Structural Change and Economic Dynamics 81 (2026) 184–197

Contents lists available at ScienceDirect

Structural Change and Economic Dynamics

journal homepage: www.elsevier.com/locate/strueco

Open-sourcing, labor productivity, and the moderating role of R&D

Sara Amoroso [a], Randolph Luca Bruno [b,c,d,e], Laura Magazzini [f,*]

[a] German Institute for Economic Research—DIW, Berlin, Germany
[b] Università Cattolica del Sacro Cuore, Economic Policy Department, Piacenza, Italy
[c] University College London, School of Slavonic and East European Studies, London, UK
[d] Rodolfo DeBenedetti Foundation, Milan, Italy
[e] IZA, Institute of Labor, Bonn, Germany
[f] Institute of Economics and L'EMbeDS, Sant'Anna School of Advanced Studies, Piazza Martiri della Libertà 33, 56127 Pisa, Italy

ARTICLE INFO

JEL classification:
C23
O32
L20

Keywords:
Publications
Open innovation
Intellectual property
Productivity
R&D investment

ABSTRACT

Scientific publication represents an observable form of selective knowledge disclosure through which firms signal capabilities, access knowledge networks, and attract talent. Building on the literature linking sectoral openness to productivity growth, we examine whether firms' propensity to publish scientific work amplifies the productivity returns to R&D investment. We find that adopting an open-sourcing strategy is positively associated with labor productivity among R&D-intensive firms, controlling for unobserved heterogeneity at the firm and sector level, hence controlling for technology. This result is consistent with theories emphasizing the role of knowledge diffusion and absorptive capacity for firm performance. Methodologically, the paper extends the Hausman–Taylor-type approach for the identification of time-invariant variables in a fixed effect framework. Monte Carlo simulations show that this approach reliably identifies the impact of these regressors on productivity.

1. Introduction

A central feature of contemporary structural change is the reorganization of knowledge production and diffusion. The growth of corporate science, open innovation, and open-science practices represents a shift in the technological and institutional patterns through which firms generate and appropriate knowledge. This structural shift matters for productivity dynamics because it changes the channels through which R&D translates into performance—via both innovation output and the diffusion/absorption of external knowledge (David, 1994; Fagerberg, 2000).

Firms publish in scientific journals as part of a strategy to balance innovation, collaboration, reputation, and intellectual property management. This practice is particularly prevalent in industries where scientific and technical advancements are key drivers of competitive advantage (Rotolo et al., 2022). Research has explored the impact of publishing on firm innovative performance. In Edler and Fagerberg's (2017) view, publications may incentivize businesses to absorb and use innovations quickly in contexts that are time and/or geographically constrained. Evidence indicates that a firm's publishing activity positively correlates with its innovative output, as demonstrated by the number of forward citations its patents receive (Arora et al., 2018; Gambardella, 1992), the breadth of these citations (Cassiman et al., 2008), and the number of patent families granted to the firm (Almeida et al., 2011). Additionally, firm publishing has been positively associated with financial performance, including market valuation (Pellens and Malva, 2017; Simeth and Cincera, 2016).

While innovation scholars have focused mainly on how knowledge spillovers from publications have an impact on firms' abilities to innovate, very few studies look at how publications influence, directly or indirectly, the production performance of firms. For example, Chen et al. (2016) find a positive contribution of scientific linkages (number of scientific publications cited by a firm's patents relative to the citations of other patents) to the total factor productivity (TFP) of a set of innovative Korean firms. Belenzon and Patacconi (2014), in a production function framework, find that the stock of publications directly and positively affects firms' sales. Cassia et al. (2009) show that public companies in UK do benefit from innovation systems underpinned by collaborations (e.g. co-publishing) with Universities.

The majority of contributions related to the role of publications for firm performance, when they are considered as *open-sourcing strategy*, is found in the literature on open innovation (OI) (Bogers et al., 2019; Ahn et al., 2015). Studies in this literature use a wide range of OI modes (number and types of collaborations, external R&D expenditure, patent licensing, spin-offs, publications, citations, etc.) and rarely focus

* Corresponding author.
E-mail address: laura.magazzini@santannapisa.it (L. Magazzini).

https://doi.org/10.1016/j.strueco.2026.09.007
Received 17 April 2026; Received in revised form 7 August 2026; Accepted 16 September 2026
Available online 21 September 2026
0954-349X/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

on labor productivity as an outcome variable, despite the recognized importance of publications as an intangible input into production.

Publication activity is not equivalent to a fully open-source model in which all knowledge assets, technical designs, or intellectual property are released without restriction. Rather, publication is a specific and observable form of openness through which firms selectively disclose codified scientific knowledge. The decision to publish implies that the firm allows part of its internally generated knowledge to circulate beyond organizational boundaries and become accessible to external scientific, technological, and industrial communities. In this sense, publication activity captures a firm's propensity to engage in a controlled form of knowledge sharing. This is particularly relevant for R&D-intensive firms, for which scientific disclosure may support reputation building, participation in knowledge networks, recruitment of highly skilled labor, and access to complementary external knowledge. At the same time, because disclosure may expose firms to imitation and knowledge spillovers, the productivity implications of such openness are unlikely to be unconditional and may depend on the firm's own absorptive capacity.

Hence, in this paper, we use the expression "open-sourcing strategy" in a broad sense, not restricted to the release of software source code. We refer instead to a firm's deliberate decision to disclose part of its internally generated scientific and technological knowledge to external audiences. Scientific publication is therefore interpreted as a form of selective knowledge disclosure: firms do not necessarily make all proprietary knowledge freely available, but they allow selected research outputs to enter the public scientific domain. This behavior is consistent with the open-innovation and selective-revealing literatures, according to which firms may benefit from sharing knowledge when disclosure enhances reputation, facilitates access to external knowledge networks, attracts scientific and technical talent, and improves their ability to absorb complementary knowledge.

We analyze whether firms' engagement in scientific publication, interpreted as an observable form of selective knowledge disclosure and, therefore, as a proxy for an open-sourcing orientation, affects the labor productivity of R&D-intensive multinational enterprises (MNEs), and whether this effect is conditional on the level of R&D investment.

The paper contributes to the literature along three main dimensions. First, it contributes to the literature on open innovation, corporate science, and firm performance by shifting the focus from innovation outputs and market-based indicators to labor productivity. Existing studies have documented that firm publication activity is associated with patent quality, patent breadth, innovative output, and firm valuation, but less is known about whether publication activity is also related to production performance. By examining labor productivity, we assess whether scientific openness is reflected in firms' capacity to transform knowledge and R&D investments into productive efficiency.

Second, the paper contributes to the literature on absorptive capacity and the dual role of R&D by explicitly modeling R&D investment as a moderator of the relationship between scientific disclosure and productivity. The empirical specification does not assume that openness is unconditionally beneficial. Rather, it tests whether the productivity returns to publication-based openness depend on the firm's internal knowledge base. This is important because firms with stronger R&D capabilities may be better able to absorb external knowledge, recombine it with internal capabilities, and benefit from the reputational, network, and learning effects generated by scientific publication. Conversely, firms with weaker absorptive capacity may be less able to appropriate the benefits of disclosure and may be more exposed to the costs of knowledge spillovers.

Third, the paper makes a methodological contribution to panel-data econometrics. Since the publication-based open-sourcing indicator is time-invariant in our empirical setting, its direct effect cannot be estimated using a standard fixed effects (FE) model. We therefore extend a Hausman–Taylor-type approach by exploiting a homogeneity assumption in the correlation between selected time-varying regressors and firm-specific unobserved heterogeneity. This allows us to identify the effect of time-invariant regressors without relying on external instruments or imposing the stronger assumption that these variables are uncorrelated with firm fixed effects (as in a random effect model).

The empirical analysis integrates firm-level data from three different sources: the EU Industrial R&D Investment Scoreboard on world top R&D investors (Grassano et al., 2021), which provides information on firms' R&D expenditures and employees, the JRC-OECD COR&DIP database (v.2, 2019) that includes information on scientific publications, and additional financial data are retrieved from the Moody's Orbis dataset . The merged longitudinal dataset spans the period 2007–2016 and includes information on about 1200 companies.

The paper proceeds as follows. Section 2 briefly reviews the existing literature on the use of publication data to proxy open-sourcing strategy. Section 3 describes the data and provides summary statistics. Section 4 introduces the estimated model and proposes a new methodology to identify the slope coefficient of time-invariant variables in a FE framework. Section 5 presents the results of the empirical analysis, and Section 6 concludes.

2. Publications as an open-sourcing strategy and productivity

In the context of this paper, open-sourcing is understood as a strategic orientation towards the controlled disclosure of knowledge rather than as the unrestricted release of all firm-specific intellectual property. Scientific publication represents one important manifestation of this orientation. When firms publish research outputs, they deliberately make selected elements of their internal knowledge base visible and accessible outside the firm. This does not imply that firms abandon appropriation strategies or disclose all commercially relevant knowledge. Rather, publication can be interpreted as a form of selective revealing, whereby firms disclose specific knowledge components while retaining control over complementary assets, tacit capabilities, downstream applications, or protected intellectual property.

This interpretation connects the analysis to the open-innovation literature. In the open-innovation framework, firms do not rely exclusively on internally generated knowledge, but combine internal R&D with external ideas, collaborations, and knowledge flows (Chesbrough, 2003; Bogers et al., 2019). Open-sourcing is therefore not opposed to appropriation, but it may be part of a broader strategy through which firms manage the boundary between internal knowledge creation and external knowledge exchange. The concept of selective revealing is particularly useful in this respect. Henkel (2006) shows that firms can voluntarily reveal selected pieces of knowledge when disclosure improves the functioning of the innovation ecosystem, reduces duplication, or supports complementary innovation. Similarly, Alexy et al. (2013) stress that knowledge revealing is selective and strategic: firms disclose some knowledge while retaining control over other knowledge assets, and the benefits of revealing depend on who is expected to use the disclosed knowledge and how the firm can appropriate indirect returns from disclosure.

The interpretation of publication as a strategic form of openness is also supported by the literature on the drivers of corporate publishing, recently systematized by Rotolo et al. (2022). The paper shows that firms' publication activity is not an incidental by-product of R&D, but responds to identifiable strategic incentives. Indeed, publishing grants access to external knowledge networks and frontier academic research (Hicks, 1995; Rosenberg, 1990; Pénin, 2007); it signals scientific capability to partners, investors, and competitors (Polidoro and Theeke, 2012; Simeth and Cincera, 2016); it attracts scientific human capital, as researchers accept lower wages in exchange for the ability to publish (Stern, 2004); it serves as a firm's intellectual-property strategy, since disclosure creates prior art that preempts rivals' patenting ("defensive publishing"; Parchomovsky, 2000); and it supports commercialization in science-based markets (Gambardella, 1992). Simeth and

Raffo (2013) show that firms' engagement in open science varies systematically with their technological environment and appropriability conditions, confirming that publishing is a deliberate strategic choice.

