9.80
Manuf. of electrical equipment	7.48	9.23
Specialized construction activity	5.92	7.28
Architectural and engineering activity; technical testing and analysis	3.89	3.20
Computer programming, consultancy and related activities	2.49	2.77
Manuf. of basic metals	2.34	2.60
Scientific research and development	0.93	0.56
Others	21.33	18.33

Size distribution

Size	Treated (%)	Controls (%)
Small	25.93	23.01
Medium	36.41	39.19
Large	25.47	24.10
Very large	12.19	13.70

This implies that the effect of LHC procurement on the HR for filing a patent for the first time takes time to build and shows up with a lag of at least 3 years.[11] Overall, these results

---

[11] The time span that separates the onset of the procurement relationship to the filing date of a patent provides useful information for computing the rates of return on research (Griliches, 1979; Pakes and Shankerman, 1984). Earlier analyses show that, while the lead from R&D expenditure to patent applications is a matter of a year (see e.g., Pakes and Griliches, 1980; Hall *et al.*, 1984; Hausman *et al.*, 1984), the time span that separates academic research from the introduction of new products in the market or the development of new processes is much longer and is in the range of 6–20 years (see (Mansfield, 1991), (Mansfield, 1998); Sternitzke, 2010; Toole, 2012). We note that none of these articles focuses on LSRI.

provide evidence for the existence of a directional association running from CERN procurement to the start of patenting activity by LHC suppliers. A wide array of robustness checks presented in Appendix C confirms and complements our main findings.[12]

The role of technological intensity, firm size, and order value. We investigate possible heterogeneity in the impact of CERN procurement on firms' hazard of filing a patent by dividing the sample along three dimensions: the firm's technological classification (i.e., hi- or low-tech), the firm's size, and the value of the order. To save space, the tables with the estimation results are reported in Appendix D.

Table D1 presents the results obtained by dividing the sample into two subsamples: high-tech firms (60% of the sample) and low-tech firms (40%). Inspection of Table D1 shows that the results in the full sample are driven by the effect for high-tech firms. This is in line with our expectations, since it is public procurement from high-tech firms (as we classify them) that is more likely to require technological advances to fulfil the contract requirements.[13]

Table D3 reports the regressions obtained when splitting the sample into small and medium firms versus large and very large firms, and shows that results in the full sample are driven by the impact on small and medium firms. This is notable, from a policy perspective, since it shows that PPI for LSRI can be particularly effective in triggering patenting exactly for those firms (the relatively small ones) for whom patenting propensity is generally low (Cohen *et al.*, 2000).

Finally, we take into account the heterogeneity in the size of the CERN orders. To do this, we split the sample of suppliers into firms that received orders above the median and those that received orders below the median of the order size distribution. Table D4 shows that the CERN procurement effect also exists for small contracts. This result is relevant because it suggests that the link between CERN procurement and the hazard of filing a patent is not due just to a "budget effect" that would allow suppliers to spend some fixed R&D costs that would otherwise be unaffordable.

5.3. The impact of CERN procurement on the hazard of patenting: quasi-experimental perspective

Working with a sample consisting solely of LHC suppliers does not allow us to make causal statements about the relationship between PPI and patenting. First, all the entities in our sample are eventually treated—that is, they get a contract from CERN. Second, assignment to treatment might be endogenously determined. In fact, the probability of gaining an LHC contract might be positively correlated with a firm's technological level, with variables capturing CERN's procurement rules (e.g., the existence of a "fair return" with respect to the member states' contributions), with the presence of a home premium towards Swiss and French contractors (Bastianin and Del Bo, 2021), or with a preference for firms that have previously worked with CERN.

To address the possible sample selection problem that could affect model (1), we used PSM to construct the counterfactual control group of firms that are not CERN suppliers and whose characteristics in the pretreatment period are similar to those of the LHC suppliers (treated group). This approach allows any differences in the hazard of filing a patent between the two groups in the post-treatment period to be attributed to the treatment with a higher degree of confidence. While this part of the analysis is relevant to address sample selection issues and should provide a causal interpretation of the association between CERN procurement and firms' hazard of filing a patent, we note that data limitations may affect the quality of the matching. In fact, only a few

---

[12] In addition, we looked at the probability that a firm files a second patent, provided that it has previously filed a patent after the beginning of CERN collaboration, as a rough measure of innovation persistence. We found that firms classified as high-tech and large firms have a higher probability of becoming "persistent innovators". The results are available upon request.

[13] Differences in patent propensity across sectors may raise concerns about these results. As discussed in Section 4, the high-tech dummy variable mostly captures technological intensity and the degree of product customization involved in the collaboration between CERN and the focal firm, rather than firm sectoral classification. Table D2 shows that software and other service firms are not predominant within the group classified as low-tech (in fact, there is a sector—manufacturing of computer, electronic, and optical products—which is the sector with the most representatives in the low-tech group and ranks second in the high-tech group). In addition, the manufacturing sectors to which CERN suppliers belong are among those for which patent propensity is relatively significant (see Cohen *et al.*, 2000—Table A1, page 50). Therefore, it seems unlikely that the result we obtain is an artefact of firm classification.

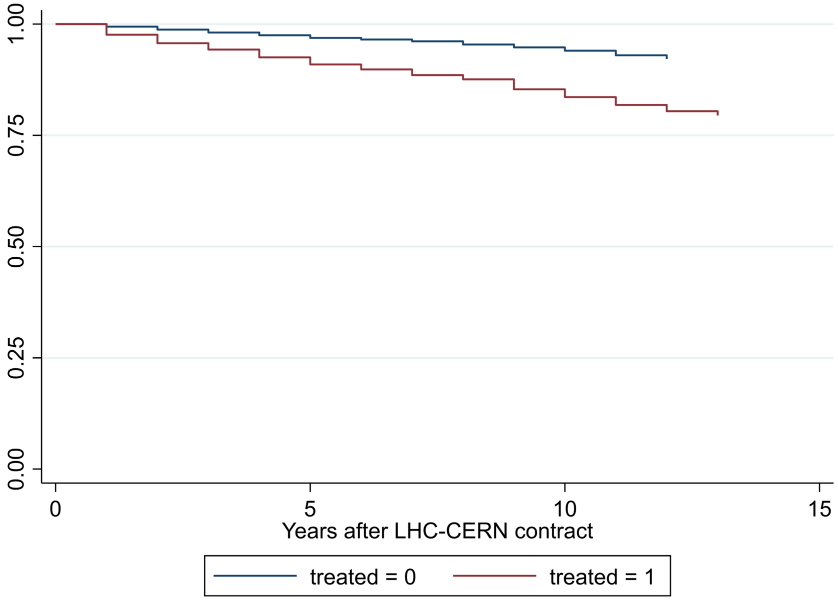


Figure 3. Estimated survival function. Notes: Unconditional survival rates for firms in the treated (treated = 1) and control group (treated = 0).

variables are available to carry out the matching between the two groups. Note that the PSM is based on the observable characteristics of a firm in the period before CERN awarded it a contract to build the LHC, which was between 1995 and 2008. The further back in time one goes, the fewer the observations that are available in the Orbis database because of missing values.

The second reason is that financial statements for the US and Japanese companies are not in the Orbis database because they are not required to publish them.

Our matching algorithm relies on both firm-level and region-level information. For the firm-level characteristics, we selected the sector, the year of incorporation and the size of the firm. Since the number of firm-level variables is small and does not include a measure of innovation capacity, we complemented this information with data for the region where the firm is located. To this end, we consider the number of regional patents per capita to control for regional innovation capabilities[14] (see e.g., Bettencourt *et al.*, 2007; Mewes and Broekel, 2022).

Data on the innovation capability of regions were sourced from the OECD Patent Statistics database using the Territorial Level 2 (TL2) geographical classification; this corresponds to NUTS 2 regions for European countries, states for the USA, and prefectures for Japan.[15]

Table 2 show the sectoral and size distribution of firms in the treated and untreated groups after the matching (for $k = 3$). Moreover, the t-test performed on the mean value of firms' age rejects the null hypothesis that there are statistically significant differences between the treated and untreated groups after the matching is implemented (P-value < 0.050).

