



Big science and innovation: gestation lag from procurement to patents for CERN suppliers

Andrea Bastianin¹ · Paolo Castelnovo¹ · Massimo Florio¹ · Anna Giunta²

Accepted: 18 March 2021 / Published online: 29 March 2021
© The Author(s) 2021

Abstract

CERN, the European Organization for Nuclear Research, is the most important laboratory for particle physics in the world. It requires cutting edge technologies to deliver scientific discoveries. This paper investigates the time span needed for technology suppliers of CERN to absorb the knowledge acquired during the procurement relation and develop it into a patent. We estimate count data models relying on a sample of CERN suppliers for the Large Hadron Collider (LHC), a particle accelerator. Firms in our sample received their first LHC-related order over a long-time span (1995–2008). This fact is exploited to estimate the time lag that separates the beginning of the procurement relationship and the filing date of patents. Becoming a supplier of CERN is associated with a statistically significant increase in the number of patent applications by firms. Moreover, such an effect requires a relatively long gestation lag in the range of five to eight years.

Keywords Big science · CERN · Innovation · Public procurement · Patents · Gestation lag

JEL Codes C21 · C23 · H57 · L39 · O31

1 Introduction

The term “Big Science” identifies the style of scientific analysis characterizing research in physics, astronomy and molecular biology after World War II (Dennis, 2017). Big Science Centers (BSC) rely on large-scale research instruments and infrastructures, have a long list of participating institutions, attract generous funding from governments and rely on public procurement to develop technologies required for scientific research.

BSC’s industrial partners face both challenges and opportunities. In fact, public procurement through BSC is a form of Public Procurement for Innovation (PPI) with its own distinctive characteristics, not shared in other contexts (Georghiou et al., 2014; Hameri, 1997; Vuola & Hameri, 2006). In the medium and long-term, innovation and entrepreneurship

✉ Paolo Castelnovo
paolo.castelnovo@unimi.it

¹ Department of Economics, Management and Quantitative Methods, University of Milan, Via Conservatorio, 7, 20122 Milan, Italy

² Roma Tre University and Rossi-Doria Centre, Rome, Italy

might emerge from BSC through three main channels: (1) technological breakthroughs leading to start-ups and spin-offs involving BSC's employees and collaborators; (2) inventions and patents filed by researchers at universities and firms collaborating with the BSC; (3) new business opportunities for BSC's industrial partners.¹

BSC's suppliers are required to deliver new products with technology specifications developed for scientific purposes and for which a market might not exist yet. Consequently, there are risks related to the specificity of such technologies and a long time might be required for suppliers to absorb the knowledge acquired during their collaboration with the BSC and to translate it into an innovation output, such as a patent application.

This paper investigates the effect of procurement through BSC on firms' innovation output. We analyze a sample of suppliers of the European Organization for Nuclear Research (CERN) with the aim of estimating the average time lag that separates the beginning of the procurement relationship with CERN and the subsequent filing of a patent, if any. We label this time span as "*gestation lag of innovation*" (or, more shortly, "*gestation lag*"). We create a unique dataset with information on firms collaborating at the development of the world biggest research infrastructure, the Large Hadron Collider (LHC). The fact that firms in our sample received the first LHC-related order over a long-time span (1995–2006) delivers a natural partition of industrial partners into "suppliers" and "not-yet-suppliers". The variation over time of firms' status is then exploited to trace the impact of CERN on patent applications and to estimate the time lag that runs from the beginning of the procurement relationship to the filing date of the patent. Our econometric analyses show that there is positive and statistically significant effect of CERN on patent applications only with a delay of 5 years from the beginning of the procurement relationship and, hence, highlight the existence of a sizeable gestation lag.

The paper contributes to two different strands of the literature. First, we refer to the literature investigating the time needed to translate private and public research into innovation output and for its commercial exploitation. It is shown that estimates of such time lag are highly heterogeneous and depend on the underlying assumptions (i.e. definition of time lag), level of analysis (e.g. country vs. industry) and industry (e.g. pharmaceutical vs computer science). However, to the best of our knowledge, this is the first paper investigating the "*gestation lag of innovation*" in BSC.

Second, our study contributes to the literature on the economic effects of BSC (Battistoni et al., 2016; Florio & Sirtori, 2016; Helmers & Overman, 2017) focusing on PPI through BSC as a channel of transmission of benefits to firms (Georghiou et al., 2014; Hameri, 1997; Vuola & Hameri, 2006). The consensus view emerging from this literature points to the existence of a positive association between PPI through BSC and firms' economic performance and innovation output (Autio et al, 2004; Castelnovo et al., 2018; Schmied, 1977). Our paper advances knowledge on the innovative outcomes generated by PPI through BSC, shedding new light on the time required for the BSC suppliers to "absorb" the technological content of the order and translate it into a patent application.

The rest of the paper is organized as follows. Section 2 reviews the literature on the gestation lag of innovation and PPI through BSC and outlines our research hypotheses. Section 3 presents the data, descriptive statistics and the econometric model, while results and robustness checks are discussed in Sect. 4. Section 5 concludes. An Appendix completes the paper.

¹ In some cases, these three channels may overlap (e.g. suppliers employing former BSC staff or that are spin-offs of universities involved in experiments at the BSC).

2 Literature review and research hypotheses

2.1 Gestation lag of innovation

In this paper we focus on the time distance that runs from the beginning of the procurement relationship with CERN to the filing date of a patent. We refer to this time lag as the “*gestation lag of innovation*”. Timing is relevant for policy makers who decide how much to invest in BSC; in fact, estimates of both the social and private rates of return to research depend on the definition and quantification of such time lags, Griliches (1979).² As shown in Table 1, while several papers investigated the time lag needed to translate private, public and academic research into innovation output or to exploit it commercially, there is no previous literature on the gestation lag of innovation that is specific to BSC.

Mansfield (1991, 1998) reported that the mean time interval between relevant academic research and the first commercial introduction of new products or processes deriving from it is 6–7 years. The estimated time span between the appearance of academic research and its effect on productivity in the form of knowledge absorbed by an industry might be even longer, on average approximately 20 years (Adams, 1990). In the case of computer science and engineering—two cases especially relevant for what concerns CERN—the time lag is 10 years (Adams, 1990). Heher (2006) showed that it can take up to 10 years for an institution, and 20 years for a country, to attain a positive rate of return from an investment in research and technology transfer. In the pharmaceutical industry, the time-lag between research, development, and commercialization can reach up to 20 years (see Sternitzke, 2010 and Toole, 2012). In contrast, earlier analyses by Pakes and Griliches (1980), Hausman et al. (1984), Hall et al. (1984) reported that about a year is necessary for translating R&D expenditure by firms into a patent application. Rapoport (1971) and Wagner (1968) distinguished between the lag from project inception to completion and the lag from project completion to commercial application, finding a total lag of around 2–3 years.

2.2 Public procurement for innovation

Providing empirical evidence about the economic effects of procurement through BSC is key for governments contributing to their budgets (Florio et al., 2018; Florio et al., 2018). In fact, BSC are financed not only because of the promise of significant scientific discoveries, but also in the hope that a radical technological innovation—yielding spillovers in different application domains (Dahlin & Behrens, 2005)—might arise as a side-effect, while attempting to advance human knowledge (Hallonsten, 2014). The search for breakthrough or general-purpose technologies is a relevant target of innovation policies (Edler & Fagerberg, 2017). Being sources of aggregate productivity growth (Crépon et al., 1998), R&D and innovation represent key drivers of short and medium-term business cycle fluctuations

² Mansfield (1968) used time lags between the investment in academic research and the industrial utilization of their findings in the computation of the social rate of return to academic research. Pakes and Schankerman (1984) provided a different definition of “gestation lag”. These authors decompose the “total R&D lag” into the lag between project inception and completion—the “gestation lag”—and the time from project completion to commercial application—the “application lag”. They showed that the time lag between the deployment of research resources and the beginning of the stream of private revenues from their commercial applications heavily influences the private rate of return to research.

Table 1 Time lag in the literature

Author	Research agent	Beneficiary of research impact	Time lag	Definition of time lag	
				From	To
Adams (1990)	Academic	Manufacturing industry	20 on avg	academic research	impact on productivity
Heher (2006)	Academic	Research institutes and national economy	10–20	academic research	commercial application
Mansfield (1991, 1998)	Academic	Manufacturing industry	6–7	academic research	commercial application
Toole (2012)	Public	Pharmaceutical industry	Up to 20	project inception	commercial application
Sternitzke (2010)	Public and private	Pharmaceutical industry	Up to 20	project inception	commercial application
Hall et al. (1984)	Private	Manufacturing industry	1 on avg	project inception	patent application
Hausman et al. (1984)	Private	Manufacturing industry	1 on avg	project inception	patent application
Pakes and Griliches (1980)	Private	Variuos industries	1 on avg	project inception	patent application
Rapoport (1971)	Private	Chemical, machinery, and electronic industry	0.8–2.1	project inception	project completion
			0.2–0.4	project completion	commercial application
			1.2–2.5	project inception	commercial application
			1.1–1.2	project inception	project completion
			1.0–1.5	project completion	commercial application
Wagner (1968)	Private	Variuos industries	2.1–2.7	project inception	commercial application

(Basu et al., 2006; Comin & Gertler, 2006; Kung & Schmid, 2015) and forces affecting long-run economic growth (Bresnahan & Trajtenberg, 1995; Schaefer et al., 2014).