Two implications of this literature are central to our empirical design. First, because the drivers of publishing (i.e., network access, hiring, reputation, IP positioning) reflect deliberate strategic choices and complementary investments, publication activity can be treated as a persistent, structural orientation of the firm (see Section 3). Second, publication propensities are strongly industry-specific (Hicks, 1995; Rotolo et al., 2022), which motivates defining our open-sourcing indicator relative to the sector average rather than in absolute terms.

Building on these drivers, we can identify the mechanisms through which publication-based openness may affect labor productivity. A first mechanism operates through knowledge access: by publishing, firms become more embedded in scientific networks involving universities, public research organizations, and other firms, gaining earlier access to frontier knowledge that can be incorporated into products and processes, thereby raising the efficiency of the firm's innovation activity. A second mechanism involves the quality of the labor input: the possibility to publish helps firms attract and retain highly skilled scientists (Stern, 2004), increasing the average output per employee. A third mechanism operates through signaling and reputation: publications credibly attest to technological capabilities to partners, customers, and investors, facilitating collaborations, access to finance, and the commercialization of science-intensive products. A fourth mechanism is learning through exposure: disclosure invites feedback, citations, and complementary research by others, from which the publishing firm can itself benefit. All these mechanisms increase the productivity payoff, provided that the firm is able to interpret, absorb, and recombine external knowledge. This capability, as we argue below, depends on the firm's own R&D.

At the same time, scientific disclosure involves costs and risks. Publishing research outputs may reveal information that competitors can use, reduce the firm's ability to appropriate knowledge exclusively, or accelerate spillovers outside the boundaries of the organization. Direct evidence that disclosure is strategically restrained when appropriation concerns dominate comes from publicly funded research consortia. Amoroso and Vannuccini (2026) find that the participation of top R&D investors in EU-funded collaborative projects is associated with significantly lower publication output, consistent with large firms exercising control over the disclosure of jointly produced knowledge and with the declining valuation of scientific publication relative to patenting among leading corporate R&D performers (Arora et al., 2018). For this reason, the productivity effect of publication-based openness is theoretically ambiguous. Openness may improve performance when the firm is able to absorb, recombine, and appropriate the knowledge flows activated by disclosure. Conversely, openness may generate limited or even negative returns when the firm lacks the internal capabilities needed to transform scientific visibility and external knowledge access into productive advantages.

This is where R&D plays a central moderating role. Following Cohen and Levinthal (1989), R&D has two faces: it generates new knowledge directly, and it builds the firm's absorptive capacity. Absorptive capacity is cumulative and cannot simply be purchased on the market, as it is a by-product of the firm's own research activity. Its relevance for firm's productivity is well documented by Griffith et al. (2004): R&D raises productivity not only through innovation but also by enhancing technology absorption, and internal R&D and external knowledge acquisition are complements rather than substitutes (Cassiman et al., 2008; Leten et al., 2022).

Firms with stronger R&D capabilities are better positioned to benefit from publication-based openness because they can combine disclosed scientific knowledge with internal research competences, interpret external scientific developments, and appropriate indirect returns from participation in knowledge networks. In this sense, R&D has a dual role: it directly contributes to productivity through internal knowledge accumulation, and it indirectly affects productivity by increasing the firm's ability to benefit from openness.

This theoretical framework implies that the relationship between scientific publication and productivity should not be expected to be uniformly positive. Rather, publication activity should generate stronger productivity benefits for firms with sufficiently high R&D intensity. For firms with weak internal knowledge capabilities, disclosure may not translate into productivity gains and may even expose the firm to the costs of openness without providing the absorptive capacity required to capture its benefits. Our empirical model therefore focuses not only on the direct association between publication-based openness and labor productivity, but also on the interaction between openness and R&D intensity.

The empirical specification in Section 5 follows directly from this theoretical argument. If publication activity captures a firm's orientation towards selective scientific disclosure, and if R&D intensity captures the internal capability required to absorb and appropriate the benefits of openness, then the relevant empirical question is whether the productivity effect of publication-based openness depends on the level of R&D investment.

Concretely, the dual-role argument predicts a positive interaction between openness and R&D intensity and a negative level effect at low R&D intensity, with the overall effect of openness turning positive only above a threshold of R&D investment. Testing the second prediction requires identifying the coefficient of a time-invariant regressor, which is precisely what standard FE estimation cannot deliver, and what motivates the methodology of Section 4.

3. Data and measures

The dataset used in the empirical analysis integrates information from three data sources. First, to measure firms' open-sourcing strategy, we rely on firms' scientific publication activity (Cockburn and Henderson, 2000; Leten et al., 2022; Mansfield, 1991). The information on publishing activity at the firm level has been retrieved from the JRC-OECD COR&DIP database (v.2, 2019),[1] including information on the top 2000 R&D performing companies worldwide. For each company, the database provides information on the number of scientific publications over the years 2014–2016. R&D expenditure values and the number of employees are also available from this source over the years 2013–2016.

Substantial heterogeneity exists in the publication activity across industrial sectors. Table 1 shows that the number of publications ranges from an average of 387.4 publications per firm in 'Aerospace & Defence' to 90.3 in 'Industrials'.

It is therefore important to account for sector-specific effects, as publication propensities differ substantially across technological domains and industries. Coherently with this view, we classify a firm as having an open-sourcing orientation when its publication activity is

Table 1
Average number of publications, by sector.

Sector	Average
Aerospace & Defence	387.4
Automobiles & other transport	113.0
Chemicals	153.0
Health industries	249.4
ICT producers	123.4
ICT services	157.8
Industrials	90.3
Others	111.4
All sectors	146.9

[1] The COR&DIP database can be accessed at https://forms.office.com/r/9bt7kSbPJ0.

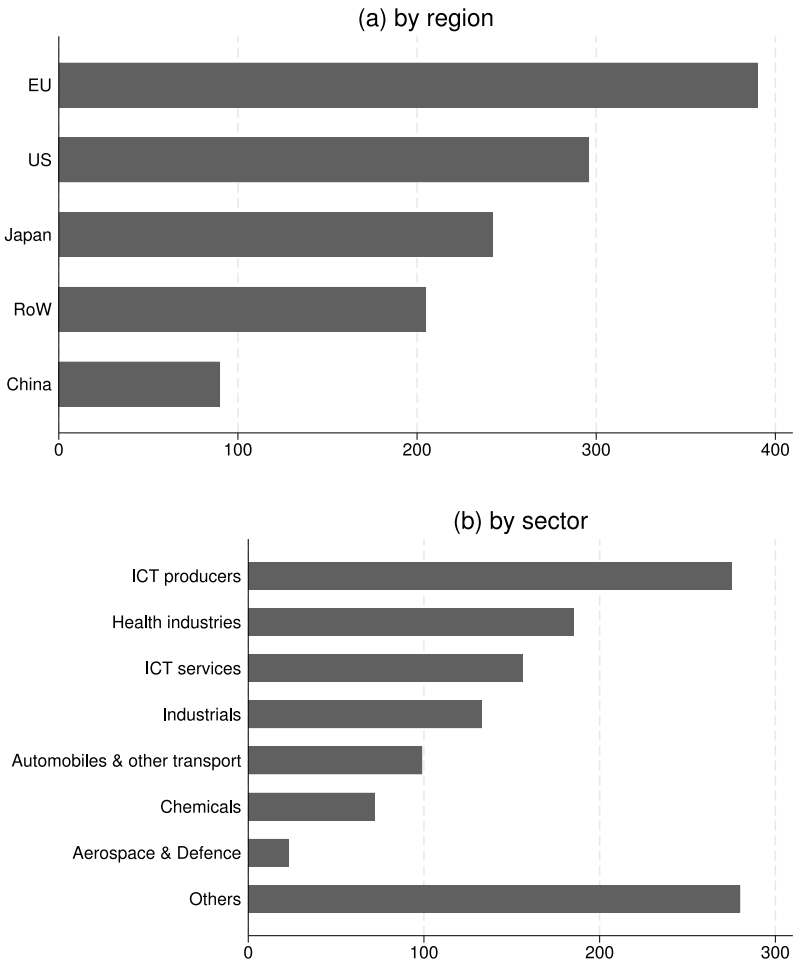


Fig. 1. Distribution of companies in the sample according to (a) region; (b) industrial sector.

high *relative* to the "publication norm" of its own sector. This makes the indicator consistent with the idea of openness as a strategic orientation rather than as a purely scale-driven outcome. Operationally, we define the open-sourcing measure to be used in the empirical analysis, S_i, as a dummy variable equal to 1 if the firm's total number of publications over the time 2014–2016 is above the average of the sector in which the firm operates, 0 otherwise.

To enhance time coverage, backward information on R&D expenditures and employees has been retrieved from various editions of the EU Industrial R&D Investment Scoreboard (Grassano et al., 2021), also reporting industry classification and country of the headquarter location of our sample of companies. Additional information on relevant financial data, namely, value added and physical capital, has been extracted from the Moody's Orbis dataset. When merging company information in the two datasets we were able to retrieve financial data for 1982 companies.

The perpetual inventory method is used to compute both physical capital and R&D expenditure stocks.[2] As financial data were not complete for some companies, in the analysis we use an unbalanced dataset of 6677 observation on 1223 companies over the period 2007–2016.

Fig. 1 shows the distribution of firms across industrial sectors and regions. Most firms are located in Europe, operating in the ICT producers sector.

Note that we also expand backward the indicator of open-sourcing strategy to the year 2007. By doing so, we are implicitly making two assumptions. First, firms' corporate structures in 2014 have not changed since 2007. Second, firms publishing during 2014–2016 are also likely to have done so in earlier years (2007–2013). Specifically, we argue that publishing is a firm's structural strategy, which is unlikely to change substantially over time. Furthermore, while a firm may change the overall number of its publications, it is unlikely to change its relative position above or below the sector average. In other words, "publishing above sector average" is a proxy for open-sourcing network-mediated diffusion not just as a firm tactic but a persistent organizational regime (a structural strategy), hence time-invariant, in line with structural change as evolving behavioral and institutional patterns.

In order to gauge the differences between the characteristics of firms that operate an open-sourcing strategy and firms that do not, Fig. 2 shows the distribution of R&D and employment for the two groups classified according to the value of the variable S_i. Firms engaging in open-sourcing strategy have a larger size, as measured by the number of employees, and they also record larger R&D intensity.

Beyond data availability, the 2007–2016 window has two features that make it appropriate for our research design. First, the sample period ends before the wave of protectionist and techno-nationalist policies that has reshaped global innovation governance since the late 2010s. Trade restrictions, export controls, investment screening, and research-security measures directly alter the incentives for cross-border scientific disclosure (Petricevic and Teece, 2019; Luo, 2022). Because our identification treats openness as a stable structural orientation of

[2] For physical capital we follow Gal (2013), whereas in the case of R&D we consider a 5% pre-sample growth rate and a rate of obsolescence of knowledge equal to 15%.