As the first step of our counterfactual analysis, we estimate the Kaplan–Meier survival function for the treated group and the control group. Figure 3 plots the unconditional survival rates of firms in the treatment and control groups, showing on the vertical axis the proportion of firms that have not yet filed their first patent in each year following the award of the first contract from CERN. It can be noted that the survival rates of firms in the control group are higher in each period, meaning that LHC suppliers are more likely to start patenting.

Table 3 illustrates the results for the difference-in-differences estimator conducted on the samples obtained through the application of the NN matching algorithm with k equal to 3 and k

---

[14] We thank an anonymous referee for this suggestion.

[15] Source: https://doi.org/10.1787/patent-data-en

Table 3. Counterfactual impact assessment

	(1) $k = 3$	(2) $k = 5$	(3) $k = 3$	(4) $k = 5$
$Supplier^{*}Post$	0.439^{***}	0.610^{*}	0.464^{**}	0.620^{**}
	(0.208)	(0.340)	(0.227)	(0.306)
Supplier	0.881^{***}	0.997^{***}	1.450^{***}	1.891^{***}
	(0.137)	(0.255)	(0.534)	(0.351)
Post	-0.285^{*}	-0.338^{**}	-0.368^{*}	-0.362^{**}
	(0.173)	(0.150)	(0.199)	(0.156)
HT_EU			−0.043	0.906^{***}
			(0.270)	(0.127)
Medium			0.008	-0.384^{*}
			(0.211)	(0.209)
Large			0.649^{***}	0.160
			(0.235)	(0.164)
V. Large			1.504^{***}	1.018^{***}
			(0.165)	(0.165)
Age			0.002	0.003
			(0.002)	(0.002)
Pat			0.002^{**}	0.001^{***}
			(0.001)	(0.000)
N	41 551	80 092	41 551	80 092
Country FE	No	No	Yes	Yes
Sector FE	No	No	Yes	Yes

Notes.
$^{*}P < 0.10$.
$^{**}P < 0.05$.
$^{***}P < 0.01$.
Cluster robust standard errors (cluster is the 2-digit NACE code) in parentheses. The table shows estimates of coefficients, not hazard ratios. The dummy variable "$Small_i$" is excluded, as small firms are used as a reference category.

equal to 5. This is presented for both model specifications, with (columns 3 and 4) and without (columns 1 and 2) control variables. The coefficient of the interaction between the variables "supplier" and "post" represents the DiD estimate of the treatment effects and captures the impact of procurement on the hazard that a firm starts filing patents after the onset of collaboration with CERN. In all columns of Table 3, the DiD estimate is positive and highly statistically significant.

The results imply that, following their appointment as LHC suppliers, firms in the treated group are more likely to apply for patents than those in the control group. This suggests that CERN procurement has a positive impact on the innovativeness of firms. Overall, the counterfactual analysis supports the results previously obtained by estimating model (1), thus allowing a causal interpretation of the link between CERN procurement and the overcoming of the innovation threshold for suppliers.

6. Conclusions

Does becoming a CERN supplier increase a firm's probability of filing a patent for the first time? Using survival models and PSM, this paper finds that the answer to this question is positive, with the effect occurring primarily for small firms involved in more technologically complex projects and with frequent interactions with CERN. Moreover, such an effect manifests itself three to seven years after the start of the collaboration, indicating a relatively slow learning process.

It is important to emphasize that the two empirical strategies we use are complementary and, although each has its limitations, their combined use allows us to obtain relevant results. On the one hand, the analysis focused on a sample composed entirely of suppliers allows us to use variables that accurately describe the technical characteristics of the orders and the intensity of the procurement relationship, and to study the timing of the effects of the relationship. On the other hand, this approach is affected by problems of sample selection. The counterfactual analysis helps to overcome this problem and provides a causal interpretation of our results, but relies on a smaller number of variables describing the characteristics of the firms and the procurement relationship. This issue may affect the quality of the matching and limit the number of control variables we can include in the DiD regressions.

Despite these limitations, our findings are relevant for two reasons. First, while previous studies (e.g., Bastianin *et al.*, 2021) showed that the procurement for LSRIs has an incremental effect on firms' patenting activity, the result obtained in this paper is in some ways stronger: the technological learning process induced by the collaboration with CERN can push firms that have never filed a patent before to change their status from non-patenting to patenting firms. From a policy point of view, public procurement by CERN turns out to have a positive impact exactly on those firms for which support is especially needed; these firms, in turn, can play a key role in generating variety and preserving novelty in the process of technological evolution. Second, the quantification of the time lag that separates R&D investment from innovation output matters for policymakers who must decide on the amount of resources to be invested in research activities; in fact, estimates of the social rate of return on research depend on such time distance.

Acknowledgments

This article has been prepared while the authors were contributors to the Future Circular Collider study: Milan and CERN (KE3044/ATS). A.B. and L.Z. acknowledge financial support within the "Fund for Departments of Excellence academic funding" provided by the Ministero dell'Università e della Ricerca (MUR), established by the Stability Law, Legge di Stabilità n.232/2016, 2017, Project of the Department of Economics, Management, and Quantitative Methods, University of Milan, CUP: G43C22004550001.

References

Aberg, S. and A. Bengtson (2015), 'Does CERN procurement result in innovation?,' *Innovation: The European Journal of Social Science Research*, **28**(3), 360–383.

Arrow, K. J. (1962), 'The economic implications of learning by doing,' *The Review of Economic Studies*, **29**(3), 155–173.

Arroyabe, M. F. and M. Schumann (2022), 'On the estimation of true state dependence in the persistence of innovation,' *Oxford Bulletin of Economics and Statistics*, **84**(4), 850–893.

Aschhoff, B. and W. Sofka (2009), 'Innovation on demand – can public procurement drive market success of innovations?,' *Research Policy*, **38**(8), 1235–1247.

Autio, E., A. P. Hameri and M. Bianchi-Streit (2003), *Technology Transfer and Technological Learning through CERN's Procurement Activity (No. CERN-2003-005)*. CERN: Geneve, Switzerland.

Autio, E., A. P. Hameri and O. Vuola (2004), 'A framework of industrial knowledge spillovers in Big-science centers,' *Research Policy*, **33**(1), 107–126.

Bartoloni, E. (2012), 'The persistence of innovation: a panel data investigation on manufacturing firms,' *International Review of Applied Economics*, **26**(6), 787–810.

Bastianin, A., P. Castelnovo, M. Florio and A. Giunta (2021), 'Big science and innovation: gestation lag from procurement to patents for CERN suppliers,' *Journal of Technological Transfer*, **47**(2), 531–555.

Bastianin, A. and C. F. Del Bo (2021), 'Public procurement in Big Science: politics or technology? The case of CERN,' *Industrial and Corporate Change*, **30**(3), 740–761.

Bastianin, A., C. F. Del Bo, M. Florio and F. Giffoni (2023), 'Projecting the socio-economic impact of a big science center: the world's largest particle accelerator at CERN,' *Applied Economics*, **55**(49), 5768–5789.

Bettencourt, L. M., J. Lobo and D. Strumsky (2007), 'Invention in the city: Increasing returns to patenting as a scaling function of metropolitan size,' *Research Policy*, **36**(1), 107–120.

Blei, D. M., A. Y. Ng and M. I. Jordan (2003), 'Latent Dirichlet allocation,' *Journal of Machine Learning Research*, **3**(Jan), 993–1022.

Bloom, N., J. Van Reenen and H. Williams (2019), 'A toolkit of policies to promote innovation,' *Journal of Economic Perspectives*, **33**(3), 163–184.

Breschi, S., F. Malerba and L. Orsenigo (2000), 'Technological regimes and Schumpeterian patterns of innovation,' *The Economic Journal*, **110**(463), 388–410.

Castelnovo, P. and M. Dal Molin (2021), 'The learning mechanisms through public procurement for innovation: The case of government-funded basic research organizations,' *Annals of Public and Cooperative Economics*, **92**(3), 411–446.