Although a fast-growing strand of the literature highlights that procurement through BSC is quite distinct from other public procurement practices (Georghiou et al., 2014; Hameri, 1997; Vuola & Hameri, 2006), our analysis can be cast in the broader debate about public procurement and PPI (see e.g. Lember et al., 2015). Public procurement entails the purchase of goods or services by governments, public institutions and state-owned enterprises; as such, it can be seen as a policy tool aimed at boosting a country's internal demand. A distinction is usually made between regular procurement and PPI. The former involves the purchase of "off-the-shelf" products and services that do not require research and development activities, while PPI refers to public authorities requiring products or services that are not yet available, but which could be developed within a reasonable timeframe (Edquist et al., 2000).

PPI represents an important demand-side innovation policy (Edler & Georghiou, 2007; Edquist & Zabala-Iturriagagoitia, 2012; Edquist et al., 2015; Uyarra & Flanagan, 2010) having the potential of fostering technical change and supporting the development of novel technologies. This is particularly true when the development of sophisticated products is required (Salter & Martin, 2001) and in industries characterized by high-risk that cannot be borne entirely by the private sector (Mazzucato, 2016). The demand-pull effects of PPI on innovation seem more effective than other policies such as R&D subsidies and tax credits (Lichtenberg, 1990; Guerzoni & Raiteri, 2015). Moreover, PPI can either be a complement or an alternative to supply-side policies (Edler & Georghiou, 2007). For instance, Aschhoff and Sofka (2009) compared four public policies to stimulate innovation: public procurement, regulation, R&D subsidies, and subsidies to universities and research institutions. They found a positive effect of public procurement on market opportunities, and no difference between public procurement and access to knowledge created by universities and research institutes, while the other channels seem to be less effective. Raiteri (2018) exploited patent data to investigate the role of PPI in supporting private innovation activities, concluding that it may lead to—or accelerate—the deployment of a new general-purpose technology.

In most cases public organizations do not take part directly in the innovation process carried out by their suppliers, on the contrary, CERN makes available to suppliers its infrastructures and know-how to carry out experiments, and often takes an active part in the design and development of the products demanded (Nielsen & Anelli, 2016). While we focus specifically on CERN procurement, it is worth mentioning that there is a growing literature that studies the impact of other BSCs procurement on firms' innovation output and economic performance.³ The evidence mainly comes from surveys to suppliers. Fernandes et al. (2014) examined the impact of the European Southern Observatory procurement, uncovering technological benefits that they impute to technical challenges posed by collaborations. Castelnovo and Dal Molin (2020) investigated the benefits for 150 suppliers of the Italian Institute of Nuclear Physics (INFN), confirming the existence of technological learning. Several studies analysed the economic impacts of Danish companies' involvement and cooperation with the European Space Agency (see e.g. Cohendet, 1997; Bach et al. 2002; Danish Agency for Science, 2008), focusing mainly on the economic return

³ For an extensive review of the effect of technology transfer and knowledge spillovers from Big Science see Scarrà and Piccaluga (2020).

(utility-expenditure ratio) generated by this intergovernmental organisation rather than on its innovation impact.

2.3 CERN procurement and its economic effects

CERN—founded after World War II with the aim of studying the basic constituents of matter—operates the largest particle physics laboratory in the world and is a leading example of BSC. CERN research is publicly funded by 23 Member States.⁴ In turn, Member States expect an industrial return proportional to their contribution to CERN annual budget. The construction cost of the LHC was more than 4 billion of Swiss Francs and involved almost 1300 firms (CERN, 2019).

CERN procurement contracts often require cutting-edge technologies and radical innovations combined with an intense collaboration with its industrial suppliers. CERN suppliers are exposed “to a highly diverse knowledge environment” (Autio et al., 2004: 110) which could positively impact on their expected future innovativeness, productivity and profitability. Procurement through CERN is thus not a form of “general public procurement” (i.e. buying off-the-shelf products), but rather a form of PPI with its own specific features.

The effects of CERN procurement on the innovation output of its industrial partners have been investigated with a variety of methodologies ranging from case studies to econometric analyses: the consensus view emerging from this strand of the literature points to a positive association between CERN procurement and firms’ innovation output. CERN’s procurement has led to technological advances in superconductivity, cryogenics, electromagnets, ultra-high vacuum, distributed computing, rad-resistance materials, and fast electronics (Evans, 2009; Giudice, 2010).

A survey of CERN suppliers showed that collaboration with CERN contributed—along with other factors—to product innovation and new R&D (Autio, 2014). Castelnovo et al. (2018) relied on a simultaneous equation model to show that, after becoming CERN industrial partners, firms generally experienced a rise in R&D, patents, productivity, and profits. Closely related contributions are those by Amaldi (2012), Nielsen and Anelli (2016), and Battistoni et al. (2016) who highlighted that, thanks to the collaboration with CERN, several firms were subsequently able to develop new products for customers in other markets. Florio et al. (2018), Florio et al. (2018) showed that CERN procurement significantly affected suppliers’ innovative performance when a relational governance is in place through cooperative relations (i.e. exchanges implying that CERN and suppliers regularly cooperate to deal with complex information that is not easily transmitted or learned). Based on evidence from 14 Swedish firms, Aberg and Bengtson (2015) pointed out that firms’ product and process innovation occurs mostly when a development project is in place (i.e. when CERN invites firms to participate in developing products that cannot be bought off-the-shelf). Vuola and Hameri (2006, p. 3) relied on nine in-depth case studies and concluded that CERN is “a most fertile ground to enable and boost industrial innovation”. Autio et al.

⁴ As of November 2019, CERN’s Member States are: Austria, Belgium, Bulgaria, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Israel, Italy, Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Spain, Sweden, Switzerland and United Kingdom. Cyprus and Slovenia are Associate Member States in the pre-stage to Membership. Croatia, India, Lithuania, Pakistan, Turkey and Ukraine are Associate Member States (see: <https://home.cern/about/who-we-are/our-governance/member-states>).

(2014), based on three in depth case-studies, discovered innovation benefits accruing to the firms involved by CERN through prototypes' mocking-up and experimentation.

2.4 Research hypotheses

Collaborating with a BSC involves a risk-return trade-off. Firms engage in contracts with a BSC—such as CERN—not mainly for an immediate profit, but also because they foresee future competitive advantages. Technological learning and reputational effects gained through CERN procurement can be exploited in the relations with other BSC, or with customers in different markets. CERN features a unique combination of both PPI and cutting-edge research. Its suppliers face entirely new technological challenges that often require advancing the technological frontier, acquiring new scientific knowledge through R&D and developing radical innovations. This is possible only in virtue of the close collaboration and frequent meetings with CERN.

However, entering in a contract with a BSC is not entirely without risk. In fact, procurement of technologies required for the purpose of scientific research is often not profitable in the short-term; moreover, how benefits and costs will be balanced in the medium and long-term is uncertain. The probability that firms can profit from their collaboration with CERN in the medium or long-term depends on their ability to enrich their absorptive capacity over time. Absorptive capacity captures firms' ability to absorb external knowledge and hence to benefit from the interaction process with CERN (Cohen & Levinthal, 1990). Lastly, because of the complexity and novelty of the requests from CERN, the process of absorption of new ideas might be relatively slow. This suggests that there might be a significant time lag that separates the beginning of the relation with a CERN and its impact on a firm's innovation output, if any.

A recent survey by Sirtori et al. (2019) collected 28 case studies with in-depth interviews to representatives of firms that have received at least one order from CERN. Most of the interviewees declared that the relation with CERN improved the technical know-how and the reputation of firms. However, they also highlighted some side-effects of PPI through BSC: it is an expensive process that involves frequent interactions with the BSC, additional R&D, financial risk, specific fixed investment and training costs. Moreover, customized solutions may not be profitable in other markets. The high degree of uncertainty about how, when and whether these costs will be balanced by additional gains has inspired our research hypotheses and the econometric design used to assess the existence of a gestation lag separating the beginning of the procurement relationship with CERN and the subsequent filing of a patent, if any.