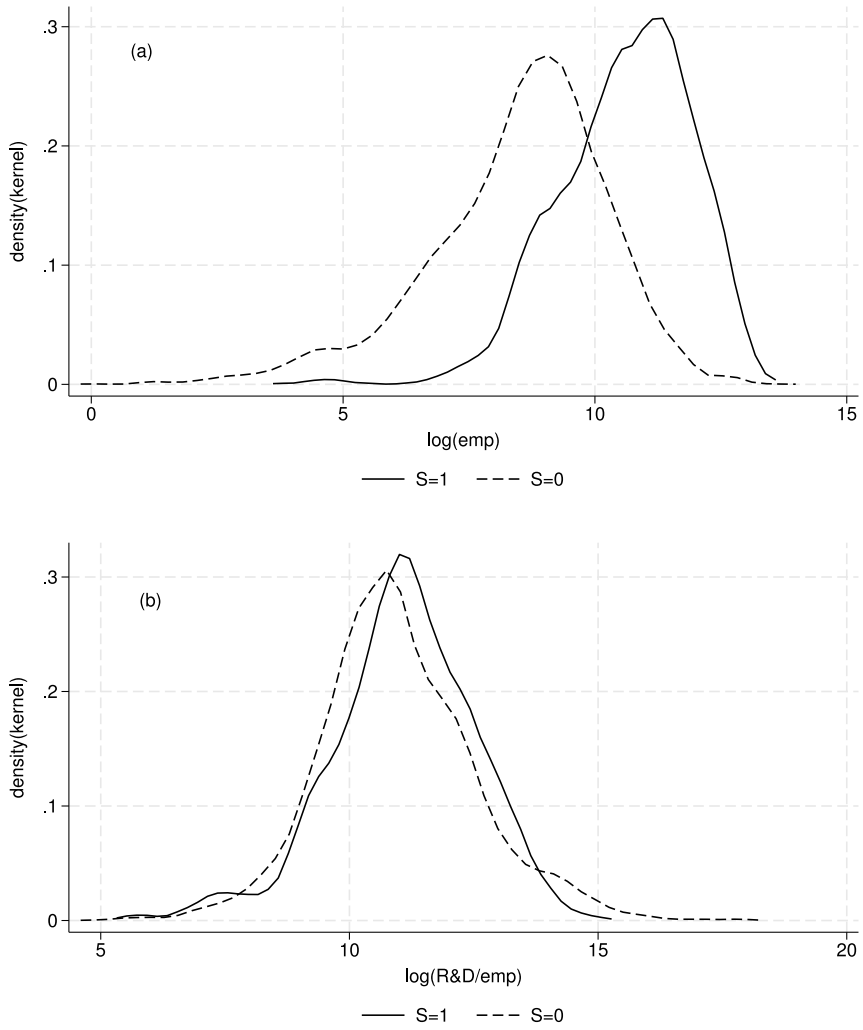


Fig. 2. Kernel distribution comparing open-sourcing firms with firms that do not operate this strategy: (a) log of employees, and (b) log of ratio of R&D stocks and employees.

the firm, a sample extending into this later period would blend firms' strategic choices with policy-induced restrictions on disclosure.

Second, although the window includes the tail of the 2008–2009 financial crisis, the evidence indicates that the crisis is unlikely to have disturbed the openness orientation our indicator captures. The crisis-induced contraction in innovation investment was concentrated among weakly innovative firms, whereas highly innovative firms, such as our sample of top R&D investors, largely maintained their innovation activities (Filippetti and Archibugi, 2011; Paunov, 2012; Archibugi et al., 2013). Moreover, the crisis is credited in the open-innovation literature with increasing firms' reliance on external knowledge channels as a cost-coping strategy (Chesbrough and Garman, 2009).

Finally, since S_i is defined relative to the sector average, even a common crisis-induced shift in publication levels within a sector would leave firms' relative openness positions, hence the indicator, unaffected, while firm, year, sector-by-year, and region-by-year effects absorb the macroeconomic impact of the crisis on productivity.

The descriptive statistics of the variables included in the analysis are reported in Table 2.

Table 2
Descriptive statistics of the variables included in the regression ($N = 6677$).

Variable	mean	std. dev.	min.	max.
$\ln(VA/L)$	10.66	1.913	−0.397	19.51
$\ln(K/L)$	10.35	2.284	−6.203	21.24
$\ln(RD/L)$	10.75	1.465	4.785	17.15
$\ln(L)$	9.218	1.625	1.386	13.35
S	0.201	0.401	0	1

4. Methodology

To explore whether an open-sourcing strategy affects labor productivity, we build our baseline regression on the model by Hall and Mairesse (1995), which is augmented by including a proxy for the firm's open-sourcing strategy, S_i:

$$\ln\left(\frac{VA_{it}}{L_{it}}\right) = \beta_1 \ln\left(\frac{K_{it}}{L_{it}}\right) + \beta_2 \ln\left(\frac{RD_{it}}{L_{it}}\right) + \beta_3 \ln(L_{it}) + \\ + \beta_4 S_i + \beta_5 \ln\left(\frac{RD_{it}}{L_{it}}\right) \times S_i + \alpha_i + \tau_t + e_{it} \quad (1)$$

where the dependent variable is firm's i labor productivity at time t, defined as the value-added per employee, VA_{it}/L_{it}. The regressors (expressed in logarithms) include the physical capital stock per employee, K_{it}/L_{it}, the stock of R&D expenditure per employee RD_{it}/L_{it} and the total number of employees, L_{it}.

As already explained, we are not allowed to use the empirical measure of open-sourcing strategy S_i as a time-varying regressor, as we would compromise the longitudinal structure of the data. This creates a

problem in the FE estimation framework, as the open-sourcing strategy variable would be automatically dropped as a firm, time-invariant effect. A FE model allows the identification of the slope coefficient of time-varying variables, even in the case of omitted variables that are correlated with the regressors of interest, provided that these unobserved effects are constant over time, so that the coefficient β_5 is identified. A notable drawback of this approach is that the specific effect of time-invariant variables, such as individuals' gender and race, firms' nationality, or, in our case, open-sourcing strategy, cannot be identified due to perfect collinearity. However, the identification and estimation of the slope coefficients of such time-invariant variables –β_4 in our setting– are of interest in applied research.

The methodological contribution of the paper addresses this identification problem. Specifically, we show that, relying on the assumption that the correlation between (a subset of) time-varying variables and the individual time-invariant effect is homogenous over time (originally suggested by Breusch et al. 1989), it is possible to achieve identification of the impact of time-invariant variables without the need to choose exogenous variables or rely on external instruments. The econometric literature has largely overlooked this result.

4.1. Identification of time-invariant variables under homogeneity

Our estimation strategy builds on the literature related to the estimation of time-invariant variables in a FE panel data model, as it is the case with S_i in (1). In a FE framework, correlation is unrestricted between the individual effect α_i and the independent variables.[3] Despite the ability to control for unobservable individual characteristics, the FE approach only allows the identification of the effect of time-varying variables, whereas the impact of time-invariant variables cannot be identified. To identify the latter, the literature has proposed to exploit additional assumptions.[4]

To the best of our knowledge, Hausman and Taylor (1981) first proposed a methodology for the identification of the effect of time-invariant variables, based on partitioning the vector of regressors (both time-varying and time-invariant) into two sets of variables based on the assumptions on the correlation between the variables and the individual effect α_i. The assumption of exogeneity for a subset of variables is the key to the definition of instrumental variables that can be exploited to achieve identification. The proposal has been refined to extend the set of legitimate instruments (i.e., orthogonality conditions) by Amemiya and MaCurdy (1986) and Breusch et al. (1989), Ahn and Schmidt (1995), Im et al. (1999).

Particularly relevant to our framework is the contribution of Breusch et al. (1989, henceforth BMS89), that also includes an assumption of homogenous correlation between the individual effect and the endogenous variables in the model.

Recently, some estimators have been proposed that relax the assumption on the exogeneity of time-varying variables but, on the other hand, they require either all time-invariant variables to be exogenous or, when endogeneity is allowed, they require the availability of *external* instrumental variables (Plümper and Troeger, 2007; Greene, 2011; Pesaran and Zhou, 2018; Chen et al., 2020), which may be problematic to retrieve.

The econometric literature has so far overlooked the implication of the BMS89 assumption in a static framework. Indeed, homogenous-correlation variables can be exploited to identify the effect of time-invariant variables, without the need to rely on an exogeneity assumption, i.e. without the need to partition the variables in the model into an endogenous and exogenous group.

To better assess the implications of the assumption, consider a simple model with one time-varying variable x_{it} and one time-invariant variable z_i $(i = 1, \dots, N; t = 1, \dots, T)$[5]:

$$y_{it} = x_{it}\beta + z_i\gamma + \alpha_i + e_{it} \tag{2}$$

Within a FE framework, correlation is unrestricted between the variables on the right-hand side and the individual FE α_i. The model is estimated by considering the within-group transformation, that is, individual-demeaned variables are considered in the estimating equation:

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)\beta + (z_i - z_i)\gamma + (\alpha_i - \alpha_i) + e_{it} - \bar{e}_i = (x_{it} - \bar{x}_i)\beta + e_{it} - \bar{e}_i$$

This approach allows consistent estimation even when arbitrary correlation with α_i is present, as the individual effect is removed from the estimating equation by the within-group transformation. However, the within-group transformation also removes the time-invariant variables z_i, so that the parameter γ is not identified/identifiable.

As noticed by Verbeek (2008, pag. 354), the within-group estimator can also be obtained by an instrumental variable regression in which the within-group transformed x_{it} (i.e., $x_{it} - \bar{x}_i$) is used as the instrument in the level Eq. (2) so that it is possible to write (see also Im et al. 1999):

$$E[(x_{it} - \bar{x}_i)(\alpha_i + e_{it})] = E[(x_{it} - \bar{x}_i)(y_{it} - \beta x_{it} - \gamma z_i)] = 0 \tag{3}$$

Still, only the parameter β can be identified as there is no information (internal instrumental variable or orthogonality condition) for the estimation of γ.

In this set up, the BMS89 assumption of homogenous correlation can be specified as

$$\text{For all } i,\ E(x_{it}\alpha_i) = constant \text{ is the same for all } t \tag{4}$$

The assumption of homogenous correlation between x_{it} and α_i would be satisfied any time the growth of x_{it} is not linked to the individual effect. In contrast, no restriction is imposed on the correlation between α_i and the level of x_{it}.

BMS89 noticed that, by assuming homogeneity over time in the correlation structure between the regressor x_{it} and the individual component α_i, further valid instruments for the level equation can be defined, and it is possible to write

$$E[(x_{is} - \bar{x}_i)(\alpha_i + e_{it})] = 0 \quad \text{for all } t \text{ and } s \tag{5}$$

The set of T instruments only exploits the within variability of x_{it}. However, increasing the number of instruments also allows identifying γ, without the need to distinguish endogenous and exogenous regressors as in the original HT81 framework. What the literature has, so far, failed to recognize is that the *additional* instrumental variables (orthogonality conditions) can be effectively employed for the identification of γ.

The estimation approach we propose in this paper builds on the BMS89 insight and suggests exploiting the $T - 1$ additional instruments spanning from the assumption of homogenous correlation in order to identify the effect of time-invariant variable(s). The additional instruments are[6]:

$$x_{i2} - \bar{x}_i,\ x_{i3} - \bar{x}_i,\ x_{i4} - \bar{x}_i,\ \dots,\ x_{iT} - \bar{x}_i \tag{6}$$

[3] The strict exogeneity assumption should also be invoked ruling out simultaneity and the presence of feedback effects.