Castelnovo, P., M. Florio, S. Forte, L. Rossi and E. Sirtori (2018), 'The economic impact of technological procurement for large-scale research infrastructures: Evidence from the Large Hadron Collider at CERN,' *Research Policy*, **47**(9), 1853–1867.

Cefis, E. (2003), 'Is there persistence in innovative activities?,' *International International Journal of Industrial Organization*, **21**(4), 489–515.

Cefis, E. and O. Marsili (2015), 'Crossing the innovation threshold through mergers and acquisitions,' *Research Policy*, **44**(3), 698–710.

Cefis, E. and L. Orsenigo (2001), 'The persistence of innovative activities: A cross-countries and cross-sectors comparative analysis,' *Research Policy*, **30**(7), 1139–1158.

Chan, L. K., Lakonishok, J. and Sougiannis T (2001), 'The Stock Market Valuation of Research and Development Expenditures,' The Journal of Finance, **56**: 2431–2456.

Clotfelter, C., E. Glennie, H. Ladd and J. Vigdor (2008), 'Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina,' *Journal of Public Economics*, **92**(5–6), 1352–1370.

Cohen, W. M., R. Nelson and J. P. Walsh (2000), 'Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not),' NBER Working paper 7552.

Cox, D. R. (1972), 'Regression models and life tables,' *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **34**, 187–220.

Czarnitzki, D. and H. Hottenrott (2010), 'Financing constraints for industrial innovation: what do we know?' Available at SSRN 1749386.

Duguet, E. and S. Monjon (2004), 'Is innovation persistent at the firm level? An econometric examination comparing the propensity score and regression methods,' Cahiers de la Maison des Sciences Economiques v04075, Paris.

Edler, J. and L. Georghiou (2007), 'Public procurement and innovation—resurrecting the demand side,' *Research Policy*, **36**(7), 949–963.

Edquist, C., L. Hommen and L. Tsipouri (Eds.), (2000), *Public Technology Procurement and Innovation*. Kluwer Academic Publishers: Boston, MA, pp. 5–70.

Edquist, C., N. S. Vonortas, J. M. Zabala-Iturriagagoitia and J. Edler (Eds.), (2015), *Public Procurement for Innovation*. Edward Elgar Publishing: Cheltenham, UK.

Edquist, C. and J. M. Zabala-Iturriagagoitia (2012), 'Public Procurement for Innovation as mission-oriented innovation policy,' *Research Policy*, **41**(10), 1757–1769.

Fernandes, A., A. Pessoa and M. R. Silva (2014), 'Does astronomy generate economic benefits? Technological innovation seen through the lens of the European Southern observatory's very large telescope,' *International Journal of Technology Policy and Management*, **14**(4), 327–345.

Flaig, G. and M. Stadler (1994), 'Success breeds success. The dynamics of the innovation process,' *Empirical Economics*, **19**(1), 55–68.

Florio, M., F. Giffoni, A. Giunta and E. Sirtori (2018), 'Big science, learning, and innovation: evidence from CERN procurement,' *Industrial and Corporate Change*, **27**(5), 915–936.

Fontana, R. and A. Vezzulli (2016), 'Technological leadership and persistence in product innovation in the local area network industry 1990–1999,' *Research Policy*, **45**(8), 1604–1619.

Gentzkow, M., B. Kelly and M. Taddy (2019), 'Text as data,' *Journal of Economic Literature*, **57**(3), 535–574.

Geroski, P. A., J. Van Reenen and C. F. Walters (1997), 'How persistently do firms innovate?,' *Research Policy*, **26**(1), 33–48.

Griffiths, T. L. and M. Steyvers (2004), 'Finding scientific topics,' *Proceedings of the National Academy of Sciences*, **101**(1), 5228–5235.

Griliches, Z. (1979), 'Issues in assessing the contribution of research and development to productivity growth,' *The Bell Journal of Economics*, **10**(1), 92–116.

Griliches, Z. (1990), 'Patent statistics as economic indicators: A survey,' *Journal of Economic Literature*, **284**, 1661–1707.

Guerzoni, M. and E. Raiteri (2015), 'Demand-side vs. supply-side technology policies: Hidden treatment and new empirical evidence on the policy mix,' *Research Policy*, **44**(3), 726–747.

Hall, B., Z. Griliches and J. Hausman (1984), 'Patents and R&D: Is there a lag?,' *International Economic Review*, **27**(2), 265–283.

Hall, B. H., G. V. Graevenitz and C. Helmers (2021), 'Technology entry in the presence of patent thickets,' *Oxford Economic Papers*, **73**(2), 903–926.

Hameri, A. P. (1996), 'Technology transfer between basic research and industry,' *Technovation*, **16**(2), 51–92.

Hausman, J., B. Hall and Z. Griliches (1984), 'Econometric models for count data with an application to the patents-R&D relationship,' *Econometrica*, **52**(4), 909–938.

Leoncini, R., Marzucchi, A., Montresor, S., Rentocchini, F. and Rizzo, U. (2019) 'Better late than never': the interplay between green technology and age for firm growth,' *Small Bus Econ* **52**, 891–904.

Lichtenberg, F. R. (1988), 'The private R and D investment response to federal design and technical competitions,' *The American Economic Review*, **78**(3), 550–559.

Lundvall, B. Å. (1992), 'User-producer relationships, national systems of innovation and internationalization,' Vol. **242** in B. A. Lundvall (1992), *Na Theory of Innovation and Interactive Learning*. London: Pinter, 47-70.

Malerba, F., R. Nelson, L. Orsenigo and S. Winter (2007), 'Demand, innovation, and the dynamics of market structure: The role of experimental users and diverse preferences,' *Journal of Evolutionary Economics*, **17**(4), 371–399.

Malerba, F. and L. Orsenigo (1995), 'Schumpeterian patterns of innovation,' *Cambridge Journal of Economics*, **19**(1), 47–65.

Malerba, F. and L. Orsenigo (1997), 'Technological regimes and sectoral patterns of innovative activities,' *Industrial and Corporate Change*, **6**(1), 83–118.

Malerba, F. and L. Orsenigo (1999), 'Technological entry, exit and survival: an empirical analysis of patent data,' *Research Policy*, **28**(6), 643–660.

Manez, J. A., M. E. Rochina-Barrachina, A. Sanchis-Llopis and J. A. Sanchis-Llopis (2015), 'The determinants of R&D persistence in SMEs,' *Small Business Economics*, **44**(3), 505–528.

Mansfield, E (1991), 'Academic research and industrial innovation,' *Research Policy*, **20**, 1–12.

Mansfield, E. (1998) 'Academic research and industrial innovation: An update of empirical findings,' Research Policy, **26**, 773–776.

Marin, G. (2014), 'Do eco-innovations harm productivity growth through crowding out? Results of an extended CDM model for Italy,' *Research Policy*, **43**(2), 301–317.

Mazzucato, M. (2016), 'From market fixing to market-creating: a new framework for innovation policy,' *Industry and Innovation*, **23**(2), 140–156.

Mewes, L. and T. Broekel (2022), 'Technological complexity and economic growth of regions,' *Research Policy*, **51**(8), 104156.

Nelson, R. R. and S. G. Winter (1982), *An Evolutionary Theory of Economic Change*. Belknap Press of Harvard University Press: Cambridge, MA.

OECD (2014), 'The Impacts of Large Research Infrastructures on Economic Innovation and on Society: Case Studies at CERN,' *OECD Technical Report*.

Pakes, A. and Z. Griliches (1980), 'Patents and R&D at the firm level: A first report,' *Economics Letters*, **5**(4), 377–381.

Pakes, A. and M. Shankerman (1984), 'The rate of obsolescence of patents, research gestation lags, and the private rate of return to research resources,' in Z. Griliches (ed.), *R&D, Patents, and Productivity*. University of Chicago Press: Chicago, pp. 73–88.

Peters, B. (2009), 'Persistence of innovation: stylised facts and panel data evidence,' *The Journal of Technology Transfer*, **34**(2), 226–243.