In the light of these considerations and evidence from earlier literature, we formulate the following research hypotheses. Firstly, firms engaged in a procurement relation with CERN experience a learning process that ultimately leads to an innovation output. Secondly, given the complexity and novelty of a Big Science Centre technological requirements, the process of absorption of new ideas might be relatively slow, leading to a significant gestation lag of innovation.

Both the research hypotheses rely on the *number of patent applications* filed by CERN suppliers as empirical proxy of innovation output. Patents are widely used as a proxy of innovation output (Aghion et al., 2013; Crépon et al., 1998; Jia et al., 2019; Raiteri, 2018), although we are aware that they may underestimate the innovation effect of BSC procurement since many firms might not patent their innovations or use alternative forms of intellectual property protection (see Hall et al., 2014 and Dziallas & Blind, 2019 for a critical

overview of innovation indicators). In fact, as it has been observed in other contexts, innovation can take a more indirect pathway, when a patent cites non-patent literature arising from scientific advances (OECD, 2009), including knowledge created at BSC (see Florio, 2019, page 140 and Catalano et al., 2020 for a recent application to a synchrotron light source facility).

3 Data, statistical methods and preliminary analyses

3.1 Data

We have collected information on a sample firms collaborating at the development of the world biggest research infrastructure: the LHC. The LHC project was approved by the CERN council in December 1994. The experiments at the LHC started in September 2008 and in July 2012 the discovery of the Higg's boson—for which François Englert and Peter Higgs were awarded with the 2013 Nobel Prize in physics—was announced to the public.

Our dataset summarizes information from three main sources. First, we identified firms that over the 1995–2006 period received at least one LHC-related order above 10,000 CHF from the database maintained by the CERN Procurement and Industrial Services Group.⁵ From this source, we also retrieved the date marking the beginning of the procurement relationship, the “activity codes” used by CERN to classify purchases from its suppliers, the number and the total amount of LHC-related orders supplied by the firm. Second, we sourced balance-sheet data from the ORBIS database maintained by Bureau van Dijk. We collected information on the geographical location, size, incorporation date and sector of activity (based on NACE 2 digits codes) of firms, as well as data on the amount of their intangible fixed assets.⁶ Lastly, we collected information on patents filed by LHC suppliers over the 1993–2006 from the PATSTAT and ORBIS Intellectual Property databases. Specifically, we obtained data about patents' application date, patent office(s) where they were filed and patents' families (in case of multiple applications). Patent families are built according to the DOCDB simple patent family concept,⁷ that is a collection of related patent applications covering the same technical content.

Over 1200 entities are recorded in the CERN procurement data in relation to the LHC construction, but most of them played a relatively minor role, as they are university departments, research institutes, small companies or other entities for which financial data are not available in ORBIS. The screening process and the careful merging of data from different sources resulted in a panel dataset of 263 firms.⁸

With count data models we aim to quantify the incremental number of patents that firms filed after the beginning of their relationship with CERN. Hence, in our preferred specification we focus on firms have filed at least one patent since their incorporation (38% of our

⁵ The rationale for excluding low value orders is that they are unlikely to generate knowledge spillovers for firms. Moreover, we notice that experimental collaborations, such as ATLAS and CMS, have some procurement autonomy, so their orders are not considered, except when they are directly managed by CERN. For additional details see: <http://procurement.web.cern.ch/procurement-strategy-and-policy>

⁶ We rely on unconsolidated financial statements.

⁷ <https://www.epo.org/searching-for-patents/data/bulk-data-sets/docdb.html>

⁸ The reduction in sample size is mainly driven by missing observations for some control variables, notably intangible assets.

Table 2 Variables used in the empirical analysis: 1993–2006

Variable	Description
p_{it}	Yearly number of patent applications (based on patent families)
$CERN_{i,t}^k$	Dummy = 1 if procurement relation with CERN started k years ago
$Hi-Tech_i$	Number of orders classified as hi-tech by CERN experts
$Medium_i$	Dummy = 1 if the company is classified as “medium-size” in the ORBIS database
$Large_i$	Dummy = 1 if the company is classified as “large” in the ORBIS database
$V Large_i$	Dummy = 1 if the company is classified as “very large” in the ORBIS database
$Avg. p_i$	Average number of patent applications per year before 1993
IFA_{it}	Logarithm of Intangible Fixed Asset
$Order_i$	Logarithm of the total amount of LHC orders received by firm i
$pct_{c,t}$	Contribution of country c to CERN budget in year t – expressed as percentage of the total contribution of Member States
α_c	Dummy = 1 if firm is in country c
α_s	Dummy = 1 if firm is in sector s
α_t	Dummy = 1 if year = t

sample) and exclude those that have never filed any patents. As a result, our main analysis is carried out on a sample of 100 firms that filed patent applications. A robustness check is performed on the larger sample of 263 companies.

Since firms have received their first LHC-related order over a long time-span, we have a natural partition of statistical units into “suppliers” and “not-yet-suppliers”. The sample period of our analyses begins in 1993 when the procurement for the LHC had not yet started (i.e. all firms were not-yet-suppliers) and ends in 2006 so that we are left with 27 entities that still have the “not-yet-supplier” status and act as a control group. Firms in this group received the first order from CERN in 2007 or 2008.

3.2 Empirical strategy

The dependent variable in our regressions is the yearly number of patent applications filed by firms. Further details about the definition of this outcome variable are provided in Sect. 3.2.1. The expected number of patents applications $p_{i,t}$ —for $i = 1, \dots, 100$ and $t = 1993, \dots, 2006$ —can be written as follows:

$$E(p_{i,t} | \mathbf{X}_{isc,t}) = \exp \left(\sum_k \gamma_k CERN_{i,t}^k + \mathbf{X}_{i,t} \boldsymbol{\beta} + \delta pct_{c,t} + \sum_s \alpha_s + \sum_c \alpha_c + \sum_t \alpha_t \right) \quad (1)$$

Table 2 describes the variables entering our main empirical specifications. All regressions include a set of dichotomous variables ($CERN_{i,t}^k$) capturing the time-varying effects of CERN procurement on firms’ patent applications. The dummy variable $CERN_{i,t}^k$ is set to one if the procurement relation with CERN started k years ago, while is zero otherwise. Estimates of the coefficients on these variables quantify the gestation lag of innovation for CERN suppliers. In fact, the time variation of the status of LHC’s suppliers can be exploited in the econometric analysis to evaluate how CERN procurement has affected their patenting activity. The date when the first order is awarded

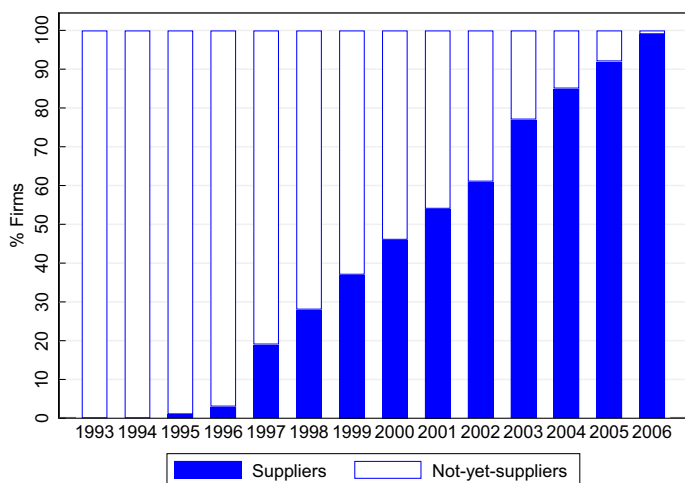


Fig. 1 Cumulative share of suppliers per year

marks the beginning of the procurement relationship, while the filing date of the patent proxies the time when the innovation output is produced.

The empirical investigation relies on a panel dataset for the 1993–2006 period. Figure 1 shows the cumulative share of suppliers per year from 1993 to 2006.

The beginning of the observation period pre-dates the start of the construction of the LHC because we are interested in the timing of the “CERN effect”. To be sure that such effect does not precedes the start of the procurement relation, our main empirical specifications include a set of dummy variables taking value one if the statistical unit will become a CERN supplier in one or two years. The coefficients of such dichotomous variables—that act as leads—should never be statistically distinguishable from zero in case of a causal interpretation of the “CERN effect” on firms’ patent applications.