[4] The FE estimator can also be obtained by adoption of the correlated random effect approach (Wooldridge, 2002), that allows to control for the presence of unobserved heterogeneity by adding group means to the estimating equation. In our context, it would not be possible to separately identify both the coefficients of the variable S_i itself and the "Mundlak correction" as the two variables would be perfectly collinear.

[5] For simplicity, time effects are not included. However, the inclusion of a full set of time dummy variables in the model would not change the validity of the approach.

[6] $x_{i1} - \bar{x}_i$ is omitted from the set of instruments to avoid multicollinearity in the full instrument set that also includes $x_{it} - \bar{x}_i$ as instrument for x_{it}.

Overall, this enlarged set of instrumental variables, spanning from the homogeneity assumption (4), allows the identification of both parameters β and γ.

Of course, besides exogeneity, the proposed set of instruments should also be "strong", such that weak instrument issues do not emerge in estimation, in other words z_i has to be related to within-deviations of x_{it}. To better gauge the implications of this condition, consider this linear combination of the full instruments set that can be interpreted in terms of variable growth (to be more easily interpreted in economic terms than the within-group transformation):

$$x_{it} - x_{i1},\ x_{it} - x_{i2},\ \dots,\ x_{it} - x_{iT} \tag{7}$$

for $t = 1, \dots, T$.[7] The correlation between z_i and x_{it} should be driven by the correlation between z_i and the growth of x_{it} (by homogeneity, unrelated with α_i), or in other words, different values of z_i should lead to different evolution (growth path) of x_{it}.

In other words, when z_i is controlled for, the growth of x_{it} is unrelated to α_i.

Summing up, the following conditions are needed for consistent estimation:

(A1) Homogeneity: For all i, $E(x_{it}\alpha_i)$ is the same for all t
(A2) When $k = g = 1$, $T \geq 2$
(A3) z_i is related to the within evolution of x_{it}: $E(z_i(x_{it} - \bar{x}_i)) \neq 0$ (or, equivalently, $E(z_i(x_{it} - x_{is})) \neq 0$, $s, t = 1, \dots, T$)

Condition (A1) is needed for consistent estimation of β and γ: if the condition does not hold, the set of instrumental variables in (6) does not satisfy the exogeneity requirement.[8]

Condition (A2) is the order condition for identification. As we estimate two parameters (β, γ) associated with two endogenous variables, at least two instrumental variables are needed. This condition can be generalized to the case of multiple regressors. In the case of $k > 1$ time-varying variables and $g > 1$ time-invariant variables, condition (A2) would be generalized to $(T-1)k \geq g$. Note, however that condition (A2) needs not to hold for *all* the variables included in the model, it should nonetheless hold for a subset of variables, $k_{(A1)}$, satisfying condition (A1). In this case, condition (A2) would be written as, when $k > 1$ and $g > 1$, $(T-1)k_{(A1)} \geq g$.

Condition (A3) requires that instruments are relevant for z: the parameter γ could only be identified when z is correlated with the instruments in (6), that is, to the within evolution of x (Stock and Yogo, 2005).[9]

The standard theory of instrumental variables applies to the estimation.

As an alternative strategy, we suggest the 'FE filtering' method with IV estimation (FEF-IV) proposed by Pesaran and Zhou (2018):

- In the first step, the within-group transformation is applied and the FE estimate of β is obtained;
- Then, residuals from the level Eq. (2), $\hat{\alpha}_i + \hat{e}_{it}$, are computed, and their group mean is regressed on z_i, using an IV strategy that exploits instruments (6).[10]

[7] The linear combination of these instruments reproduces the within-group transformation as: $(x_{it} - x_{i1} + x_{it} - x_{i2} + \dots + x_{it} - x_{iT})/T = x_{it} - \bar{x}_i$. The instruments in (6) can be obtained as $x_{it} - x_{ij} = (x_{it} - \bar{x}_i) - (x_{ij} - \bar{x}_i)$ for $j = 2, \dots, T$.

[8] Note that this assumption is customarily exploited when estimating dynamic panel data model with endogenous or predetermined variables using the system-GMM estimator (Blundell and Bond, 1998; Blundell et al., 2001).

[9] Note however that, in case of weak instruments, robust methodologies could be employed, such as computation of Anderson–Rubin (AR) or conditional likelihood ratio (CLR) confidence intervals (Andrews et al., 2019).

[10] Pesaran and Zhou (2018) also propose a FEF strategy that uses OLS estimation in the second stage. This approach would produce consistent estimation of γ in the case of exogenous z_i, an assumption that we would prefer to avoid in our setting (see Pesaran and Zhou, 2018).

The proposed framework allows residual serial correlation and/or heteroskedasticity. Extension to unbalanced panel datasets is straightforward too. However, note that, in the case of unbalanced panel datasets, using all available instruments in the estimation procedure would imply using only the balanced subset of the data. It may therefore be wise to limit the number of instruments to maximize estimates' precision and/or instruments' strength.

The performance of the proposed estimator is assessed in the Monte Carlo experiments, reported in Appendix A.

5. Estimation results

The empirical analysis proceeds in two steps, mirroring the two predictions derived in Section 2. We first estimate Eq. (1) by standard FE, which identifies the interaction between R&D intensity and the open-sourcing indicator (β_5), but not the level effect of openness (β_4). We then apply the estimation strategy proposed in Section 4, which recovers β_4 and allows us to test the full dual-role of R&D hypothesis, that is, publication-based openness raises productivity only for firms whose internal R&D provides the absorptive capacity to exploit it.

Results from the FE estimation of Eq. (1) are reported in Table 3. Column (1) reports the baseline estimation as in Hall and Mairesse (1995) where there is no assumption of constant returns to scale (which is not rejected in our sample as the coefficient of $\ln(L)$ is not statistically different from zero).

Column (2) introduces the open-sourcing variable S_i and its interaction with the R&D stock. Because S_i is time-invariant, its main effect cannot be separately identified in a fixed-effects specification. Columns (3)-(5) progressively extend the model by adding sector-by-year and region-by-year interaction terms to capture time-varying heterogeneity across sectors and regions (e.g. specific regional or industrial policies). Finally, Column (6) imposes constant returns to scale by omitting $\ln(L)$ from the regression. The estimated coefficients for physical capital and R&D stocks are statistically significant and consistent in magnitude with those typically reported in the literature.

To address potential concerns about the strict exogeneity assumption, the lower panel of Table 3 reports the p-value from a test of strict exogeneity based on an augmented regression that includes the forward (lead) values of $\ln(K/L)$, $\ln(RD/L)$, and $\ln(L)$ (see Wooldridge, 2002, p.285). The reported p-value corresponds to the null hypothesis that the coefficients on the lead variables ($t+1$) are jointly equal to zero.

The regression coefficient of the open-sourcing strategy S_i is not estimated with the traditional FE approach, as it would be treated just like an unobserved time-invariant individual effect, whereas the interaction between S_i and R&D stock can be identified. The interaction term is positive and statistically significant at 10% significance level, indicating that firms with an above-sector average open-sourcing strategy have much higher returns from R&D expenditures in terms of labor productivity.

This result provides direct support for the first prediction of the dual-role framework ($\beta_5 > 0$): openness and internal R&D are complements in the production of value added. Through the lens of the mechanisms discussed in Section 2, firms that publish above their sector average are more deeply embedded in external knowledge networks and more exposed to feedback and complementary research. A larger internal knowledge stock increases the rate at which external knowledge is identified, assimilated, and recombined into productive use (Cohen and Levinthal, 1989). The estimate is also in line with the evidence that internal R&D and external knowledge acquisition reinforce each other rather than substitute for one another (Cassiman et al., 2008; Griffith et al., 2004).

To test the second and more relevant prediction of our framework, it is essential to identify the coefficient β_4 of the open-sourcing strategy S_i, as different values of β_4 have different implications for interpreting our results. A positive β_4 would suggest that an open-sourcing strategy always benefits firms' productivity, regardless of the R&D capability,

Table 3
FE estimator of the productivity-publications relationship.

Dep.var: VA/L	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(K/L)$	0.611***	0.613***	0.613***	0.605***	0.605***	0.605***
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
$\ln(RD/L)$	0.319***	0.276***	0.295***	0.241**	0.257**	0.294***
	(0.099)	(0.094)	(0.094)	(0.104)	(0.105)	(0.080)
$\ln(L)$	−0.025	−0.014	−0.031	−0.050	−0.067	–
	(0.091)	(0.091)	(0.092)	(0.101)	(0.102)	
S		(omitted)	(omitted)	(omitted)	(omitted)	(omitted)
$\ln(RD/L)\times S$		0.395	0.411	0.417	0.424*	0.429*
		(0.282)	(0.270)	(0.267)	(0.256)	(0.253)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Sector × Year FE	No	No	Yes	No	Yes	Yes
Region × Year FE	No	No	No	Yes	Yes	Yes
N	6677	6677	6677	6677	6677	6677
p strict exo.	0.872	0.847	0.917	0.813	0.848	0.632

Note: Statistical significance $^{***}p<0.01$, $^{**}p<0.05$, $^{*}p<0.1$.

Table 4
FEF (S_i treated as exogenous) and FEF-IV (S_i treated as endogenous) estimate of model (1); instrument set built by assuming the variable $\ln(K/L)$ as the homogenous variable.

Variable	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Method	FEF	FEF-IV	FEF-IV	IV	FEF	FEF-IV	FEF-IV	IV	FEF-IV	IV
Instruments	–	All	2012,15	2012,15	–	All	2012,15	2012,15	2012,15	2012,15
$\ln(K/L)$	0.603***	0.673***	0.618***	0.618***	0.605***	0.674***	0.620***	0.620***	0.620***	0.620***
	(0.027)	(0.031)	(0.029)	(0.029)	(0.027)	(0.031)	(0.029)	(0.029)	(0.029)	(0.029)
$\ln(RD/L)$	0.306***	0.336***	0.352***	0.413***	0.257**	0.303**	0.292**	0.367***	0.294***	0.330***
	(0.110)	(0.131)	(0.131)	(0.141)	(0.104)	(0.137)	(0.124)	(0.135)	(0.097)	(0.107)
$\ln(L)$	−0.076	−0.023	−0.029	0.041	−0.066	−0.007	−0.004	0.084		
	(0.101)	(0.125)	(0.115)	(0.115)	(0.101)	(0.121)	(0.117)	(0.121)		
$\ln(RD/L)\times S$	–	–	–	–	0.424*	0.262*	0.549*	0.589*	0.550*	0.574*
					(0.254)	(0.255)	(0.321)	(0.332)	(0.315)	(0.324)
S	0.148	−0.096	−0.765	−0.607	−4.542	−3.014	−7.331*	−7.293*	−6.906*	−6.870*
	(0.222)	(0.694)	(0.784)	(0.574)	(2.862)	(2.879)	(3.783)	(3.930)	(3.616)	(3.667)
N obs.	6677	3157	4752	4572	6677	3157	4752	4752	4752	4752
N.firms	1223	318	587	587	1223	318	587	587	587	587
Hansen-J	–	–	–	0.359	–	–	–	0.297	–	0.310
J p-value				0.549				0.586		0.578
KP F	–	2.867	13.16	6.236	–	2.867	13.16	9.354	13.16	23.52

Note: Firm, time, sector-time and region-time FE included in all regressions.
Statistical significance $^{***}p<0.01$, $^{**}p<0.05$, $^{*}p<0.1$.

for example through pure reputation effects that demand no complementary internal capability. On the contrary, the level effect of openness should be negative at low R&D intensity ($\beta_4<0$), and the overall effect of openness should turn positive only beyond a threshold of R&D investment, supporting the hypothesis that R&D plays a crucial moderating role within the firm. For example, disclosing research outputs to attract partners, investment, or skilled labor may not improve productivity for firms with low R&D capacity, as they lack the absorptive capabilities to benefit fully from knowledge spillovers. Discriminating between these two mechanisms requires identifying the coefficient of a time-invariant regressor, which motivates the methodology proposed in Section 4.