Raiteri, E. (2018), 'A time to nourish? Evaluating the impact of public procurement on technological generality through patent data,' *Research Policy*, **47**(5), 936–952.

Raymond, W., P. Mohnen, F. Palm and S. S. Van Der Loeff (2010), 'Persistence of innovation in Dutch manufacturing: Is it spurious?,' *The Review of Economics and Statistics*, **92**(3), 495–504.

Roper, S. and N. Hewitt-Dundas (2008), 'Innovation persistence: Survey and case-study evidence,' *Research Policy*, **37**(1), 149–162.

Rosenberg, N. (1976), *Perspectives on Technology*. Cambridge University Press: Cambridge.

Salter, A. J. and B. R. Martin (2001), 'The economic benefits of publicly funded basic research: a critical review,' *Research Policy*, **30**(3), 509–532.

Scarrà, D. and A. Piccaluga (2022), 'The impact of technology transfer and knowledge spillover from Big Science: A literature review,' *Technovation*, **116**, 102165.

Sternitzke, C. (2010), 'Knowledge sources, patent protection, and commercialization of pharmaceutical innovations,' *Research Policy*, **39**(6), 810–821.

Toole, A. A. (2012), 'The impact of public basic research on industrial innovation: Evidence from the pharmaceutical industry,' *Research Policy*, **41**(1), 1–12.

Uyarra, E. and K. Flanagan (2010), 'Understanding the innovation impacts of public procurement,' *European Planning Studies*, **18**(1), 123–143.

Weitzman, M. L. (1998), 'Recombinant growth,' *The Quarterly Journal of Economics*, **113**(2), 331–360.

Appendix A.

Dataset: further details

- *Age*: the variable firm age is constructed starting from the variable "incorporation year" in Orbis. Specifically, it built as the difference between the year of beginning of the collaboration with CERN and firms' incorporation year.

- *IFA*: Intangible fixed assets recorded in the Orbis database include all intangible assets such as formation expenses, research expenses, goodwill, development expenses, and all other expenses with a long-term effect.
- *High-tech*: dummy variable that takes value 1 if the supplier has delivered a high-tech order according to the technological classification developed by CERN experts. Such classification is based on the following five-point technological-intensity scale, designed to capture differences in both product specificity and closeness of the supplier's collaboration with CERN:

Class 1: most likely "off-the-shelf" orders of low-technological intensity;

Class 2: off-the-shelf orders with average technological intensity;

Class 3: mostly off-the-shelf but usually high-tech and requiring some careful specification;

Class 4: high-tech orders with moderate to high intensity of specification activity to customise products for the LHC;

Class 5: products at the technological frontier, with intensive customization and co-design involving CERN staff.

We defined high-tech codes as Classes 3, 4, and 5 and then divided the LHC suppliers into two broad groups, according to type of order delivered.

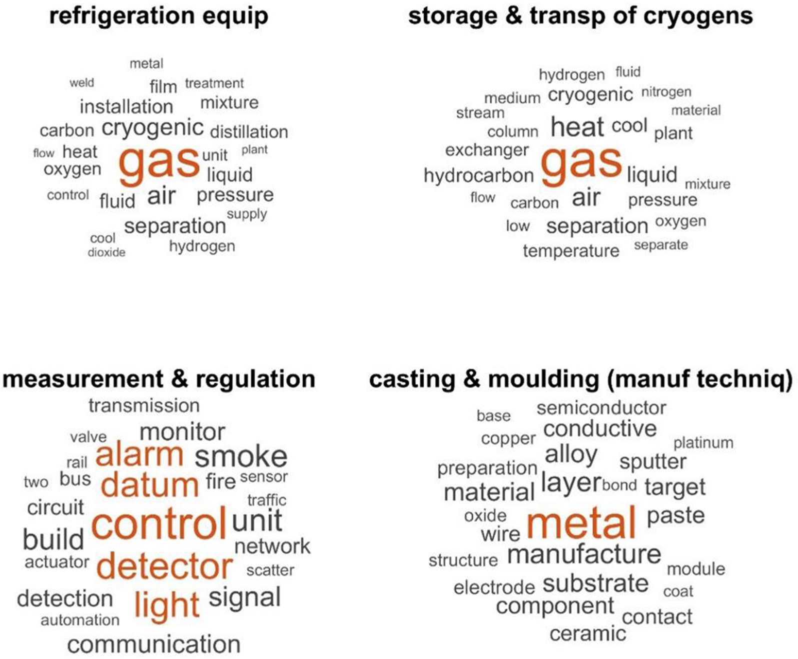
- *Order*: this variable is built as the sum of the monetary amount of all the LHC-related orders received by a supplier during its collaboration with CERN.
- *pct*: this variable represents the yearly contribution to the CERN budget of the country where the supplier is located. It is expressed as percentage of the total contribution of Member States. The information is taken form CERN annual reports.
- *Size dummies*: these dummy variables are constructed starting from the "size" variable provided in the Orbis database, which classifies companies as very large, large, medium and small on the basis of number of employees, total assets, and operating revenue.
- *pat*: number of regional patents per inhabitant pat_r, where the subscript r indicates the region where firm i is located. Sourced from the OECD Patent Statistics database using the Territorial Level 2 (TL2) geographical classification; this corresponds to NUTS 2 regions for European countries, states for the USA, and prefectures for Japan.

Appendix B.

Text analysis: further details and results

Data preparation. To implement Latent Dirichlet Allocation (LDA)—also known as topic modeling analysis—we sourced from the ORBIS Intellectual Property database the title of approximately 19 200 patents filed, in any patent office, by CERN suppliers, as well as the patents' technological classification provided by the World Intellectual Property Organization (WIPO). In the case of multiple orders, we classify each firm using the CERN activity code associated to its first LHC-related order. We note that the number of firms whose orders are assigned to multiple activity codes is very small. LDA involves some preliminary steps aimed cleaning the text, reducing the number of features analysed and focusing only on words that are diagnostic of the content of patents. After such data cleaning process, we are left with a vocabulary of 7152 words which occurred a total of 117 988 times in 19 200 patents' titles. More precisely, we dropped non-English titles and removed any punctuation, numbers, common words, also known as "stop words" (e.g., "the," "of," "is"), as well as very rare words (i.e., words that occur less than 10 times). As a final step, we apply the so-called Porter stemmer to replace words with their roots (e.g., "statistic," "statistics," "statistical" are all replaced by the stem "statistic").

LDA implementation. In our implementation of LDA, we assume that each patent's title is the result of multiple latent topics. Viewing titles as a probabilistic mixture of topics and topics as probabilistic distribution of words, we can use LDA to discover the topics from our collection of patents' titles. See Blei *et al.* (2003), Gentzkow *et al.* (2019) and Griffiths and Steyvers (2004) for further methodological details. After fitting a LDA model, we computed the predicted topic probabilities: these are the probabilities that topic t appears in document d. These are collected in the $(D \times T)$ matrix ψ, where D is the patents' titles and T is the number of topics in the

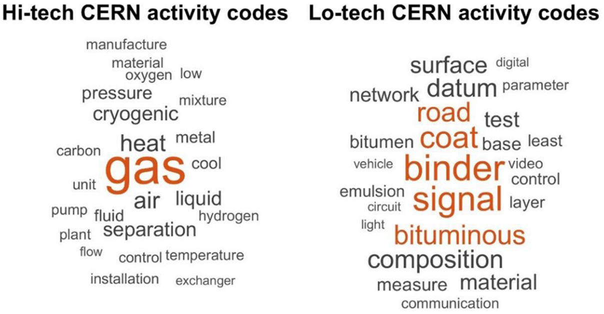

refrigeration equip

storage & transp of cryogens

measurement & regulation

casting & moulding (manuf techniq)

Figure B1. Word clouds based on the title of patents for the four most frequent CERN activity codes. Notes: the figure shows the most recurring words in patents' titles for the top-four most frequent CERN activity codes. The font size is proportional to the number of times the word is found in the title of patents.