Our empirical design needs to deal with some specific features of the CERN procurement process. First, the transition from the status of “not-yet-supplier” to that of “CERN supplier” does not happen in a single year common to all firms. Second, the highly specialized nature of firms collaborating with CERN and the characteristics of goods and services they provide makes it impossible to find a meaningful control group of entities that are structurally similar to CERN suppliers in our sample. In fact, observable covariates, such as firms’ financial and technological profile are based on broad classifications that would lead only to a spurious match with CERN’s suppliers. Lastly, we are mainly interested in estimating the average time lag that separates the beginning of the procurement relationship with CERN for each firm and the subsequent filing of a patent, therefore the usual DiD approach would not allow to estimate the gestation lag. We thus draw our empirical strategy from the approach developed by Stevenson and Wolfers (2006). For our purposes, such approach entails estimating the coefficients on a set of dichotomous variables measuring the years from the start of the collaboration with CERN. This is a natural strategy to be adopted in our setting. Further details are provided in the next sections.

3.2.1 Dependent and control variables

3.2.1.1 Dependent variable The dependent variable is the yearly number of patent applications filed by LHC's suppliers. There are different ways of counting patent applications (see OECD, 2009; Oldham, 2019): we rely on patent families, specifically on the DOCDB simple patent family concept. Hence, our dependent variable counts either patent family applications covering the same technical content or individual patents that do not belong to any family. Focusing on patent families allows both to avoid over-counting patents that are members of the same family (i.e. patents referring to the same or similar inventions), and to associate to each patent family a unique priority filing (i.e. the oldest priority date) which should in principle be the one closer to the invention.

Descriptive statistics highlight that most companies have filed a limited number of patents over the time span considered: 23% of suppliers filed just 1 patent, 53% up to 5 patents, while only 24% filed more than 10 patents. This implies that patents are quite homogeneously distributed among suppliers, hence results should not be driven by few large companies that filed most of patents in our sample.

3.2.1.2 Control variables We consider several control variables in our regressions. The vector $\mathbf{X}_{i,t}$ includes firm-level variables used to capture factor that might affect the absorptive capacity of firms (Cohen & Levinthal, 1990). Following Autio et al. (2004), we posit that industrial learning effects of BSC are generated by their dyads with firms (i.e. a “dyad” is a relationship between organizations). Whether firms are able to exploit knowledge spillovers will depend both on the firms' absorptive capacity and on the absorptive capacity of the dyad (Lane & Lubatkin, 1998), namely the sum of relation-specific assets that facilitate both knowledge disclosure and knowledge communication within the dyad.

The effect of firm size on the probability to patent is captured by dichotomous variables classifying firms into small (i.e. that is the reference category, accounting for 6% of the total), medium (21%), large (47%) and very-large (26%).⁹ Firm size is expected to be positively associated with the probability to patent since large firms can exploit economies of scale, have access to a broader pool of highly qualified collaborators (Fernández-Olmos & Ramírez-Alesón, 2017), have more financial resources to afford the costly process of patent application (Blind et al., 2006; Block et al., 2015; Leiponen & Byrna, 2009, among the others).

Next, we include a discrete variable (*Hi-tech_i*) that specifies the number of orders classified as hi-tech by CERN experts received by each firm.¹⁰ Firms receiving high-tech orders might have higher absorptive capacity and be more capable of translating hi-tech orders from BSC into marketable innovations (Castelnovo et al., 2018; Edquist & Hommen, 2000; Hameri & Vuola, 1996). In our sample, firms receiving at least one high-tech order account for 70% of the total; high-tech orders account for more than one-half in 65% of the firms, and 53% of the firms are entirely involved in high-tech orders.

We also consider the (logarithm of) total amount of LHC orders received by each firm (*Order_i*) as a proxy of the involvement and continuity of the procurement relationship with CERN (Åberg & Bengtson, 2015). Firms in our sample have supplied to CERN up to nine

⁹ This classification—provided in the ORBIS database—exploits information on the amount of total assets, operating revenues and the number of employees.

¹⁰ Details on the construction of this variable are presented in the Appendix A1.

LHC-related orders with median value above 67,000 CHF. In fact, as previously mentioned, long-lasting collaborations for hi-value orders often involve repeated interactions with CERN that might boost the learning effects of procurement (Autio et al., 2004; Florio, Bastianin et al., 2018; Florio et al., 2018).

Moreover, following Blundell et al. (1999), we use the mean pre-sample patent count (i.e. $\text{Avg. } p_i$, the average number of patent applications per year before 1993) to capture firms' unobserved propensity to patent. $\text{Avg. } p_i$ measures unobservable firm-specific fixed effects, reflecting any permanent differences in the level of innovation across firms which are independent of CERN procurement (see e.g. Hausman et al., 1984 and Aghion et al., 2013).

Last but not least, firm-level variables include (the logarithm of) Intangible Fixed Asset¹¹ ($IFA_{i,t}$) as a proxy of expenditure in R&D (Chan et al., 2001; Leoncini et al., 2017; Marin, 2014). R&D expenditure is a key control variable for analysing patent activity (Aghion et al., 2013; Gurmu & Pérez-Sebastián, 2008; Hall et al., 1984; Hausman et al., 1984), but, unfortunately, our proxy features many missing observations. To include $IFA_{i,t}$, while maintaining the size of the panel dataset reasonably large, we excluded firms reporting more than 4 missing values.¹² Due to the presence of a large number of zeros (approximately 21% of the observations), the sampling distribution of $IFA_{i,t}$ is highly skewed with sample average equal to 1170 thousand Euros and sample median equal to 43 thousand Euros.

To control for country heterogeneity, we include the contribution of country c to CERN budget in year t —expressed as percentage of the total contribution of Member States— $pct_{c,t}$. 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. In fact, CERN procurement rules are designed to try to balance orders across its Member States. Additional factors capturing unobservable country-specific heterogeneity are modelled with country dummy variables, α_c . Similarly, we include sector specific dummy variables, α_s , whose aim is to allow for heterogeneity at the sector-level, and add time dummies (α_t) to control for common macroeconomic shocks hitting all firms in the sample. In fact, macroeconomic conditions are expected to affect the profitability of firms, the business environment where they operate and the relationship between technology collaboration networks and innovation performance (Fernández-Olmos & Ramírez-Alesón, 2017).

3.3 Preliminary analyses

3.3.1 A first look at the gestation lag

Figure 2 displays the number of patent applications per firm over “relative years”, denoted as “ k ”. A relative year is the number of years after (positive distance) or before (negative distance) the procurement event, which is set, firm by firm, as year zero. Therefore, $k > 0$ for a specific firm indicates that the first LHC order was received k years ago in our data.

¹¹ 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.

¹² 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 regressions.

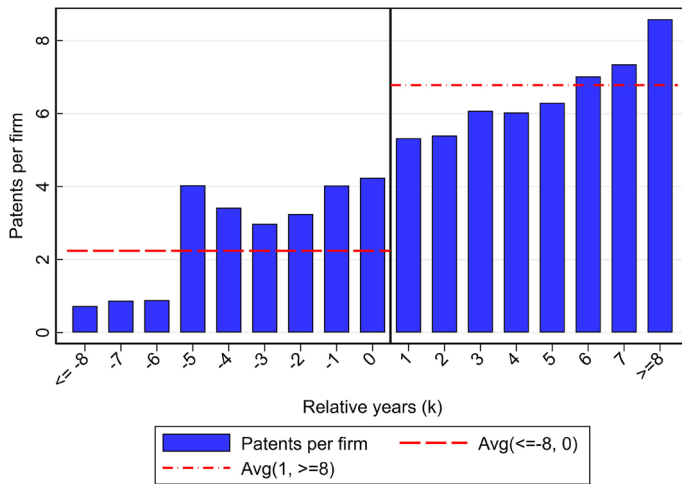


Fig. 2 Patent applications per firm: relative years Notes “Relative years (k)” measures the time distance from the first LHC order. Therefore, $k > 0$ indicates that the first order was received k years ago and $k < 0$ indicates that the order will be received in k years. For each k the figure reports the number of patents per firm. The dashed line denoted as “Avg($-11, 0$)” represents average number of patents per firm over relative years $k = -11, \dots, 0$, that is before the start of the LHC procurement. Similarly, the dash-dotted line denoted as “Avg($1, 11$)” represents average number of patents per firm over relative years $k = 1, \dots, 11$, that is after the start of the LHC procurement. On the x-axis we use “ > 8 ” (“ < -8 ”) to denote relative years $k = 9, 10, 11$ ($k = -9, -10, -11$)

The two horizontal lines in Fig. 1 are sample averages for $k \leq 0$ and $k > 0$. Before the beginning of the relationship with CERN (i.e. $k \leq 0$), the sample average is 2.2 patent applications per firm, while after receiving the first LHC order (i.e. $k > 0$) the sample average rises to 6.8. Lastly, we emphasize that the figure exhibits three sharp peaks in correspondence of $k = 6, 7, \geq 8$ thus suggesting that a sizeable gestation lag separates the procurement event and any sizeable impact on the number of patents. All in all, the figure provides evidence that a positive association between CERN procurement and innovation shows up after some years from the beginning of the relationship with CERN.