Table 4 reports the estimates obtained using the two-step procedure described in Section 4, applying both OLS and IV estimation in the second step. Although the OLS coefficients are likely to be biased, they are reported for comparison. The IV estimates rely on the instrument set proposed in this paper, which enables identification of the effect of the time-invariant variable S_i. All results are obtained under the assumption that capital is the homogeneous variable.[11]

[11] The R&D variable was found to be too weak to yield reliable results. Moreover, R&D expenditures are expected to be more volatile than investments in physical capital, making the latter more suitable to satisfy the homogeneity assumption.

Table 4 presents the estimated coefficients from alternative specifications of the productivity model. Models (7)–(10) exclude the interaction term between R&D intensity and the open-sourcing indicator, $\ln(RD/L)\times S_i$. In these regressions, the coefficient on S_i is not statistically significant. When the model incorporates the moderating role of S_i on the effect of R&D (Models 11–14), the coefficient on S_i becomes negative and statistically different from zero.

In Models (8) and (12), however, the KP-F statistic reveals that the instruments are extremely weak. To strengthen identification, we restrict the instrument set to the subset that maximizes the first-stage F-statistic—that is, the regression of z_i on all possible combinations of instruments in the set (6). Models (9) and (13), based on this stronger subset of instruments (among all the ones that are available), yield an insignificant coefficient on S_i when the interaction is excluded (Model 9), but a negative and statistically significant coefficient when the interaction with R&D is included (Model 13). Joint IV estimates are also reported in Models (10) and (14) and provide additional support for the results.

The contrast between the two sets of specifications is itself informative about the theory. When the moderating role of R&D is ignored (Models 7–10), the coefficient on S_i is small and not statistically significant. This is because the unconditional association between openness and productivity has a positive effect among R&D-intensive firms and a negative effect among the others, and the two largely offset. Once the interaction is included (Models 11–14), the two mechanisms are

revealed: the negative and significant estimate of β_4, and the positive estimate of β_5. Firms that disclose scientific knowledge without a sufficiently large internal knowledge base bear the costs of openness, as they lack the absorptive capacity required to generate offsetting gains.

Finally, Models (15) and (16) exclude $\ln(L)$ (and its demeaned counterpart from the instrument set), since this variable is never statistically significant in any specification. Estimated coefficients are coherent across the two estimators (FEF-IV and joint IV estimation), supporting the presence of a differential productivity effect of R&D depending on firms' open-sourcing propensity. We argue that estimates obtained via the FEF-IV two-step approach may be preferable, as this method ensures that first-stage estimates (in our case, the fixed-effects estimation of time-varying variables) are not affected by potential misspecifications in the assumptions required to identify time-invariant regressors in the second stage (Kripfganz and Schwarz, 2019). However, the two estimators provide very similar coefficients of the variables of interest.

Based on the estimates in Table 4, we can compute the "turning point", i.e. the level of (log) R&D expenditure at which the advantage of open-sourcing firms appears compared to the case $S_i = 0$. The difference between the level of productivity for firms with $S_i = 1$ as compared to firms with $S_i = 0$ is given by

$$\beta_4 + \beta_5 \ln(RD/L)$$

so the difference becomes positive for $\ln(RD/L) = -\beta_4/\beta_5$. According to the estimates in Model (15), the turning point is about 12.56, corresponding to about the 90th percentile in the distribution of $\ln(RD/L)$ observed in our sample. Using the IV estimates from Model (16), the turning point decreases to 11.97, which is slightly below the 85th percentile of the same distribution. The IV estimator also allows computation of the 95% confidence interval of the turning point, that is (10.24, 13.69).

Fig. 3 plots the predicted values of labor productivity, $\ln(VA/L)$, based on the estimates from Model (16), as a function of $\ln(RD/L)$ for the two cases $S_i = 1$ and $S_i = 0$. The first, second, and third quartiles of the distribution of $\ln(RD/L)$ are also shown in the figure. Consistent with the interpretation of the estimated coefficients, the productivity advantage of being linked to science arises only for firms with higher levels of R&D investment; for firms with lower R&D intensity, the effect of openness is negative. These results support the view that prior knowledge (i.e., absorptive capacity, Cohen and Levinthal, 1989) is necessary to benefit from scientific inputs, implying that publication activity at the firm level should be complemented by investments that strengthen absorptive capacity (Rosenberg, 1990; Cohen and Levinthal, 1989).

It is worth noting that our sample is drawn from the world's top corporate R&D investors, yet the estimated threshold lies between the 85th and 90th percentiles of the within-sample distribution of R&D intensity. In other words, even among leading R&D performers, only firms in the upper tail obtain a net productivity gain from publication-based openness. For the majority, the costs of disclosure are not offset by absorptive gains. Such a demanding threshold is consistent with the cumulative nature of absorptive capacity, which cannot be purchased on the market but must be accumulated through sustained internal research. It also connects to the evidence that leading corporate R&D performers increasingly restrain scientific disclosure when appropriation concerns dominate. Arora et al. (2018) document a declining valuation of publications relative to patents at the firm level, and Amoroso and Vannuccini (2026) find that the participation of top R&D investors in publicly funded consortia significantly reduces project publication output. If disclosure yields net gains only above a high capability threshold, a selective and declining engagement in corporate science is what one would expect from firms weighing its costs and benefits.

Table 5
Robustness checks: alternative measure of open-sourcing and samples; FEF-IV (S_i treated as endogenous) estimate of model (1); instrument set built by assuming the variable $\ln(K/L)$ as the homogenous variable.

Variable	(17)	(18)	(19)
Method	FEF-IV	FEF-IV	IV
Instruments	All	2011,14,15	2011,14,15
$\ln(K/L)$	0.675***	0.632***	0.631***
	(0.031)	(0.030)	(0.030)
$\ln(RD/L)$	0.197**	0.253**	0.204*
	(0.111)	(0.107)	(0.114)
$\ln(RD/L) \times S$	0.502***	0.520**	0.454**
	(0.189)	(0.207)	(0.212)
S	−5.606***	−6.128***	−5.352**
	(2.021)	(2.297)	(2.294)
N obs.	3157	4177	4177
N.firms	318	457	457
Hansen-J			4.340
J p-value			0.114
KP F	3.16		10.44

Firm, time, sector-time and region-time FE included in all regressions.
Statistical significance $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

5.1. Robustness checks

In this Section we explore the robustness of our results to changes in the definition of the open-sourcing variable and to sample composition.

First, we consider a modified definition of the open-sourcing, in which we use as a benchmark the third quartile rather than the mean. Results are reported in Table 5, broadly confirming our baseline results. The interaction terms between R&D and open-sourcing remain positive and statistically different from zero, and the estimated effect of the open-sourcing variable remains negative. The turning point computed on the basis of these results is 11.78 when IV results of Model (19) are considered (95% confidence interval 9.56–14.00). When computed according to FEF-IV results in Model (18), the figure is 11.79, reinforcing previous results on the role of absorptive capacity as a key determinant of gains from knowledge disclosure.

Next, we explore the stability of the results of Model (16) in Table 4 to sample composition by removing regions and sectors from the analysis (Figs. 4 and 5 - tables listing the full set of results are reported in Appendix B). The Figures report the estimated coefficients of our main variables along with 90% and 95% confidence intervals. Overall, results remain stable across the various specifications. Statistical significance is, in some instances, mitigated, and this happens when we remove the regions or sectors with the largest share of observations, suggesting that lack of power is the main motivation for this result.

6. Discussion and conclusions

This paper examines whether firms' engagement in scientific publication, interpreted as a form of selective knowledge disclosure, and hence as a proxy for an open-sourcing long-term strategy, is associated with the labor productivity gains, and whether this association depends on the intensity of firms' own R&D. The question fits into the discussion on structural change in knowledge production. Indeed, the boundaries between corporate and open science are blurry (David, 1994; Fagerberg, 2000; Arora et al., 2018), and the productivity consequences of openness become an empirical matter of how disclosure interacts with firms' internal capabilities, rather than an unconditional assumption that diffusion is unconditionally beneficial.

Three findings emerge from the empirical analysis. First, publication-based openness and internal R&D are complements: the productivity returns to R&D are significantly higher for firms that publish above their sector norm. Second, the level effect of openness is negative, so that disclosure is associated with lower productivity among firms with weak internal knowledge bases. Third, the two effects jointly

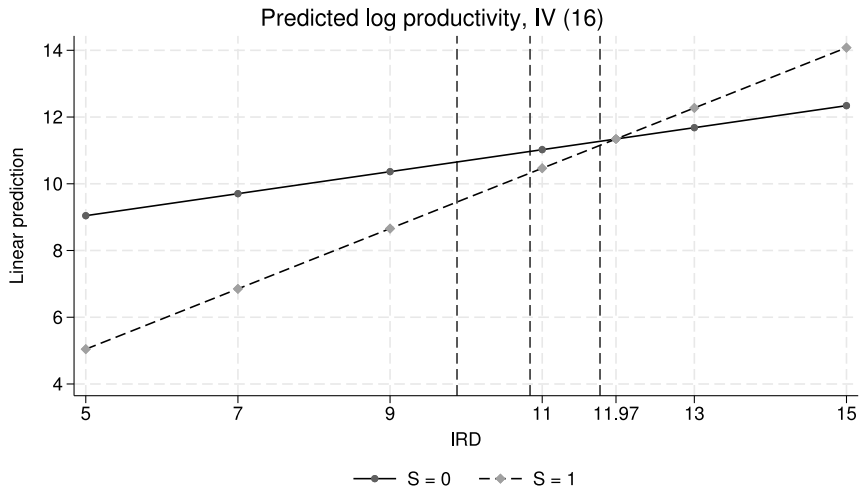


Fig. 3. Predicted productivity from Model (16) in Table 4, as a function of R&D, $S = 0$ and $S = 1$; vertical lines identify Q25, median, and Q75 in the distribution of $\ln(RD/L)$.