Hi-tech CERN activity codes

Lo-tech CERN activity codes

Figure B2. Word clouds based on the title of patents for high- and low-tech CERN activity codes. Notes: see notes to Figure B1.

LDA model. Each row of ψ sums to one. For each CERN activity code, we computed the mean predicted topic probability over patents' titles filed by firms assigned to that activity code ψ_j. Then, we defined the most *diagnostic topic* for each activity code as the topic t for which the ratio of ψ_j to the sum of ψ_j across all other activity codes was maximized. This is the procedure underlying Figure 1, where in the upper panel we show the predicted topic probability for each of the 14 most recurrent CERN activity codes reported on the vertical axis,[16] where darker shades indicate higher probability. Word clouds in the lower panel display the five most probable words in each topic.

---

[16] In the case of multiple orders, we classify each firm using the CERN activity code associated to its first LHC-related order. We note that the number of firms whose orders are assigned to multiple activity codes is very small.

Table B1. Top 10 patent's IPC codes by firm technological classification

High-tech firms (no. of patents = 18,580)		
IPC codes (3 digits)	**%**	**IPC description**
F25	11.35	Refrigeration or cooling; combined heating and refrigeration systems; heat pump systems; manufacture or storage of ice; liquefaction or solidification of gases
B01	9.77	Physical or chemical processes or apparatus in general
H01	8.06	Basic electric elements
G01	7.26	Measuring; testing
F17	6.08	Storing or distributing gases or liquids
B23	3.90	Machine tools; metal-working not otherwise provided for
C01	3.54	Inorganic chemistry
F04	3.54	Positive-displacement machines for liquids; pumps for liquids or elastic fluids
F16	3.16	Engineering elements or units; general measures for producing and maintaining effective functioning of machines or installations; thermal insulation
C07	3.02	Organic chemistry
Low-tech firms (no. of patents = 612)		
H04	19.77	Electric communication technique
G01	14.87	Measuring; testing
E01	11.44	Construction of roads, railways, or bridges
C08	10.13	Organic macromolecular compounds; their preparation or chemical working-up; compositions based thereon
C04	5.72	Cements; concrete; artificial stone; ceramics; refractories
C12	5.07	Biochemistry; microbiology; enzymology; mutation or genetic engineering
H03	4.41	Basic electronic circuitry
G06	3.27	Computing; calculating or counting
C09	3.27	Dyes; paints; polishes; natural resins; adhesives; compositions not otherwise provided for; applications of materials not otherwise provided for
A61	2.12	Medical or veterinary science; hygiene

Further results. The dataset used for text analysis also allows to identify the most recurring words in patents' titles for the top-four most frequent CERN activity codes. In Figure B1, the font size in word clouds is proportional to the number of times a word is found in the title of patents filed by firms that have started filing patents after the start of their relationship with CERN. As we can see comparing the titles of the word clouds and their content, CERN activity codes and patents' titles are largely compatible.

Moreover, the text analysis can be used to validate our partition of firms as high- and low-tech using their CERN activity codes. This exercise is presented in Figure B2 and shows the most recurring words in patents' titles for high- and low-tech activity codes. While low-tech activity codes are associated with words pertaining to construction activities, hi-tech activity codes are associated with words the cryogenic system at LHC.[17]

In Table B1, we further explore the degree of consistency between the codes assigned to the LHC contracts and the content of the patents subsequently filed. We use the codes in the International Patent Classification (IPC) to group the patents filed by LHC's suppliers and show the 10 most common classes. Comparing Table B1 with Figures B1 and B2, we can confirm that there is a large overlap between contract type and patent content.

---

[17] Cryogenics is a field of physics dealing with the production and effects of very low temperatures. The LHC is the largest cryogenic system in the world. Some of LHC's main magnets operate at a temperature of −271.3°C. As a reference, the temperature in the outer space is −270.5°C.

Appendix C.

Hazard models: robustness checks

We present some robustness checks of the baseline specification in Equation (1) as well as further results to complement our main findings.

High-tech classification and number of orders. The fact that the coefficient on the dichotomous variable used to classify statistical units into high- and low-tech firms is never statistically distinguishable from zero is somehow surprising, given that this variable captures some aspects related with the absorptive capacity of firms To verify that the large standard errors associated with these coefficients are not due to measurement error, we attempt to proxy the degree of technological intensity of firms with a continuous variable that captures the share of orders classified as hi-tech, which is used in place of the "$Hi\text{-}tech_i$" dichotomous variable. The findings of Table 1 are unaffected when using this alternative measure of firms' technological intensity whose coefficient remains statistically indistinguishable from zero. Similarly, we replace the total amount of LHC orders with the (logarithm of the) orders count to consider an alternative proxy of the involvement and continuity of the procurement relationship. Our main results remain unchanged. See Tables C1 and C2.

Member Country Effect. Then, we extend the baseline specification with additional control variables. As a further extension of the baseline specification, we add a variable that controls for country heterogeneity: the percent yearly contribution of each country to CERN budget.[18] This variable captures the fact that firms located in countries contributing more, might have a higher probability of receiving an order and hence to benefit from knowledge spillovers. While, as expected, the coefficient on this variable is positive, it is statistically undistinguishable from zero and the main results are confirmed. See Table C3.

R&D Expenditure. We also add a proxy of R&D expenditure that represents a key control variable for analyzing patent activity (Hausman *et al.*, 1984). However, since this information on R&D expenditure is mostly missing in the Orbis database, we add intangible assets as a proxy (see (Chan *et al.*, 2001); (Leoncini *et al.*, 2019); Marin, 2014). To have a balanced panel dataset, missing observations have been substituted with zeros and a dummy variable taking value one for such observations is included in the regressions. As expected, the coefficient of our R&D proxy is positive and highly statistically significant, and main results are confirmed. See Table C4.

Age of firms. Lastly, we include firms' age to account for the possibility that start-ups display a superior innovative output. Such variable is built as the difference between the year of beginning of the collaboration with CERN and firms' incorporation year. This control variable turns out to be statistically undistinguishable from zero and its inclusion does not affect our main findings. Overall, we conclude that the inclusion of all these additional control variables—one at the time or jointly—does not affect our main findings. See Table C5.

Appendix D.

Further results

In this Appendix, we provide the tables for the results discussed in Section 5.2.

- Table D1 presents the results obtained by dividing the sample in two subsamples: high-tech firms (60% of the sample) and low-tech firms (40%).
- Table D2 shows that software and other service firms are not predominant within the group classified as low-tech (in fact, there is a sector—manufacturing of computer, electronic and optical products—which is the most represented sector within the low-tech group and second in the ranking within the high-tech).
- Tables D3 reports regressions obtained splitting the sample into small-medium firms vs large-very large firms.
- In Table D4, we split the sample of suppliers into firms that received orders above or below the median of the order size distribution.

---

[18] This information is taken form CERN annual reports.

Table C1. Alternative hi-tech classification

	(1) $k = 0$	(2) $k = 1$	(3) $k = 2$	(4) $k = 3$	(5) $k = 4$	(6) $k = 5$	(7) $k = 6$	(8) $k = 7$	(9) $k = 8$	(10) $k \geq 9$
$CERN^{k}_{i,t}$	0.034	0.044	0.047*	0.054**	0.067***	0.080***	0.088***	0.070*	−0.013	0.026
	(0.046)	(0.036)	(0.028)	(0.021)	(0.022)	(0.021)	(0.022)	(0.041)	(0.059)	(0.072)
$\%\ HTO_i$	0.170	0.172	0.176	0.175	0.179	0.185	0.190	0.184	0.169	0.171
	(0.223)	(0.223)	(0.224)	(0.226)	(0.228)	(0.229)	(0.229)	(0.226)	(0.223)	(0.223)
$Order_i$	−0.005	−0.009	−0.010	−0.012	−0.016	−0.019	−0.017	−0.007	0.008	0.005
	(0.071)	(0.070)	(0.068)	(0.065)	(0.066)	(0.066)	(0.065)	(0.061)	(0.062)	(0.062)
$Medium_i$	0.515*	0.516*	0.518*	0.522*	0.530*	0.533*	0.530*	0.523*	0.520*	0.520*
	(0.274)	(0.273)	(0.273)	(0.273)	(0.272)	(0.274)	(0.275)	(0.275)	(0.280)	(0.279)
$Large_i$	1.159***	1.157***	1.156***	1.161***	1.167***	1.171***	1.163***	1.159***	1.168***	1.164***
	(0.395)	(0.393)	(0.393)	(0.393)	(0.394)	(0.397)	(0.401)	(0.407)	(0.408)	(0.408)
$V\ Large_i$	2.069***	2.072***	2.078***	2.088***	2.106***	2.120***	2.114***	2.089***	2.064***	2.067***
	(0.368)	(0.363)	(0.362)	(0.359)	(0.359)	(0.358)	(0.361)	(0.364)	(0.379)	(0.378)
α_c	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
α_s	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

country (α_c) and sector (α_s) fixed effects are included in all the specifications.