3.3.2 The relationship between CERN orders and patents

For the internal validity of our results, a crucial point is to assess to what extent patents filed by firms in our sample are related with the orders that CERN awarded to them. We conduct such a check in several ways. First, we check if the patent applications of firms in our sample cite as prior-art patents filed by the CERN itself or other CERN contractors. We searched in Orbis IP patents filed by CERN. These are just 331 because CERN’s policy is generally against the use of patents. These 331 patents have been cited in 1210 other patents; however only 8 firms in our database are the owner of patents citing CERN’s patents. Second, we check if the activity code of the CERN contract is compatible in terms of technological area to the patent application considered for that contract. For this purpose, we collected detailed data about patent applications for the sample of firms used in our empirical analysis. We restrict the analysis to approximately 19,200 patent applications by firms that started filing patents after they were awarded the first LHC-order and

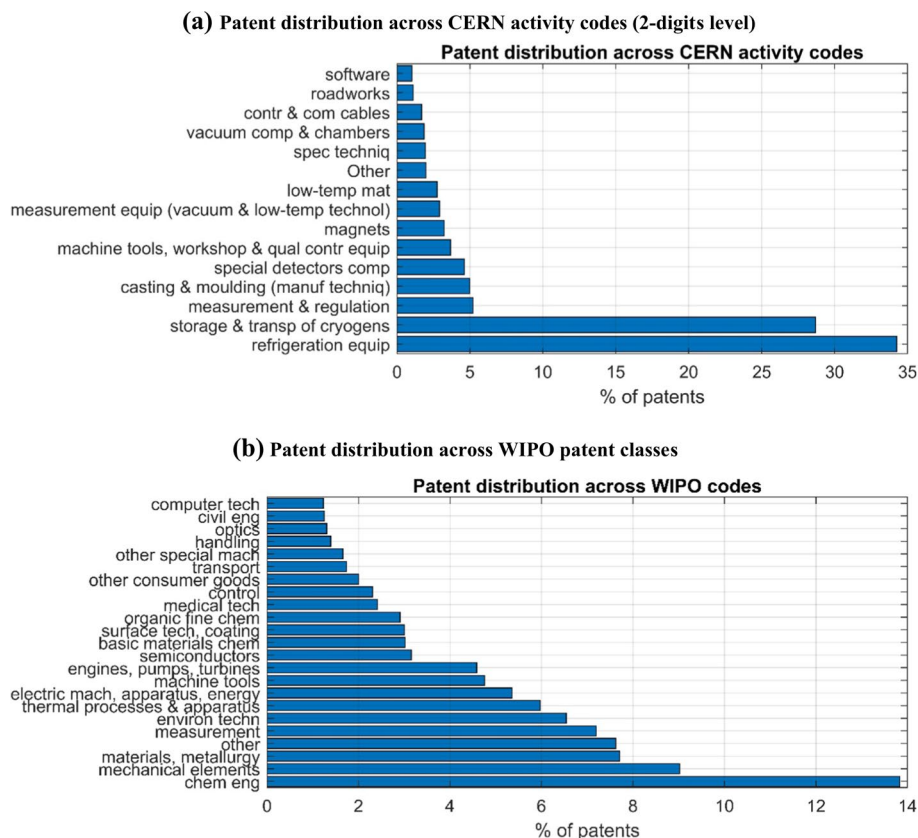


Fig. 3 Distribution of patents over CERN activity codes and WIPO codes *Notes* both in panel (a) and (b) of the figure the category labelled as “Other” is a residual class that collects codes where there is less than 1% of the total number of patents (approx. 19,000)

consider applications until 2019.¹³ We sourced the title of patents filed in any patent office and the patents’ technological classification—as provided by the World Intellectual Property Organization (WIPO)—from the ORBIS Intellectual Property database. Moreover, we classify each firm using the CERN activity code associated to its first LHC related order.¹⁴

Figure 2 displays the distribution of patents across CERN activity codes (Panel a) as well as across WIPO codes (Panel b). Relying on activity codes provided by CERN, it can be seen from Fig. 3a that the three most frequent classes are associated with orders related to “refrigeration equipment”, “storage and transportation of cryogenes” and “measurement and regulation”. Focusing on Fig. 3b, we see that the four most frequently observed WIPO codes are “chemical engineering”, “mechanical elements”, “materials—metallurgy” and

¹³ While this might somehow speak to a causal interpretation of our results, we stress that there might be confounding factors that we do not control for. A more extensive analysis of the causal nexus between orders and patent applications is beyond the scope of the paper.

¹⁴ Note that the number of firms whose orders are assigned to more than one activity code is small. See Appendix A1.

“measurement”. In terms of hi- vs lo-tech classification of firms, descriptive statistics show that 97% of the patents are associated to firms classified as hi-tech. Comparison of the two panels shows that LHC orders are technologically related with the patent applications filed by firms in our sample. A summary of the association between WIPO codes and CERN activity codes is provided by the Cramér’s V measure for categorical variables. This statistic shows the strength of the association between two categorical variables. The value of the Cramér’s V between WIPO codes and CERN activity codes is 39% (p -value=0.000) and is statistically distinguishable from zero at common significance levels, thus confirming our finding.¹⁵

4 Results

4.1 Is there an innovation gestation lag?

The existence of a gestation lag of innovation for CERN suppliers is investigated in Tables 3, where we regress the number of patent applications per year on a set of dummy variables that track the timing of CERN impact on patents, while controlling for several covariates that might act as confounding factors. Since we want to be sure that the change in firms’ patenting activity post-dated the beginning of the collaboration with CERN, all specifications also include the leads of the dummy variable marking the beginning of the procurement relationship. More precisely, the variable denoted as $CERN_{i,t}^{(-2,-1)}$ indicates that firm i will receive the first order from CERN in at most a couple of years.¹⁶

Negative Binomial models reported in Table 3 are well suited for capturing the count nature of the dependent variable. In fact, empirical evidence presented at the bottom of Table 3 highlights that our data might be over-dispersed and therefore violate the assumption of equi-dispersion underlying the Poisson regression model. The Poisson model would imply that the variance of the number of patents in each period is equal to its expected value during the same time frame (i.e. equi-dispersion). The Negative Binomial distribution includes the Poisson as a special case and allows for both under- and over-dispersion. Over-dispersed variables have variance greater than the expected value.

Column 1 reports our baseline specification. The coefficients of $CERN_{i,t}^k$ are statistically significant at the 95% confidence level only for $k \geq 5$. This suggests that there is a considerable time lag separating LHC procurement from changes in the number of patent applications. Moreover, the coefficient on $CERN_{i,t}^{(-2,-1)}$ is never statistically distinguishable from zero. This is consistent with the expectation that any impact of LHC procurement on firms’ patents should post-date the start of the relationship with CERN and strengthen the causal interpretation of our results. The remaining columns include robustness checks and are commented in Sect. 4.2.

Potential selection bias arising from a not randomized sample have been mitigated with a large set of control variables including firm, sector and time-fixed effects. For instance,

¹⁵ As pointed out by one of the reviewers the standard Cramér’s V measure could overestimate the level association and a correction might thus be needed to address such bias. This correction is inversely proportional to the number of instances (more than 17,000 in our example). The corrected Cramér’s V measure is equal to 33%.

¹⁶ Results in the paper aggregate leads and lags in blocks of two or more years. For instance, $CERN_{i,t}^{(1,2)}$ identifies firms that received LHC one or two years before time t . Results available from the authors upon request highlight that estimates of the gestation lag are not affected by such aggregation and remain valid even when using dummy variables for each lead and lag.