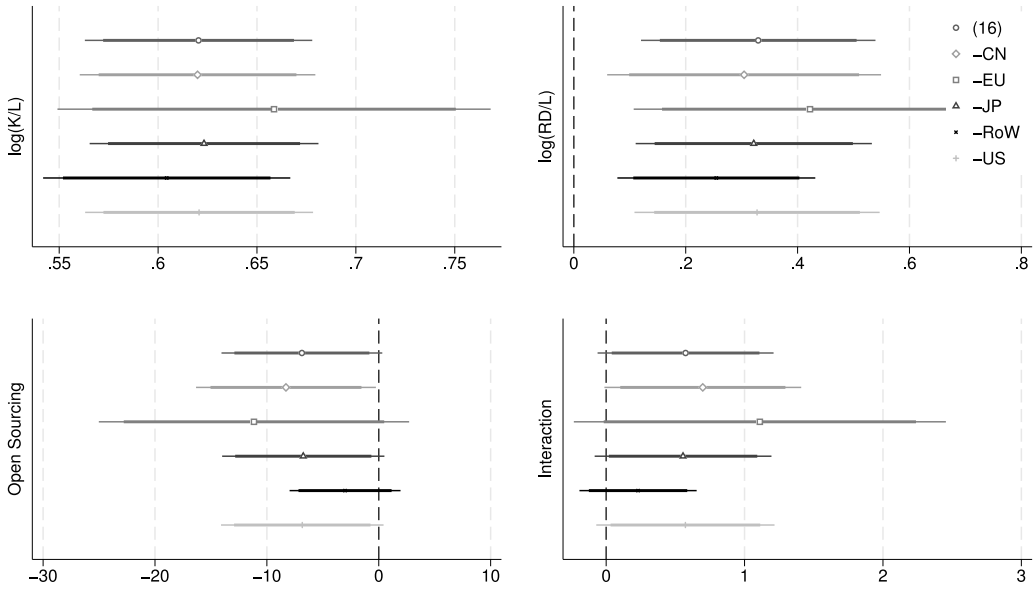


Fig. 4. Stability of Model (16) with respect to sample composition, by region: estimated coefficients from Model (16) in the sample excluding one region at a time.

imply a turning point located in the upper tail (85th–90th percentile) of the R&D-intensity distribution. Even within a sample of top corporate R&D investors, only the most research-intensive firms convert openness into a net productivity advantage. Taken together, these results support the dual-role view of R&D (Cohen and Levinthal, 1989; Griffith et al., 2004). Internal research both generates knowledge and builds the absorptive capacity without which the knowledge-access, talent, signaling, and learning mechanisms unlocked by publication cannot be turned into productive efficiency. The other side of the coin is that a firm which is "open without sufficient capacity" bears the appropriation costs of disclosure while capturing few of its benefits.

The paper also makes a methodological contribution. Because our openness indicator is time-invariant, its level effect cannot be recovered by standard FE estimation. We show that the homogeneous-correlation assumption of Breusch et al. (1989), so far exploited only for efficiency purposes, generates internal instruments that identify the coefficients of endogenous time-invariant regressors without partitioning the regressors into exogenous and endogenous sets and without external instruments. Monte Carlo evidence confirms that the estimator is reliable when the assumptions hold, and that standard diagnostics detect their failure. The approach is applicable well beyond the present setting, wherever the effects of time-invariant characteristics (e.g., gender, nationality, institutional features, founding conditions) must be identified within a FE framework.

Our results have implications for both firm strategy and policy. For firms, openness is not a substitute for internal research but a complement. The decision to disclose scientific knowledge should be evaluated jointly with the investments that determine the capacity to absorb the knowledge flows disclosure activates. For policy, initiatives that encourage or mandate open-science practices among firms, such as conditions linked to publicly funded research, implicitly assume that diffusion benefits all participants. Our estimates suggest a more nuanced picture: the gains from openness are concentrated among firms with strong absorptive capacity, and openness requirements imposed on firms below the capability threshold may weaken rather than strengthen their performance. Policies promoting knowledge diffusion are therefore likely to be more economically effective when coupled with instruments that build firms' internal research capabilities, a complementarity that echoes the broader lesson of the structural-change

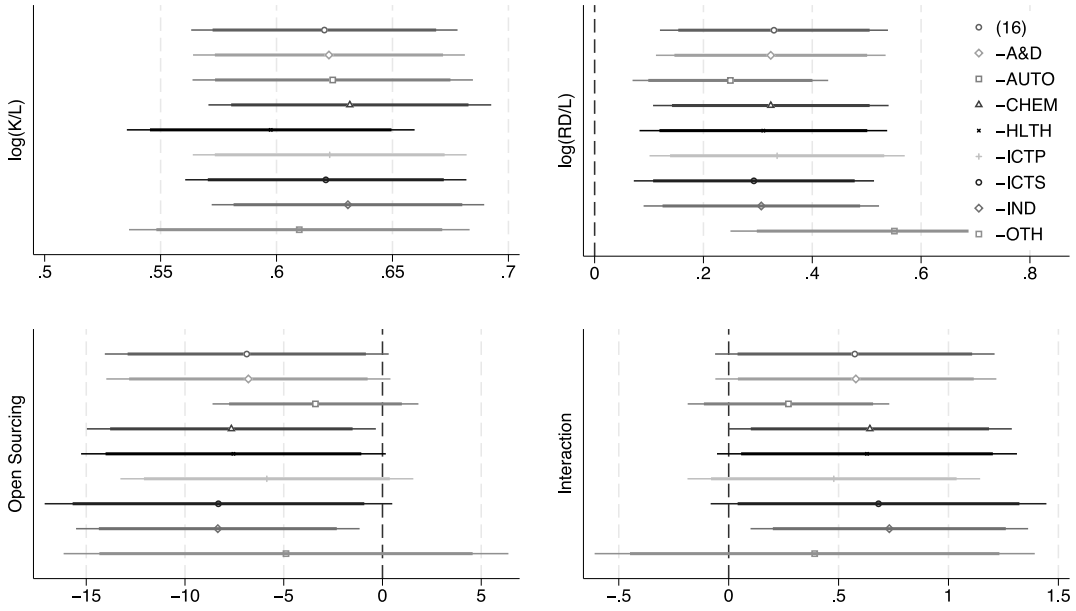


Fig. 5. Stability of Model (16) with respect to sample composition, by sector: estimated coefficients from Model (16) in the sample excluding one sector at a time.

literature that the diffusion of new knowledge regimes is mediated by the absorptive conditions of adopters (Fagerberg, 2000).

Our study has limitations that could offer directions for future work. First, our data do not allow us to discriminate empirically among the specific mechanisms through which publication-based openness operates, whether it is through knowledge access, talent attraction, signaling, or learning effects. Identifying the relative contribution of each channel is a promising avenue for future research. Second, the openness indicator is measured over 2014–2016 and treated as a persistent structural orientation of the firm; while this treatment is consistent with the strategic drivers of corporate publishing discussed in Section 2, longer publication panels would permit testing whether transitions in openness status carry productivity consequences. Third, the analysis concerns the world's top R&D investors, and the estimated threshold should not be extrapolated to the broader firm population, where absorptive-capacity constraints are likely to bind even more tightly. Finally, although our identification strategy addresses the endogeneity of the time-invariant openness indicator, the estimates are best read as conditional associations consistent with the proposed mechanisms rather than as definitive causal effects.

CRediT authorship contribution statement

Sara Amoroso: Writing – review & editing, Writing – original draft, Conceptualization, Identification. **Randolph Luca Bruno:** Writing – review & editing, Writing – original draft, Conceptualization, Identification . **Laura Magazzini:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization, Identification.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used ChatGPT and CoPilot for English editing. The authors reviewed and edited the output as needed and take full responsibility for the content of the published article.

Appendix A. Monte Carlo experiment

Our Monte Carlo setup is inspired by the experiments of Pesaran and Zhou (2018), henceforth referred to as PZ18, and adapted to our framework.

We consider the following data-generating process with one endogenous time-varying variable x_{it} and one endogenous time-invariant variable z_i $(i = 1, \dots, N; t = 1, \dots, T)$:

$$y_{it} = 1 + \alpha_i + \beta x_{it} + \gamma z_i + \epsilon_{it} \tag{8}$$

with $\beta = \gamma = 1$. The focus will be placed on the estimation of γ.

We consider $N = 1000$ to resemble the sample size in the empirical application in Section 5, and $T = 4, 8$. In all experiments, both x_{it} and z_i are correlated with α_i, but uncorrelated with ϵ_{is} at all periods $(s = 1, \dots, T)$ so that the strict exogeneity assumption is satisfied, and the FE estimation (within-group transformation) provides consistent estimation of β. In particular, we set:

$$z_i = 1 + \alpha_i + \xi_i \tag{9}$$

with $\xi_i \sim IIDN(0, 1)$, and

$$x_{it} = \alpha_i g_t + \xi_i h_t + w_{it} \tag{10}$$

with g_t and h_t generated as $U(a, b)$ and kept fixed across all replications.

The terms g_t and h_t drive the validity of assumptions (A1) and (A3). In particular

- Assumption (A1) holds when $g_t = g_1$ for all t;
- Condition (A3) is satisfied when h_t is changing over time, making the within-variability of x_{it}, i.e. the instruments in (6), relevant for z_i; the length of the interval (a, b) drives the instrument's strength.

In all experiments, z_i is endogenous, i.e. correlated with α_i. When the assumptions (A1)-(A3) hold, the proposed methodology allows identification of γ, by exploiting the homogeneity assumption (A1) and the correlation of z_i and the growth rate ("within evolution") of x_{it} (A3). The strength of the correlation between the instruments and the variable z_i is driven by the width of the interval (a, b) used for the generation of the term h_t. In particular, the experiments with strong identification are obtained by setting $a = 0.1$ and $b = 1.9$. The case of

Table 6
Results of Monte Carlo experiments: mean of $\hat{\beta}$ and $\hat{\gamma}$, standard deviation in parenthesis. DGP with (A1) homogenous x_{it}; and (A3) z_i related to within deviation of x_{it} ($a = 0.1; b = 1.9$) – order condition (A2) satisfied in all experiments.

	$T = 4$		$T = 8$	
	FEF-IV	IV	FEF-IV	IV
	Homoskedastic ϵ_{it}			
$\hat{\beta}$	1.001	1.003	.999	1.001
	(.021)	(.022)	(.012)	(.013)
$\hat{\gamma}$	1.003	1.002	1.009	1.008
	(.058)	(.058)	(.042)	(.042)
5% rej J	.049	.049	.056	−057
avg. KP	71.10	>100	77.45	>100
	Heteroskedastic (uncorrelated) ϵ_{it}			
$\hat{\beta}$	.999	1.001	1.000	1.002
	(.022)	(.022)	(.012)	(.013)
$\hat{\gamma}$	1.002	1.001	1.008	1.006
	(.059)	(.059)	(.040)	(.040)
5% rej J	.059	.059	.045	.044
avg. KP	71–63	>100	77.99	>100
	Heterosk., serially correlated ϵ_{it}			
$\hat{\beta}$	.999	1.002	1.001	1.002
	(.018)	(.019)	(.011)	(.013)
$\hat{\gamma}$	1.003	1.002	1.007	1.006
	(.065)	(.065)	(.045)	(.045)
5% rej J	.036	.037	.041	.040
avg. KP	70.57	>100	77.16	>100

Table 7
Results of Monte Carlo experiments: mean of $\hat{\beta}$ and $\hat{\gamma}$, standard deviation in parenthesis. DGP with non-homogenous x_{it}: (A1) not satisfied.

