forthcoming, International Economic Review

TECHIES AND FIRM LEVEL PRODUCTIVITY *

James Harrigan[†] Ariell Reshef[‡] Farid Toubal[§]

June 24, 2026

Abstract

We study the impact of techies—engineers and other technically trained workers—on firm-level productivity. We first report new facts on the role of techies in the firm using French administrative data. Techies are STEM-skill intensive and are associated with innovation, as well as with technology adoption, management, and diffusion within firms. Using structural econometric methods, we then estimate the effect of techies on firm-level Hicks-neutral productivity in both manufacturing and non-manufacturing industries. We find that techies raise firm-level productivity, and that this effect goes beyond the employment of R&D workers, extending to ICT and other techies. In non-manufacturing firms, the impact of techies on productivity operates mostly through ICT and other techies, not R&D workers.

JEL codes: D2, D24, F1, F16, F6, F66, J2, J23, J24, O52.

Keywords: productivity, R&D, ICT, techies, STEM skills.

---

[*] We thank Daron Acemoglu, Mary Amiti, Zsofia Barany, Eric Bartelsman, Flora Bellone, Andrew Bernard, Tibor Besedes, Esther Ann Bøler, Davin Chor, Chiara Criscuolo, Jan De Loecker, Alon Eizenberg, Georg Graetz, Giammario Impullitti, Beata Javorcik, Guy Michaels, Jean-Marc Robin, and John Van Reenen for helpful comments. The paper benefited from comments from seminar participants at Aarhus University, American University, AMSE, Bar Ilan University, Ben Gurion University, CREST, Dresden University, Groningen University, France Stratégie, Hebrew University, University of Le Mans, LSE, Nottingham University, New York University, SAIS, and the Paris Trade Seminar, as well as workshop participants at the CEPR, DEGIT, ENEF, ERWIT, IFN (Stockholm), SETC, and TRISTAN (Dresden). This research was supported by the Bankard Fund for Political Economy at the University of Virginia and by the Agence National de la Recherche under grant numbers ANR-16-CE60-0001-01 and ANR-10-EQPX-17 (Investissements d'Avenir) and by CEPREMAP.

[†] University of Virginia and NBER: james.harrigan@virginia.edu.

[‡] Université de Paris 1, Panthéon-Sorbonne, Paris School of Economics, CEPII and CESifo: ariell.reshef@psemail.eu.

[§] Université Paris-Dauphine- PSL, CEPII, CESifo and CEPR: farid.toubal@dauphine.psl.eu.

1 Introduction

Technically trained workers—engineers, ICT specialists, and other technical staff who we will call *techies*—are thought to play a central role in shaping productivity. Their importance is implicit in modern growth theory, where the accumulation and application of technical knowledge is the key engine of long-run growth (Romer, 1990). Empirically, allocation of more talent to technical occupations is associated with faster growth (Murphy et al., 1991), and historical studies show that the availability of technical skills shaped the timing and diffusion of major growth episodes.[1] Despite the prominent perceived role of techies in productivity-enhancing activities, evidence at the firm level is limited. Our paper provides just such evidence.

We estimate the contribution of techies to firm-level productivity using matched survey and administrative data from France. Our analysis shows that techies increase firm-level productivity, and that this effect is not limited to techies that are engaged in research and development (R&D). Techies working in ICT and other technology-related functions also contribute significantly. Their effects are pervasive, extending beyond manufacturing to non-manufacturing industries.

We identify techie workers using the French occupational classification. INSEE (2003) clearly distinguishes techies from other occupations: their tasks are characterized by the installation, management, maintenance, and support of ICT, product and process design, R&D activities, as well as other technology-related tasks. In line with this, we show that techies differ from other workers by their STEM qualifications, skill profiles, and experience. Survey data also show that techies are strongly associated with innovation efforts and

---

[1] For example, Kelly et al. (2014) and Ben Zeev et al. (2017) emphasize the role of the apprentice system in supplying the skills necessary for technology adoption during the British Industrial Revolution. Kelly et al. (2023) show that industrialization began in areas with abundant technically trained mechanics, while Hanlon (2022) highlights the emergence of professional engineers. Maloney and Valencia Caicedo (2017) document spatial patterns in engineer intensity across the Americas in the 19th century and relate them to long-run income. An early discussion of techie labor markets appears in Blank and Stigler (1957).

outcomes (e.g., patents), and are involved in adopting and diffusing technology within firms.

Techies are not homogeneous. Using information from INSEE (2003), we classify them into three groups based on the precise tasks they perform: R&D, ICT, and Other technical occupations. This classification allows us to estimate their distinct contributions to firm-level productivity. While R&D techies are concentrated in manufacturing, ICT techies are more prevalent in non-manufacturing. Therefore, focusing only on R&D techies or only on manufacturing understates the broader role of techies in the economy.

To quantify the contribution of techies to productivity, we construct an unbalanced panel of French firms from 2011 to 2019, merging administrative records on revenues, capital, materials, and detailed occupational labor input (hours). We estimate structural models of firm-level Hicks-neutral total factor productivity (TFP), where firm-level productivity is a function of lagged productivity and lagged expenditure on techies. This methodology allows us to identify the impact of techies on firm-level productivity.

Our identification strategy relies on three assumptions. First, techies affect firm productivity with a lag. Second, they do not contribute directly to current output. Third, productivity follows a controlled Markov process. Our controlled Markov assumption, which is standard in the productivity estimation literature, implies that we can estimate the effect of techies on productivity without estimating a model of optimal techie employment.[2] We employ different production function estimators that deal with identification challenges in different ways. We implement this approach across a range of specifications, including flexible nonlinear productivity processes and alternative classifications of techies. We control for exporting status throughout, as exporting is known to correlate with firm productivity (De Loecker, 2013; Barrows et al., 2025). We also show in Section 6.3.3 that our findings are not sensitive to the inclusion of managers as an additional determinant of firm-level productivity (Bloom et al., 2017).

---

[2] Starting with Olley and Pakes (1996), applications of the controlled Markov methodology to estimating the drivers of firm-level productivity include Levinsohn and Petrin (2003), Doraszelski and Jaumandreu (2013), Ackerberg et al. (2015), and many more.

We find that firms that employ techies experience a substantial increase in future productivity: 4–5 percent higher productivity one year later, with a cumulative effect of over 30 percent over ten years. While R&D techies drive this pattern in manufacturing, ICT and other techies also contribute significantly. In non-manufacturing, only ICT and other techies enhance productivity, while R&D techies have no significant effect. Disaggregating techies into engineers and technicians, we show that both increase productivity across sectors, with larger effects for engineers.

Our assumption that techies do not affect current output but do affect future productivity is key to our research design. We examine the validity of this assumption. We reject the null hypothesis that techies are no different than other workers that contribute only to current output directly in a standard way, in favor of the alternative that our assumption about their contribution to firm output only through firm productivity is a better fit to the data.

Related research. A growing literature examines the role of technically trained workers in shaping firm-level outcomes such as productivity, employment structure, and output. Tambe and Hitt (2014) motivate this line of inquiry by noting that "the technical know-how required to implement new IT innovations is primarily embodied within the IT workforce." Similarly, Deming and Noray (2018) argue that "STEM jobs are the leading edge of technology diffusion in the labor market." While there is an extensive literature on the firm-level returns to overall IT investment and R&D expenditure, these papers do not study the productivity implications of the workers who install, manage, and diffuse these technologies inside firms. In addition, papers that use R&D expenditure to study their effect on firm productivity are at risk of double counting, since, as we show below, about 75 percent of R&D expenditure is on labor.

A key reason for this gap is the lack of matched firm-occupation data in most administrative or survey datasets. An exception is Barth et al. (2017), who link individual workers to manufacturing plants in U.S. data, which allows them to match occupations to workplaces. They show that 80 percent of scientists and engineers work outside R&D occupations, and that the share of these workers is related to plant-level revenue per worker.[3]. Another exception is Harrigan et al. (2021), who use detailed occupational data for the entire French private sector (1994–2007) and show that firms with more techies experience faster employment growth and within-firm skill upgrading. Earlier studies by Lichtenberg (1995) and Tambe and Hitt (2012) estimate a positive output elasticity of IT labor, while Brynjolfsson and Hitt (1996) consider IT spending. Tambe and Hitt (2014), using a novel dataset on IT worker mobility, interpret job-switching patterns as evidence of inter-firm knowledge spillovers. More recently, Brynjolfsson et al. (2024) construct firm-level IT usage measures from online job postings.

However, none of these studies structurally estimate the effect of techies on productivity, nor do they distinguish between different types of techies—for example, those working in R&D versus ICT. Hall et al. (2013) distinguish between R&D and ICT investments using a sample of Italian firms, but they study impacts on sales per worker rather than productivity *per se*. Like us, they examine how R&D and ICT correlate with innovation, but they do not capture the broader role of technical workers beyond these categories, such as our "Other techies" group. Moreover, they consider only manufacturing firms, while we also study the impact of techies in other industries outside of manufacturing.

Hsieh and Rossi-Hansberg (2023) provide related evidence from the service sector. They show that R&D and ICT employment is associated with market expansion, which they attribute to greater productivity, and they emphasize the role of ICT in what they call the "industrialization of services." We share their view that techies contribute to output through productivity-enhancing activities. Their findings provide additional motivation to our analysis of productivity effects beyond manufacturing and to examining different classes of techies.

The broader literature on productivity measurement recognizes the substantial heterogeneity in firm-level TFP, but the determinants of this heterogeneity remain poorly under-

---

[3] Barth et al. (2017) study "Science and engineering workers", a category that has substantial but incomplete overlap with what we call "techies"

stood. As emphasized in the survey by Syverson (2011), identifying what drives productivity differences is an open question. De Loecker and Syverson (2021) argue that few papers tackle this challenge in a structural framework, which is needed to estimate jointly productivity and what drives it. This is precisely what we do.

Two pioneering studies on what drives firm-level productivity are De Loecker (2013), who examines the effects of exporting, and Doraszelski and Jaumandreu (2013), who study R&D expenditure in manufacturing. These two papers jointly estimate TFP levels along with the drivers of TFP, and we build on their frameworks. Crepon et al. (1998) also study the effect of R&D on productivity, but rely on a cross-sectional setting and different methodology.[4]

The rest of the paper is organized as follows. In Section 2 we describe the sources and construction of our datasets. In Section 3 we discuss the role of techies within firms, highlighting their technical expertise and their crucial role in adopting, mediating, and diffusing technology at the firm level. Section 4 presents a simple model of how techies affect productivity and analyzes optimal techie employment. Section 5 describes our econometric methodology. In Section 6 we present our empirical results, including various sensitivity checks to assess the robustness of our baseline findings. We conclude in Section 7 with a summary of our key results and a discussion of their implications for policymakers.

2 Data

We construct a firm-level panel for the French private sector over 2011–2019 by merging three confidential administrative datasets.[5] The estimation sample covers firms in 17 manufacturing and non-manufacturing industries (Table A1).[6] We complement these data with

---

[4] Crepon et al. (1998) emphasize the endogeneity problems in estimating this relationship (selection and simultaneity). The insights in De Loecker (2013) and Doraszelski and Jaumandreu (2013)—which we rely on—address these issues in ways that we discuss in Section 5.

[5] The year 2011 is the first in which occupation codes at the 4-digit level are comprehensively reported and 2019 the last pre-pandemic year.

[6] *Coke and refined petroleum* (NAF 19) is excluded due to negligible hours shares. *Transportation and storage* (NAF 49–53) is excluded because the GLZ production function estimator fails to converge. *Computers and electronics* (NAF 26) is dropped because of its very high intensity in techie workers. When nearly all workers in a sector are techies, there is insufficient variation to identify the effect of techies on productivity. *IT and R&D consulting sectors*, specifically Computer Programming, consultancy, and related activities

surveys that help characterizing techies and their roles within firms. Firms are matched across datasets using a common identifier (SIREN). Additional details are provided in Appendix A.

2.1 The composition of labor within firms

Employment data come from the DADS.[7] All firms with employees report wages, paid hours, occupation, and 2-digit sector. Our baseline labor input is hours paid. As a robustness check, we construct quality-adjusted labor input following Fox and Smeets (2011).[8]

DADS reports nearly 500 detailed 4-digit PCS occupation codes. We use the detailed definitions of these codes from (INSEE, 2003) to select the 56 4-digit occupations that we classify as techies. These are reported in Table A2. These occupations involve installation, management, maintenance, and support of ICT; product and process design; longer-term R&D; and other technology-related tasks. In short, techie employment therefore measures firms' investment in technology. The 4-digit PCS codes allow us to classify techie occupations along two dimensions: (i) *technical managers and engineers* versus *technicians*; and (ii) technological orientation – *R&D techies*, *ICT techies*, and *Other techies* (Table A2).

The INSEE (2003) documentation shows that techies perform tasks distinct from other occupations (see Appendix A). For example, technical managers and engineers (PCS 38) differ from other managers (PCS 37) because the former prioritize "the scientific or technical aspect takes precedence over the administrative or commercial aspect", whereas for the latter "the administrative or commercial aspect prevails". Similar distinctions apply to technicians and other occupations.[9] Beyond occupational titles (reported in Table A2), the INSEE

(NAF 62), Data Processing, Hosting, and related activities (NAF 63.1), and Scientific R&D (NAF 72), are dropped because these firms' main output is techie services. In our framework, techies affect productivity by adopting, adapting, and diffusing technology within the firm. For IT consultancies and R&D service providers, techie labor is the primary input directly producing output, not a factor shaping how other inputs are combined. The production function interpretation is therefore fundamentally different in these sectors. The exclusion of high-techie-intensity sectors is analogous to dropping sectors where all firms export when estimating export premia: identification requires variation in the treatment variable.

[7] *Déclaration Annuelle de Données Sociales.*

[8] We multiply the hours of lower-paid, less-qualified workers by the ratio of their average wage to that of higher-paid, highly qualified workers, as in Gandhi et al. (2020).

[9] See pages 191, 221, and 343 of INSEE (2003).

documentation also makes clear that techies perform tasks that *support* production but are not production or fabrication tasks *per se*. This grounds our assumption that techies raise productivity rather than contribute contemporaneously to output.

R&D occupations explicitly reference "research and development," while ICT occupations reference "information technology," "computer science," or "telecommunications."

A close look at the detailed INSEE (2003) descriptions of the Other techies category yields two observations. First, this group exhibits heterogeneity in their composition, comprising engineers, technical executives, and technicians involved in the adoption and dissemination of technologies and production methods within their firms (unrelated to R&D). A case in point are the engineers and managers of production method (PCS 387c), who are responsible for adapting and optimizing manufacturing methods in the private sector. Second, while being notably distinct from production and fabrication occupations, they *optimize* the productivity of workers in those fields.

We assume that non-techie occupations contribute directly to current output.

2.2 Balance sheets and exporting

We use firm balance sheet information from the FARE dataset for 2011–2019 on including revenues, input expenditures, and the variables required to construct capital stocks.[10] Appendix A details the methodology for constructin firm-level capital. We use French Customs data to construct a firm-year export status indicator.

2.3 Survey data

We use three surveys to provide additional information on techies and to characterize their role within firms. First, the 2015 Training and Professional Qualification (TPQ) survey provides information on STEM education (Science, Technology, Engineering, Mathematics, including Computer Science) and post-degree training. It reports individuals' field of highest degree and subsequent training received. Second, the Annual Survey on the Means

---

[10] *Fichier Approché des Résultats Ésane.* The source of information is firms' tax declarations.

dedicated to Research and Development (R&D survey) provides data on firms' internal and external R&D expenditures. It reports the labor cost component of R&D spending, the share outsourced, and firms' innovation activities. Third, the Information and Communication Technology survey (ICT survey) documents ICT training and technology diffusion within firms.

3 Facts about techies

Using DADS and survey data, this section documents techies' education and training and their role in technology adoption and diffusion within firms. We present the main descriptive results here and report additional evidence and details in Appendix B.

3.1 Fact 1. The incidence of Techies across industries

Table 1 reports techie wage bill shares by category in our sample and across French manufacturing and non-manufacturing industries.

Table 1: Wage bill shares of techies by categories (2019)

	Overall (1)	Manufacturing (2)	Non-Manufacturing (3)	% techie wage bill in manufacturing (4)
All Techies	18.4	31.8	10.9	62.3
R&D	3.5	8.3	0.7	87.0
ICT	2.2	2.3	2.1	37.9
Other	12.7	21.1	8.0	59.8
Engineers	12.0	19.9	7.5	60.0
Technicians	6.4	11.8	3.4	66.5

Source: DADS. Columns (1), (2) and (3) report the wage bill share of Techies or their subcategories in the private sector overall, within manufacturing and within non-manufacturing industries, respectively. For these columns the R&D, ICT and Other rows sum to All Techies; similarly, the Engineer and Technicians rows sum to All Techies. Column (4) reports the share of the Techie wage bill (or sub-categories thereof) that is in manufacturing. The wage bill share of manufacturing in 2019 in our sample is 36 percent.

In column (1), techies account for 18% of the private-sector wage bill, with a higher share in manufacturing (32%, column 2) than in non-manufacturing (11%, column 3). Across sectors, Other techies represent a larger share of the techie wage bill than R&D or ICT techies, motivating analysis beyond R&D activities.[11] Column (4) shows that 62% of techie expenditures are in manufacturing and more than one-third in non-manufacturing. This justifies studying productivity in both sectors rather than restricting attention to manufacturing, as is common in the literature.

Disaggregating by type reveals further patterns. Eighty-seven percent of R&D techie expenditures are in manufacturing (column 4), and manufacturing is substantially more R&D-techie intensive than non-manufacturing (columns 2 and 3). The productivity effects of R&D techies can therefore be studied largely within manufacturing. In contrast, 62% of ICT techie expenditures are in non-manufacturing, while ICT intensity is similar across sectors. This underscores the need to include non-manufacturing when assessing the productivity effects of techies.

Table 1 also reports wage bill shares by occupation. Engineers account for roughly twice the techie wage bill share of technicians.

3.2 Fact 2. Techies have more STEM education and training than other occupations

We use the TPQ survey to classify degrees and training and build an indicator for STEM(see Appendix B). The survey includes 26,861 individuals with valid observations, of whom 5.4% are Engineers (PCS 38) and 5.1% are Technicians (PCS 47), shares close to those in the DADS data.

Table B1 shows that techies have substantially more STEM education and training than other occupations. Overall, 63% of techies hold a STEM degree and/or training, and about one-fifth have both a STEM degree and additional STEM training. STEM degrees are more common among engineers (55%) than technicians (41%).

By contrast, STEM education is rare in other PCS occupations: 11% hold a STEM degree,

---

[11] Barth et al. (2017) report that 80 percent of U.S. private-sector scientists and engineers worked outside R&D occupations in 2013, close to the 18.5% share of R&D techies in our techie wage bill.

fewer than one-fifth have either a STEM degree or training, and only 2% have both. This also applies to administrative and commercial managers. These patterns are consistent with the distinctions in the PCS documentation (INSEE 2003) and support the view that techies perform technologically oriented roles distinct from other workers, including managers.

3.3 Fact 3. Most R&D spending is on wages and occurs in-house

In the structural analysis, we use the techie wage bill share to measure firm resources devoted to productivity improvement. Here we compare this measure to total R&D expenditures.

Total R&D spending includes materials and capital goods, which can generate double counting in production function estimation. First, materials used for R&D are included in total materials inputs and cannot be separately identified. Second, R&D capital expenditures are part of total investment, which we use to construct capital stocks. In addition, capital spending often occurs in "spikes", overstating productivity effort in investment years and understating it otherwise. Focusing on R&D wages and excluding techies from current labor input avoids double counting and provides a more stable indication of firms' R&D effort.