Table 3 Main results and robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)
Baseline		Larger sample: 5 missing in IFA allowed	Zero-inflated Negative Binomial	No leads	Age proxy: ln(firm age)	Age proxy: startup
$CERN^{(-2,-1)}_{it}$	0.299 (0.190)	0.281 (0.227)	0.196 (0.276)		0.294 (0.185)	0.243 (0.184)
$CERN^0_{it}$	0.180 (0.237)	0.351 (0.282)	0.583 (0.374)	0.022 (0.214)	0.186 (0.233)	0.124 (0.231)
$CERN^{(1,2)}_{it}$	0.339 (0.268)	0.346 (0.256)	0.637** (0.320)	0.157 (0.235)	0.333 (0.259)	0.231 (0.255)
$CERN^{(3,4)}_{it}$	0.418 (0.277)	0.493* (0.267)	0.799*** (0.305)	0.195 (0.240)	0.444 (0.273)	0.315 (0.264)
$CERN^{(5,6)}_{it}$	0.620* (0.320)	0.744** (0.316)	0.805*** (0.304)	0.369 (0.279)	0.639** (0.313)	0.501* (0.304)
$CERN^{(7,8)}_{it}$	0.951*** (0.368)	0.952*** (0.354)	1.140*** (0.404)	0.672** (0.323)	0.976*** (0.360)	0.816** (0.353)
$CERN^{>8}_{it}$	1.036** (0.507)	1.118** (0.495)	0.746* (0.385)	0.729 (0.460)	1.021** (0.494)	0.860* (0.478)
<i>Hi-Tech_i</i>	0.017*** (0.003)	0.018*** (0.003)	0.013*** (0.003)	0.017*** (0.002)	0.023*** (0.003)	0.020*** (0.002)
<i>Medium_i</i>	2.852*** (0.987)	2.810*** (1.001)		2.783*** (0.983)	2.847*** (1.040)	2.582** (1.032)
<i>Large_i</i>	3.035*** (0.980)	3.016*** (0.993)		2.981*** (0.975)	2.981*** (1.032)	2.879*** (1.031)
<i>V Large_i</i>	5.790*** (0.986)	5.813*** (1.001)		5.711*** (0.980)	5.536*** (1.031)	5.410*** (1.033)
<i>Avg. P_i</i>	0.012*** (0.004)	0.011*** (0.004)		0.012*** (0.004)	0.010** (0.004)	0.012*** (0.004)

Table 3 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline	Larger sample: 5 missing in IFA allowed	Zero-inflated Negative Binomial	No leads	Age proxy: ln(firm age)	Age proxy: startup
IFA_{it}	0.020 (0.028)	-0.001 (0.028)		0.018 (0.029)	0.039 (0.027)	0.038 (0.027)
$Order_i$	-0.162*** (0.046)	-0.163*** (0.047)	-0.124 (0.091)	-0.151*** (0.045)	-0.237*** (0.056)	-0.146*** (0.045)
$pcr_{c,t}$	0.003 (0.027)	0.024 (0.027)	0.061*** (0.012)	0.004 (0.026)	0.001 (0.027)	0.001 (0.026)
Age					-0.800*** (0.264)	
Startup						-1.204*** (0.322)
N	1400	1568	3682	1400	1400	1400
Overdispersion(α)	2.0341	1.6559	1.1050	1.5647	1.4864	1.4277
a_c	0.0000	0.0000		0.0000	0.0000	0.0000
a_s	0.0000	0.0000		0.0000	0.0000	0.0000
a_t	0.0000	0.0380		0.0686	0.0332	0.0505

Country (α_c), sector (α_s), year (α_t) fixed effects have been included in all the specifications. * p -value < 0.10, ** p -value < 0.05, *** p -value < 0.01. Cluster robust standard errors (i.e. cluster is the 2 digits NACE code) in parentheses. The reference category are Belgian small-sized firms operating in the Other Manufacturing sector in 1993. The regression also includes a dummy variable equal to one when there is a missing value in IFA. "Over-dispersion (α)" denotes the estimate of the variance inflation factor in the Negative Binomial model where the variance is assumed to be equal to: $\text{var} = \alpha \times \text{mean}$. The over dispersion test for $H_0: \alpha = 1$ is: 298.13 (0.0000), 274.01 (0.0000), 271.27 (0.0000), where p -values are shown in ()

sector- and time-fixed effects should absorb the effect of non-random selection of firms by CERN depending on the LHC project schedule. Similarly, firm-level controls and fixed-effects might help absorbing any bias due to the non-random selection of firms by CERN depending on their expected ability to fulfil its requests.

4.2 Robustness checks

In this section we consider robustness checks that confirm the main results from our baseline specification (Table 3, columns 2–6).

4.2.1 Number of missing observations

Baseline results in Column 1 rely on a sample design that excludes firms reporting more than 4 missing values in IFA_{it} . As a robustness check, in column 2 we changed this arbitrary threshold from 4 up to 5 missing values per firm. Setting the exclusion rule to 5 missing observations yields a sample of 112 firms (or 1568 observations). The main findings on the timing of the CERN effect on patents are unaffected by this change.

4.2.2 Zero-inflated model

Our main results use a sample of firms that have filed at least one patent since their incorporation. We now rely on a larger sample that also includes firms that have never filed a patent. This increases the number of firms to 263 and the sample size to 3682 units, but also introduces a greater number of zero observations in our regressions. To handle the fact that in this case the number of “zeros” in the data might exceed those that can be predicted with a Negative Binomial model, we rely on a Zero-Inflated Negative Binomial (ZINB) specification. In this case the number of patents is predicted supplementing the Negative Binomial distribution with a logit model that describes the probability to observe a zero. A nil observation can thus occur either as a realization of the count density or as a realization of the binary process. The ZINB involves estimating two equations one for the binary model and one of the count specification. Since the number of parameters is quite large, the model might be subject to curse of dimensionality and maximization of the likelihood function quickly becomes hard for numerical routines. For these reasons, we deliberately kept the binary model simple, positing that the probability of observing a zero is influenced only by firm size and IFA_{it} .¹⁷ These variables are excluded from the negative binomial estimation step. Note that also the variable measuring the pre-sample patent count ($Avg. p_i$) is excluded from the set of regressors because it corresponds to the value (i.e. zero) of the dependent variable for companies that have never filed any patent. Column 3 of Table 3 shows that our main conclusions concerning timing of CERN procurement on patent applications are not qualitatively different, although the effect starts being statistically distinguishable from zero a little earlier than in the main specification.

4.2.3 Removing leads

In column 4 we exclude the set of dummy variables acting as leads that were introduced to be sure that the “CERN effect” post-dates the beginning of the procurement relations. Results are unaffected by this model modification.

¹⁷ The estimate of the selection equation is reported in Appendix A2 Table 4.

4.2.4 Start-up firms

Lastly, we add to our baseline specification two different proxies of whether a firm can be considered a young start-up or not. We control for the age of firms with a variable that counts the years running from the incorporation date to the year of beginning of the relationship with CERN (see column 5). In the second specification in column 6 we consider a dichotomous variable that is equal to one for firms whose age is less than 15 years. On average, the age of firms in our samples is 28 years, with 30% of firms being less than 15 years old when they received the first LHC-related order. Results in columns 5 and 6 highlight that none of these variables affect our main findings.

5 Conclusions

This paper contributes to the literature on PPI through BSCs studying the impact of CERN on the innovation output of its suppliers. We have exploited a unique dataset to empirically test the impact of CERN procurement on the innovative output of its industrial partners. Patent applications have been used as a proxy of innovation output in count data models that provide estimates of the timing of CERN procurement impact on the number of patents filed by LHC's suppliers. Since collaborating firms in the sample have received their first order from CERN over a long time span, we have a natural partition of statistical units into "suppliers" and "not-yet-suppliers" that allows to investigate the timing of the "CERN effect".

Our results show that a "CERN effect" on patent applications does exist, it is positive and statistically significant with a delay of at least 5 years from the beginning of the procurement relationship. The existence of such a lag between procurement from BSC and innovation might signal that learning from technology at the frontier of science and translating such knowledge into commercial applications is a medium-long run process.

There is an important consideration in this perspective. Sometimes the R&D is implemented by the CERN in the first place and, during the procurement relationship, firms absorb radically new concepts and solutions required for the scientific purposes. After the end of the procurement relation, firms reconsider what they have learnt and have to understand whether and where new market opportunities may arise in the future. Only at the end of this "learning by interacting process", BSC industrial partners' own R&D for the production of an innovation output—with possible commercial applications—may start. The key point is that this process is quite different from the one driving the established R&D-patents correlation in the usual business environment (such as in the seminal paper by Hausman et al. 1984); in fact, it requires patient firms to invest to for producing an innovation output from the new knowledge acquired through BSC procurement. What makes procurement for innovation at the frontier of science special is that it poses new technological challenges, triggering a sort of 'surprise' learning mechanism for firms, in the meaning of Solow (1997). This has potentially interesting implications as a complement to other, more established, innovation policies.

The existence of a sizeable time lag between procurement from BSC and subsequent benefits for collaborating firms is in line with the findings in related strands of the literature. Griliches (1979), in his survey of econometric analyses of the R&D-productivity

nexus, highlighted that a bell-shaped lag structure connects firm R&D to changes in productivity; such a shape is due to the fact that it takes time before research can be fruitfully exploited by firms. This points to an absorption mechanism requiring protracted learning and adaptation. As underlined by Hameri and Vuola (1996: 131), while incremental innovation may easily access the market, “*the incubation times for revenues from a new technological application range from several years up to a decade, depending on the novelty of the solution*”. In this perspective, the Big Science context is an ideal testing ground of this hypothesis, as it often poses extreme challenges to firms.