	$T = 4$		$T = 8$	
	FEF-IV	IV	FEF-IV	IV
	Homoskedastic ϵ_{it}			
$\hat{\beta}$	1.000	1.174	1.000	1.039
	(.020)	(.039)	(.011)	(.031)
$\hat{\gamma}$	1.156	1.073	1.473	1.442
	(.160)	(.137)	(.067)	(.069)
5% rej J	1.00	1.00	.999	.999
avg. KP	31.48	24.42	53.30	>100
	Heteroskedastic (uncorrelated) ϵ_{it}			
$\hat{\beta}$	1.000	1.171	1.000	1.040
	(.020)	(.040)	(.011)	(.031)
$\hat{\gamma}$	1.157	1.074	1.472	1.440
	(.157)	(.135)	(.071)	(.072)
5% rej J	1.00	1.00	1.00	1.00
avg. KP	31.47	24.35	53.30	>100
	Heterosk., serially correlated ϵ_{it}			
$\hat{\beta}$	1.000	1.176	1.000	1.038
	(.017)	(.040)	(.011)	(.031)
$\hat{\gamma}$	1.157	1.072	1.474	1.444
	(.158)	(.139)	(.069)	(.071)
5% rej J	1.00	1.00	1.00	1.00
avg. KP	30.85	23.80	53.05	>100

weak instruments is obtained by letting $a = 0.75$ and $b = 1.25$ when $T = 4$, and $a = 0.8$ and $b = 1.2$ when $T = 8$. The limiting case $a = b = 1$ would correspond to the case of irrelevant instruments, i.e. the case in which assumption (A3) is not satisfied.

The stochastic component of x_{it}, w_{it}, is generated as in PZ18, as an heterogeneous stationary AR(1) process:

$$w_{it} = \mu_i(1 - \rho_{w,i}) + \rho_{w,i} w_{it-1} + \sqrt{1 - \rho_{w,i}^2}\epsilon_{it}$$

with $\epsilon_{it} \sim IIDN(0, \sigma_{\epsilon,i}^2)$ for all i, $\sigma_{\epsilon,i}^2 \sim 0.5(1 + 0.5IID\chi_2^2)$, $w_{i0} \sim IIDN(\mu_i, \sigma_{\epsilon,i}^2)$, $\rho_{w,i} \sim IIDU(0, 0.98)$, $\mu_i \sim IIDN(0, 2)$.

Table 8
Results of Monte Carlo experiments: mean of $\hat{\beta}$ and $\hat{\gamma}$, standard deviation in parenthesis. DGP with weak association between z_i and within deviation in x_{it} condition (A3) satisfied with weak instruments.

	$T = 4$		$T = 8$	
	FEF-IV	IV	FEF-IV	IV
	Homoskedastic ϵ_{it}			
$\hat{\beta}$	.999	1.001	.999	1.000
	(.025)	(.025)	(.014)	(.015)
$\hat{\gamma}$	1.019	1.017	1.047	1.046
	(.177)	(.177)	(.098)	(.098)
5% rej J	.053	.053	.057	.058
avg. KP	8.340	6.286	8.687	7.663
	Heterosk. (uncorrelated) ϵ_{it}			
$\hat{\beta}$	1.000	1.001	1.001	1.002
	(.023)	(.024)	(.014)	(.015)
$\hat{\gamma}$	1.019	1.018	1.046	1.045
	(.181)	(.182)	(.100)	(.100)
5% rej J	.058	.057	.076	.076
avg. KP	8.320	6.268	8.655	7.635
	Heterosk., serially correlated ϵ_{it}			
$\hat{\beta}$	1.000	1.002	1.000	1.001
	(.018)	(.020)	(.013)	(.015)
$\hat{\gamma}$	1.025	1.024	1.047	1.046
	(.197)	(.197)	(.115)	(.115)
5% rej J	.063	.063	.060	.060
avg. KP	8.275	6.237	8.732	7.705

The FE α_i are generated as $0.5(\chi_2^2 - 2)$ for $i = 1, \ldots, N$. As for ϵ_{it}, as in PZ18, and consider three different designs:

- Homoskedastic ϵ_{it}:

  $\epsilon_{it} \sim IIDN(0, 1), i = 1, \ldots, N; t = 1, \ldots, T$

- Heteroskedastic ϵ_{it}:

  $\epsilon_{it} \sim IIDN(0, \sigma_i^2), i = 1, \ldots, N; t = 1, \ldots, T$

  with $\sigma_i^2 \sim 0.5(1 + 0.5\chi_2^2)$ for all i.

- Serially correlated and heteroskedastic ϵ_{it}:

  $$\epsilon_{it} = \rho_{\epsilon,i}\epsilon_{it-1} + \sqrt{1 - \rho_{\epsilon,i}^2}u_{it}$$

  for $t = -49, -48, \ldots, T$ (the first value of ϵ_{it} is set to 0 for all i), $u_{it} \sim IIDN(0, \sigma_{ui}^2)$ for all i and t, $\sigma_{ui}^2 \sim 0.5(1 + 0.5IID\chi_2^2)$, and $\rho_{\epsilon,i} \sim IIDU(0, 0.98)$ for all i. We discard the first 50 observations, using the remaining T observations in the experiments.

Results of Monte Carlo experiments, based on 1000 replications, are reported in Tables 6–8, that show Monte Carlo mean and standard deviation of the FEF-IV estimator, and the IV estimates of β and γ obtained by estimating the equation in levels and pooling the instrument set, i.e. using within-group transformation of the variables as instruments together with the instruments (6) spanning from the homogeneity assumption on x_{it}. The intercept is estimated but not reported.

To compute FEF-IV estimates, the FE estimator is first obtained. Residuals of the within-group estimation are then computed and their group average is considered the dependent variable in the second stage. This entails an IV estimation with independent variable z_i (and the intercept) and the BMS variables (6) used as instruments.[12] The tables also report, in the IV case, the 5% rejection rate of the Hansen J test

[12] Estimates are obtained using Stata. FEF and FEF-IV estimates are obtained using the *xtfef* command available at https://qiankunzhou.weebly.com/research.html. IV estimates are obtained with the *ivreg2* command Baum et al. (2010).

Table 9
Robustness checks: stability of results with respect to sample composition – IV estimates (Model 16 in Table 4) estimated by removing one region at a time. Instrument set built by assuming the variable $\ln(K/L)$ as the homogenous variable.

Excluded region	China	EU	Japan	RoW	US
$\ln(K/L)$	0.620***	0.659***	0.623***	0.605***	0.621***
	(0.030)	(0.0559)	(0.0295)	(0.032)	(0.029)
$\ln(RD/L)$	0.304**	0.422***	0.322***	0.2547***	0.327***
	(0.125)	(0.161)	(0.108)	(0.090)	(0.112)
$\ln(RD/L)\times S$	0.699*	1.111	0.556*	0.231	0.573*
	(0.363)	(0.686)	(0.326)	(0.216)	(0.329)
S	−8.297**	−11.15	−6.748*	−3.015	−6.838*
	(4.101)	(7.076)	(3.700)	(2.529)	(3.706)
N obs.	4271	2062	4539	3895	4241
N.firms	532	301	529	490	496
Hansen-J	0.275	0.449	0.301	0.652	0.316
J p-value	0.600	0.503	0.583	0.419	0.574
KP F	23.98	15.98	21.27	8.355	21.69

Firm, time, sector-time and region-time FE included in all regressions.
Statistical significance $^{***}p<0.01$, $^{**}p<0.05$, $^{*}p<0.1$.

Table 10
Robustness checks: stability of results with respect to sample composition – IV estimates (Model 16 in Table 4) estimated by removing one sector at a time. Instrument set built by assuming the variable $\ln(K/L)$ as the homogenous variable. Sector abbreviations: (A&D) Aerospace & Defence; (AUTO) Automobiles & other transport; (CHEM) Chemicals; (HLTH) Health industries; (ICTP) ICT producers; (ICTS) ICT services; (IND) Industrials; (OTH) Others.

	Excluded sector							
	A&D	AUTO	CHEM	HLTH	ICTP	ICTS	IND	OTH
$\ln(K/L)$	0.623***	0.6242***	0.632***	0.597***	0.623***	0.621***	0.631***	0.610***
	(0.0299)	(0.031)	(0.031)	(0.032)	(0.030)	(0.031)	(0.020)	(0.037)
$\ln(RD/L)$	0.324***	0.249***	0.324***	0.310***	0.335***	0.293***	0.306***	0.551***
	(0.108)	(0.092)	(0.110)	(0.116)	(0.120)	(0.113)	(0.110)	(0.154)
$\ln(RD/L)\times S$	0.579*	0.272	0.643*	0.629*	0.478	0.682*	0.731**	0.391
	(0.326)	(0.234)	(0.329)	(0.348)	(0.340)	(0.390)	(0.322)	(0.511)
S	−6.790*	−3.393	−7.652**	−7.546*	−5.854	−8.305*	−8.337**	−4.884
	(3.671)	(2.659)	(3.732)	(3.936)	(3.784)	(4.490)	(3.663)	(5.746)
N obs.	4623	4311	4465	4022	3900	4225	4108	3610
N.firms	574	530	551	487	474	515	513	455
Hansen-J	0.448	0.259	0.213	0.419	0.721	0.105	0.028	0.548
J p-value	0.503	0.611	0.644	0.518	0.395	0.746	0.868	0.459
KP F	18.56	20.37	18.94	6.938	28.81	10.97	33.85	305.3

Firm, time, sector-time and region-time FE included in all regressions.
Statistical significance $^{***}p<0.01$, $^{**}p<0.05$, $^{*}p<0.1$.

of overidentifying restriction,[13] and the average Kleibergen–Paap (KP) F test for assessing instrument strength.

Table 6 is our baseline estimation framework, in which all the assumptions needed for the application of the proposed methodology are satisfied, i.e. correlation between α_i and x_{it} is homogenous over time (i.e. A1 is satisfied when generating x_{it}, $g_t=1$ for all t), and z_i is related to within deviations of x_{it} (A3 satisfied, $h_t=t$ in Eq. (10)). As $T>2$, and we consider the case of one endogenous time-varying variable x_{it} and one endogenous time-invariant variable z_i, condition (A2) is satisfied in all experiments.

Results in Table 6 show that identification of γ is achieved, and similar results emerge with joint IV estimation as compared to the FEF-IV estimation (Pesaran and Zhou, 2018). The J-test of overidentifying restrictions has generally the correct size. The KP F statistic is large, pointing to the instruments' strength.

In Table 7, x_{it} is generated such that the homogeneity condition (A1) is not satisfied, i.e. the parameter g_t varies over time. In this case, FEF-IV has the advantage of providing a consistent estimator of β, whereas the estimate of γ is biased. Note that the J-test exhibits a 100% rejection rate.