The R&D survey reports labor costs associated with R&D, as well as how much of the firm's R&D budget is spent in-house, particularly on wages related to R&D. Table B3 shows that wages account for most of R&D spending, especially when R&D is done within the firm. For example, the median share of externally-sourced R&D services is zero, while the mean is only 9 percent. For the average and median firm wage costs are 67 percent of total R&D spending, and 74 percent of in-house R&D. These findings are consistent with those of Saunders and Brynjolfsson (2016) in a sample of U.S. firms, where they find that more than half of all spending on IT was on IT-related techies.[12] Similarly, Schweitzer (2019) finds that in 2014, labor costs accounted for 60 percent of aggregate R&D spending in France.[13]

A potential concern is that firms may purchase ICT, R&D, or related consulting ser-

---

[12] Saunders and Brynjolfsson (2016) find that for a sample of 127 large publicly traded US firms from 2003 to 2006, half of all spending on IT is for "Internal IT Services (e.g., custom software, design, maintenance, administration)". Including IT training services brings the share to 0.54.

[13] The remainder 40 percent are split into 6 percent capital expenditures and 34 percent "other current expenses".

vices, which would appear as purchased services rather than internal productivity-enhancing activity. Table B3 indicates that this is not a large concern, since expenditure on R&D is overwhelmingly spent within the firm, with the median firm spending nothing on external R&D. Moreover, less than 3 percent of techie hours are in the IT and R&D consulting sectors in 2019, which implies that over 97 percent of the hourly services supplied by techies are obtained in-house rather than purchased from consultants.[14]

3.4 Fact 4. Techies are positively associated with the diffusion of ICT skills within firms

The ICT survey reports whether firms provide training to develop or improve ICT skills. ICT training is uncommon: 18 percent of firms offer it (Table B9). After merging the ICT survey with DADS, we estimate linear probability models of the likelihood of offering ICT training, controlling for firm size and including sector and year fixed effects (Table B10).[15]

We find a strong association between the likelihood of offering ICT training and the employment of techies, even after controlling for firm size. This association is particularly strong for ICT techies, and is much weaker for other techie categories. We interpret ICT training as an investment in worker skills that affects future performance, and techies are closely associated with this investment activity.

3.5 Fact 5. Techies are positively associated with patenting and innovation

The R&D survey reports firm-level patent filings and product and process innovation. As is observed elsewhere, patenting is rare. The firm at the 75^{th} percentile of the patenting distri-

---

[14] We refer to the IT and R&D consulting sectors as industry codes 62 (Computer Programming, consultancy, and related activities), 631 (Data Processing, Hosting, and related activities; web portals), and 72 (Scientific R&D) in the NAF classification. These are dropped from our analysis.

[15] Controlling for firm size captures the ability of firms to overcome fixed costs more generally. Thus, our regressions pick up the Techie-specific association with ICT training, over and above the higher propensity of larger firms to offer training, a fixed cost activity. In practice, controlling for size does not influence our results. Controlling for firm size accounts for fixed-cost considerations. The estimates therefore capture the techie-specific association with ICT training. Results are unchanged when size is excluded.

bution files no patents, and the 95^{th} percentile firm files only 4 patents. The 99^{th} percentile firm files 26 patents, and the top four firms file around 2,000. In contrast, innovation is quite common: only a quarter of firms report no process or product innovations in the past year, while half report having both (Table B11). We match the survey outcomes with the information on techies from the DADS.

Patenting is positively correlated with all forms of R&D expenditure—internal and external, wages and other spending (Table B12). The strongest correlations are with R&D wages and internal R&D. The techie wage bill is also positively associated with patenting (Table B13), particularly for R&D techies.

Techies are positively related to both product and process innovation (Table B14). Interestingly, the R&D and ICT techie wage bills are similarly correlated with product innovation (although in non-manufacturing the relationship for ICT techies is not statistically significant). In contrast, Other techies are uncorrelated with product innovation. The R&D and Other techie wage bills are positively related to process innovation (although in non-manufacturing the relationship for R&D techies is not statistically significant). In contrast, ICT techies are not associated with process innovation.

Overall, techie wage expenditures are systematically related to patenting and innovation. The patterns suggest differentiated roles: R&D techies are associated with both product and process innovation; ICT techies with product innovation; and Other techies with process innovation. These findings are consistent with Hall et al. (2010), who link R&D to both types of innovation.[16] Arora et al. (2017) show that corporate research in the U.S. increases innovation and patenting, with productivity effects linked to researcher quality, plausibly reflected in wages.

---

[16] As in related work, our methodology does not separately identify product and process channels in productivity estimation.

4 Why only some firms employ techies: An illustrative model

Despite the potential productivity gains, relatively few firms employ techies. This fact mirrors a well-documented fact in trade: although exporting is associated with higher productivity, only a subset of firms export. Following Melitz (2003), the standard explanation is the presence of fixed or variable costs that make exporting profitable only for high-productivity firms. In this section we develop a simple model of firm decision-making that applies the same logic to techie employment. The model rationalizes why only some firms choose to employ techies, and motivates our structural estimation strategy.

For maximum simplicity, suppose there are only two periods. Firm f takes the demand, costs, and initial period log productivity ω_{ft-1} as given and has to choose optimal techie employment T_{ft-1} to maximize profits. The relationship between techies and changes in productivity is

$$\omega_{ft} = \omega_{ft-1} + Max\left[\beta ln T_{ft-1}, 0\right], \quad \beta \geq 0\,.$$

Fixed costs of employing positive techies are κ_f and the wage of techies is r, so the cost of hiring techies is $rT_{ft-1} + \kappa_f$. With heterogeneity in the costs κ_f not all firms will employ techies, and we derive the following very intuitive conclusions in Appendix C. First, the optimal amount of techies is more likely to be positive when demand and/or initial productivity are higher. Conversely, the optimal amount of techies is more likely to be zero when their fixed costs are high. Second, the optimal amount of techies may be zero even if the fixed cost of employing techies is zero. This happens when the marginal cost of techies exceeds their marginal effect on revenue. Finally, when the optimal amount of techies is positive, it is increasing in initial productivity.

These predictions are consistent with findings in Brynjolfsson et al. (2023), who find larger incidence of IT investment in larger firms, who also benefit more from it (we estimate a similar pattern below). A further implication of this framework is that since firms that export will have a higher demand level, they will also be more likely to employ techies. This prediction aligns with our empirical strategy, where we control for exporting. It also echoes the findings of Aw et al. (2011), who show that firms endogenously select into both exporting and productivity-enhancing investments, such as R&D, when returns justify the fixed and sunk costs.

5 Empirical strategy

We estimate the effect of techies on firm-level productivity using structural productivity methods. Our setting is neither experimental nor a design with random assignment of techies. We interpret our estimates as the productivity estimation literature has always interpreted consistently estimated production function parameters: given firm choices of inputs, what are the resulting effects on output? Answering such questions does not require estimating systems of factor demands. Our estimation of techie effects on productivity is very similar to the way other authors have estimated the effects of R&D (Doraszelski and Jaumandreu, 2013) and exporting (De Loecker, 2013).

Our empirical strategy addresses two challenges. First, we observe revenue rather than physical output.[17] Second, productivity is unobserved to the econometrician but observed by firms when they make input decisions. We follow Grieco et al. (2016) (GLZ) and Gandhi et al. (2020) (GNR), which provide solutions to these challenges under different assumptions.

We model firm revenue using a standard CES demand system and a Hicks-neutral production technology. Let output be $Q_{ft} = \Omega_{ft} F(\mathbf{x}_{ft})$, where Ω_{ft} is firm-level productivity and $\mathbf{x}_{ft}$ is a vector of inputs, that comprises capital, labor (excluding techies), and materials. Taking logs, we write

$$q_{ft} = \omega_{ft} + f(\mathbf{x}_{ft}). \tag{1}$$

Demand for the firm's output takes the form $Q_{ft} = B_t P_{ft}^{-\eta}$, where B_t is an industry demand

---

[17] For a discussion of the challenges that such a data environment poses for estimation, see De Loecker and Goldberg (2014). Even when quantities are known, there remain difficult methodological issues in relating inputs to outputs in multi-product, multi-input firms, which cannot be overcome without strong assumptions. See Barrows et al. (2025) for a recent treatment of this challenge.

shifter, P_{ft} is the price that the firm charges and $\eta = 1/(1-\rho)$ is the elasticity of demand, with $\rho \in (0,1)$. Taking logs:

$$q_{ft} = b_t - \eta p_{ft}. \tag{2}$$

We do not observe physical output or prices. Instead, we observe revenue r_{ft}, which in logs equals $q_{ft} + p_{ft} + u_{ft}$, where u_{ft} is an ex post firm-specific demand shock. Combining equations (1) and (2), we obtain the firm-level "revenue production function":

$$r_{ft} = (1-\rho)b_t + \rho\omega_{ft} + \rho f(\mathbf{x}_{ft}) + u_{ft}. \tag{3}$$

Productivity evolves according to a controlled Markov process:

$$\omega_{ft} = g(\omega_{ft-1}, \mathbf{z}_{ft-1}) + \xi_{ft}, \tag{4}$$

where $\mathbf{z}_{ft-1}$ includes lagged techie employment and export status. The unobserved components ξ_{ft} and u_{ft} differ in timing: ξ_{ft} is known by the firm when input choices are made, while u_{ft} is realized ex post. This timing structure underpins identification in both the GLZ and GNR estimators. In this framework, the function g represents the firm's expectation of productivity in period t, which is based on the information set in $t-1$. This function can be written as the result of a dynamic programming problem, as in Olley and Pakes (1996).

Techies enter only in (4), not (3), reflecting our key identifying assumption: techies affect productivity with a lag and do not enter the production function as an input. This is consistent with how R&D or investment decisions are typically modeled in the productivity literature (Doraszelski and Jaumandreu, 2013).[18] We assess in Section 6.3 the robustness of our results to allowing techies to affect current output through their contribution to labor input.[19]

---

[18] Beaudry et al. (2016) use a similar framework, where cognitive labor contributes to future organizational capital but not current output.

[19] In our methodology the firm decision that takes into account lagged productivity, here, lagged techies, is identified through its role in the Markov process, not through a coefficient in the production function.

While we present equation (4) in a linear form, our results are robust to richer specifications of the productivity process. In Appendix H1, we report estimates from a third-order polynomial in lagged productivity. These are consistent with our baseline findings. We also report below results when we allow for non-linear effects of techies by interacting techies with lagged productivity.

Finally, in order to identify the production function we must assume that either the demand shocks u_{ft} are i.i.d. (as required by the GLZ estimator), or that they share the same persistence process as productivity ω_{ft} (the GNR estimator can have it either way).[20]

5.1 Controlled Markov approach

Equation (4) generalizes the productivity process in the control function literature, notably Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg et al. (2015) (OP/LP/ACF). It separates expected productivity, $g(\omega_{ft-1}, \mathbf{z}_{ft-1})$, from the innovation ξ_{ft}. The inclusion of lagged firm decisions, such as techie employment, in $\mathbf{z}_{ft-1}$ allows us to estimate the impact of those decisions, under the maintained assumption that ξ_{ft} is mean independent of past choices. This can be justified if firms are forward-looking and make decisions based on the expected outcomes of their actions.

De Loecker (2013) adopts this structure and emphasizes two implications. First, the lag structure implies that productivity innovations are realized after $\mathbf{z}_{ft-1}$ is chosen. Second, persistence in productivity is accounted for by including ω_{ft-1} as a state variable. The coefficient on $\mathbf{z}_{ft-1}$ in equation (4) thus captures the incremental effect of firm choices—such

---

Placing techies in the production function as a separate variable input violates the moment conditions that identify the controlled Markov. Doraszelski and Jaumandreu (2013) and De Loecker (2013) make precisely this choice for precisely this reason. We note that a large share of firm-years have techies = 0. Standard Inada conditions would make this nonsensical (CES, Cobb-Douglas, translog production functions, or the non-parametric approach of GNR). More importantly, adding techies as a separate input to the production function would require dropping every firm-year with no techies, which selects exactly on the extensive margin of adoption that the paper is built to characterize.

[20] This assumption is not directly verifiable. As shown in Melitz and Levinsohn (2006), when only revenue is observed (and not output quantity), in the presence of both persistent demand and persistent productivity processes, one can identify the production function as long as the persistence of both processes is the same. This is also acknowledged in De Loecker (2011). In this case of common persistence, what we call productivity is really "revenue productivity". This does not change the interpretation of the effect of techies (or any other control variable) on productivity in the controlled Markov.

as techie hiring—on future productivity.

Importantly, this framework does not require a structural model of firm decision-making. As in Doraszelski and Jaumandreu (2013), the estimation strategy identifies the impact of firm-level choices on productivity evolution from cross-sectional differences in outcomes between firms that do and do not hire techies, conditional on observables.

While keeping the controlled Markov, we do not apply the OP/LP/ACF control function estimator for two reasons. First, Gandhi et al. (2020) show that it suffers from weak instruments when there is insufficient input price variation. Second, Ackerberg et al. (2023) show that the associated GMM objective often exhibits multiple global minima, which can make estimates sensitive to starting values in the numerical search for a minimum. The GLZ and GNR estimators do not suffer from these issues, and are better suited to our data and identification problem.

5.2 The GLZ estimator

The estimator of Grieco et al. (2016) addresses the challenge of unobserved material input quantities by leveraging theoretical restrictions on firm behavior when only expenditures on materials are observed. It assumes a constant elasticity of substitution (CES) production function with constant returns to scale, and monopolistic competition with CES demand, as in Klette and Griliches (1996).

In order to overcome the challenge of unobserved material input quantities the GLZ estimator relies on the existence of at least one flexible input in addition to materials that adjusts flexibly after productivity is observed. These “static” inputs contrast with dynamic inputs such as capital. In practice, we assume that labor is such a static input. With these assumptions, the estimator identifies the production function and the demand elasticity by nonlinear least squares, without requiring instruments or assumptions on the productivity process. Once this is done we compute firm-level productivity $\{\widehat{\omega}_{ft}^{GLZ}\}$ for each firm and year, up to an industry-by-year demand shifter.

To estimate the impact of techies on productivity we regress $\widehat{\omega}_{ft}^{GLZ}$ on its lag and lagged firm-level techie decisions:

$$\widehat{\omega}_{ft}^{GLZ} = \theta_{it} + \lambda\widehat{\omega}_{ft-1}^{GLZ} + \boldsymbol{\beta}\mathbf{z}_{ft-1} + \xi_{ft}, \tag{5}$$

where the fixed effects θ_{it} capture industry-year demand shifts. We estimate this specification separately for manufacturing and non-manufacturing firms. Standard errors are computed by bootstrapping the full two-step procedure and clustering at the firm level.

5.3 The GNR estimator

The estimator of Gandhi et al. (2020) imposes no functional form on the production function and allows for arbitrary returns to scale. While GNR's baseline estimator assumes all physical quantities are observed, they propose an extension that relies only on revenues, building on Klette and Griliches (1996). Unlike GLZ, GNR identifies the full revenue production function (3) jointly with the productivity process via the moment conditions implied by the controlled Markov equation (4).

A limitation of the GNR approach in our setting is that the demand elasticity curvature parameter ρ is identified only from time-series variation. With only nine years of data (2011–2019), our estimates of ρ are imprecise. We therefore identify productivity scaled by ρ, $\widehat{\rho\omega}_{ft}^{GNR}$, industry by industry, and estimate:

$$\widehat{\rho\omega}_{ft}^{GNR} = \theta_{it} + \lambda\widehat{\rho\omega}_{ft-1}^{GNR} + (\beta\overline{\rho})\mathbf{z}_{ft-1} + \xi_{ft}, \tag{6}$$

where $\overline{\rho}$ is the average elasticity across industries (or just ρ_i if estimated for only one industry i). As expected, the estimated coefficients (reported below) in equation (6) are scaled-down relative to equation (5), which is consistent with demand elasticities being greater than unity ($\rho \in (0, 1)$), and which is what we estimate using the GLZ methodology. We implement the GNR estimator separately for manufacturing and non-manufacturing, and compute standard errors using a bootstrap clustered at the firm level.

5.4 Comparing GLZ to GNR

While GNR does not impose any functional form on the production function, this flexibility comes at a cost in our context. First, the parameters β in equation 6 are identified only up to a scale factor $\overline{\rho}$. Nonetheless, the signs and relative magnitudes of the elements of $\beta\overline{\rho}$ remain informative. Second, GNR assumes that input quantities are observed, while our data report only expenditures. To address this, we deflate material expenditures using industry-specific price indices. As noted by Grieco et al. (2016), this approach may bias measures of productivity dispersion.

Unlike GLZ, GNR does not require labor to be static, which is relevant in the French labor market context. French firms face a dual labor market, with permanent and short-term contracts and large firing costs, which introduce rigidity in adjustment. Accordingly, we implement GNR under two alternative assumptions: one treating labor and materials as static inputs (as in GLZ), and another allowing labor to be dynamic and slow-adjusting.

6 Results

We first discuss our baseline results using the GLZ methodology, and then report results from the GNR methodology. Section 6.3 reports extensive sensitivity analysis of our baseline results. Our measure of techies in the controlled Markov process (4) is a a dummy for positive values and the wage bill share of techies for each firm. These are defined either for all techies or by sub-group.

Quantification of the controlled Markov estimates requires descriptive statistics for different categories of techies, separately in manufacturing and non-manufacturing industries. Table (2) reports the percentage of observations with positive values for each techie category, the percentiles of the techie wage bill shares for observations that have positive values, as well as the 75^{th}-25^{th} percentile difference (the inter-quartile range or IQR). As explained in Section 2.1 above, overall techies are subdivided in two different ways: as R&D, ICT, or Other techies, and alternatively as Engineers or Technicians.

Table 2: Descriptive statistics for estimation sample

	Percent with positive values	Mean conditional on positive values	Percentiles of techie wage bill shares on positive support, percent					
			10	25	50	75	90	IQR
Manufacturing								
Techies	71.1	22.4	6.3	11.2	18.9	30.1	43.6	18.9
R&D techies	35.3	7.3	1.2	2.5	5.1	9.6	16.0	7.1
ICT techies	22.8	3.5	0.6	1.0	1.9	3.5	6.9	2.4
Other techies	69.0	18.2	5.4	9.4	15.5	24.1	34.4	14.7
Engineers	60.1	14.3	4.2	7.1	11.9	18.8	27.8	11.6
Technicians	60.2	12.1	2.5	5.0	9.4	16.1	25.0	11.1
Non-Manufacturing								
Techies	19.6	16.6	2.1	5.3	12.0	23.2	37.9	17.8
R&D techies	1.3	5.3	0.3	0.9	2.5	6.4	13.4	5.5
ICT techies	5.0	10.3	0.6	1.6	3.9	10.6	30.9	9.0
Other techies	17.9	15.0	2.0	5.0	11.1	21.1	33.6	16.1
Engineers	13.6	13.7	2.0	4.7	10.0	18.7	30.1	14.0
Technicians	13.4	10.4	1.1	2.8	6.6	13.6	25.0	10.8

Table 2 shows that techies are much more prevalent in manufacturing firms (71% of the observations) than in non-manufacturing firms (20% of the observations). Table 2 also shows that the wage bill shares of different types of techies vary across sectors. While Other techies have the highest wage bill shares on average in both manufacturing and non-manufacturing sectors, R&D techies have higher wage bill shares in manufacturing and ICT techies have a higher wage bill share in non-manufacturing. This pattern is even more pronounced for firms with the highest wage bill shares.

In addition to Table 2, Table D1 presents descriptive statistics for the main variables used in estimation separately for manufacturing and non-manufacturing firms.