Lastly, we point out that this paper leaves some related interesting questions unanswered. While we have analyzed the impact of BSCs on the innovation activity focusing on the number of patents, it would be equally interesting to assess whether BSCs affect the content of patents filed by firms. A metric based on the quality of the patents—such as the citations—could be used for this purpose. We leave this investigation for future research.

Second, as discussed by Bastianin and Del Bo (2020) other EU BSCs—including the European Space Agency and the European Molecular Biology Laboratory—have procurement rules inspired by the same principles as those enforced by CERN. Moreover, much like CERN they have a long list of industrial partners with which they strictly collaborate to develop frontier technologies required for research purposes. Therefore, another issue that is worth investigating relates with the external validity of our results. Whether our results can be generalized to other BSCs across the world and in other research area, as we may suggest, is another interesting question for a future research agenda.

Appendix

See Table 4.

Table 4 Inflate equation in the zero-inflated negative binomial model

	Zero patents
<i>Medium_i</i>	−2.730** (1.177)
<i>Large_i</i>	−2.586** (1.164)
<i>V Large_i</i>	−8.902*** (2.205)
<i>IFA_{it}</i>	0.005 (0.060)
<i>Sector F.E</i>	Yes
<i>Country F.E</i>	Yes
<i>Year F.E</i>	Yes
<i>N</i>	3682

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Classification of hi- and lo-tech orders

In the original database CERN orders are classified by an “activity code” identifying each product type with a highly detailed 3-digit level. We used the 2-digit classification, which covers around 100 items and was sufficiently detailed for our purposes. In some cases, we also inspected the 3-digit classification to better interpret the technological content.

After a preliminary analysis of the overall distribution of order codes, we followed Florio et al. (2016) in identifying the specific activity codes most likely to be associated with high-tech goods and services for the construction of the LHC. In some instances, the code descriptors were generic (“28-Electrical engineering,” say, or “45-Software”). To minimise classification errors, we sampled 300 orders for a more in-depth analysis. These orders were placed with 207 different suppliers, 16% of all those who received at least one order for the LHC during the period under analysis. The orders thus sampled were then evaluated in detail by CERN experts and classified, according to their technological intensity, along a five-point 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 customisation and co-design involving CERN staff.

We defined high-tech codes as Classes 3, 4 and 5. The data indicate that the first order is generally a good predictor of the technological intensity of subsequent ones.

The vast majority of firms in our sample—about 86.9%—has supplied to CERN either only hi-tech (55.7%) or only lo-tech (31.2%) orders. Moreover, an additional 3.5% the sample has received at least 80% of the orders classified as hi-tech. Similarly, an additional 1.9% (of firms has received at least 80% of the orders classified as lo-tech.

Acknowledgements We thank the editor and two anonymous referees for their constructive comments. The authors also thank participants at: the workshop on “The economic impact of CERN colliders: Technological spillovers from LHC to HL-LHC and beyond” at the CERN FCC Week held in Berlin in May 2017, at the XXX Conference of the Italian Society of Public Economics (SIEP) held in September 2018 at the University of Padua, and at the XIV Workshop of the Italian Society of Industrial Policy and Economics (SIEPI) held in January 2019 at the Roma Tre University. Moreover, we thank Caterina Liberati and Emanuela Sirtori for useful comments on a previous version of the paper. Financial support from the FCC study in the frame of the Collaboration Agreement between the University of Milan and CERN (KE3044/ATS) is also gratefully acknowledged.

Funding Open access funding provided by Università degli Studi di Milano within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the

material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Åberg, S., & Bengtson, A. (2015). Does CERN procurement result in innovation? *Innovation The European Journal of Social Science Research*, 28(3), 360–383.
- Adams, J. D. (1990). Fundamental stocks of knowledge and productivity growth. *Journal of Political Economy*, 98(4), 673–702.
- Aghion, P., Van Reenen, J., & Zingales, L. (2013). Innovation and institutional ownership. *American Economic Review*, 103(1), 277–304.
- Amaldi, U. (2012). *Sempre Più Veloci*. Zanichelli.
- Aschhoff, B., & Sofka, W. (2009). Innovation on demand-can public procurement drive market success of innovations? *Research Policy*, 38(8), 1235–1247.
- Autio, E., (2014). "Innovation from Big Science: Enhancing Big Science Impact Agenda". Imperial College Business School, London, In: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/288481/bis-14-618-innovation-from-big-science-enhancing-big-science-impact-agenda.pdf.
- Autio, E., Hameri, A. P., & Vuola, O. (2004). A framework of industrial knowledge spillovers in big-science centers. *Research Policy*, 33(1), 107–126.
- Bach, L., Cohendet, P., & Schenk, É. (2002). Technological transfers from the European space programs: A dynamic view and comparison with other R&D projects. *The Journal of Technology Transfer*, 27(4), 321–338.
- Bastianin, A. and Del Bo, C. (2020) "Public procurement in Big Science: Politics or technology? The case of CERN", *Industrial and Corporate Change*, forthcoming.
- Bastianin, A., & Florio, M. (2019). "Initial guidelines for a social cost-benefit analysis of the FCC programme", CERN Technical Report, CERN-ACC-2019–0037. Available online at: <http://cds.cern.ch/record/2666742>.
- Basu, S., Fernald, J. G., & Kimball, M. S. (2006). Are technology improvements contractionary? *American Economic Review*, 96(5), 1418–1448.
- Battistoni, G., Genco, M., Marsilio, M., Pancotti, C., Rossi, S., & Vignetti, S. (2016). Cost-benefit analysis of applied research infrastructure. Evidence from health care. *Technological Forecasting Social Change*, 112, 79–91.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Blind, K., Edler, J., Frietsch, R., & Schmoch, U. (2006). Motives to patent: Empirical evidence from Germany. *Research Policy*, 35, 655–672.
- Block, J. H., Fisch, C. O., Hahn, A., & Sandner, P. G. (2015). Why do SMEs file trademarks? Insights from firms in innovative industries. *Research Policy*, 44, 1915–1930.
- Blundell, R., Griffith, R., & Van Reenen, J. (1999). Market share, market value and innovation in a panel of British manufacturing firms. *Review of Economic Studies*, 66(3), 529–554.
- Bresnahan, T., & Trajtenberg, M. (1995). General purpose technologies: Engines of growth? *Journal of Econometrics*, 65(1), 83–108.
- Castelnovo P. and Dal Molin M. (2020). The learning mechanisms through public procurement for innovation: the case of government-funded basic research organizations. Forthcoming in: *Annals of Public and Cooperative Economics*. Available online at: <https://onlinelibrary.wiley.com/doi/abs/10.1111/apce.12311>
- Castelnovo, P., Florio, M., Forte, S., Rossi, L., & Sirtori, E. (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.
- Catalano G., Florio M., Pancotti C., Vignetti S., Sánchez Grueso A. and García López G. (2020). The pathways from experiments to innovation impacts: evidence from ALBA Synchrotron Light Facility. Report ALBA-CSIL. Available at: https://www.albasynchrotron.es/en/industry/csil-alba_report_final.pdf
- CERN (2019). "LHC the guide". Available online at Retrieved from March 2019 <http://press.cern/press-kit>.
- Chan, L. K. C., Lakonishok, J., & Sougiannis, T. (2001). The stock market valuation of research and development expenditures. *The Journal of Finance*, 56(6), 2431–2456.
- Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35(1), 128–152.