Finally, Table 8 reports the results in the case when (A3) is only weakly satisfied, i.e. $h_t\sim U(a,b)$ when generating x_{it} in (10), with $a=0.75$ and $b=1.25$ when $T=4$, and $a=0.8$ and $b=1.2$ when $T=8$[14]: within-evolution of x_{it} is only weakly related to z_i, so that the BMS instruments are weak. We believe this is an interesting case as it resembles the results of our empirical application. Again, a consistent estimate of β is obtained with FEF-IV, and the estimate of γ exhibits little bias. Minor differences emerge when comparing FEF-IV and joint IV estimation. In this context, the KP statistic is small (as expected), allowing to recognize the problem related to instruments' strength in this setting, so that statistical inference robust to weak instruments can be chosen by the researcher.

Appendix B. Robustness checks - details

This Appendix reports the details of the estimated equations shown in Figs. 4 and 5. In particular, Table 9 shows the detailed results of the estimation of Model (16) in Table 4 that are obtained by removing one region at a time, whereas in Table 10 we remove one sector at a

[13] See Kripfganz and Schwarz (2019) for applying the J-test in a context analogous to the one of this paper. Being the number of replications 1000, the 95% confidence interval for the test size at that 5% level is 3.65%–6.35% (Morris et al., 2019).

[14] Setting $h_t=1$ would result in irrelevant instruments, i.e. within-evolution of x_{it} is not dependent upon z_i, so that the BMS instruments are irrelevant. Results are not shown to save space, however a consistent estimate of β is obtained with FEF-IV, and the estimate of γ is biased. In this context, the KP statistic is extremely small, pointing to a problem related to instruments' strength in this setting. Results are available from the authors upon request.

time. The set of estimates aims at assessing whether the results in our analysis are driven by one particular region or one particular sector. Overall, the results remain stable across the various specifications.

Data availability

The authors do not have permission to share data.

References

Ahn, J.M., Minshall, T., Mortara, L., 2015. Open innovation: a new classification and its impact on firm performance in innovative SMEs. J. Innov. Manag. 3 (2), 33–54.

Ahn, S.C., Schmidt, P., 1995. Efficient estimation of models for dynamic panel data. J. Econometrics 68 (1), 5–27.

Alexy, O., George, G., Salter, A.J., 2013. Cui bono? The selective revealing of knowledge and its implications for innovative activity. Acad. Manag. Rev. 38 (2), 270–291.

Almeida, P., Hohberger, J., Parada, P., 2011. Individual scientific collaborations and firm-level innovation. Ind. Corp. Chang. (ISSN: 0960-6491) 20 (6), 1571–1599.

Amemiya, T., MaCurdy, T.E., 1986. Instrumental-variable estimation of an error-components model. Econ.: J. Econ. Soc. 54 (4), 869–880.

Amoroso, S., Vannuccini, S., 2026. Teaming up with large R&D investors: Good or bad for knowledge production and diffusion? J. Technol. Transf. Forthcoming.

Andrews, I., Stock, J.H., Sun, L., 2019. Weak instruments in instrumental variables regression: Theory and practice. Annu. Rev. Econ. 11 (1), 727–753.

Archibugi, D., Filippetti, A., Frenz, M., 2013. Economic crisis and innovation: Is destruction prevailing over accumulation? Res. Policy 42 (2), 303–314.

Arora, A., Belenzon, S., Patacconi, A., 2018. The decline of science in corporate RandD. Strat. Manag. J. 39 (1), 3–32.

Baum, C., Schaffer, M., Stillman, S., 2010. ivreg2: Stata module for extended instrumental variables/2sls, gmm and ac/hac, liml and k-class regression.

Belenzon, S., Patacconi, A., 2014. How does firm size moderate firms' ability to benefit from invention? evidence from patents and scientific publications. Eur. Manag. Rev. 11 (1), 21–45.

Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. J. Econometrics 87 (1), 115–143.

Blundell, R., Bond, S., Windmeijer, F., 2001. Estimation in dynamic panel data models: improving on the performance of the standard gmm estimator. In: Nonstationary Panels, Panel Cointegration, and Dynamic Panels. In: Advances in Econometrics, Emerald Group Publishing, Limited.

Bogers, M., Chesbrough, H., Heaton, S., Teece, D.J., 2019. Strategic management of open innovation: A dynamic capabilities perspective. Calif. Manage. Rev. 62 (1), 77–94.

Breusch, T.S., Mizon, G.E., Schmidt, P., 1989. Efficient estimation using panel data. Econometrica (ISSN: 00129682, 14680262) 57 (3), 695–700.

Cassia, L., Colombelli, A., Paleari, S., 2009. Firms' growth: Does the innovation system matter? Struct. Change Econ. Dyn. (ISSN: 0954-349X) 20 (3), 211–220, Special Issue on Regional Dimensions of Innovation.

Cassiman, B., Veugelers, R., Zuniga, P., 2008. In search of performance effects of (in) direct industry science links. Ind. Corp. Chang. 17 (4), 611–646.

Chen, J.-R., Kan, K., Tung, I.-H., 2016. Scientific linkages and firm productivity: Panel data evidence from taiwanese electronics firms. Res. Policy (ISSN: 0048-7333) 45 (7), 1449–1459.

Chen, J., Yue, R., Wu, J., 2020. Testing for individual and time effects in the two-way error component model with time-invariant regressors. Econ. Model. 92 (C), 216–229.

Chesbrough, H.W., 2003. Open Innovation: The New Imperative for Creating and Profiting from Technology. Harvard Business School.

Chesbrough, H.W., Garman, A.R., 2009. How open innovation can help you cope in lean times. Harv. Bus. Rev. 87 (12), 68–76.

Cockburn, I.M., Henderson, R.M., 2000. Publicly funded science and the productivity of the pharmaceutical industry. Innov. Policy Econ. 1, 1–34.

Cohen, W.M., Levinthal, D.A., 1989. Innovation and learning: the two faces of r & d. Econ. J. 99 (397), 569–596.

David, P.A., 1994. Why are institutions the 'carriers of history'?: Path dependence and the evolution of conventions, organizations and institutions. Struct. Change Econ. Dyn. (ISSN: 0954-349X) 5 (2), 205–220.

Edler, J., Fagerberg, J., 2017. Innovation policy: what, why, and how. Oxf. Rev. Econ. Policy (ISSN: 0266-903X) 33 (1), 2–23.

Fagerberg, J., 2000. Technological progress, structural change and productivity growth: a comparative study. Struct. Change Econ. Dyn. (ISSN: 0954-349X) 11 (4), 393–411.

Filippetti, A., Archibugi, D., 2011. Innovation in times of crisis: National systems of innovation, structure, and demand. Res. Policy 40 (2), 179–192.

Gal, P.N., 2013. Measuring total factor productivity at the firm level using oecd-orbis. OECD Economics Department Working Papers, No. 1049.

Gambardella, A., 1992. Competitive advantages from in-house scientific research: The us pharmaceutical industry in the 1980s. Res. Policy (ISSN: 0048-7333) 21 (5), 391–407.

Grassano, N., Hernandez, H., Fako, P., Tuebke, A., Amoroso, S., Georgakaki, A., Napolitano, L., Pasimeni, F., Rentocchini, F., Compano, R., et al., 2021. The 2021 eu industrial r&d investment scoreboard. Technical report, Joint Research Centre (Seville site).

Greene, W., 2011. Fixed effects vector decomposition: A magical solution to the problem of time-invariant variables in fixed effects models? Political Anal. 19 (2), 135–146, (ISSN: 10471987, 14764989).

Griffith, R., Redding, S., Reenen, J.V., 2004. Mapping the two faces of r&d: Productivity growth in a panel of oecd industries. Rev. Econ. Stat. 86 (4), 883–895.

Hall, B.H., Mairesse, J., 1995. Exploring the relationship between r&d and productivity in french manufacturing firms. J. Econometrics 65 (1), 263–293.

Hausman, J.A., Taylor, W.E., 1981. Panel data and unobservable individual effects. Econometrica (ISSN: 00129682, 14680262) 49 (6), 1377–1398.

Henkel, J., 2006. Selective revealing in open innovation processes: The case of embedded linux. Res. Policy 35 (7), 953–969.

Hicks, D., 1995. Published papers, tacit competencies and corporate management of the public/private character of knowledge. Ind. Corp. Chang. 4 (2), 401–424.

Im, K.S., Ahn, S., Schmidt, P., Wooldridge, J., 1999. Efficient estimation of panel data models with strictly exogenous explanatory variables. J. Econometrics 93 (1), 177–201.

Kripfganz, S., Schwarz, C., 2019. Estimation of linear dynamic panel data models with time-invariant regressors. J. Appl. Econometrics 34 (4), 526–546.

Leten, B., Kelchtermans, S., Belderbos, R., 2022. How does basic research improve innovation performance in the world's major pharmaceutical firms? Ind. Innov. 29 (3), 396–424.

Luo, Y., 2022. Illusions of techno-nationalism. J. Int. Bus. Stud. 53, 550–567.

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

Morris, T.P., White, I.R., Crowther, M.J., 2019. Using simulation studies to evaluate statistical methods. Stat. Med. 38 (11), 2074–2102.

Parchomovsky, G., 2000. Publish or perish. Mich. Law Rev. 98 (4), 926–952.

Paunov, C., 2012. The global crisis and firms' investments in innovation. Res. Policy 41 (1), 24–35.

Pellens, M., Malva, A. Della, 2017. Corporate science, firm value, and vertical specialization: evidence from the semiconductor industry. Ind. Corp. Chang. (ISSN: 0960-6491) 27 (3), 489–505.

Pénin, J., 2007. Open knowledge disclosure: An overview of the evidence and economic motivations. J. Econ. Surv. 21 (2), 326–347.

Pesaran, M.H., Zhou, Q., 2018. Estimation of time-invariant effects in static panel data models. Econometric Rev. 37 (10), 1137–1171.

Petricevic, O., Teece, D.J., 2019. The structural reshaping of globalization: Implications for strategic sectors, profiting from innovation, and the multinational enterprise. J. Int. Bus. Stud. 50 (9), 1487–1512.

Plümper, T., Troeger, V.E., 2007. Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects. Political Anal. 15 (2), 124–139.

Polidoro, F., Theeke, M., 2012. Getting competition down to a science: the effects of technological competition on firms' scientific publications. Organ. Sci. 23 (4), 1135–1153.

Rosenberg, N., 1990. Why do firms do basic research (with their own money)? Res. Policy 19 (2), 165–174.

Rotolo, D., Camerani, R., Grassano, N., Martin, B.R., 2022. Why do firms publish? a systematic literature review and a conceptual framework. Res. Policy (ISSN: 0048-7333) 51 (10), 104606.

Simeth, M., Cincera, M., 2016. Corporate science, innovation, and firm value. Manag. Sci. 62 (7), 1970–1981.

Simeth, M., Raffo, J.D., 2013. What makes companies pursue an open science strategy? Res. Policy (ISSN: 0048-7333) 42 (9), 1531–1543.

Stern, S., 2004. Do scientists pay to be scientists? Manag. Sci. 50 (6), 835–853.

Stock, J., Yogo, M., 2005. Testing for Weak Instruments in Linear IV Regression. Cambridge University Press, New York, pp. 80–108.

Verbeek, M., 2008. A Guide to Modern Econometrics. John Wiley & Sons.

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press.