6.1 Production function estimates

The GLZ production function estimates and implied elasticities are reported in Table 3. We report industry-by-industry estimates of the production function parameters and the demand elasticity. All of our estimates of the elasticity of substitution across inputs σ and of the demand elasticity η are greater than one and in all industries we can reject the null hypotheses that $\sigma = 1$ and $\eta = 1$ at conventional levels of statistical significance. Rejecting $\sigma = 1$ is important for identification in the GLZ estimator. This is because the expression for materials input quantities (as a function of expenditures on materials, the wage bill and labor input in quantities) is not defined for the knife-edge case of $\sigma = 1$ (i.e., a Cobb-Douglas production function; see Grieco et al. (2016)). Additionally, finite profits require $\eta > 1$.

Overall, our estimates of the production function and demand elasticities are very plausible. For example, we find particularly large elasticities in Wholesale and Retail, which is consistent with low profit margins in these industries. In contrast, elasticites of demand are estimated to be much lower in industries that exhibit greater product differentiation. Estimates of the distribution parameters α_N, α_M and α_K reflect the relative importance of each input in production in ways that are in line with what one might expect.[21]

We relegate the estimates of the revenue production function using the GNR methodology to Appendix F. Despite using quite different methodologies, the estimates from the two methodologies are broadly in line with each other. For example, the relative importance of materials, labor and capital are quite similar (the levels are not comparable because we do not identify ρ in GNR).

6.2 Controlled Markov

6.2.1 Baseline results

We capture the effect of lagged techies on productivity along two margins. The first is the "extensive techie margin", measured by an indicator for whether the firm employs techies, either overall or separately for each category of techies, $I(T_{ft-1} > 0)$. The second is the "intensive techie margin", measured by the techie wage bill share T_{ft-1}, either overall or by category of techies. We always control for the extensive margin when examining the intensive margin, which identifies the impact of techie-intensity over and above the extensive

---

[21] The GLZ estimator ensures that the distribution parameters are equal to output elasticities at the geometric mean of the data.

Table 3: GLZ Production function estimates

Industries	α_N	α_M	α_K	σ	η	# Obs.	#Firms
Food, beverage, tobacco	0.226	0.602	0.173	2.677	5.267	27378	4481
	(0.003)	(0.007)	(0.009)	(0.222)	(0.256)		
Textiles, wearing apparel	0.347	0.566	0.086	1.728	2.788	7798	1160
	(0.007)	(0.012)	(0.019)	(0.305)	(0.085)		
Wood, paper products	0.280	0.414	0.306	1.388	4.280	15636	2312
	(0.006)	(0.009)	(0.015)	(0.072)	(0.271)		
Chemical products	0.155	0.563	0.283	1.595	4.460	6713	858
	(0.003)	(0.012)	(0.015)	(0.083)	(0.297)		
Pharmaceutical products	0.175	0.447	0.378	1.504	3.398	1530	202
	(0.015)	(0.038)	(0.053)	(0.193)	(0.638)		
Rubber and plastic	0.228	0.536	0.236	1.673	3.901	14071	1895
	(0.004)	(0.009)	(0.013)	(0.102)	(0.185)		
Basic metal and fabricated metal	0.297	0.395	0.308	1.469	3.558	25523	3519
	(0.004)	(0.005)	(0.009)	(0.051)	(0.105)		
Electrical equipment	0.194	0.549	0.258	1.638	3.968	4390	587
	(0.007)	(0.020)	(0.027)	(0.170)	(0.386)		
Machinery and equipment	0.197	0.559	0.244	1.577	3.407	9788	1290
	(0.005)	(0.015)	(0.021)	(0.169)	(0.193)		
Transport equipment	0.179	0.542	0.279	1.817	5.441	5786	790
	(0.006)	(0.018)	(0.024)	(0.219)	(0.624)		
Other manufacturing	0.334	0.420	0.245	1.596	2.879	21557	3260
	(0.006)	(0.008)	(0.014)	(0.089)	(0.085)		
Construction	0.392	0.391	0.217	1.444	2.688	109935	21055
	(0.004)	(0.004)	(0.008)	(0.033)	(0.042)		
Wholesale	0.119	0.737	0.144	1.298	8.940	172332	25760
	(0.000)	(0.002)	(0.003)	(0.020)	(0.197)		
Retail	0.131	0.797	0.072	1.818	5.967	240473	37978
	(0.000)	(0.002)	(0.002)	(0.079)	(0.067)		
Accommodation and food services	0.399	0.267	0.334	1.883	5.486	109782	21476
	(0.006)	(0.004)	(0.010)	(0.057)	(0.302)		
Publishing and broadcasting	0.382	0.060	0.558	1.233	2.255	14906	2565
	(0.019)	(0.003)	(0.022)	(0.023)	(0.123)		
Administrative and support activities	0.465	0.069	0.466	1.712	3.344	29470	5471
	(0.015)	(0.002)	(0.017)	(0.045)	(0.190)		

Notes. The CES production function is $Q_{ft} = e^{\omega_{ft}}(\alpha_N N_{ft}^{\gamma} + \alpha_K K_{ft}^{\gamma} + \alpha_M M_{ft}^{\gamma})^{1/\gamma}$, where Q_{ft} is the quantity of output produced using labor N_{ft}, intermediate inputs M_{ft} and capital K_{ft}. As discussed by GLZ, it is important for identification to normalize each data series by its geometric mean, which we do. The elasticity of substitution across inputs σ is determined by γ, where $\gamma = (\sigma - 1)/\sigma$, and η is the elasticity of demand.

margin, while allowing for separate effects of each margin.[22] Our estimating equations have

---

[22] Using the inverse hyperbolic sine transformation of T_{ft-1} or terciles on the positive support of T_{ft-1} yield virtually identical results. These results are available upon request.

the general form

$$\widehat{\omega}_{ft}^{GLZ} = \theta_{it} + \lambda \widehat{\omega}_{ft-1}^{GLZ} + \beta_0 I(T_{ft-1} > 0) + \beta_1 T_{ft-1} + \xi_{ft}, \tag{7}$$

θ_{it} is an industry-year fixed effect, which implies that the parameters are identified using variation across firms within an industry-year cell. We estimate (7) by OLS, with productivity computed from the industry-by-industry GLZ estimates of equation (3) reported in Table 3 above. As discussed above, we report bootstrapped standard errors that are clustered by firm.

We report our baseline controlled Markov estimates of (7) in Table 4. We report the effects of techies on firm-level productivity in the samples of manufacturing industries (columns 1 to 6) and non-manufacturing industries (columns 7 to 12).

Table 4 columns (1) and (7) show that firms that employ techies have higher productivity one year later than firms without techies. The effect is sizable at 4.1 log points in manufacturing industries and 5.6 log points in non-manufacturing industries. While the estimated effects of employing techies are similar in the two sectors, as shown in Table 2 the incidence of techies is 3.5 times higher in manufacturing, so the overall effect of techies on within-industry productivity dispersion is estimated to be higher in manufacturing.

Table 4 columns (2) and (8) include the techie wage bill share in addition to the techie indicator. We find statistically significant effects of techies on productivity along the intensive margin. The coefficients on the techie indicator remain statistically significant, but are more than halved in both sectors. This shows that the presence of even a small number of techies raises future productivity and that the effect increases with greater techie employment.

Columns (3), (4), (9), and (10) in Table 4 report estimates when techie workers are broken down by their detailed job descriptions. We find that both the presence and the intensity of R&D techies have a large impact on productivity in manufacturing. These findings corroborate the results of Doraszelski and Jaumandreu (2013), indicating that R&D

Table 4: Impact of techies on productivity – GLZ estimates

	Manufacturing						Non-Manufacturing					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I\left(T_{ft-1}>0\right)$	0.041***	0.018***					0.056***	0.023***				
	(0.003)	(0.003)					(0.003)	(0.003)				
T_{ft-1}		0.120***						0.206***				
		(0.008)						(0.012)				
$I\left(T_{ft-1}^{RD}>0\right)$			0.016***	0.011***					0.011*	-0.001		
			(0.002)	(0.003)					(0.006)	(0.007)		
$I\left(T_{ft-1}^{ICT}>0\right)$			0.023***	0.016***					0.026***	0.015***		
			(0.002)	(0.003)					(0.003)	(0.004)		
$I\left(T_{ft-1}^{OTH}>0\right)$			0.030***	0.013***					0.051***	0.018***		
			(0.003)	(0.003)					(0.003)	(0.003)		
T_{ft-1}^{RD}				0.065***						0.155*		
				(0.024)						(0.087)		
T_{ft-1}^{ICT}				0.109**						0.131***		
				(0.045)						(0.023)		
T_{ft-1}^{OTH}				0.108***						0.234***		
				(0.010)						(0.015)		
$I\left(T_{ft-1}^{38}>0\right)$					0.032***	0.013***					0.047***	0.013***
					(0.002)	(0.003)					(0.003)	(0.003)
$I\left(T_{ft-1}^{47}>0\right)$					0.016***	0.007**					0.033***	0.020***
					(0.002)	(0.003)					(0.003)	(0.003)
T_{ft-1}^{38}						0.149***						0.253***
						(0.014)						(0.018)
T_{ft-1}^{47}						0.084***						0.125***
						(0.013)						(0.016)
$I\left(x_{ft-1}>0\right)$	0.009***	0.007***	0.001	0.003	0.004*	0.005*	0.008***	0.007**	0.006**	0.006**	0.005*	0.005*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
$\hat{\omega}_{ft-1}$	0.910***	0.912***	0.907***	0.910***	0.909***	0.912***	0.872***	0.873***	0.872***	0.874***	0.872***	0.873***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Obs.			116,088						484,541			
No. firms			19,571						100,087			

Notes. The table reports estimates of equation (7) in the text. The dependent variable is $\hat{\omega}_{ft}$, log estimated productivity. $I(.)$ is the indicator function. T is the techie wage bill share, superscripts $\{RD, ICT, OTH, 38, 47\}$ denote R&D, ICT, other techies, engineers and technicians respectively, x is the value of firm exports. Industry-year fixed effects included in all columns. Bootstrap standard errors clustered by firm in parentheses. *** denotes p-value ≤ 0.01, ** p-value ≤ 0.05, * p-value ≤ 0.10

expenditures, most of which are accounted for by techie wage bills, play an important role in explaining the differences in productivity across manufacturing firms.

However, techies' positive impact on productivity is not limited to R&D techie workers. In columns (3) and (9), we also find positive impacts of the presence of ICT and other techie workers on the productivity of both manufacturing and non-manufacturing firms.

We find that R&D techies have larger effects in manufacturing, while Other techies are more important in non-manufacturing and ICT techies have similar effects across sectors, being only slightly larger effects in non-manufacturing. We offer two, related interpretations. First, the nature of productivity improvement differs across sectors. Manufacturing firms improve efficiency through process innovation, product design, and automation—activities that are concentrated in R&D occupations. Non-manufacturing firms (e.g., in wholesale, retail, business services) improve efficiency through inventory management, logistics optimization, customer relationship systems, and data analytics—activities that are captured by Other techies, and also rely heavily on ICT. Second, the scope for technology adoption differs. Manufacturing firms often operate near the technological frontier in their product space, which R&D pushes outward. Non-manufacturing firms are more likely to adopt existing technologies to reorganize production. Other and ICT techies facilitate this adoption and integration.

Columns (5), (6), (11), and (12) in Table 4 display the estimates when we distinguish between engineers (PCS 38) and technicians (PCS 47). Both engineers and technicians positively affect productivity, although the engineers effect is larger both at the extensive and intensive margins. This makes sense, as engineers have more technical education and training than technicians.

Turning to the effect of exporting, we find a positive impact on productivity in columnbs (1), (2), (7) and (8), with similar effects in manufacturing and in non-manufacturing. We note that only 11.5% of non-manufacturing firms in our sample are exporters (primarily in wholesale, publishing, and broadcasting).[23] This suggests that exporting is not a significant factor accounting for the variability of productivity in non-manufacturing. These finding is in line with De Loecker (2013), who argues that investments in technology partly drive the impact of exports on productivity. We estimate smaller impacts of exporting on productivity

---

[23] In our estimation sample 49% of wholesale firms export, and 22.6% of publishing and broadcasting firms export.

when we employ more flexible specifications for techies, distinguishing them by their tasks or occupation types, such as engineers versus technicians.

How important are techies in explaining productivity? To answer this question we combine the estimates just discussed in Table 4 with the variation in techies reported in Table 2. These calculations are reported in Table 5. The columns labeled "0-$p50$" report the productivity effect of techies on a firm with the median amount of techies compared to a firm with no techies. The columns labeled "*IQR*" report the productivity effect for firms with a lot of techies (75th percentile) compared to those with not many techies (25th percentile).

Table 5: Impact of techies on productivity – Magnitude of the baseline estimates (percent)

	Manufacturing		Non-Manufacturing	
	$0-p50$	*IQR*	$0-p50$	*IQR*
A. Impact effects				
Techies	4.15	2.29	4.89	3.76
R&D techies	1.44	0.46	0.29	0.86
ICT techies	1.82	0.27	2.03	1.19
Other techies	3.02	1.60	4.50	3.84
Engineers	3.12	1.76	3.90	3.61
Technicians	1.50	0.94	2.87	1.36
B. 10-year effects				
Techies	31.78	16.63	32.05	23.97
R&D techies	10.29	3.21	1.70	5.11
ICT techies	12.87	1.84	12.43	7.11
Other techies	22.36	11.37	29.46	24.76
Engineers	23.08	12.50	25.00	22.92
Technicians	10.73	6.59	17.97	8.22

Notes. Units are percent points. IQR stands for interquartile range, the difference between the 25th and 75th percentiles. We use the statistics on the median and IQR from the descriptive statistics in Table 2 and the estimated parameters from columns (2), (4), (6), (8) (10) and (12) in Table 4 to compute the impact and 10-year effects of the baseline specification. For instance, when comparing a firm with no techies to a firm with the median intensity of techies, the estimated impact effect of techies is equal to $\widehat{\beta}_{T_{ft-1}} + \widehat{\beta}_{I(T_{ft-1}>0)} \times p50$. The 10-year effects are computed by multiplying the impact effects by $(1-\widehat{\lambda}^{10})/(1-\widehat{\lambda})$, where $\widehat{\lambda}$ is the estimated coefficient on lagged productivity, $\hat{\omega}_{ft-1}$, reported in Table 4.

The one-year effects reported in Panel A of Table 4 are economically important: compared to a firm with no techies, a firm with the median wage bill share of techies will have more than 4 percent higher productivity one year later. Compared to a firm with a low techie share, manufacturing firms with a high techie share will have 2.3 percent higher productivity a year later, while non-manufacturing firms will have 3.8 percent higher productivity. R&D techies are more important in manufacturing since they are concentrated in that sector, while ICT and Other techies are of comparable importance in both sectors. Engineers are about twice as important as Technicians in both sectors.

The results just discussed are one-year effects. Because our 8-year panel is relatively short, we have no way to directly estimate longer run effects. However the AR(1) structure of equation (5) and the associated estimates of the AR(1) parameter λ can be used to calculate implied 10-year cumulated effects. We report these in Panel B of Table 5, and they are large: after 10 years, firms with the median level of techies will have nearly one-third higher productivity than firms with no techies.[24]

6.2.2 Generalized controlled Markov

The results reported in Table 5 are calculated from estimates of equation (7), which is a simple linear AR(1) version of the general controlled Markov process given by equation (4). We next consider a more general specification of (4) which allows the effect of techies to differ across the distribution of lagged productivity,

$$\begin{aligned} \widehat{\omega}_{ft}^{GLZ} &= \theta_{it} + \lambda \widehat{\omega}_{ft-1}^{GLZ} + \beta_0 I\left(T_{ft-1} > 0\right) + \beta_1 T_{ft-1} \\ &+ \beta_{0\omega}\left[I\left(T_{ft-1} > 0\right) \times \widehat{\omega}_{ft-1}^{GLZ}\right] + \beta_{1\omega}\left[T_{ft-1} \times \widehat{\omega}_{ft-1}^{GLZ}\right] + \xi_{ft}, \end{aligned} \tag{8}$$

---

[24] We validate these projections using a simulation exercise. Starting from initial conditions in 2011, we simulate the productivity path for each firm using our estimated AR(1) process with the baseline "$\mathbf{1}(T > 0)$, T" specification, then compare predicted productivity changes to actual changes observed for firms that remain in the sample through 2019. The correlation between predicted changes and realized productivity changes is 0.40, and regressing actual changes on predicted changes yields a coefficient of 0.47 with an R^2 of 0.16 (Results are similar when we include industry fixed effects: the coefficient is 0.48 with a within-R^2 of 0.16.) While this confirms that our model does not fully capture long-run dynamics, it suggests that the estimated productivity process has meaningful predictive content beyond the one-year horizon.

Compared to a firm with no techies, the productivity effect of moving to employing techies at the p^{th} percentile for a firm with productivity at the q^{th} percentile is then $\beta_0 + \beta_1 T_p + \beta_{0\omega} \left[\widehat{\omega}_q^{GLZ}\right] + \beta_{1\omega} \left[T_p \times \widehat{\omega}_q^{GLZ}\right]$. We report estimates of this quantity for $p, q \in \{25, 50, 75\}$ in Table 6.[25]

We find that the effect of techies on future productivity is positive along the entire distribution. The marginal effect of techies increases with the percentile of the techie wage bill (moving down the rows in Table 6). For example, at the median level of lagged productivity, the effect on productivity is 0.9 percent (year-on-year) greater at the median techie intensity compared to the 25th percentile; the effect is 1.3 percent greater at the 75th percentile compared to the median techie intensity. In contrast, the techie effect increases linearly across different percentiles of lagged productivity (moving across the columns in Table 6), except for very high techie intensities.

Table 6: Impact of techies on productivity – General specification

	Percentile of lagged ω		
	25	50	75
Percentile of lagged Techies			
Manufacturing			
25	1.70	3.00	4.29
50	3.16	3.89	4.60
75	5.32	5.19	5.06
Non-manufacturing			
25	0.83	2.53	4.49
50	3.14	4.13	5.26
75	7.12	6.86	6.57

Notes. Units are percent points.

6.2.3 Sectoral heterogeneity

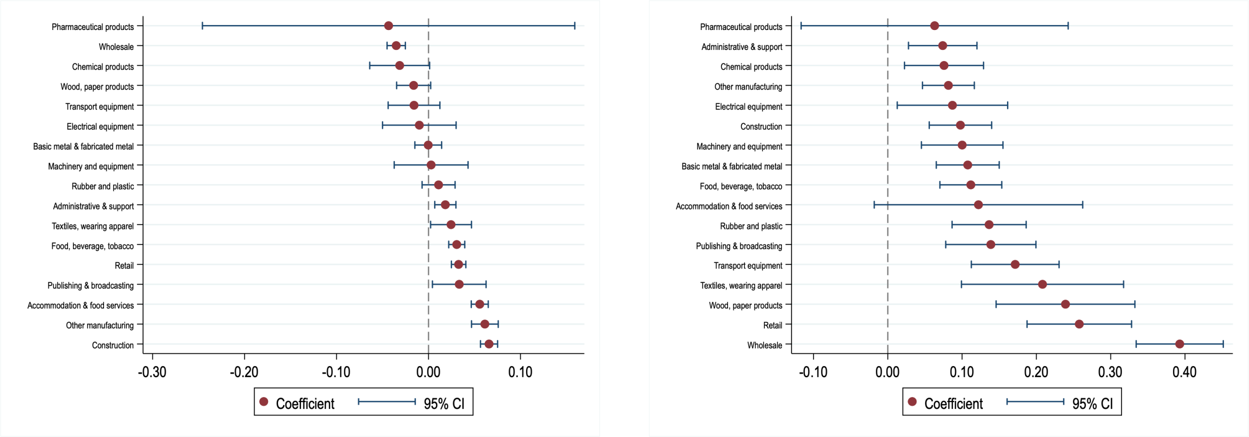
We estimate the controlled Markov specification separately for each two-digit A38 industry. Figure 1 reports the estimates for the extensive margin, $\mathbf{1}(T_{ft-1} > 0)$, and the intensive

---

[25] Table E1 in the appendix presents the estimates of equation (8), with a short analysis of their implications.

margin, T_{ft-1}.