$^{*}P < 0.10$.

$^{**}P < 0.05$.

$^{***}P < 0.01$.

Cluster robust standard errors (i.e., cluster is the 2 digits NACE code) in parentheses. The table shows estimates of the coefficients in Equation (1) and not hazard ratios. The dummy variable "$Small_i$" is excluded and small firms are used as a reference category.

Table C2. No. orders in place of total order amount

	(1) $k = 0$	(2) $k = 1$	(3) $k = 2$	(4) $k = 3$	(5) $k = 4$	(6) $k = 5$	(7) $k = 6$	(8) $k = 7$	(9) $k = 8$	(10) $k \geq 9$
$CERN_{i,t}^{k}$	0.039	0.048	0.051^{*}	0.058^{***}	0.071^{***}	0.085^{***}	0.092^{***}	0.074^{*}	−0.009	0.032
	(0.046)	(0.036)	(0.028)	(0.021)	(0.022)	(0.021)	(0.023)	(0.041)	(0.059)	(0.072)
$Hi\text{-}tech_i$	0.151	0.151	0.153	0.152	0.154	0.159	0.167	0.167	0.157	0.160
	(0.178)	(0.177)	(0.177)	(0.177)	(0.177)	(0.177)	(0.178)	(0.179)	(0.179)	(0.179)
$\#Order_i$	−0.083	−0.091	−0.092	−0.097	−0.104	−0.111	−0.105	−0.086	−0.058	−0.064
	(0.109)	(0.108)	(0.105)	(0.100)	(0.102)	(0.103)	(0.103)	(0.098)	(0.096)	(0.097)
$Medium_i$	0.505^{*}	0.507^{*}	0.508^{*}	0.514^{*}	0.523^{*}	0.527^{*}	0.523^{*}	0.514^{*}	0.508^{*}	0.509^{*}
	(0.277)	(0.277)	(0.276)	(0.276)	(0.276)	(0.278)	(0.278)	(0.278)	(0.284)	(0.283)
$Large_i$	1.183^{***}	1.181^{***}	1.181^{***}	1.186^{***}	1.194^{***}	1.198^{***}	1.189^{***}	1.184^{***}	1.191^{***}	1.187^{***}
	(0.381)	(0.378)	(0.378)	(0.377)	(0.377)	(0.380)	(0.384)	(0.391)	(0.396)	(0.395)
$V\ Large_i$	2.117^{***}	2.120^{***}	2.126^{***}	2.137^{***}	2.157^{***}	2.172^{***}	2.165^{***}	2.136^{***}	2.111^{***}	2.113^{***}
	(0.355)	(0.349)	(0.348)	(0.345)	(0.346)	(0.344)	(0.347)	(0.351)	(0.368)	(0.366)
α_c	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
α_s	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Country (α_c) and sector (α_s) fixed effects are included in all the specifications.

$^{*}P < 0.10$.

$^{***}P < 0.01$.

Cluster robust standard errors (i.e., cluster is the 2 digits NACE code) in parentheses. The table shows estimates of the coefficients in Equation (1) and not hazard ratios. The dummy variable "$Small_i$" is excluded and small firms are used as a reference category.

Table C3. Controlling for the percent contribution of country to CERN budget

	(1) $k = 0$	(2) $k = 1$	(3) $k = 2$	(4) $k = 3$	(5) $k = 4$	(6) $k = 5$	(7) $k = 6$	(8) $k = 7$	(9) $k = 8$	(10) $k \geq 9$
$CERN_{i,t}^{k}$	0.034	0.043	0.046	0.054^{**}	0.067^{***}	0.080^{***}	0.088^{***}	0.071^{*}	−0.009	0.034
	(0.047)	(0.037)	(0.029)	(0.022)	(0.023)	(0.022)	(0.023)	(0.040)	(0.059)	(0.073)
$Hi\text{-}Tech_i$	0.133	0.131	0.132	0.130	0.130	0.135	0.143	0.147	0.140	0.143
	(0.176)	(0.175)	(0.174)	(0.175)	(0.175)	(0.174)	(0.175)	(0.175)	(0.175)	(0.175)
$Order_i$	0.001	−0.003	−0.004	−0.006	−0.009	−0.012	−0.009	−0.000	0.013	0.010
	(0.068)	(0.066)	(0.064)	(0.061)	(0.062)	(0.062)	(0.061)	(0.057)	(0.058)	(0.058)
$Medium_i$	0.515^{*}	0.516^{*}	0.517^{*}	0.522^{*}	0.529^{*}	0.533^{*}	0.529^{*}	0.523^{*}	0.519^{*}	0.520^{*}
	(0.273)	(0.273)	(0.272)	(0.272)	(0.272)	(0.274)	(0.274)	(0.275)	(0.279)	(0.278)
$Large_i$	1.164^{***}	1.162^{***}	1.162^{***}	1.166^{***}	1.173^{***}	1.178^{***}	1.170^{***}	1.164^{***}	1.172^{***}	1.168^{***}
	(0.398)	(0.396)	(0.395)	(0.395)	(0.396)	(0.400)	(0.404)	(0.409)	(0.410)	(0.410)
$V\ Large_i$	2.055^{***}	2.057^{***}	2.062^{***}	2.071^{***}	2.090^{***}	2.103^{***}	2.096^{***}	2.071^{***}	2.050^{***}	2.051^{***}
	(0.370)	(0.363)	(0.362)	(0.359)	(0.360)	(0.359)	(0.362)	(0.367)	(0.379)	(0.378)
$pct_{c,t}$	0.002	0.002	0.001	0.001	0.002	0.001	0.002	0.002	0.002	0.002
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
α_c	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
α_s	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Country (α_c) and sector (α_s) fixed effects are included in all the specifications.
$^{*}P < 0.10$.
$^{**}P < 0.05$.
$^{***}P < 0.01$.
Cluster robust standard errors (i.e., cluster is the 2 digits NACE code) in parentheses. The table shows estimates of the coefficients in Equation (1) and not hazard ratios. The dummy variable "$Small_i$" is excluded and small firms are used as a reference category.