- Cohendet, P. (1997). Evaluating the industrial indirect effects of technology programmes: the case of the European Space Agency (ESA) programmes. In *Policy evaluation in innovation and technology: towards best practices*, Paris, ISBN 9264156976, pp. 189–223. Available at: <http://www.oecd.org/sti/innopolicevaluationininnovationandtechanologytowardsbestpractices.htm>
- Comin, D. A., & Gertler, M. (2006). Medium-term business cycles. *American Economic Review*, 96(3), 523–551.
- Crépon, B., Duguet, E., & Mairesse, J. (1998). Research, innovation and productivity: An econometric analysis at the firm level. *Economics of Innovation and New Technology*, 7(2), 115–158.
- Dahlin, K. B., & Behrens, D. M. (2005). When is an invention really radical? Defining and measuring technological radicalness. *Research Policy*, 34(5), 717–737.
- Danish Agency for Science. (2008). *Evaluation of Danish industrial activities in the European Space Agency (ESA)*. Assessment of the economic impacts of the Danish ESA membership.
- Dennis, M. A. (2017). Encyclopedia Britannica, s.v. “Big Science”. Available online at: <https://www.britannica.com/science/Big-Science-science>
- Dziallas, M., & Blind, K. (2019). Innovation indicators throughout the innovation process: An extensive literature analysis. *Technovation*, 80–81, 3–29.
- Edler, J., & Fagerberg, J. (2017). Innovation policy: What, why, and how. *Oxford Review of Economic Policy*, 33(1), 2–23.
- Edler, J., & Georghiou, L. (2007). Public procurement and innovation: Resurrecting the demand side. *Research Policy*, 36(7), 949–963.
- Edquist, C., & Hommen, L. (2000). Public technology procurement and innovation theory. In C. Edquist, & L. Hommen & Tsipouri L. (Eds.), *Public technology procurement and innovation* (pp. 5–70). Aarhus, Denmark: Kluwer Academic Publishers, pp. 323.
- Edquist, C., Hommen, L., & Tsipouri, L. (2000). *Public technology procurement and innovation*. . Kluwer Academic Publishers.
- Edquist, C., Vonortas, N.S., Zabala-Iturriagagoitia, J.M., & Edler, J. (2015). *Public procurement for innovation* (ESFRI (2016), Strategy Report on Research Infrastructures.). Cheltenham, UK: Edward Elgar. Retrieved from <http://www.esfri.eu/roadmap-2016>.
- Edquist, C., & Zabala-Iturriagagoitia, J. M. (2012). Public procurement for innovation as mission-oriented innovation policy. *Research Policy*, 41(10), 1757–1769.
- Evans, L.R., (2009). “The Large Hadron Collider: A Marvel of Technology”. EPFL Press, pp. 312
- Fernandes, A., Pessoa, A., & Silva, M. R. (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 Manag.*, 14(4), 327–345.
- Fernández-Olmos, M., & Ramírez-Alesón, M. (2017). How internal and external factors influence the dynamics of SME technology collaboration networks over time. *Technovation*, 64, 16–27.
- Florio, M. (2019). *Investing in science*. . MIT Press.
- Florio, M., Bastianin, A., & Castelnovo, P. (2018a). The socio-economic impact of a breakthrough in the particle accelerators’ technology: A research agenda. *Nuclear Instruments and Methods in Physics Research Section A*, 909, 21–26.
- Florio, M., Giffoni, F., Giunta, A., & Sirtori, E. (2018). Big science, learning, and innovation: Evidence from CERN procurement. *Industrial and Corporate Change*, 27(5), 1915–936.
- Florio, M., & Sirtori, E. (2016). Social benefits and costs of large scale research infrastructures. *Technological Forecasting and Social Change*, 112, 65–78.
- Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3), 535–574.
- Georgiou, L., Edler, J., Uyarrar, E., & Yeow, J. (2014). Policy instruments for public procurement of innovation: Choice, design and assessment. *Technological Forecasting & Social Change*, 86, 1–12.
- Giudice, G. F. (2010). *A Zeptospace Odyssey. A journey into the physics of the LHC*. . Oxford University Press.
- Griffiths, T. L., & Steyvers, M. (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. *Bell Journal of Economics*, 10(1), 92–116.
- Guerzoni, M., & Raiteri, E. (2015). Demand-side versus supply-side technology policies: Hidden treatment and new empirical evidence on the policy mix. *Research Policy*, 44(3), 726–747.
- Gurmu, S., & Pérez-Sebastián, F. (2008). Patents, R&D and lag effects: evidence from flexible methods for count panel data on manufacturing firms. *Empirical Economics*, 35(3), 507–526.
- Hall, B., Griliches, Z., & Hausman, J. (1984). Patents and R&D: Is there a lag? *International Economic Review*, 27(2), 265–283.

- Hall, B., Helmers, C., Rogers, M., & Sena, V. (2014). The choice between formal and informal intellectual property: A review. *Journal of Economic Literature*, 52(2), 375–423.
- Hallonsten, O. (2014). How expensive is Big Science? Consequences of using simple publication counts in performance assessment of large scientific facilities. *Scientometrics*, 100(2), 483–496.
- Hameri, A. P. (1997). Innovating from big science research. *Journal of Technology Transfer*, 22(3), 27–35.
- Hameri, A. P., & Vuola, O. (1996). Using basic research as a catalyst to exploit new technology based innovations. A case study. *Technovation*, 16(10), 531–539.
- Hausman, J., Hall, B., & Griliches, Z. (1984). Econometric models for count data with an application to the patents-R&D Relationship. *Econometrica*, 52(4), 909–938.
- Heher, A. D. (2006). Return on investment in innovation: Implications for institutions and national agencies. *The Journal of Technology Transfer*, 31(4), 403–414.
- Helmers, C., & Overman, H. G. (2017). My precious! The location and diffusion of scientific research: Evidence from the synchrotron diamond light source. *Economic Journal*, 127(604), 2006–2040.
- Jia, N., Huang, K., & Zhang, C. M. (2019). Public governance, corporate governance, and firm innovation: An examination of state-owned enterprises. *Academy of Management Journal*, 62(1), 220–247.
- Kung, H., & Schmid, L. (2015). Innovation, growth, and asset prices. *Journal of Finance*, 70(3), 1001–1037.
- Lane, P. J., & Lubatkin, M. (1998). Relative absorptive capacity and interorganizational learning. *Strategic Management Journal*, 19(5), 461–477.
- Leiponen, A., & Byma, J. (2009). If you cannot block, you better run: Small firms, cooperative innovation, and appropriation strategies. *Research Policy*, 38, 1478–1488.
- Lember, V., Kattel, R., & Kalvet, T. (2015). Quo vadis public procurement of innovation? *Innovation The European Journal of Social Science Research*, 28(3), 403–421.
- Leoncini, R., Marzucchi, A., Montresor, S., Rentocchini, F., & Rizzo, U. (2017). “Better late than never: the interplay between green technology and age for firm growth.” *Small Business Economics*, 1–14.
- Lichtenberg, F. R. (1990). US government subsidies to private military R&D investment: The defense department’s independent R&D policy. *Defence Economics*, 1(2), 149–158.
- Mansfield, E. (1968). *Industrial research and technological innovation: An econometric analysis*. Norton.
- Mansfield, E. (1991). Academic research and industrial innovation. *Research policy*, 20(1), 1–12.
- Mansfield, E. (1998). Academic research and industrial innovation: An update of empirical findings. *Research policy*, 26(7–8), 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.
- Nielsen, V., & Anelli, G. (2016). Knowledge transfer at CERN. *Technology Forecasting and Social Change*, 112, 113–120.
- OECD (2009). Patent Statistics Manual. Available at: <http://www.oecd.org/sti/inno/oecdpatentstatisticsmanual.htm>
- Oldham, P. (2019). The WIPO Patent Analytics Handbook. Available at: <https://wipo-analytics.github.io/handbook/>
- Pakes, A., & Griliches, Z. (1980). Patents and R&D at the firm level: A first report. *Economics Letters*, 5(4), 377–381.
- Pakes, A., & Shankerman, M. (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*. (pp. 73–88). University of Chicago Press.
- 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.
- Rapoport, J. (1971). The anatomy of the product-innovation process: Cost and time. In E. Mansfield (Ed.), *Research and innovation in the modern corporation*. (pp. 110–135). Norton.
- Salter, A. J., & Martin, B. (2001). The economic benefits of publicly funded basic research: A critical review. *Research Policy*, 30(3), 509–532.
- Scarrà D. and Piccaluga A. (2020). The impact of technology transfer and knowledge spillover from Big Science: A literature review. Forthcoming in: *Technovation*. Available online at: <https://doi.org/10.1016/j.technovation.2020.102165>
- Schaefer, A., Schiess, D., & Wehrli, R. (2014). Long-term growth driven by a sequence of general purpose technologies. *Economic Modelling*, 37, 23–31.
- Schmied, H. (1977). A study of economic utility resulting from CERN contracts. *IEEE Transactions on Engineering Management*, 4, 125–138.

- Sirtori, E., Catalano, G., Giffoni, F., Pancotti, C. Caputo, A., and Florio, M. (2019). "Impact of CERN procurement actions on industry: 28 illustrative success stories." Available online at Retrieved April 2019 <https://cds.cern.ch/record/2670056?ln=it>
- Solow, R. M. (1997). *Learning from learning by doing: Lessons for economic growth*. . Stanford University Press.
- Sternitzke, C. (2010). Knowledge sources, patent protection, and commercialization of pharmaceutical innovations. *Research Policy*, 39(6), 810–821.
- Stevenson, B., & Wolfers, J. (2006). Bargaining in the shadow of the law: Divorce laws and family distress. *The Quarterly Journal of Economics*, 121(1), 267–288.
- 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., & Flanagan, K. (2010). Understanding the innovation impacts of public procurement. *European Planning Studies*, 18(1), 123–143.
- Vuola, O., & Hameri, A. P. (2006). Mutually benefiting joint innovation process between industry and big-science. *Technovation*, 26(1), 3–12.
- Wagner, L. U. (1968). Problems in estimating research and development investment and stock. In *Proceedings of the business and economic statistics section*, 189–98. Washington, D.C.: American Statistical Association.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.