Figure 1: Industry-level estimates by margin



(a) Extensive Margin ($\mathbf{1}(T_{ft-1} > 0)$) (b) Intensive Margin (T_{ft-1})

Notes: For each two-digit A38 industry, the points plot the coefficient from a controlled Markov equation estimated separately. Horizontal bars represent 95% confidence intervals.

Three facts stand out. First, the extensive-margin coefficient displays substantial heterogeneity across industries: point estimates vary in sign and precision. Precision is lowest in sectors where techie adoption is close to one or close to zero, because within-industry variation in $\mathbf{1}(T_{ft-1} > 0)$ is mechanically limited. Second, the intensive-margin coefficient is positive in all industries. Estimates are more precise in sectors with intermediate adoption rates and less precise at the extremes. Third, in industries where the extensive-margin estimate is negative (e.g., Wholesale), the intensive-margin estimate remains positive and is typically larger and more precisely estimated.

These patterns mirror heterogeneity in techie adoption across industries (see Table D2 in the Appendix). This motivates our baseline specifications that pool industries into broad manufacturing and non-manufacturing groups, which increases statistical power while preserving the main sectoral differences.

We interpret the negative extensive margin coefficients as follows. These coefficients capture the effect of going from zero to an infinitesimal stock of techies at a one-year horizon (which is very rare in the data). Such a marginal adopter has probably not made the complementary organizational changes that are needed to translate techies into measured productivity.[26] The intensive margin, the effect of adding more techies once a firm already has some, is positive in all sectors. In particular, in the only sector for which the extensive margin is negative and statistically significant, Wholesale, the positive intensive margin is the largest.

In order to improve our understanding of the cross-industry heterogeneity, we estimate a specification of the controlled Markov in which we drop the continuous techie measure $T_{f,t-1}$ and keep the extensive-margin indicator $I(T_{f,t-1} > 0)$ (the rest remains as above). In this specification, reported in Appendix Figure D1, the extensive-margin coefficient is positive and significant at the 5% level in 12 of the 17 retained sectors, and no estimate is statistically significantly negative. In particular, Wholesale, which has a negative extensive margin effect in Figure 1, exhibits an overall positive and statistically significant effect for $I(T_{f,t-1} > 0)$ in Figure D1.

6.3 Sensitivity analysis

In this section we report sensitivity analysis in several dimensions. As discussed in section 5.1, the logic of our methodology is that current techies do not affect current productivity. Nonetheless, it is informative to look at estimates that violate this logic. In Appendix Table G1 we re-estimate the controlled Markov equation, including current techies, $I(T_{ft} > 0)$, either alone or jointly with lagged techies, $I(T_{ft-1} > 0)$. Current techies are insignificant when included alone. When both are included, current techies receive a negative coefficient while lagged techies remain positive and precisely estimated. Note that adding current techies is a violation of the assumptions that are required for identification. With this caveat in mind, we cautiously interpret these results as being consistent with techies affecting output through future productivity, rather than contemporaneously.[27]

Our baseline controlled Markov estimates of equation (5) reported in Table 4 include

---

[26] This is reminiscent of the productivity J-curve mechanism in Brynjolfsson et al. (2021): complementary investments take time to build, so the early returns to a new general-purpose technology can be flat or negative even when the long-run returns are large.

[27] These results hold in a subsample of "techie-switchers", where we drop firms that always have techies and firms that never have any. This is shown in Table G2.

lagged techies entering linearly. In Appendix Table H1, we report estimates where we add $\omega_{f,t-1}^2$ and $\omega_{f,t-1}^3$. The results using this more elaborate specification of the Markov process are not materially different from the baseline results. In Appendix Table I1 we report the results that consider the quality of labor inputs and show that our baseline results are qualitatively unchanged.

Below we report sensitivity analysis in two additional dimensions. We explore how our results change when we modify the way techies enter the analysis. We then consider management in the controlled Markov. Finally, we report results using the GNR estimator.

6.3.1 Alternative assumption: techies in the production function

Central to our methodology is that we assume that techies affect output only through their effect on future productivity and not through any contemporaneous contribution to factor services that affect current output. This assumption is analogous to the standard assumption that investment in $t-1$ does not affect output in $t-1$, but raises output in t through its contribution to capital in time t. One way to check if this assumption makes sense is to compare it to an alternative where techies are no different from other workers. To do so, we estimate the production functions and associated Hicks neutral productivity series with techies included in the definition of labor. If techies only contribute to production, then they should not affect productivity when we estimate the controlled Markov specification for productivity with techies.

Table 7 reports the results of this exercise. The full results are reported in Appendix Table J1. The null hypothesis that the effects are zero is easily rejected. We thus conclude that the data reject the model that techies affect output only through a contemporaneous effect on output. Of course, under our baseline model, the results in Table 7 are inconsistent, so they should not be compared to our baseline results in Table 4. This is because the GLZ production function estimator requires labor to be a static input, and the specification reported in Table 7 is not consistent with this.

Table 7: Allocating techies to production – GLZ estimates

	Manufacturing		Non-Manufacturing	
	(1)	(2)	(3)	(4)
$I(T_{ft-1} > 0)$	0.023***	0.007**	0.028***	0.010***
	(0.003)	(0.004)	(0.003)	(0.003)
T_{ft-1}		0.089***		0.117***
		(0.010)		(0.011)
$I(x_{ft-1} > 0)$	0.009***	0.008***	0.024***	0.023***
	(0.003)	(0.003)	(0.003)	(0.003)
$\hat{\omega}_{ft-1}$	0.917***	0.915***	0.879***	0.878***
	(0.003)	(0.003)	(0.002)	(0.002)
Other controls	Yes		Yes	
Obs.	115,115		486,240	
No. firms	19,456		100,111	

Notes. The table reports estimates of equation (7). The dependent variable is $\hat{\omega}_{ft}$, log estimated productivity. $I(.)$ is the indicator function. T is the techie wage bill share and x firm exports. Industry-year fixed effects included. Bootstrap standard errors clustered by firm in parentheses. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

6.3.2 Alternative assumption: Other techies in production, not in the controlled Markov equation

Considering the heterogeneity of the occupations that we group into Other techies, it is possible that not all of them satisfy our assumption that techies contribute to output only through their effect on future productivity. To address this, here we make the opposite assumption and allocate Other techies to general labor. We then estimate the effects of R&D and ICT techies on productivity estimated with this alternative treatment of Other techies.

Table 8 reports results of this specification. Comparing Table 8 to our baseline results in Table 4, the most important comparison is the estimated effects of R&D and ICT techies reported in columns (3), (4), (9) and (10) in the two tables. The estimated effects at both the intensive and extensive margins are substantially larger in Table 8, which is to be expected, since the incidence of Other techies is correlated with R&D and ICT techies. This means that when we take Other techies out of the controlled Markov, more of the explanatory power of techies is shifted onto R&D and ICT techies. This interpretation should be taken with a grain of salt, since according to our methodology, if Other techies belong in the controlled Markov, as the baseline results indicate, then they should be excluded from production.

Our conclusion from this exercise is that our baseline conclusions about the importance of R&D and ICT techies for productivity are not sensitive to the treatment of Other techies.

Table 8: Allocating Other techies to production – GLZ estimates

	Manufacturing						Non-Manufacturing					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(T_{ft-1} > 0)$	0.038***	0.021***					0.057***	0.039***				
	(0.002)	(0.003)					(0.004)	(0.004)				
T_{ft-1}		0.220***						0.192***				
		(0.026)						(0.021)				
$I(T^{RD}_{ft-1} > 0)$			0.027***	0.013***					0.033***	0.022***		
			(0.003)	(0.003)					(0.007)	(0.008)		
$I(T^{ICT}_{ft-1} > 0)$			0.027***	0.020***					0.055***	0.038***		
			(0.003)	(0.003)					(0.004)	(0.004)		
T^{RD}_{ft-1}				0.223***						0.249**		
				(0.031)						(0.099)		
T^{ICT}_{ft-1}				0.182***						0.188***		
				(0.054)						(0.022)		
$I(T^{38}_{ft-1} > 0)$					0.029***	0.013***					0.048***	0.030***
					(0.003)	(0.003)					(0.005)	(0.005)
$I(T^{47}_{ft-1} > 0)$					0.022***	0.017***					0.036***	0.026***
					(0.003)	(0.003)					(0.004)	(0.005)
T^{38}_{ft-1}						0.261***						0.219***
						(0.039)						(0.032)
T^{47}_{ft-1}						0.125***						0.124***
						(0.043)						(0.037)
$I(x_{ft-1} > 0)$	0.001	0.002	-0.000	0.001	-0.001	0.001	0.020***	0.021***	0.020***	0.021***	0.020***	0.021***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
$\hat{\omega}_{ft-1}$	0.913***	0.914***	0.913***	0.913***	0.913***	0.913***	0.876***	0.876***	0.876***	0.876***	0.876***	0.876***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Obs.	115,563						485,984					
No. firms	19,489						100,108					

Notes. The dependent variable is $\hat{\omega}_{ft}$. $I(.)$ denotes the indicator function. T is the techie wage bill share; superscripts $\{RD, ICT, 38, 47\}$ denote R&D, ICT, engineers, and technicians, respectively. x denotes firm exports. Industry-year fixed effects included. Bootstrap standard errors clustered by firm in parentheses. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

6.3.3 Alternative assumption: managers in the controlled Markov equation?

As discussed in Section 5, a core element of our methodology is that techies are the only workers in the firm who affect output with a lag, through their effect on future productivity, rather than contemporaneously. In other words, no workers other than techies belong in the second-stage controlled Markov given by equation (4). This treatment of techies is motivated by a careful study of the tasks that techies do (Section 2.1 above) as well as their qualifications (Section 3.2) and their associations with innovative and productivity-enhancing activities (Sections 3.4 and 3.5). In contrast, we treat managers as part of general labor, whose contributions to output are contemporaneous. In Table 9 we examine whether including lagged managerial workers (PCS code 37) in the second stage controlled Markov changes our inference on the role of techies, along the lines of (Bloom et al., 2017). In this table we exclude managers from production, as we treat them in the same way as techies. Columns (1) and (3) reproduce our baseline specification, while columns (2) and (4) add lagged managerial labor to the controlled Markov equation. Note that in columns (1) and (3) the controlled Markov is misspecified, since managers are missing.

The results in Table 9 indicate that including lagged managers does not materially affect our inference on the role of techies in increasing productivity: both the indicator for positive techie employment and the techie wage bill share remain positive and precisely estimated, although the magnitude is somewhat smaller. Lagged managerial labor enters positively and significantly when included in columns (2) and (4), but hardly affects the coefficients to techies. This would imply that management has a separate effect on future productivity that is almost orthogonal to that of techies.

We see in Table 9 that excluding managers from production results in large, statistically significant negative estimates of the impact of exporting in non-manufacturing (columns 3 and 4). We take this as a sign of mis-specification of the allocation of managers in this configuration. These results could arise if exporting is associated with a large increase in

Table 9: Adding Managers to the Controlled Markov – GLZ estimates

	Manufacturing		Non-Manufacturing	
	Baseline (1)	Managers (2)	Baseline (3)	Managers (4)
$I(T_{ft-1} > 0)$	0.028***	0.023***	0.033***	0.024***
	(0.003)	(0.003)	(0.003)	(0.003)
T_{ft-1}	0.079***	0.084***	0.093***	0.100***
	(0.008)	(0.009)	(0.012)	(0.011)
$I(M_{ft-1} > 0)$		0.007***		0.017***
		(0.003)		(0.002)
M_{ft-1}		0.073***		0.151***
		(0.013)		(0.008)
$I(x_{ft-1} > 0)$	0.011***	0.006***	-0.019***	-0.032***
	(0.002)	(0.002)	(0.002)	(0.002)
$\hat{\omega}_{ft-1}$	0.917***	0.919***	0.860***	0.869***
	(0.003)	(0.003)	(0.002)	(0.002)
Obs.	112,577		486,073	
No. firms	19,070		100,477	

Notes. The dependent variable is $\hat{\omega}_{ft}$. $I(.)$ denotes the indicator function. T is the techie wage bill share, M is the managers (PCS37) wage bill share, and x is firm exports. Industry-year fixed effects included in all columns. Bootstrap standard errors clustered by firm. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

management input.

Alternatively, we can examine the impact of adding management to the controlled Markov specifications in Table 4 without removing them from production. This is similar to our specification test in Table 7, applied to management. To be clear, this implies keeping management in production and adding it to the controlled Markov, in a way that is not consistent with the necessary assumptions needed for identification. We report these results in Table 10.

The results in Table 10 indicate that management has a strong, negative intensive margin effect on future productivity. This is not consistent with a positive role for management in both current output and future productivity. As in Table 9, adding management hardly changes the estimates of the effect of techies on future productivity. We conclude that our estimates on the role of techies are not materially influenced by how we treat management, and that management does not likely contribute to future productivity in our sample.

Table 10: Adding Managers to the Controlled Markov while keeping Managers in production – GLZ estimates

	Manufacturing		Non-Manufacturing	
	Baseline (1)	Managers (2)	Baseline (3)	Managers (4)
$I(T_{ft-1} > 0)$	0.018***	0.018***	0.023***	0.016***
	(0.003)	(0.003)	(0.003)	(0.003)
T_{ft-1}	0.120***	0.117***	0.206***	0.202***
	(0.008)	(0.009)	(0.013)	(0.013)
$I(M_{ft-1} > 0)$		0.001		0.029***
		(0.003)		(0.002)
M_{ft-1}		-0.055***		-0.023***
		(0.011)		(0.006)
$I(x_{ft-1} > 0)$	0.007***	0.009***	0.007***	0.004*
	(0.002)	(0.002)	(0.002)	(0.002)
$\hat{\omega}_{ft-1}$	0.912***	0.913***	0.873***	0.871***
	(0.003)	(0.003)	(0.002)	(0.002)
Obs.	116,088		484,541	
No. firms	19,571		100,087	

Notes. The dependent variable is $\hat{\omega}_{ft}$. $I(.)$ denotes the indicator function. T is the techie wage bill share, M is the managers (PCS37) wage bill share, and x is firm exports. Industry-year fixed effects included in all columns. Bootstrap standard errors clustered by firm. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

6.3.4 Alternative estimator: results using the GNR estimator.

All the results discussed so far have been computed using the GLZ estimator. Here we consider how our results change using the GNR estimator, for two reasons. The first is simply a general robustness check. The second is that the GNR estimator allows us to relax the assumption that labor is a static input, which is an important consideration given that there are large firing costs in the French labor market. Table 11 reports the results when labor is assumed to be "static" (like materials, and as we assumed when implementing the GLZ estimator), and Table 12 reports the results for when labor is assumed to be "predetermined" (with slow adjustment, like capital).

Recall that the estimates here are not directly comparable to our GLZ estimates because GNR does not separately identify the coefficients β from the demand parameter ρ in equation (6). In other words, the numbers we report in Tables 11 are estimates of $\beta\overline{\rho}$, not β. In both tables the estimated effects of the control variables are generally lower than those reported in Table 4, which is consistent with $\rho < 1$ and with the demand elasticities that we estimate using the GLZ estimator (see Table 3).

Table 11: Impact of techies on productivity – GNR estimates assuming labor to be static

	Manufacturing						Non-Manufacturing					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(T_{ft-1} > 0)$	0.040***	0.031***					0.024***	0.014***				
	(0.002)	(0.003)					(0.001)	(0.001)				
T_{ft-1}		0.042***						0.052***				
		(0.004)						(0.003)				
$I(T^{RD}_{ft-1} > 0)$			0.014***	0.013***					0.007***	0.007***		
			(0.001)	(0.002)					(0.002)	(0.002)		
$I(T^{ICT}_{ft-1} > 0)$			0.015***	0.013***					0.011***	0.007***		
			(0.002)	(0.002)					(0.001)	(0.001)		
$I(T^{OTH}_{ft-1} > 0)$			0.033***	0.028***					0.022***	0.013***		
			(0.002)	(0.003)					(0.001)	(0.001)		
T^{RD}_{ft-1}				0.016						-0.022		
				(0.013)						(0.023)		
T^{ICT}_{ft-1}				0.020						0.037***		
				(0.019)						(0.008)		
T^{OTH}_{ft-1}				0.029***						0.058***		
				(0.006)						(0.003)		
$I(T^{38}_{ft-1} > 0)$					0.030***	0.027***					0.021***	0.014***
					(0.002)	(0.003)					(0.001)	(0.001)
$I(T^{47}_{ft-1} > 0)$					0.023***	0.020***					0.013***	0.009***
					(0.002)	(0.002)					(0.001)	(0.001)
T^{38}_{ft-1}						0.027***						0.049***
						(0.007)						(0.005)
T^{47}_{ft-1}						0.019***						0.041***
						(0.006)						(0.005)
$I(x_{ft-1} > 0)$	0.016***	0.016***	0.014***	0.014***	0.014***	0.014***	0.007***	0.006***	0.006***	0.006***	0.005***	0.005***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\hat{\omega}_{ft-1}$	0.912***	0.913***	0.910***	0.911***	0.908***	0.909***	0.937***	0.937***	0.937***	0.937***	0.937***	0.937***
	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Obs.	118,796						549,881					
No. firms	19,789						107,649					

Notes. The table reports estimates of equation (6). The dependent variable is $\widehat{\rho\omega}_{ft}$. $I(.)$ denotes the indicator function. T is the techie wage bill share. Superscripts $\{RD, ICT, OTH, 38, 47\}$ denote R&D, ICT, other techies, engineers, and technicians. x is firm exports. Industry-year fixed effects included. Bootstrap standard errors clustered by firm. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

Despite differences in methodologies, including assumptions on the response of labor to innovations to productivity and on returns to scale, the results in Tables 11 and 12 are consistent with those using the GLZ estimator that are reported in Table 4. In particular, we find that techies cause higher productivity both via the extensive and the intensive margins, both in manufacturing and non-manufacturing industries—more so in the former than in the latter. We also identify effects of techies on productivity that extend beyond their

Table 12: Impact of techies on productivity – GNR estimates assuming labor to be predetermined

	Manufacturing						Non-Manufacturing					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(T_{ft-1} > 0)$	0.031***	0.020***					0.010***	0.008***				
	(0.003)	(0.003)					(0.001)	(0.001)				
T_{ft-1}		0.053***						0.014***				
		(0.007)						(0.003)				
$I(T^{RD}_{ft-1} > 0)$			0.003*	-0.001					0.004**	0.005***		
			(0.002)	(0.002)					(0.001)	(0.002)		
$I(T^{ICT}_{ft-1} > 0)$			0.010***	0.004**					0.000	-0.001		
			(0.002)	(0.002)					(0.001)	(0.001)		
$I(T^{OTH}_{ft-1} > 0)$			0.028***	0.018***					0.010***	0.008***		
			(0.003)	(0.003)					(0.001)	(0.001)		
T^{RD}_{ft-1}				0.036**						-0.025		
				(0.015)						(0.017)		
T^{ICT}_{ft-1}				0.084***						0.019**		
				(0.025)						(0.009)		
T^{OTH}_{ft-1}				0.055***						0.014***		
				(0.008)						(0.003)		
$I(T^{38}_{ft-1} > 0)$					0.020***	0.014***					0.007***	0.006***
					(0.002)	(0.002)					(0.001)	(0.001)
$I(T^{47}_{ft-1} > 0)$					0.018***	0.012***					0.007***	0.006***
					(0.002)	(0.002)					(0.001)	(0.001)
T^{38}_{ft-1}						0.045***						0.009**
						(0.009)						(0.004)
T^{47}_{ft-1}						0.045***						0.013***
						(0.010)						(0.005)
$I(x_{ft-1} > 0)$	0.029***	0.028***	0.028***	0.028***	0.027***	0.028***	0.006***	0.006***	0.006***	0.006***	0.005***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\hat{\omega}_{ft-1}$	0.670***	0.668***	0.671***	0.669***	0.671***	0.669***	0.840***	0.840***	0.841***	0.840***	0.840***	0.840***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.020)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Obs.	118,782						540,153					
No. firms	19,787						105,495					

Notes. The table reports estimates of equation (6). The dependent variable is $\hat{\rho\omega}_{ft}$. $I(.)$ denotes the indicator function. T is the techie wage bill share. Superscripts $\{RD, ICT, OTH, 38, 47\}$ denote R&D, ICT, other techies, engineers, and technicians. x is firm exports. Industry-year fixed effects included. Bootstrap standard errors clustered by firm. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

involvement in R&D. The impact of R&D on productivity in manufacturing is stronger and more tightly identified than in non-manufacturing. Overall, the impact of ICT and Other techies is greater than that of R&D. Finally, we find that engineers have a greater impact than technicians on the extensive and intensive productivity margins in both manufacturing and non-manufacturing industries.