Table C4. Controlling for R&D effort (intangible assets)

	(1) $k = 0$	(2) $k = 1$	(3) $k = 2$	(4) $k = 3$	(5) $k = 4$	(6) $k = 5$	(7) $k = 6$	(8) $k = 7$	(9) $k = 8$	(10) $k = 9$
$CERN_{i,t}^{k}$	0.039	0.048	0.050^{*}	0.058^{***}	0.069^{***}	0.084^{***}	0.091^{***}	0.075^{*}	−0.010	0.026
	(0.045)	(0.036)	(0.027)	(0.020)	(0.022)	(0.021)	(0.022)	(0.040)	(0.060)	(0.073)
$Hi\text{-}Tech_i$	0.156	0.155	0.156	0.154	0.155	0.162	0.170	0.175	0.166	0.168
	(0.161)	(0.160)	(0.159)	(0.160)	(0.159)	(0.159)	(0.160)	(0.160)	(0.162)	(0.161)
$Order_i$	0.003	−0.001	−0.001	−0.003	−0.006	−0.010	−0.008	0.002	0.017	0.014
	(0.073)	(0.072)	(0.070)	(0.067)	(0.067)	(0.067)	(0.066)	(0.063)	(0.064)	(0.064)
$Medium_i$	0.501^{*}	0.501^{*}	0.502^{*}	0.508^{*}	0.516^{*}	0.520^{*}	0.520^{*}	0.513^{*}	0.509^{*}	0.509^{*}
	(0.283)	(0.283)	(0.282)	(0.283)	(0.283)	(0.285)	(0.286)	(0.287)	(0.290)	(0.289)
$Large_i$	1.046^{***}	1.041^{***}	1.041^{***}	1.046^{***}	1.055^{***}	1.059^{***}	1.054^{***}	1.048^{**}	1.062^{**}	1.058^{**}
	(0.399)	(0.397)	(0.398)	(0.400)	(0.401)	(0.404)	(0.408)	(0.413)	(0.412)	(0.412)
$V\ Large_i$	1.775^{***}	1.779^{***}	1.785^{***}	1.794^{***}	1.808^{***}	1.821^{***}	1.816^{***}	1.792^{***}	1.770^{***}	1.774^{***}
	(0.391)	(0.386)	(0.386)	(0.383)	(0.384)	(0.383)	(0.384)	(0.388)	(0.401)	(0.400)
$Int.\ Assets_i$	0.177^{***}	0.178^{***}	0.177^{***}	0.178^{***}	0.178^{***}	0.181^{***}	0.181^{***}	0.179^{***}	0.173^{***}	0.173^{***}
	(0.036)	(0.036)	(0.035)	(0.035)	(0.035)	(0.035)	(0.034)	(0.033)	(0.035)	(0.035)
α_c	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
α_s	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Country α_c) and sector (α_s) fixed effects are included in all the specifications.
$^{*}P < 0.10$.
$^{**}P < 0.05$.
$^{***}P < 0.01$.
Cluster robust standard errors (i.e., cluster is the 2 digits NACE code) in parentheses. The table shows estimates of the coefficients in Equation (1) and not hazard ratios. The dummy variable "$Small_i$" is excluded and small firms are used as a reference category.

Table C5. Controlling for firms' age

	(1) $k = 0$	(2) $k = 1$	(3) $k = 2$	(4) $k = 3$	(5) $k = 4$	(6) $k = 5$	(7) $k = 6$	(8) $k = 7$	(9) $k = 8$	(10) $k = 9$
$CERN^k_{i,t}$	0.032	0.037	0.042	0.046**	0.062***	0.079***	0.078***	0.061	−0.088	−0.005
	(0.047)	(0.036)	(0.028)	(0.020)	(0.023)	(0.026)	(0.027)	(0.049)	(0.079)	(0.077)
$Hi\text{-}Tech_i$	0.234	0.234	0.234	0.232	0.231	0.235	0.244	0.249	0.240	0.245
	(0.185)	(0.185)	(0.184)	(0.185)	(0.184)	(0.186)	(0.188)	(0.188)	(0.188)	(0.188)
$Order_i$	−0.003	−0.006	−0.007	−0.008	−0.013	−0.017	−0.010	−0.000	0.013	0.009
	(0.082)	(0.080)	(0.077)	(0.073)	(0.074)	(0.073)	(0.071)	(0.068)	(0.070)	(0.070)
$Medium_i$	0.576**	0.579**	0.580**	0.581**	0.587**	0.587**	0.580**	0.572**	0.574**	0.572**
	(0.239)	(0.237)	(0.236)	(0.236)	(0.235)	(0.237)	(0.237)	(0.237)	(0.238)	(0.238)
$Large_i$	1.321***	1.319***	1.316***	1.318***	1.319***	1.313***	1.305***	1.306***	1.340***	1.325***
	(0.464)	(0.461)	(0.458)	(0.459)	(0.457)	(0.459)	(0.465)	(0.478)	(0.475)	(0.473)
$V\ Large_i$	2.240***	2.245***	2.254***	2.262***	2.287***	2.305***	2.284***	2.251***	2.215***	2.223***
	(0.416)	(0.407)	(0.405)	(0.400)	(0.401)	(0.396)	(0.394)	(0.392)	(0.413)	(0.410)
Age_i	0.072	0.071	0.069	0.069	0.067	0.068	0.069	0.071	0.078	0.076
	(0.128)	(0.129)	(0.128)	(0.127)	(0.127)	(0.126)	(0.126)	(0.128)	(0.132)	(0.131)
α_c	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
α_s	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Country (α_c) and sector (α_s) fixed effects are included in all the specifications.
$^{**}P < 0.05$.
$^{***}P < 0.01$.
Cluster robust standard errors (i.e., cluster is the 2 digits NACE code) in parentheses. The table shows estimates of the coefficients in Equation (1) and not hazard ratios. The dummy variable "$Small_i$" is excluded and small firms are used as a reference category.

Table D1. Cox proportional hazard models by firm technological class

	(1) $k = 0$	(2) $k = 1$	(3) $k = 2$	(4) $k = 3$	(5) $k = 4$	(6) $k = 5$	(7) $k = 6$	(8) $k = 7$	(9) $k = 8$	(10) $k = 9$
High-tech firms										
$CERN_{i,t}^{k}$	0.066*	0.069*	0.038	0.059***	0.090***	0.096***	0.094***	0.066	−0.032	0.041
	(0.040)	(0.039)	(0.029)	(0.023)	(0.022)	(0.023)	(0.032)	(0.048)	(0.073)	(0.070)
$Order_i$	−0.009	−0.012	−0.002	−0.009	−0.019	−0.019	−0.012	−0.001	0.014	0.008
	(0.091)	(0.092)	(0.091)	(0.088)	(0.088)	(0.088)	(0.087)	(0.082)	(0.084)	(0.086)
$Medium_i$	0.115	0.121	0.105	0.118	0.136	0.144	0.132	0.108	0.079	0.089
	(0.369)	(0.368)	(0.366)	(0.366)	(0.366)	(0.365)	(0.359)	(0.350)	(0.360)	(0.360)
$Large_i$	0.895**	0.895**	0.902**	0.896**	0.894**	0.902**	0.905**	0.907**	0.920**	0.914**
	(0.423)	(0.422)	(0.423)	(0.424)	(0.426)	(0.430)	(0.432)	(0.434)	(0.431)	(0.429)
$V\ Large_i$	1.515***	1.524***	1.522***	1.533***	1.559***	1.574***	1.566***	1.543***	1.508***	1.519***
	(0.388)	(0.383)	(0.388)	(0.389)	(0.388)	(0.387)	(0.388)	(0.381)	(0.394)	(0.393)
Low-tech firms										
$CERN_{i,t}^{k}$	−0.028	−0.004	0.053	0.043	0.025	0.052	0.076*	0.071	0.007	−0.012
	(0.075)	(0.055)	(0.058)	(0.038)	(0.044)	(0.041)	(0.041)	(0.048)	(0.083)	(0.94)
$Order_i$	−0.049	−0.054	−0.073	−0.069	−0.061	−0.067	−0.071	−0.066	−0.056	−0.053
	(0.077)	(0.076)	(0.070)	(0.073)	(0.078)	(0.076)	(0.079)	(0.086)	(0.082)	(0.083)
$Medium_i$	1.099**	1.079**	1.047**	1.065**	1.075**	1.081**	1.090**	1.093**	1.077**	1.070**
	(0.458)	(0.458)	(0.440)	(0.449)	(0.458)	(0.458)	(0.459)	(0.470)	(0.466)	(0.459)
$Large_i$	1.266*	1.268*	1.286*	1.291*	1.284*	1.303*	1.302*	1.275*	1.268*	1.276*
	(0.765)	(0.753)	(0.727)	(0.731)	(0.749)	(0.740)	(0.735)	(0.739)	(0.747)	(0.746)
$V\ Large_i$	2.951***	2.961***	2.981***	2.990***	2.984***	3.017***	3.035***	3.003***	2.964***	2.954***
	(0.838)	(0.826)	(0.819)	(0.817)	(0.833)	(0.826)	(0.817)	(0.820)	(0.823)	(0.819)

Sample: 441 high-tech firms; 300 low-tech firms.
Country (D_c) and sector (D_s) fixed effects are included in all the specifications.
$^{*}P < 0.10$.
$^{**}P < 0.05$.
$^{***}P < 0.01$.
Cluster robust standard errors (i.e., cluster is the 2 digits NACE code) in parentheses. The table shows estimates of coefficients, not hazard ratios. The dummy variable "$Small_i$" is excluded, as small firms are used as a reference category.