Some differences with Table 4 are apparent. For example, in Table 11 we do not identify a statistically significant impact of ICT in the intensive margin in manufacturing. And in Table 12, we find that the extensive margin of ICT techies in non-manufacturing industries is nil, although the intensive margin is very large. However, these differences do not undermine the main conclusions from the baseline analysis. Broadly, the two sets of GNR estimates are consistent with those in the main analysis, for example, in the relative magnitudes of the effects of R&D, ICT and Other techies.

7 Conclusion and implications

Our paper demonstrates the key role of techies in raising firm-level productivity. This conclusion holds for both manufacturing and non-manufacturing firms in the French economy from 2011 to 2019. An important contribution of our paper is to separately estimate the role of techies who work in R&D from those who work in ICT and other technical occupations. R&D techies are more common and more important to productivity in manufacturing, while ICT techies are more important in non-manufacturing, which is the bulk of the private sector in all advanced economies.

Economists often conceive R&D as improving the technological frontier, and our results are consistent with this interpretation. However, it is likely that attaining the frontier is at least as important to productivity as expanding it, and this is where ICT and other techies are likely to be crucial. Our results on ICT and other techies challenge the view that focusing solely on R&D can fully capture the impact of firms' choices and investments in improving their productivity.

This carries implications for policymakers concerned with promoting economic growth. Capital accumulation and R&D are rightly central to achieving this goal. Our findings about the key role of ICT and other techies suggest that educational, training and other policies that enhance the supply of non-R&D techies will also have positive effects on growth.

References

Abowd, John M, Francis Kramarz, and David N Margolis, "High wage workers and high wage firms," *Econometrica*, 1999, *67* (2), 251–333.

Ackerberg, Daniel A, Garth Frazer, Kyoo il Kim, Yao Luo, and Yingjun Su, "Under-identification of structural models based on timing and information set assumptions," *Journal of Econometrics*, 2023, *237* (1), 105463.

_ , Kevin Caves, and Garth Frazer, "Identification properties of recent production function estimators," *Econometrica*, 2015, *83* (6), 2411–2451.

Arora, Ashish, Sharon Belenzon, and Lia Sheer, "Back to basics: why do firms invest in research?," 2017. NBER Working Paper No. 23187.

Aw, Bee Yan, Mark J Roberts, and Daniel Yi Xu, "R&D investment, exporting, and productivity dynamics," *American Economic Review*, 2011, *101* (4), 1312–1344.

Barrows, Geoffrey, Hélène Ollivier, and Ariell Reshef, "Production Function Estimation with Multi-Destination Firms," 2025. CESifo Working Paper No. 10716.

Barth, Erling, James C. Davis, Richard B. Freeman, and Andrew J. Wang, "The Effects of Scientists and Engineers on Productivity and Earnings at the Establishment Where They Work," Working Paper 23484, National Bureau of Economic Research June 2017.

Beaudry, Paul, David A Green, and Benjamin M Sand, "The great reversal in the demand for skill and cognitive tasks," *Journal of Labor Economics*, 2016, *34* (S1), S199–S247.

Ben Zeev, Nadav, Joel Mokyr, and Karine Van Der Beek, "Flexible supply of apprenticeship in the british industrial revolution," *The Journal of Economic History*, 2017, *77* (1), 208–250.

Blank, David M and George J Stigler, *The Demand and Supply of Scientific Personnel*, NBER, 1957.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen, "Management as a Technology?," 2017. NBER Working Paper No. 22327.

Bonleu, Antoine, Gilbert Cette, and Guillaume Horny, "Capital utilization and retirement," *Applied Economics*, 2013, *45* (24), 3483–3494.

Brynjolfsson, Erik and Lorin Hitt, "Paradox lost? Firm-level evidence on the returns to information systems spending," *Management science*, 1996, *42* (4), 541–558.

_ , Daniel Rock, and Chad Syverson, "The Productivity J-Curve: How Intangibles Complement General Purpose Technologies," *American Economic Journal: Macroeconomics*, 2021, *13* (1), 333–372.

_ , Wang Jin, and Sebastian Steffen, "Do IT Capabilities Still Drive Productivity and Innovation in the Digital Age?," 2024. Available at SSRN 4765508.

_ , _ , and Xiupeng Wang, "Information Technology, Firm Size, and Industrial Concentration," 2023. NBER Working Paper No. 31065.

Cette, Gilbert, Nicolas Dromel, Rémy Lecat, and Anne-Charlotte Paret, "Production factor returns: the role of factor utilization," *Review of Economics and Statistics*, 2015, *97* (1), 134–143.

Crepon, Bruno, Emmanuel Duguet, and Jacques Mairesse, "Research, Innovation and productivity: an econometric analysis at the firm level," *Economics of Innovation and new Technology*, 1998, *7* (1), 115–158.

Deming, David J and Kadeem L Noray, "STEM Careers and Technological Change," Working Paper 25065, National Bureau of Economic Research September 2018.

Doraszelski, Ulrich and Jordi Jaumandreu, "R&D and productivity: Estimating endogenous productivity," *The Review of Economic Studies*, 2013, *80* (4), 1338–1383.

Fox, Jeremy T and Valérie Smeets, "Does input quality drive measured differences in firm productivity?," *International Economic Review*, 2011, *52* (4), 961–989.

Gandhi, Amit, Salvador Navarro, and David A Rivers, "On the identification of gross output production functions," *Journal of Political Economy*, 2020, *128* (8), 2973–3016.

Grieco, Paul LE, Shengyu Li, and Hongsong Zhang, "Production function estimation with unobserved input price dispersion," *International Economic Review*, 2016, *57* (2), 665–690.

Hall, Bronwyn H, Francesca Lotti, and Jacques Mairesse, "Evidence on the impact of R&D and ICT investments on innovation and productivity in Italian firms," *Economics of Innovation and New Technology*, 2013, *22* (3), 300–328.

_ , Jacques Mairesse, and Pierre Mohnen, "Measuring the Returns to R&D," in "Handbook of the Economics of Innovation," Vol. 2, Elsevier, 2010, pp. 1033–1082.

Hanlon, W Walker, "The rise of the engineer: Inventing the professional inventor during the Industrial Revolution," 2022. NBER Working Paper No. 29751.

Harrigan, James, Ariell Reshef, and Farid Toubal, "The March of the Techies: Job Polarization Within and Between Firms," *Research Policy*, 2021.

Hsieh, Chang-Tai and Esteban Rossi-Hansberg, "The industrial revolution in services," *Journal of Political Economy Macroeconomics*, 2023, *1* (1), 3–42.

INSEE, "Nomenclature des professions et catégories socioprofessionnelles PCS," 2003.

_ , "Nomenclature des professions et catégories socioprofessionnelles (PCS 2020): Guide analytique," 2024. Collecte 2026.

Kelly, Morgan, Joel Mokyr, and Cormac Ó Gráda, "Precocious Albion: a new interpretation of the British industrial revolution," *Annu. Rev. Econ.*, 2014, *6* (1), 363–389.

_ , _ , and Cormac Ó Gráda, "The Mechanics of the Industrial Revolution," *Journal of Political Economy*, 2023, *131* (1), 59–94.

Klette, Tor Jakob and Zvi Griliches, "The inconsistency of common scale estimators when output prices are unobserved and endogenous," *Journal of applied econometrics*, 1996, *11* (4), 343–361.

Levinsohn, James and Amil Petrin, "Estimating production functions using inputs to control for unobservables," *The Review of Economic Studies*, 2003, *70* (2), 317–341.

Lichtenberg, Frank R, "The output contributions of computer equipment and personnel: A firm-level analysis," *Economics of innovation and new technology*, 1995, *3* (3-4), 201–218.

Loecker, Jan De, "Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity," *Econometrica*, 2011, *79* (5), 1407–1451.

_ , "Detecting learning by exporting," *American Economic Journal: Microeconomics*, 2013, *5* (3).

_ **and Chad Syverson**, "An industrial organization perspective on productivity," in "Handbook of industrial organization," Vol. 4, Elsevier, 2021, pp. 141–223.

_ **and Pinelopi Koujianou Goldberg**, "Firm Performance in a Global Market," *Annual Review of Economics*, 2014, *6*, 201–27.

Maloney, William F and Felipe Valencia Caicedo, "Engineering growth: innovative capacity and development in the Americas," 2017. CESifo Working Paper No. 6339.

Melitz, Marc, "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 2003, *71*, 1695–1725.

_ **and James A Levinsohn**, "Productivity in a differentiated products market equilibrium," 2006. Mimeo, Princeton University.

Murphy, Kevin M, Andrei Shleifer, and Robert W Vishny, "The Allocation of Talent: Implications for Growth," *The Quarterly Journal of Economics*, 1991, *106*, 503–530.

Olley, G Steven and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 1996, pp. 1263–1297.

Romer, Paul M, "Endogenous technological change," *Journal of political Economy*, 1990, *98* (5, Part 2), S71–S102.

Saunders, Adam and Erik Brynjolfsson, "Valuing Information Technology Related Intangible Assets.," *MIS Quarterly*, 2016, *40* (1).

Schweitzer, Camille, "L'enquête R&D: mesurer l'effort de R&D des entreprises, au-delà du crédit d'impôt recherche," 2019. INSEE Références, édition 2019 – Dossier.

Syverson, Chad, "What determines productivity?," *Journal of Economic literature*, 2011, *49* (2), 326–365.

Tambe, Prasanna and Lorin M Hitt, "The productivity of information technology investments: New evidence from IT labor data," *Information Systems Research*, 2012, *23* (3-part-1), 599–617.

_ **and** _ , "Job hopping, information technology spillovers, and productivity growth," *Management Science*, 2014, *60* (2), 338–355.

Varian, Hal R, *Microeconomic Analysis*, Norton New York, 1992.

Appendices

A Data definitions and construction

Here we discuss in detail the three administrative and survey datasets used in our paper, as well as details on supplementary publicly available data.

A key feature of the French statistical system is that establishments are identified by a unique number, the SIRET, used by all data sources. The first 9 digits of an establishment's SIRET comprise the SIREN of the firm to which the establishment belongs. This makes it easy to aggregate from establishments to firms.

A.1 Industries.

The French industry classification system (Nomenclature d'Activités Française, NAF) is documented at https://www.insee.fr/fr/information/2406147. Our data use the NAF rev.2 classification at the A38 aggregation level. Table A1 lists all industries represented in our sample with their corresponding A38 codes.

Table A1: Industries in Sample

NAF Code	Industry	Firm-years	Firms
Manufacturing			
10–12	Food, beverage, tobacco	29,277	4,721
13–15	Textiles, wearing apparel, leather	8,936	1,312
16–18	Wood, paper, printing	17,384	2,543
20	Chemical products	7,380	941
21	Pharmaceutical products	1,703	222
22–23	Rubber, plastic, non-metallic minerals	16,100	2,143
24–25	Basic metals, fabricated metal products	30,407	4,148
27	Electrical equipment	5,094	675
28	Machinery and equipment	11,526	1,502
29–30	Transport equipment	6,465	873
31–33	Other manufacturing, repair & installation	24,178	3,601
Non-manufacturing			
41–43	Construction	119,766	22,417
45–46 (*w/o* 45.32)	Wholesale trade	188,565	27,882
45.32 & 47	Retail trade	258,474	40,393
55–56	Accommodation and food services	116,511	22,411
58–60	Publishing, audiovisual, broadcasting	15,771	2,680
77–82	Administrative and support activities	31,177	5,707

Notes: Industries classified according to NAF rev. 2 (Nomenclature d'Activités Française, révision 2). Full documentation available at https://www.insee.fr/fr/information/2406147. Sample period is 2011–2019.

A.2 Workers: DADS Poste.

Our source for information on workers is the DADS Poste, which is based on mandatory annual reports filed by all firms with employees, so our data includes all private-sector French workers except the self-employed.[28] The DADS Poste is an INSEE database compiled from the mandatory firm-level DADS reports. For each worker, the DADS Poste reports gross and net wages, hours paid, occupation, tenure, gender and age. There is no information about workers' education or overall labor market experience. The data do not include worker identifiers, so we can not track workers over time, but this is of no concern to us given our focus on firm-level rather than individual outcomes.[29] Our unit of analysis is a firm-year observation.

When building our sample from the raw DADS files we eliminate observations that are duplicated ("doublons"). Doublons arise due to the way that the DADS employment data is organized. Some workers report a residence commune and a workplace commune that are each located in different départements. The DADS records these workers twice: once in the département-level file of residence, and a second time in the département-level file of the workplace. Naïvely aggregating employment across files causes double counting because such individuals appear in two different département-level files—although their workplace is located only in one. We avoid such double counting.

The DADS reports detailed 4-digit occupational codes, almost 500 in total, beginning in 2009, which determines the first year of our sample. We use the French occupational classification PCS-ESE and the exhaustive definition of tasks for each occupation provided by the INSEE (2003) to identify techie workers precisely. The INSEE (2003) classification is valid throughout our sample period. More importantly, the 2020 revision of the PCS explicitly preserved the two-digit categories we use (CS 38 and CS 47) to ensure temporal comparability. As stated in the PCS-2020 guide: "the groups and socio-professional categories remain unchanged in their scope and content to allow for long-term analyses, since 1982" (INSEE, 2024, p. 9). The conceptual definitions are identical. The 2020 revision only reorganized the detailed 4-position codes while preserving the aggregate structure we employ. The French PCS system explicitly distinguishes techies from other workers based on the nature of their tasks. For engineers and technical managers (PCS 38) versus administrative and commercial managers (PCS 37), the official documentation states: *"l'aspect scientifique ou technique l'emporte sur l'aspect administratif ou commercial"* (the scientific or technical aspect takes precedence over the administrative or commercial aspect). For technicians (PCS 47), the documentation emphasizes that they have *"un rôle d'étude, d'assistance, de conseil ou d'expertise"* (a role of study, assistance, advice, or expertise), which is explicitly distinguished from production and fabrication tasks. The 2003 PCS revision specifically created new categories to capture *"fonctions transversales aux différentes activités industrielles (méthodes, contrôle-qualité, logistique)"* (cross-cutting functions across industrial activities: methods, quality control, logistics), functions that support production rather than constitute it.

We distinguish between three types of techie workers: ICT, R&D, and other techies. Table A2 reports our classification. For completeness, Table A3 presents the remaining two-digit PCS occupations.

---

[28] All employers and their employees are covered by the DADS declaration with the exception of self-employed and government bodies, domestic services (section 97-98 of NAF rev. 2) and employees in businesses outside French territory (section 99 of NAF rev. 2). However, local authorities and public-employed hospital staff are included since 1992. Public institutions of industrial and commercial nature are also included.

[29] A related dataset, made famous by Abowd et al. (1999), is the DADS Panel. This sample from of the DADS data does include worker identifiers.

Table A2: Classification of Techie Occupations

PCS-ESE	Description	Category
Panel A: Engineers and Technical Managers (PCS 38)		
380a	Technical directors of large companies	Other
381a	Engineers and management staff in agriculture, fishing, water, and forestry studies and operations	Other
382a	Engineers and management staff in building and public works studies	Other
382b	Architects	Other
382c	Engineers, site managers, and construction supervisors (managers) in building and public works	Other
382d	Technical sales engineers and managers in building and public works	Other
383a	Engineers and R&D managers in electricity and electronics	R&D
383b	Manufacturing engineers and managers in electrical and electronic equipment	Other
383c	Technical sales engineers and managers in professional electrical or electronic equipment	Other
384a	Engineers and R&D managers in mechanics and metalworking	R&D
384b	Manufacturing engineers and managers in mechanics and metalworking	Other
384c	Technical sales engineers and managers in professional mechanical equipment	Other
385a	Engineers and R&D managers in the transformation industries	R&D
385b	Manufacturing engineers and managers in transformation industries	Other
385c	Technical sales engineers and managers in intermediate goods transformation industries	Other
386a	Engineers and R&D managers in other industries	R&D
386d	Production and distribution engineers and managers in energy and water	Other
386e	Manufacturing engineers and managers in other industries	Other
387a	Industrial purchasing and procurement engineers and managers	Other
387b	Logistics, planning, and scheduling engineers and managers	Other
387c	Production method engineers and managers	Other
387d	Quality control engineers and managers	Other
387e	Maintenance and new works technical engineers and managers	Other
387f	Technical engineers and managers in the environment	Other
388a	Engineers and R&D managers in computer science	ICT
388b	Engineers and managers in administration, maintenance, support, and user services in computer science	ICT
388c	IT project managers and IT managers	ICT
388d	Technical sales engineers and managers in IT and telecommunications	Other
388e	Engineers and specialist managers in telecommunications	ICT
389a	Technical engineers and managers in transport operations	Other
389b	Technical and commercial navigating officers and managers of civil aviation	Other
389c	Technical navigating officers and managers of merchant navy	Other
Panel B: Technicians (PCS 47)		
471a	Technical experts and consultants in agriculture, water, and forestry studies	Other
471b	Technical experts in operation and production control in agriculture, water, and forestry	Other
472a	Building and civil engineering draftsmen	Other
472b	Surveyors and topographers	Other
472c	Quantity surveyors and various building and civil engineering technicians	Other
472d	State and local government public works technicians	Other
473a	Electrical, electromechanical, and electronic draftsmen	Other
473b	R&D technicians and manufacturing methods technicians in electricity, electromechanics, and electronics	R&D
473c	Electrical, electromechanical, and electronic production and quality control technicians	Other
474a	Mechanical and metal construction draftsmen	Other
474b	R&D technicians and manufacturing methods technicians in mechanical construction and metalworking	R&D
474c	Mechanical and metal construction production and quality control technicians	Other
475a	R&D technicians and production methods technicians in the transformation industries	R&D
475b	Production and quality control technicians in the transformation industries	Other
476a	Technical assistants, printing and publishing technicians	Other
476b	Soft materials, furniture, and wood industry technicians	Other

Continued on next page

Table A2 – *Continued from previous page*

PCS-ESE	Description	Category
477a	Logistics, planning, and scheduling technicians	Other
477b	Installation and maintenance technicians for industrial equipment	Other
477c	Installation and maintenance technicians for non-industrial equipment	Other
477d	Environmental and pollution treatment technicians	Other
478a	Computer design and development technicians	ICT
478b	Computer production and operation technicians	ICT
478c	Computer installation, maintenance, support, and user services technicians	ICT
478d	Telecommunications technicians and network IT technicians	ICT
479a	Public research or teaching laboratory technicians	Other
479b	Independent expert technicians of various levels	Other

Source: INSEE (2003): https://www.insee.fr/fr/information/2400059. Own classification. The PCS (*Professions et Catégories Socioprofessionnelles*) system of occupational codes is used to classify all workers in France. Engineers and technical managers (PCS 38) are distinguished from administrative and commercial managers (PCS 37) as follows: for the former, "the scientific or technical aspect takes precedence over the administrative or commercial aspect." Technicians (PCS 47) are characterized by "a role of study, assistance, advice, or expertise," distinct from production workers.

The "Other techies" group is diverse. Their tasks are mostly related to adopting and spreading new technologies and production methods within their firms. Unlike workers directly contributing to current output, such as sales personnel, Other techies also aim to boost productivity. Their main role is to support production processes rather than directly engage in fabrication tasks. However, the tasks performed by technicians and engineers in this category are often less clearly defined than those of R&D and ICT techies. This is why we present results reallocating Other techies to ordinary workers contributing to current output. The results on ICT and R&D are qualitatively similar.

Table A3: Other PCS two-digit categories

PCS-ESE	Description
10	Farmers, foresters, fishermen and fish farmers
21	Craft workers
22	Shopkeepers and related occupations
23	Business owners with more than 10 employees
31	Liberal professions
33	Administrative and technical executives in the civil service
34	Professors and higher scientific professions
35	Information, arts and entertainment professions
37	Administrative and commercial executives in firms
42	Primary and vocational school teachers
43	Intermediate professions in health and social work
45	Intermediate professions in the civil service (administration, security)
46	Administrative and commercial intermediate professions in firms
48	Supervisors (excluding administrative supervisors)
52	Civil service administrative employees, service staff and health auxiliaries
53	Police officers, military personnel, firefighters, private security guards
54	Administrative employees in firms
55	Sales employees
56	Personal service workers
62	Skilled industrial workers
63	Skilled craft workers
64	Drivers of transport vehicles, delivery drivers, couriers
65	Machine operators, forklift operators, warehouse workers and transport workers
67	Low-skilled industrial workers
68	Low-skilled craft workers

Continued on next page

Table A3 – *Continued from previous page*

PCS (2-digit)	Description
69	Agricultural, forestry, fishing and aquaculture workers

Source: INSEE (2003): https://www.insee.fr/fr/information/2400059.

A.3 Balance sheet data: FARE

Firm-level balance sheet information is reported in an INSEE dataset called FARE. The balance sheet variables used in our empirical analysis include revenue, expenditure on materials, and the book value of capital. We do not use balance sheet data on employment or the wage bill, because the DADS Poste data is more detailed, but the FARE wage bill and employment data are extremely highly correlated with the corresponding DADS Poste data.

We begin constructing capital stocks starting from the book value of tangible fixed assets recorded in FARE. Let i denote the firm, s the 2-digit sector, and t the year. These stocks are measured at historical cost, denoted $KBV_{i,s,t}$, and must be converted into real capital stocks $K_{i,s,t}$ valued at current prices. Following Bonleu et al. (2013) and Cette et al. (2015), we deflate book values using an investment price index that corresponds to the inferred vintage of the capital stock.

The procedure is as follows. Let $DKBV_{i,s,t}$ denote accumulated depreciation, and let $\Delta DKBV_{i,s,t}$ denote its year-on-year change. We first compute each firm's implied average age of capital,

$$a_{i,s,t} = \frac{KBV_{i,s,t}}{\Delta DKBV_{i,s,t}}.$$

To reduce the noise induced by lumpy investment, we replace $a_{i,s,t}$ with the sector-specific median,

$$\tilde{A}_s = \text{median}_{i \in s}\{a_{i,s,t}\}.$$

We then compute the corrected age of capital as

$$T_{i,s,t} = \frac{DKBV_{i,s,t}}{KBV_{i,s,t}} \times \tilde{A}_s,$$

which maps the share of depreciated capital into an age estimate. Negative or implausibly low values are set equal to one.

The corresponding vintage year of capital is

$$v_{i,s,t} = t - T_{i,s,t},$$

truncated at the earliest year for which sectoral investment deflators are available.

Finally, real capital is obtained by deflating book values using the investment price index associated with the inferred vintage:

$$K_{i,s,t} = \frac{KBV_{i,s,t}}{P^I_{s,v_{i,s,t}}}.$$

This method aligns historical-cost book values with the prices of investment goods prevailing when capital was installed, thereby yielding comparable firm-level real capital stocks across sectors and over time.

A.4 Trade data: Douanes

Data on bilateral exports of firms located in France are provided by French Customs. For each observation, we know exporting status of the firm. We use the firm-level SIREN identifier to match the trade data to other sources. This match is not perfect: we fail to match about 11 percent of imports and exports to firms. The imperfect match is because there are SIRENs in the trade data for which there is no corresponding SIREN in our other data sources. This is likely to lead to a particular type of measurement error: for some firms, we will observe zero trade even when true trade is positive. This is not a big concern because most of the missing values are in the oil refining industry, which we drop from our sample.

A.5 Survey data

The data is taken from four French surveys related to R&D, ICT, patent and innovation activities at the level of the firm and individual information on techies' vocational training.

- The Annual Survey on the Means dedicated to Research and Development (R&D survey: Enquête R&D Entreprises) provides information on the means devoted to R&D by firms in terms of in-house and external expenditure and the number of researchers and research support personnel. The survey is exhaustive for firms that have conducted in-house R&D expenditures for a level greater than or equal to 400k€and firms that have newly declared in-house R&D expenditures during the year of the survey. These "new" firms in terms of R&D are taken from administrative sources (the Research Tax Credit (RTC) database, the Young Innovative Companies (YIC) database, companies created via public incubators, i-Lab competition winners) or from the Innovation Capacity and Strategy (ICS) survey. The survey is completed with a sample of firms whose in-house R&D expenditure is strictly smaller than 400k€. We focus on the period from 2010 to 2019 to match the period of analysis in the DADS data. The survey provided pooled cross-sectional data on about 10,000 firm-level observations each year. For our purposes, we are mostly interested in how much of the firm's R&D budget is spent on internal R&D wages. Moreover, the survey asks firms if they filed patents and had any process or product innovations in the past year. We are also interested to see if internal R&D spending and employment of techies is related to patents or innovation.

- The Information and Communication Technology survey (ITC survey: Enquête sur les technologies de l'information et de la communication et le commerce électronique – TIC entreprises) provides information on the computerization and the diffusion of information and communication technologies in firms. The survey is exhaustive for firms with more than 500 employees or having the highest turnover – about 2,800 firms in the sample. It is complemented by the ICT information of smaller firms. We collected data on a pooled cross-sectional sample of about 10,000 firm-level observations per year from 2012 to 2018. For our purpose, the survey provides useful information on the relationship between ICT training and the diffusion of technology within a firm.

- The Training and Professional Qualification survey (TPQ survey: Enquête formation et qualification professionnelle) provides information on professional mobility, initial training, continuing education, social origin, and work income. Every ten years, the INSEE collects detailed information on 45,000 individuals aged 21 to 64 and residing in France. We use the 2015 edition of the survey. It gives a precise account of the specialty of the highest degree obtained by the individual and whether and which training after the highest degree he/she received. The survey provides a detailed classification of specialties and training that allows us to classify the individual's skills as STEM. It also provides characteristics such as the individual's occupation. Table A4 provides information on the list of diplomas and training that we group to identify individuals with education and training in science, technology, engineering, and math (STEM). .

Table A4: Mapping diplomas' specialties into STEM skills

French National Code	Title
Diploma	
110	Multi-science specialties
111	Physical chemistry
112	Chemistry, Biology, Biochemistry
113	Natural Sciences (Biology, Geology)
114	Mathematics, statistics
115	Physics
116	Chemistry
117	Earth Sciences
118	Life Sciences
200	Basic industrial technologies
201	Automation, robotics, industrial process control
230	Civil engineering, construction, wood
240	Multi-technology specialties in flexible materials
250	Multi-technology specialties mechanics-electricity
253	Aeronautics and space mechanics
255	Electricity, electronics
326	Computer science, information processing, networks
Training	
420	Life Sciences
440	Physical Sciences
460	Mathematics and Statistics
481	Computer Science
482	Computer use
500	Engineering, processing and production

Source: TPQ, 2015. French classifications of diploma and vocational training.

Each firm in the survey has the same identifier as in the administrative dataset. We show below that the information provided in the survey correlates well with the information in the DADS dataset.

B Facts on Techies

B.1 Fact 1. Techies have more STEM education and training than other occupations

We argue that techie workers are engineers and technicians with skills and experience in STEM. We use the TPQ survey to analyze whether techies have more STEM education and more STEM training than other occupations. We find 26,861 individuals with valid observations, among which 5.4% are Engineers (PCS 38) and 5.1% are Technicians (PCS 47). These shares are similar to the shares in the DADS administrative data.

Table B1 reports the results. We show that around 60 percent of techies have a degree and/or training in STEM, with about a fifth having a STEM degree and further STEM training. STEM degrees are more common among engineers (PCS 38, 55%) than technicians (PCS 47, 41%). By contrast, STEM education is quite uncommon in all other PCS codes, with only 11% having a STEM degree and less than a fifth having a degree or training. These results show that techies have more STEM education and more STEM training than other occupations.

Table B1: STEM education share by occupation

	Degree	Training	Degree or Training	Degree and Training
Techies				
Engineers	0.55	0.27	0.64	0.19
Technicians	0.41	0.35	0.59	0.18
Other occupations				
Average	0.11	0.09	0.18	0.02
Upper managers	0.12	0.09	0.19	0.02
Middle managers	0.09	0.08	0.16	0.01
Other office workers	0.04	0.07	0.11	0.01
Skilled industrial workers	0.19	0.22	0.36	0.05

Source: TPQ, 2015 .

Table B1 gives some additional details on STEM degrees and training for large non-techie occupations. Less than a fifth of upper managers have any STEM education, a share that is even lower among middle managers and clerical workers. By contrast, over a third of skilled industrial workers have some STEM education. However, the degrees earned by these workers are primarily general and technical high school degrees rather than university degrees. More than two-thirds of skilled industrial workers have either a professional baccalaureate (14%), a vocational school certificate (in French, CAP, 29%), or a certificate of vocational proficiency (in French, BEP, 15%).

B.2 Fact 2. Techies across industries

Table B2 reports the techie wage bill shares by category in France and the French manufacturing and non-manufacturing sectors. Our analysis indicates that most of the expenditure on techie workers takes place in manufacturing, accounting for roughly two-thirds of the total techie wage bill. We also observe interesting patterns when we break down techie workers into different categories (ICT, R&D, and other tech workers). The share of R&D workers in the manufacturing wage bill is considerably higher at 87.3% compared to the share of ICT workers, which is only 38.0%. The wage bill share of other techies workers is similar to the aggregate pattern.

Techies represent 18% of the French private sector's wage bill share, with a larger share in manufacturing than in non-manufacturing. Overall and across sectors, other techie workers are a larger share of the techie wage bill than the shares of R&D and ICT workers. The share of R&D techies is much more prominent in manufacturing, while the share of ICT techies is almost identical across sectors. Table B2 also reports the wage bill shares of engineers and technicians. Engineers are twice as large a share of the techie wage bill than technicians.

Regarding the presence of both R&D and ICT techie workers in manufacturing and non-manufacturing firms, we observe that 47% of manufacturing firms that employ R&D techies also have ICT techies. In contrast, the corresponding figure for non-manufacturing firms is 44%.

When considering the co-existence of R&D and other techie workers in manufacturing and non-manufacturing firms, we find that many manufacturing firms with R&D techies also

Table B2: Wage bill shares of techies by categories (2019)

	Overall	Manufacturing	Non-Manufacturing	% techie wage bill in manufacturing
Techies	18.3	31.5	10.8	62.6
R&D	3.4	8.2	0.7	87.3
ICT	2.2	2.3	2.1	38.0
Other	12.7	21.1	8.0	60.2
Engineers (PCS 38)	11.9	19.7	7.4	60.3
Technicians (PCS 47)	6.5	11.9	3.4	66.9

employ other techies.

Specifically, 96% of such manufacturing firms have other techies on their payroll. In non-manufacturing firms, this proportion is slightly lower, with 84% of firms with R&D techies also employing other techies.

B.3 Fact 3. Most R&D spending is on wages

The R&D survey provides detailed information on firms with positive internal R&D expenditures, which are the amounts spent on R&D that originate within the firm's control. The survey distinguishes between internal and external R&D expenditures, which are spent outside the control of the firm. We show in Table B3, that expenditure on R&D is overwhelmingly spent within the firm, with the median firm spending nothing on external R&D. We conclude that conditional on reporting positive internal R&D, most R&D expenditures originate within the control of the firm.

Table B3: External R&D and wage bill shares

	Mean	Median	P_{90}	P_{10}
External share of total	0.09	0.00	0.32	0.00
Wage bill share:				
– Total R&D	0.67	0.67	1.0	0.35
– Internal R&D	0.74	0.72	1.0	0.48

Source: R&D survey .

The R&D survey is interesting for our purpose because it gives the labor costs of those workers who effectively do R&D. It is important because we cannot assume that all labor costs in the firm's R&D department are for R&D activities. We use the R&D survey to analyze how much of the firm's R&D budget is spent on in-house R&D wages. We show in Table B3 that R&D spending is mainly spending on wages, especially when R&D is done within the firm.

In Table B4, we show that the external share of R&D spending is weakly correlated with overall R&D spending and strongly negatively correlated with the wage bill share of total R&D. We conclude that firms indirectly hire some R&D workers through external R&D spending, but not many: most R&D workers are employed by the firm paying for the R&D, and their wages make up the bulk of firm R&D spending.

Table B4: Correlations

	External Share of total R&D	Wage bill share of total R&D	Total R&D Expenditures
External share of total R&D	1		
Wage bill share of total R&D	-0.60	1	
Total R&D expenditures	0.08	-0.08	1

Source: R&D survey.

Our main data analysis uses information on various types of techies from the DADS data to explain firm-level productivity. In Table B5, we show that the wage bills of techies in the administrative data are highly correlated with different measures of R&D workers in the survey data. We show that the strength of the correlation is about the same whether we measure R&D workers in the survey by wage bill, headcount or FTEs. Reassuringly, the correlations are highest for R&D techies.

Table B5: Correlations between techie measures in the R&D survey and wage bills in DADS

		R&D survey		
		Wage bill	Headcount	FTEs
DADS	All techies	0.72	0.83	0.79
	R&D techies	0.82	0.88	0.84
	ICT techies	0.60	0.56	0.55
	Other techies	0.49	0.65	0.61

Source: R&D survey matched with DADS data.

B.4 Fact 4. Techies are positively associated with the diffusion of ICT within firms

We use the ICT survey to understand better the relationship between techies and the diffusion of technology within firms. For our purpose, we exploit three questions in the questionnaires received by the firms.

1. In 2018, was training in developing or improving skills in ICT offered by the firm to...
    - ... specialists in ICT?
    - ... other employees?
2. Does the firm employ specialists in ICT?

Table B6 shows that only 20 percent of firms surveyed offer ICT training. However, firms that employ ICT workers are six times more likely (0.66/0.11) to offer ICT training. About 11 percent of firms offer ICT training even though they do not employ ICT workers. This fact suggests a role for ICT training from outside the firm.

Table B7 shows further detail on the exposure of different types of workers on ICT training. We distinguish between ICT workers, non-ICT workers, and both categories. The table shows that firms that employ ICT workers are four times as likely to train non-ICT

Table B6: ICT workers and ICT training

		Offer ICT training?	
		No	Yes
Employ ICT workers?	No	0.89	0.11
	Yes	0.34	0.66
	Mean	0.80	0.20

Source: ICT survey.

workers in ICT. To see this, note that the first row reports that only 11 percent of firms that don't employ ICT workers train non-ICT workers in ICT. In contrast, the second row shows that among firms that do employ ICT workers, about half train non-ICT workers in the use of ICT.[30]

Table B7: Exposure to ICT training

		Which workers get ICT training?			
		None	Only ICT	Only non-ICT	ICT & non-ICT
Employ ICT workers?	No	0.89	0.00	0.11	0.00
	Yes	0.34	0.18	0.12	0.35
	Mean	0.80	0.03	0.11	0.06

Source: ICT survey.

We match the ICT survey to the DADS sample. We find very small discrepancies between the information in the DADS and ICT datasets. In particular, 10 percent of firms have ICT techies from the DADS, and 12 percent have ICT workers from the survey, a small difference. We check how having ICT workers in the survey is related to having ICT techies (both and others) in the DADS. Both panels A and B of Table B8 show that the answer is that the two are closely related. The left panel shows that the conditional probability of having ICT workers in the survey given that a firm has ICT techies in the DADS is 0.62, which is 9 times the conditional probability of having ICT workers in the survey given no ICT techies in the DADS (0.07). The right panel of Table B8 shows that the conditional probability of having ICT workers in the DADS given that a firm has ICT techies in the survey is 0.49, which is 12 times the conditional probability of having ICT workers in the DADS given no ICT techies in the survey (0.04).

We next ask if ICT techies are associated with training of workers in ICT. To answer this question, Table B9 repeats the analysis of Table B6 on the matched ICT survey and DADS sample. However, we now examine crosstabs of training with ICT techies from the DADS rather than ICT workers from the survey. Not surprisingly, the inferences are similar: firms that have ICT techies are $\frac{0.49}{0.14} = 3.5$ times likely to offer ICT training.

Next, we ask what firm characteristics are associated with ICT training, using linear probability regressions for the training dummy from the survey. All regressions include industry × year fixed effects, and the log wage bill excluding techies, named "Ex-techies", as a control for firm size.

---

[30] $0.12 + 0.35 = 0.47$ which is about half.

Table B8: ICT workers in the ICT survey and DADS dataset

		Panel A				Panel B	
		ICT workers in survey?				ICT techies in DADS?	
		No	Yes			No	Yes
ICT techies in DADS?	No	0.93	0.07	ICT workers in survey?	No	0.96	0.04
	Yes	0.38	0.62		Yes	0.51	0.49
	Mean	0.88	0.12		Mean	0.90	0.10

Source: ICT survey.

Table B9: ICT workers and ICT training

		Offer ICT training?	
		No	Yes
Employ ICT techies? (DADS information)	No	0.86	0.14
	Yes	0.51	0.49
	Mean	0.82	0.18

Source: Matched dataset.

Table B10 shows that there is a strong association between the likelihood of having techies and offering ICT training, even after controlling for firm size. To interpret the effect sizes, keep in mind that ICT training is uncommon, with only 18 percent of firms offering training (Table B9). Columns (1)-(3) use indicator variables to measure techie presence, and the results are clear: firms with techies are substantially more likely to offer training. Column (1) shows that firms with any techies are 6 percent more likely to offer ICT training. This effect is driven by ICT techies, as shown in columns (2) and (3): the coefficient on the dummy for ICT techies is 0.20, while R&D (0.06) and other techies (0.04) have a smaller albeit positive effect. Columns (4)-(6) are restricted to firms that have positive techies, and we see that the intensive margin effect is large: firms with 10 percent more expenditure on techies have a 5 percentage point higher likelihood of offering ICT training, an effect that is driven by ICT techies.

To summarize what we have found in this sub-section, measures of ICT employment in the survey are closely associated with the presence of ICT and other techies in the DADS. In addition, firms with ICT techies are much more likely to offer ICT training to their ICT and non-ICT workers.

B.5 Fact 5. Techies are positively associated with patenting and innovations

We describe the relationship between R&D spending, techies and patents, and innovation outcomes. The R&D survey provides information on whether the firm has introduced technologically new or improved products or services on the market or implemented new or improved production processes as a result of the R&D activity. It also gives the number of patents filed during the year as a result of R&D activity. We make no attempt to estimate the effects of R&D or techies on these measures of innovation, but the reduced form correlations are informative.