Table D2. Sectoral distribution of patenting firms in the hi- and low-tech groups

Hi-tech		Low-tech	
Description	**%**	**Description**	**%**
Man. of electrical equipment	18.9	Man. of computer, electronic, and optical products	32.1
Man. of computer, electronic, and optical products	17.0	Man. of machinery and equipment	11.3
Man. of machinery and equipment	13.2	Computer programming, consultancy, and related act.	5.66
Man. of fabricated metal products, except machinery and equipment	9.43	Man. of electrical equipment	3.77
Man. of basic metals	5.66	Civil engineering	3.77
Wholesale trade, except of motor vehicles and motorcycles	5.66	Specialized construction act.	3.77
Architectural and engineering act.; technical testing and analysis	4,72	Wholesale trade, except of motor vehicles and motorcycles	3,77
Act. of extraterritorial organizations	4.72	Act. of extraterritorial organizations	3.77
Man. of chemicals and chemical products	3.77	Architectural and engineering act; technical testing and analysis	3.77
Scientific research and development	2.83	Education	3.77
Others	14.1	Others	24.5
Total	**100**	**Total**	**100**

Note: the table provides the sectoral distribution of hi- and low-tech firms that have filed at least one patent.

Table D3. Cox proportional hazard models by firm size

	(1) ***k*** = 0	(2) ***k*** = 1	(3) ***k*** = 2	(4) ***k*** = 3	(5) ***k*** = 4	(6) ***k*** = 5	(7) ***k*** = 6	(8) ***k*** = 7	(9) ***k*** = 8	(10) ***k*** = 9
Small and medium firms										
$CERN^k_{i,t}$	0.017	0.016	0.052	0.082***	0.085**	0.099***	0.117***	0.136**	–0.014	–0.097
	(0.058)	(0.037)	(0.034)	(0.032)	(0.034)	(0.026)	(0.031)	(0.065)	(0.086)	(0.065)
$Hi\text{-}Tech_i$	0.152	0.153	0.171	0.190	0.192	0.201	0.207	0.203	0.145	0.143
	(0.275)	(0.272)	(0.271)	(0.270)	(0.270)	(0.274)	(0.275)	(0.283)	(0.278)	(0.277)
$Order_i$	–0.085	–0.085	–0.093	–0.097	–0.097	–0.098	–0.098	–0.095	–0.081	–0.080
	(0.116)	(0.117)	(0.114)	(0.110)	(0.111)	(0.109)	(0.108)	(0.109)	(0.116)	(0.116)
Large and very large firms										
$CERN^k_{i,t}$	0.023	0.042	0.022	0.014	0.029	0.042	0.040	–0.009	–0.044	0.028
	(0.056)	(0.053)	(0.042)	(0.027)	(0.026)	(0.031)	(0.030)	(0.045)	(0.074)	(0.072)
$Hi\text{-}Tech_i$	0.256	0.242	0.256	0.261	0.250	0.246	0.256	0.276	0.276	0.274
	(0.228)	(0.224)	(0.219)	(0.215)	(0.213)	(0.210)	(0.204)	(0.205)	(0.209)	(0.208)
$Order_i$	0.015	0.006	0.014	0.018	0.012	0.007	0.010	0.025	0.029	0.020
	(0.074)	(0.073)	(0.069)	(0.065)	(0.064)	(0.064)	(0.065)	(0.056)	(0.059)	(0.060)

Sample: 474 firms small and medium firms; 267 large and very large firms Country (D_c) and sector (D_s) fixed effects are included in all the specifications.
*$P < 0.10$.
**$P < 0.05$.
***$P < 0.01$.
Cluster robust standard errors (i.e., cluster is the 2 digits NACE code) in parentheses. The table shows estimates of coefficients, not hazard ratios.

Table D4. Cox proportional hazard models by total monetary amount of the orders received

	(1) $k=0$	(2) $k=1$	(3) $k=2$	(4) $k=3$	(5) $k=4$	(6) $k=5$	(7) $k=6$	(8) $k=7$	(9) $k=8$	(10) $k=9$
Total order amount above the median										
$CERN^k_{i,t}$	−0.003	0.037	0.044	0.049	0.039	0.059**	0.091***	0.075*	−0.043	0.054
	(0.037)	(0.046)	(0.033)	(0.032)	(0.026)	(0.027)	(0.026)	(0.038)	(0.072)	(0.062)
$Hi\text{-}tech_i$	0.407	0.373	0.367	0.363	0.372	0.367	0.367	0.393	0.402	0.405
	(0.367)	(0.360)	(0.352)	(0.340)	(0.341)	(0.332)	(0.330)	(0.332)	(0.340)	(0.340)
$Medium_i$	0.214	0.205	0.209	0.224	0.234	0.244	0.242	0.221	0.213	0.212
	(0.407)	(0.397)	(0.393)	(0.386)	(0.385)	(0.383)	(0.377)	(0.387)	(0.408)	(0.400)
$Large_i$	1.355***	1.335***	1.332***	1.335***	1.348***	1.350***	1.336***	1.323***	1.365***	1.345***
	(0.423)	(0.417)	(0.417)	(0.410)	(0.407)	(0.401)	(0.396)	(0.405)	(0.425)	(0.417)
$V\ Large_i$	2.259***	2.215***	2.215***	2.226***	2.253***	2.267***	2.269***	2.247***	2.265***	2.249***
	(0.481)	(0.478)	(0.470)	(0.456)	(0.450)	(0.438)	(0.428)	(0.447)	(0.474)	(0.463)
Total order amount equal to or below the median										
$CERN^k_{i,t}$	0.058	0.048	0.047	0.058*	0.108***	0.113***	0.070	0.023	0.029	−0.039
	(0.068)	(0.041)	(0.039)	(0.034)	(0.034)	(0.032)	(0.046)	(0.075)	(0.102)	(0.093)
$Hi\text{-}tech_i$	0.228	0.230	0.233	0.229	0.236	0.234	0.234	0.228	0.228	0.227
	(0.317)	(0.318)	(0.320)	(0.316)	(0.305)	(0.307)	(0.320)	(0.325)	(0.330)	(0.331)
$Medium_i$	0.554	0.561	0.560	0.562	0.580	0.598	0.579	0.567	0.570	0.561
	(0.564)	(0.568)	(0.568)	(0.565)	(0.565)	(0.582)	(0.585)	(0.584)	(0.593)	(0.583)
$Large_i$	0.892	0.898	0.898	0.899	0.909	0.933	0.914	0.915	0.916	0.914
	(0.750)	(0.760)	(0.760)	(0.755)	(0.748)	(0.771)	(0.786)	(0.787)	(0.794)	(0.783)
$V\ Large_i$	1.943***	1.949***	1.946***	1.957***	2.030***	2.046***	1.958***	1.909**	1.907**	1.889**
	(0.742)	(0.743)	(0.748)	(0.735)	(0.727)	(0.741)	(0.750)	(0.744)	(0.764)	(0.749)

Sample: 370 firms with total order amount above the median; 371 firms with total order amount below or equal to the median. Country (D_c) and sector (D_s) fixed effects are included in all the specifications.

*$P < 0.10$.

**$P < 0.05$.

***$P < 0.01$.

Cluster robust standard errors (i.e., cluster is the 2 digits NACE code) in parentheses. The table shows estimates of coefficients, not hazard ratios.

Industrial and Corporate Change, 2025, 00, 1–29, DOI: https://doi.org/10.1093/icc/dtaf004, Original Article