Table B10: Explaining ICT training

	(1)	(2)	(3)	(4)	(5)	(6)
I(techies $>$ 0)	0.061*** (0.006)					
I(ICT techies $>$ 0)		0.203*** (0.009)	0.188*** (0.009)			
I(R&D techies $>$ 0)			0.063*** (0.009)			
I(Other techies $>$ 0)			0.037*** (0.006)			
Wage bill (log):						
– Techie				0.048*** (0.003)		
– ICT techies					0.063*** (0.005)	0.035*** (0.007)
– R&D techies						0.024*** (0.006)
– Other techies						0.015 (0.011)
– Ex-techies	0.087*** (0.002)	0.074*** (0.002)	0.065*** (0.002)	0.068*** (0.004)	0.083*** (0.005)	0.083*** (0.011)
Obs.	47,363	47,363	47,363	30,859	15,720	8,727

Dependent variable is an indicator for whether the firm offers ICT training to any of its workers. Regressions include industry×year fixed effects, with robust standard errors in parentheses.*** denotes p-value $\leq$ 0.01, ** p-value $\leq$ 0.05, * p-value $\leq$ 0.10.

We find that the distribution of patents is extremely skewed: the 75^{th} percentile firm-year files no patents, and the 95^{th} percentile files only 4. The 99^{th} percentile firm files 26, and the top four firm-year observations are around 2,000. Responses to questions related to innovations are much less skewed, as seen in Table B11: only a quarter of firms say that they had no process or product innovations in the past year, while half had both.

Table B11: Innovation activity, share of firms

		Process innovation?	
		No	Yes
Product innovation?	No	0.24	0.10
	Yes	0.19	0.47

Source: R&D survey.

Next, we analyze the relationship between patenting, R&D spending, and techies. We proceed in two steps. First, we analyze the patenting and innovation activities of firms using the R&D variables from the R&D survey. Second, we match the R&D survey with the administrative DADS data to correlate the wage bill of techies with the firms' patenting and innovation activities. Both samples are restricted to firm-year observations with positive

R&D expenditures. We use a negative binomial model as the dependent variable is the number of patents filed by the firm and a linear probability model to analyze innovation activities. The estimates have the interpretation of elasticities as the right-hand side variables are taken in logs. In the two sets of regressions, we include the firm's non-techie wage bill as a control for size, which turns out to be unimportant. Industry and year-fixed effects are included in all regressions.

In Table B12, we report the results of the analysis of the R&D survey.

Table B12: Number of patents (Results using the R&D survey)

	Patent			Innovation			Product Innovation			Process Innovation		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Total R&D	0.609***			0.084***			0.045***			0.039***		
	(0.015)			(0.002)			(0.001)			(0.001)		
R&D Wage Bill		0.592***	0.333***		0.083***	0.066***		0.047***	0.045***		0.037***	0.021***
		(0.016)	(0.051)		(0.002)	(0.003)		(0.001)	(0.002)		(0.001)	(0.002)
R&D ex-wage bill			0.271***			0.014***			-0.001			0.015***
			(0.053)			(0.003)			(0.002)			(0.002)
Obs.	87,393	86,339	76,297	87,393	86,339	76,297	87,393	86,339	76,297	87,393	86,339	76,297

Notes: Dependent variable is firm-level patent count from R&D survey data. All explanatory variables are in logs. Industry and year-fixed effects are included in all regressions, with robust standard errors in parentheses.*** denotes p-value ≤ 0.01, ** p-value ≤ 0.05, * p-value ≤ 0.10.

The results presented in columns (1) and (2) suggest that there is a positive relationship between R&D spending and the number of patents, with an elasticity of around 0.60. This elasticity hardly changes when we use the R&D wage bill in column (2). When we break down R&D spending into wage and non-wage components in column (3), we still find a positive correlation between patenting activity and R&D expenditures. This indicates the importance of labor in producing R&D services.

Moving on to columns (4) to (12), we find a strong positive correlation between R&D spending and the likelihood of innovation in both products and processes. Interestingly, the elasticity of the R&D techie wage bill to innovation is almost five times greater than that of the R&D ex-wage bill. This underscores the importance of R&D workers in driving product innovation.

We now study the results in the matched sample in Tables B13 and B14. We include the firm's non-techie wage bill as a control for size, which turns out to be unimportant.

In Table B13, we report the results from the matched R&D and DADS datasets on the impact of techies on the number of patents.

In column (1), we estimate the impact of techies and observe a striking similarity to the effect of total Research and Development (R&D) spending presented in Table B12. We then split techies into their three subgroups by function in columns (2) to (4). We find a larger correlation between patenting and R&D techies than with ICT techies. The correlation of Other techies with patenting is much smaller and not well identified. It is noteworthy that the results on R&D and ICT techies hold across both manufacturing and non-manufacturing sectors.

Our last statistical exercise in this section reports linear probability models for the three innovation outcome indicator variables. The parameter estimates reported in Table B14 have the interpretation of semi-elasticities. Overall, Techies have a statistically significant positive relationship with the likelihood of innovation. This suggests that techies can lead to increased innovation in product development or process improvement.

R&D techies have a statistically significant positive relationship with both process and product innovation, in both manufacturing and non-manufacturing industries—except that when we focus on process innovation in non-manufacturing firms, this correlation vanishes.

Table B13: Number of patents (results using the matched dataset)

			Manufacturing	Non-Manufacturing
	(1)	(2)	(3)	(4)
Wage bill (log):				
– Techies	0.787***			
	(0.067)			
– R&D techies		0.433***	0.465***	0.321***
		(0.039)	(0.046)	(0.047)
– ICT techies		0.186***	0.152***	0.221***
		(0.040)	(0.043)	(0.066)
– Other techies		0.096	0.238***	-0.127
		(0.079)	(0.063)	(0.112)
Obs.	18,155	18,155	16,070	2,085

Source: Matched dataset.
Notes: Dependent variable is firm-level patent count from R&D survey data. All explanatory variables are in logs. Firm's non-techie wage bill and industry and year-fixed effects are included in all regressions, with robust standard errors in parentheses.*** denotes p-value ≤ 0.01, ** p-value ≤ 0.05, * p-value ≤ 0.10.

Table B14: Innovation (Results using the R&D survey)

	Innovation				Product Innovation				Process Innovation			
			Manuf.	Non-Manuf.			Manuf.	Non-Manuf.			Manuf.	Non-Manuf.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Wage bill (log):												
– Techies	0.102***				0.028***				0.074***			
	(0.011)				(0.006)				(0.007)			
– R&D techies		0.041***	0.041***	0.030***		0.017***	0.015***	0.017***		0.025***	0.026***	0.013
		(0.008)	(0.009)	(0.015)		(0.005)	(0.005)	(0.009)		(0.005)	(0.006)	(0.009)
– ICT techies		0.017**	0.017**	0.019		0.015***	0.015***	0.014		0.002	0.002	0.005
		(0.007)	(0.008)	(0.016)		(0.004)	(0.005)	(0.011)		(0.004)	(0.005)	(0.011)
– Other techies		0.037***	0.031**	0.048**		-0.001	-0.003	0.006		0.038***	0.034***	0.042***
		(0.011)	(0.013)	(0.022)		(0.007)	(0.008)	(0.014)		(0.007)	(0.008)	(0.013)
Obs.	18,305	18,305	16,209	2,096	18,305	18,305	16,209	2,096	18,305	18,305	16,209	2,096

Source: Matched dataset.
Notes: Dependent variables indicators for innovation. All explanatory variables are in logs. Firm's non-techie wage bill and industry and year-fixed effects are included in all regressions, with robust standard errors in parentheses. *** denotes p-value ≤ 0.01, ** p-value ≤ 0.05, * p-value ≤ 0.10.

This suggests that while R&D techies are beneficial for innovation outcomes in general, their impact on process innovation in non-manufacturing industries may be limited.

In addition, we find that ICT techies have a positive relationship with product innovation in the manufacturing industry, but they are not associated with product innovation in non-manufacturing industries. This implies that the presence of ICT techies may be particularly beneficial for product innovation in the manufacturing industry, but may not have a significant impact on product innovation in other industries. Interestingly, ICT techies have no impact on process innovation, regardless of the industry considered.

Finally, we show that Other techies have a positive relationship with process innovation across industries. In contrast, Other techies are not associated with product innovation. This suggests that having techies with expertise not specifically related to R&D or ICT can still contribute to innovation outcomes, but their impact may be more important in process innovation, in both manufacturing and non-manufacturing industries.

C Firm choice of techies

In this section, we describe a very simple model of a firm's choice of how many techies to employ. The purpose is to give intuition about why some but not all firms choose to hire techies. We describe the firm's optimal choice of techies, given a simple function from current techies to future productivity. A two-period model is sufficient to illustrate the mechanisms at work. Firm f faces an inverse demand curve given by

$$P_{ft} = A_f Y_{ft}^{\frac{-1}{\eta}}, \quad \eta > 1. \tag{9}$$

The relationship from techies to changes in log productivity is

$$\omega_{ft} = \omega_{ft-1} + Max\left[\beta \ln T_{ft-1}, 0\right], \quad \beta \geq 0. \tag{10}$$

Fixed costs of employing positive techies are κ_f, and techies are paid r per unit. The production function is

$$Y_{ft} = \Omega_{ft} L_{ft}$$

where L_f is a bundle of inputs available at cost w, and $\Omega_{ft} = e^{\omega_{ft}}$. By equation (9), revenue is

$$R_{ft} = A_f \left[\Omega_{ft} L_{ft}\right]^{\frac{\eta-1}{\eta}}.$$

Let labor be the numeraire. The static profit-maximizing input choice is

$$L_{ft} = \Omega_{ft}^{\eta-1} \left[\frac{\eta-1}{\eta} A_f\right]^{\eta}.$$

Plugging this back into the expression for revenue gives optimized revenue for given productivity,

$$R_{ft} = B_f \Omega_{ft}^{\eta-1}, \qquad B_f = A_f^{\eta} \left(\frac{\eta-1}{\eta}\right)^{\eta-1}.$$

With no discounting, the firm chooses T_{ft-1} to maximize two-period profits,

$$\Pi_f = B_f \Omega_{ft-1}^{\eta-1} + B_f \Omega_{ft}^{\eta-1} - \left(r T_{ft-1} + \kappa_f\right) I\left(T_{ft-1} > 0\right)$$

where $I\,()$ is the indicator function. There will be two solutions, one the corner solution with $T_{ft-1} = 0$ and the other an interior optimum with $T_{ft-1} > 0$. When $T_{ft-1} > 0$, equation (10) implies $\Omega_{ft} = \left(T_{ft-1}\right)^{\beta} \Omega_{ft-1}$. Substituting this into the expression for profits gives

$$\Pi_f^T = B_f \Omega_{ft-1}^{\eta-1} + B_f \left(\left(T_{ft-1}\right)^{\beta} \Omega_{ft-1}\right)^{\eta-1} - r T_{ft-1} - \kappa_f \tag{11}$$

At the interior solution, the firm chooses T_{ft-1} to maximize (11). The solution of this problem is

$$T_{ft-1}^{opt} = \left[r^{-1} \beta \left(\eta - 1\right) \Omega_{ft-1}^{\eta-1}\right]^{\frac{1}{1-\beta(\eta-1)}} \tag{12}$$

For high enough values of β, the second order condition of the profit maximization problem doesn't hold and optimal techie employment is infinite. To rule this out we assume

$\beta(\eta-1) < 1$. Plugging the solution (12) back into the expression for Ω_{ft} and defining the constants $\nu = \beta(\eta-1) < 1$ and $\mu = \frac{1}{1-\beta(\eta-1)} > 1$ gives

$$\Omega_{ft}^{opt} = \left[\frac{r}{\nu}\right]^{-\beta\mu} \Omega_{ft-1}^{\mu} \tag{13}$$

This equation establishes the intuitive result that optimized Ω_{ft} is decreasing in the cost of techies r and increasing in Ω_{ft-1}.

To figure out whether $T_{f1} = 0$ or $T_{f1} > 0$ yields higher profits, the firm computes maximized profits in each case. Profits at the corner solution $T_{f1} = 0$ are

$$\Pi_f^C = 2B_f\Omega_{f1}^{\eta-1}$$

To compute profits at the interior solution, substitute (12) and (13) into (11) to obtain

$$\Pi_f^T = B_f\Omega_{ft-1}^{\eta-1} + \left(\Omega_{ft-1}^{\eta-1} r^{-\nu}\nu\right)^{\mu}\left[B_f\nu^{\nu} - 1\right] - \kappa_f$$

Thus the difference between the two profit levels is

$$\Pi_f^T - \Pi_f^C = \left(\Omega_{ft-1}^{\eta-1} r^{-\nu}\nu\right)^{\mu}\left[B_f\nu^{\nu} - 1\right] - \kappa_f$$

A necessary condition for this to be positive is that the term in brackets is positive. This will be more likely when demand (captured by B_f) is higher. If the term in brackets is positive, the whole expression is more likely to be positive the smaller are κ_f and r and the larger is Ω_{ft-1}. If the term in brackets is negative, then $\Pi_f^T - \Pi_f^C < 0$ even if $\kappa_f = 0$, which shows that fixed costs are not a necessary condition for zero techies to be optimal.

The lessons from this exercise are intuitive:

- The optimal amount of techies is more likely to be positive when demand and/or initial productivity are higher.
- The optimal amount of techies is more likely to be zero when fixed costs of techies are high.
- The optimal amount of techies may be zero even if the fixed cost of employing techies is zero.
- When the optimal amount of techies is positive, it is increasing in initial productivity.

D Production function and productivity estimation methodology

We refer the reader to Grieco et al. (2016) for their methodology. We do not deviate from it. Here we provide complete details on our implementation of GNR.

GNR start with a production function (within some industry)

$$Q_{ft} = A_{ft}F(X_{ft}), \tag{14}$$

for some input vector X and Hicks-Neutral productivity A. Taking logs this becomes

$$q_{ft} = \ln Q_{ft} = \ln[A_{ft}F(e^{\ln X_{ft}})] = \ln A_{ft} + \ln[F(e^{x_{ft}})] = a_{ft} + f(x_{ft}), \tag{15}$$

where all lower case letters denote logs of upper case variables and functions. Let

$$a_{ft} = \omega_{ft} + u_{ft}, \tag{16}$$

where ω is the part of the productivity shifter that the firm observes before making input demand decisions and u is the unexpected part. While both ω and u affect output, the important distinction is that ω is be correlated with variable input choices, while u is not.

Assume that ω_{ft} follows a 1^{st} order controlled Markov (CM) process, and for purposes of exposition, let it be a simple AR(1),

$$\omega_{ft} = \text{const} + \lambda\omega_{ft-1} + \beta\mathbf{z}_{ft-1} + \xi_{ft}, \tag{17}$$

where $\mathbf{z}_{ft-1}$ is a vector that includes firm choices (techies, exporting, etc.) and ξ_{ft} is an orthogonal innovation.

We do not observe quantities. Therefore we adjust the basic GNR model. We assume that—as in GLZ—firms face an industry-specific downward sloping demand curve, with elasticity $\eta = 1/(1-\rho) > 1$, $\rho \in (0,1)$, *á la* Klette and Griliches (1996), as in GNR's Appendix O6-4 "Revenue Production Functions".

A firm that sets price P_{ft} sells quantity

$$Q_{ft} = B_t\left(\frac{P_{ft}}{\Pi_t}\right)^{-\eta}, \tag{18}$$

where Π_t is the aggregate price index and B_t is aggregate demand. Alternatively, write

$$P_{ft} = Q_{ft}^{-1/\eta}B_t^{1/\eta}\Pi_t = Q_{ft}^{-1+\rho}B_t^{1-\rho}\Pi_t. \tag{19}$$

Therefore, revenue is

$$R_{ft} = P_{ft}Q_{ft} = Q_{ft}^{\rho}B_t^{1-\rho}\Pi_t. \tag{20}$$

Given an aggregate price index Π_t we have deflated revenues

$$\widetilde{R}_{ft} = \frac{R_{ft}}{\Pi_t} = Q_{ft}^{\rho}B_t^{1-\rho}. \tag{21}$$

The theory-consistent measure of B_t is given by

$$B_t^{\rho} = \sum_{f\in\Theta_t} Q_{ft}^{\rho} = \sum_{f\in\Theta_t}\widetilde{R}_{ft}B_t^{-1+\rho} \quad\Longrightarrow\quad B_t = \sum_{f\in\Theta_t}\widetilde{R}_{ft} = \frac{1}{\Pi_t}\sum_{f\in\Theta_t}R_{ft}, \tag{22}$$

i.e., the sum of deflated revenues, where Θ_t is the set of all firms that serve the (single) market. Taking logs of (20) we have

$$r_{ft} = \rho q_{ft} + (1-\rho)\ln B_t + \ln\Pi_t, \tag{23}$$

and using the production function and rearranging we have the deflated "revenue production function"

$$\widetilde{r}_{ft} = \ln\frac{R_{ft}}{\Pi_t} = (1-\rho)\ln B_t + \rho f(\cdot) + \rho\omega_{ft} + \rho u_{ft}. \tag{24}$$

In principle, time variation in B_t can identify ρ, which can be used to "unpack" the production function from the "revenue production function"—but since we have only a few years we will take a different route. We absorb $(1-\rho)\ln B_t$ in time fixed effects (see below), so that in practice we don't need to deflate revenues, which is inconsequential for the results.

Firms are price takers on input markets. Firms maximize expected profits (the value of u is not in their current information set). By manipulating the FONC with respect to any static input j that is chosen without frictions, we obtain the associated first step factor share equation

$$s_{ft}^{j} = \ln\left[E(e^{u'})\rho\epsilon^{j}(x_{ft})\right] - u'_{ft}, \tag{25}$$

where s_{ft}^{j} is the log of the cost share of input j in revenue (potentially greater than 1, if the firm is hit by a large enough negative u shock), $\epsilon^j(x_{ft}) = \partial \ln f(x_{ft})/\partial \ln j$ is the output elasticity w.r.t. input j, and $u'_{ft} = \rho u_{ft}$.

We estimate (25) by NLLS, using some parametric assumption on $\epsilon^j(x_{ft})$. Once $E(e^{u'})\rho\epsilon^j(x_{ft})$ is identified, we use the residual to estimate $E(e^{u'})$, which allows identifying $\rho\epsilon^j(x_{ft})$. In order to identify $\epsilon^j(x_{ft})$ we need an estimate of ρ, which can be obtained in the second step. However, since our panel is too short to precisely identify ρ, we stay with $\rho\epsilon^j(x_{ft})$.

In (25) $u'_{ft} = \rho u_{ft}$ because u contributes directly to output. Unlike GLZ, the surprise shock is not a demand shock. We can assume that, like in GLZ, $a = \omega$ and that u is an *ex post* demand shock. In that case the same equation (25) arises, with the only difference that there is no ρ in the residual, i.e., $u'_{ft} = u_{ft}$. All this is inconsequential for what follows, so henceforth we drop the superscript in u'_{ft}.

In Section 5 of their paper, GNR use in the first step share equation a "complete" second-order polynomial in m, l and k plus a term that combines all three ($m \times l \times k$). They then integrate this w.r.t. m. They subtract this integral from q, and estimate the second step, in which there are only second-order terms in l and k. We adapt this to the case in which output quantities are not observed, while only revenue is.

We entertain two assumptions on labor, L_{ft}:

1. L_{ft} is "predetermined", i.e., it does not respond to the innovation to productivity ξ_{ft}, conditional on ω_{ft-1} (like K).

2. L_{ft} is "static", i.e., it responds to the innovation to productivity ξ_{ft}, conditional on ω_{ft-1}, and the static FONC holds (like M).

These are described in the following subsections.

D.1 Single static input M, both L and K predetermined

Assume that, as in GNR, material inputs are static and frictionless, and that both L and K are dynamic and predetermined. The first step share equation is

$$s^m_{ft} = \ln S^m_{ft} = \ln\left[E(e^u)\rho\epsilon^m(x_{ft})\right] - u_{ft}, \tag{26}$$

where we drop the "prime" on u because, as noted above, this is inconsequential. Denote

$$\begin{aligned} E(e^u)\rho\epsilon^m(x_{ft}) &= \gamma'(x_{ft}) \\ \rho\epsilon^m(x_{ft}) &= \gamma^m(x_{ft}) \ . \end{aligned}$$

Estimate (26) by NLLS: choose the vector γ' to minimize

$$\sum_{ft}[s^m_{ft} - \ln\begin{pmatrix} \gamma'_0 + \gamma'_m m_{ft} + \gamma'_l l_{ft} + \gamma'_k k_{ft} + \gamma'_{mm} m^2_{ft} + \gamma'_{ll} l^2_{ft} + \gamma'_{kk} k^2_{ft} \\ +\gamma'_{ml} m_{ft} l_{ft} + \gamma'_{mk} m_{ft} k_{ft} + \gamma'_{lk} l_{ft} k_{ft} + \gamma'_{mlk} m_{ft} l_{ft} k_{ft} \end{pmatrix}]^2. \tag{27}$$

Once γ' is estimated, we recover γ^m by dividing through all point estimates by $(1/N)\sum_{ft}(e^{u_{ft}})$.

Integrating $\gamma^m(x_{ft})$ yields

$$\begin{aligned} \int_0^{m_{ft}} \gamma^m(m, l_{ft}, k_{ft})dm &= \int_0^{m_{ft}} \begin{pmatrix} \gamma_0 + \gamma_m m + \gamma_l l_{ft} + \gamma_k k_{ft} + \gamma_{mm} m^2 + \gamma_{ll} l^2_{ft} + \gamma_{kk} k^2_{ft} \\ +\gamma_{ml} m l_{ft} + \gamma_{mk} m k_{ft} + \gamma_{lk} l_{ft} k_{ft} + \gamma_{mlk} m l_{ft} k_{ft} \end{pmatrix} dm \\ &= \begin{pmatrix} \gamma_0 + \frac{1}{2}\gamma_m m_{ft} + \gamma_l l_{ft} + \gamma_k k_{ft} + \frac{1}{3}\gamma_{mm} m^2_{ft} + \gamma_{ll} l^2_{ft} + \gamma_{kk} k^2_{ft} \\ +\frac{1}{2}\gamma_{ml} m_{ft} l_{ft} + \frac{1}{2}\gamma_{mk} m_{ft} k_{ft} + \gamma_{lk} l_{ft} k_{ft} + \frac{1}{2}\gamma_{mlk} m_{ft} l_{ft} k_{ft} \end{pmatrix} m_{ft} \end{aligned}$$

The lower bound for integration implies a normalization on the production function parameters and is inconsequential.

The second step equation is

$$\begin{aligned} y_{ft} &= \widetilde{r}_{ft} - u_{ft} - \int_0^{m_{ft}} \gamma^m(m, l_{ft}, k_{ft}) dm \\ &= \rho\omega_{ft} + (1-\rho)\ln B_t - \mathcal{C}\left(l_{ft}, k_{ft}\right) \\ &= \omega'_{ft} + \alpha_l l_{ft} + \alpha_{ll} l_{ft}^2 + \alpha_k k_{ft} + \alpha_{kk} k_{ft}^2 + \alpha_{lk} l_{ft} k_{ft}, \end{aligned} \tag{28}$$

where we absorb $(1-\rho)\ln B_t$ in

$$\omega'_{ft} = \rho\omega_{ft} + (1-\rho)\ln B_t.$$

For any guess of the vector of coefficients α we can compute $\widehat{\omega}'(\alpha)_{ft}$ as a residual from (28). Now invoke the Markov assumption (17), and estimate

$$\widehat{\omega}'(\alpha)_{ft} = \mathrm{FE}_t + \lambda\widehat{\omega}'(\alpha)_{ft-1} + \rho\beta\mathbf{z}_{ft-1} + \xi'_{ft}, \tag{29}$$

where $\xi'_{ft} = \rho\xi_{ft}$ and the time fixed effects FE_t absorb $(1-\rho)\ln B_t$. Here we can only identify $\rho\beta$, not β. The estimated $\widehat{\xi}'(\alpha)_{ft}$ are orthogonal to $\left(l_{ft}, l_{ft}^2, k_{ft}, k_{ft}^2, l_{ft}k_{ft}\right)$ because they are predetermined by assumption. Use this to build a GMM estimator based on the following moment conditions:

$$E\left\{\widehat{\xi}\left(\alpha_l, \alpha_{ll}, \alpha_k, \alpha_{kk}, \alpha_{lk}\right)_{ft}\left(l_{ft}, l_{ft}^2, k_{ft}, k_{ft}^2, l_{ft}k_{ft}\right)'\right\} = 0\ . \tag{30}$$

Once we have estimates of α we can compute one last time $\widehat{\omega}'(\alpha)_{ft}$ and regress (29) to obtain estimates of λ and $\rho\beta$.

Finally, we compute the revenue elasticities w.r.t. L and K :

$$\begin{aligned} \gamma^l(x_{ft}) &= \alpha_l + 2\alpha_{ll}l_{ft} + \alpha_{lk}k_{ft} + \gamma_l m_{ft} + 2\gamma_{ll}l_{ft}m_{ft} + \frac{1}{2}\gamma_{ml}m_{ft}^2 + \gamma_{lk}k_{ft}m_{ft} + \frac{1}{2}\gamma_{mlk}m_{ft}^2 k_{ft} \\ \gamma^k(x_{ft}) &= \alpha_k + 2\alpha_{kk}k_{ft} + \alpha_{lk}l_{ft} + \gamma_k m_{ft} + 2\gamma_{kk}k_{ft}m_{ft} + \frac{1}{2}\gamma_{mk}m_{ft}^2 + \gamma_{lk}l_{ft}m_{ft} + \frac{1}{2}\gamma_{mlk}m_{ft}^2 l_{ft}, \end{aligned}$$

where, as above, the true output elasticities $\epsilon^l(x_{ft}) = \gamma^l(x_{ft})/\rho$ are not identified without information on ρ.

D.2 Two static inputs M and L, K is predetermined

We estimate the first step share equations for M and L using the same procedure as above. The first step share equations are

$$s_{ft}^m = \ln\left[E(e^u)\gamma^m(x_{ft})\right] - u_{ft}^m \tag{31}$$

$$s_{ft}^l = \ln\left[E(e^u)\gamma^l(x_{ft})\right] - u_{ft}^l\ . \tag{32}$$

Here we obtain two residuals: $u_{ft}^m = u_{ft} + \psi_{ft}^m$ and $u_{ft}^l = u_{ft} + \psi_{ft}^l$. The additional ψ_{ft}^j terms account for the fact that the residuals do not coincide. They are assumed to be orthogonal to u_{ft} and x_{ft}. GNR discuss this in their Appendix O6-3 "Multiple Flexible Inputs". An efficient way to consistently estimate u is to use the average $(u_{ft}^m + u_{ft}^l)/2$. With some abuse of notation, let $u_{ft} = (u_{ft}^m + u_{ft}^l)/2$. We estimate (31) and (32) separately by NLLS, and use

u_{ft} to build $(1/N)\sum_{ft}(e^{u_{ft}})$ and to obtain $\gamma^m(x_{ft})$ and $\gamma^l(x_{ft})$ in (31) and (32), respectively.

Denote the coefficients from the M share equation γ^m and those from the L share equation γ^l. Using the result from Varian (1992) we compute the integral

$$I^{(m,l)} = \int_{m_0}^{m_{ft}} \gamma^m\left(m, l_0, k_{ft}\right) dm + \int_{l_0}^{l_{ft}} \gamma^l\left(m_{ft}, l, k_{ft}\right) dl \ . \tag{33}$$

This sum of integrals equals

$$\begin{array}{rcl} I^{(m,l)} & = & \left(\begin{array}{c} \gamma_0^m + \frac{1}{2}\gamma_m^m m_{ft} + \gamma_l^m l_0 + \gamma_k^m k_{ft} + \frac{1}{3}\gamma_{mm}^m m_{ft}^2 + \gamma_{ll}^m l_0^2 + \gamma_{kk}^m k_{ft}^2 \\ +\frac{1}{2}\gamma_{ml}^m m_{ft} l_0 + \frac{1}{2}\gamma_{mk}^m m_{ft} k_{ft} + \gamma_{lk}^m l_0 k_{ft} + \frac{1}{2}\gamma_{mlk}^m m_{ft} l_0 k_{ft} \end{array}\right) m_{ft} \\ & & -\left(\begin{array}{c} \gamma_0^m + \frac{1}{2}\gamma_m^m m_0 + \gamma_l^m l_0 + \gamma_k^m k_{ft} + \frac{1}{3}\gamma_{mm}^m m_0 + \gamma_{ll}^m l_0^2 + \gamma_{kk}^m k_{ft}^2 \\ +\frac{1}{2}\gamma_{ml}^m m_0 l_0 + \frac{1}{2}\gamma_{mk}^m m_0 k_{ft} + \gamma_{lk}^m l_0 k_{ft} + \frac{1}{2}\gamma_{mlk}^m m_0 l_0 k_{ft} \end{array}\right) m_0 \\ & & +\left(\begin{array}{c} \gamma_0^l + \gamma_m^l m_{ft} + \frac{1}{2}\gamma_l^l l_{ft} + \gamma_k^l k_{ft} + \gamma_{mm}^l m_{ft}^2 + \frac{1}{3}\gamma_{ll}^l l_{ft}^2 + \gamma_{kk}^l k_{ft}^2 \\ +\frac{1}{2}\gamma_{ml}^l m_{ft} l_{ft} + \gamma_{mk}^l m_{ft} k_{ft} + \frac{1}{2}\gamma_{lk}^l l_{ft} k_{ft} + \frac{1}{2}\gamma_{mlk}^l m_{ft} l_{ft} k_{ft} \end{array}\right) l_{ft} \\ & & -\left(\begin{array}{c} \gamma_0^l + \gamma_m^l m_{ft} + \frac{1}{2}\gamma_l^l l_0 + \gamma_k^l k_{ft} + \gamma_{mm}^l m_{ft}^2 + \frac{1}{3}\gamma_{ll}^l l_0^2 + \gamma_{kk}^l k_{ft}^2 \\ +\frac{1}{2}\gamma_{ml}^l m_{ft} l_0 + \gamma_{mk}^l m_{ft} k_{ft} + \frac{1}{2}\gamma_{lk}^l l_0 k_{ft} + \frac{1}{2}\gamma_{mlk}^l m_{ft} l_0 k_{ft} \end{array}\right) l_0 \end{array}$$

We choose the lower integration limits so that there is no constant. Choosing $(m_0, l_0) = (0, 0)$ does the trick and yields

$$\begin{array}{rcl} I^{(m,l)} & = & \int_0^{m_{ft}} \epsilon_{ft}^m\left(m, 0, k_{ft}\right) dm + \int_0^{l_{ft}} \epsilon_{ft}^l\left(m_{ft}, l, k_{ft}\right) dl \\ & = & \left(\gamma_0^m + \dfrac{1}{2}\gamma_m^m m_{ft} + \gamma_k^m k_{ft} + \dfrac{1}{3}\gamma_{mm}^m m_{ft}^2 + \gamma_{kk}^m k_{ft}^2 + \dfrac{1}{2}\gamma_{mk}^m m_{ft} k_{ft}\right) m_{ft} \\ & & +\left(\begin{array}{c} \gamma_0^l + \gamma_m^l m_{ft} + \frac{1}{2}\gamma_l^l l_{ft} + \gamma_k^l k_{ft} + \gamma_{mm}^l m_{ft}^2 + \frac{1}{3}\gamma_{ll}^l l_{ft}^2 + \gamma_{kk}^l k_{ft}^2 \\ +\frac{1}{2}\gamma_{ml}^l m_{ft} l_{ft} + \gamma_{mk}^l m_{ft} k_{ft} + \frac{1}{2}\gamma_{lk}^l l_{ft} k_{ft} + \frac{1}{2}\gamma_{mlk}^l m_{ft} l_{ft} k_{ft} \end{array}\right) l_{ft} \\ & = & \left(\gamma_0^m + \dfrac{1}{2}\gamma_m^m m_{ft} + \gamma_k^m k_{ft} + \dfrac{1}{3}\gamma_{mm}^m m_{ft}^2 + \gamma_{kk}^m k_{ft}^2 + \dfrac{1}{2}\gamma_{mk}^m m_{ft} k_{ft}\right) m_{ft} \\ & & +\left(\gamma_0^l + \dfrac{1}{2}\gamma_l^l l_{ft} + \gamma_k^l k_{ft} + \dfrac{1}{3}\gamma_{ll}^l l_{ft}^2 + \gamma_{kk}^l k_{ft}^2 + \dfrac{1}{2}\gamma_{lk}^l l_{ft} k_{ft}\right) l_{ft} \\ & & +\left(\gamma_m^l m_{ft} + \gamma_{mm}^l m_{ft}^2 + \dfrac{1}{2}\gamma_{ml}^l m_{ft} l_{ft} + \gamma_{mk}^l m_{ft} k_{ft} + \dfrac{1}{2}\gamma_{mlk}^l m_{ft} l_{ft} k_{ft}\right) l_{ft} \ . \end{array}$$

This ensures that each of the 17 unique variables in the polynomial gets a coefficient that is identified from only one first step equation.

The second step equation is

$$y_{ft} = \widetilde{r}_{ft} - u_{ft} - I^{(m,l)} = \rho\omega_{ft} + (1-\rho)\ln B_t - \mathcal{C}\left(k_{ft}\right) = \omega'_{ft} + \alpha_k k_{ft} + \alpha_{kk} k_{ft}^2, \tag{34}$$

where again we absorb $(1-\rho)\ln B_t$ in

$$\omega'_{ft} = \rho\omega_{ft} + (1-\rho)\ln B_t \ .$$

For any guess of α we can compute $\widehat{\omega}'\left(\alpha\right)_{ft}$ as a residual from (34). Now invoke the Markov assumption for ω_{ft} (17), and estimate

$$\widehat{\omega}'\left(\alpha\right)_{ft} = \text{FE}_t + \lambda\widehat{\omega}'\left(\alpha\right)_{ft-1} + \rho\beta e_{ft-1} + \xi'_{ft}, \tag{35}$$

where $\xi'_{ft} = \rho\xi_{ft}$ and the time fixed effects FE_t absorb $(1-\rho)\ln B_t$. As above, we can only identify $\rho\beta$, not β. The estimated $\widehat{\xi'}\left(\alpha\right)_{ft}$ are orthogonal to $\left(k_{ft}, k_{ft}^2\right)$ because they are predetermined by assumption. Use this to build a GMM estimator based on the following moment conditions:

$$E\left\{\widehat{\xi}'\left(\alpha_k, \alpha_{kk}\right)_{ft}\left(k_{ft}, k_{ft}^2\right)'\right\} = 0\ . \tag{36}$$

Once we have estimates of α we can compute one last time $\widehat{\omega}'\left(\alpha\right)_{ft}$ and regress (29) to obtain estimates of λ and $\rho\beta$.

Now compute the revenue elasticity w.r.t. K :

$$\begin{aligned}\gamma_{ft}^{k}(\cdot) &= \alpha_k + 2\alpha_{kk}k_{ft}\\ &\quad +\gamma_k^m m_{ft} + 2\gamma_{kk}^m m_{ft}k_{ft} + \frac{1}{2}\gamma_{mk}^m m_{ft}^2\\ &\quad +\gamma_k^l l_{ft} + 2\gamma_{kk}^l l_{ft}k_{ft} + \frac{1}{2}\gamma_{lk}^l l_{ft}^2\\ &\quad +\gamma_{mk}^l m_{ft}l_{ft} + \frac{1}{2}\gamma_{mlk}^l m_{ft}l_{ft}l_{ft}.\end{aligned}$$

D.3 Pooling firms across industries for the controlled Markov

We estimate the controlled Markov in a pooled sample of firms across industries i. This implies estimating

$$\widehat{\rho}_i\widehat{\omega}\left(\alpha\right)_{ift} = \text{FE}_{it} + \lambda\widehat{\rho}_i\widehat{\omega}\left(\alpha\right)_{ift-1} + \beta e_{ift-1} + \xi'_{ift}\ . \tag{37}$$

The estimator of λ is consistent for a weighted average of λ_i across industries. The estimator of β is consistent for a weighted average of $\rho_i\beta_i$ across industries—not a weighted average of β_i. Thus, the estimator of β conflates cross-industry variation in demand curvature ρ_i and industry-specific impacts in the controlled Markov process β_i.

D.4 Additional descriptive statistics for estimation sample

Table D1: Inputs and Output

	Mean	SD	p10	p25	p50	p75	p90
A. Manufacturing firms							
Revenues	21,661	91,493	414	1,221	4,348	15,114	46,507
Labor expenditure	1,979	4,334	113	271	705	1,863	4,727
Hours worked	105,650	210,399	7,204	15,793	40,149	103,872	259,549
Materials expenditure	10,702	52,352	86	292	1,402	6,151	21,731
Deflated materials	10,314	51,337	81	279	1,356	5,939	20,801
B. Non-Manufacturing firms							
Revenues	5,028	14,899	241	576	1,447	3,483	11,115
Labor expenditure	545	1,187	54	96	195	498	1,214
Hours worked	32,892	75,218	3,644	6,104	11,869	29,466	72,491
Materials expenditure	2,940	10,274	45	147	657	1,693	6,074
Deflated materials	2,868	10,042	40	136	630	1,644	5,927

Notes: Sample period is 2011–2019. Revenues, labor expenditure, and materials expenditure in thousands of euros. Hours worked is total annual hours across all employees. Deflated materials is $M = (E_M/p_M) \times 100$, where E_M is materials expenditure and p_M is a two-digit industry price index with base year 2010 ($p_M = 100$ in 2010), expressing materials in constant 2010 prices. Non-manufacturing: 484,541 observations. Manufacturing: 116,088 observations.

Table D2: Industry-level shares for the treatment indicators

Industry	$I(T_{ft-1} > 0)$	$I(T^{RD}_{ft-1} > 0)$	$I(T^{ICT}_{ft-1} > 0)$	$I(T^{OTH}_{ft-1} > 0)$	$I(T^{38}_{ft-1} > 0)$	$I(T^{47}_{ft-1} > 0)$	$I(x_{ft-1} > 0)$
Food, beverage, tobacco	42.38	18.45	11.41	41.44	38.06	32.98	32.23
Textiles, wearing apparel	60.07	13.94	14.48	58.54	50.54	42.54	74.91
Wood, paper products	65.14	12.35	15.28	63.70	54.52	45.81	44.59
Chemical products	88.61	67.39	38.61	85.49	84.17	77.58	87.41
Pharmaceutical products	93.33	85.13	64.10	90.26	90.26	87.18	89.23
Rubber and plastic	77.39	38.96	21.11	75.63	65.90	64.62	60.13
Basic metal and fabricated metal	76.09	40.25	18.64	73.51	61.62	65.43	48.59
Electrical equipment	90.28	65.80	38.89	88.02	81.42	82.64	79.17
Machinery and equipment	92.78	68.68	37.03	91.91	84.30	87.15	83.98
Transport equipment	89.38	66.84	41.06	88.21	83.68	80.31	74.48
Other manufacturing	63.16	24.72	18.82	60.50	51.01	52.45	38.09
Construction	33.19	1.43	1.97	32.87	24.05	22.52	1.92
Wholesale	42.13	5.01	11.94	39.07	31.88	26.67	46.26
Retail	5.22	0.21	1.60	4.25	2.71	3.53	5.27
Accommodation and food services	2.51	0.07	0.47	2.17	1.26	1.52	0.38
Publishing and broadcasting	44.37	2.42	35.44	28.07	35.34	33.52	20.95
Administrative and support activities	40.53	5.59	15.24	36.00	28.54	28.40	8.68

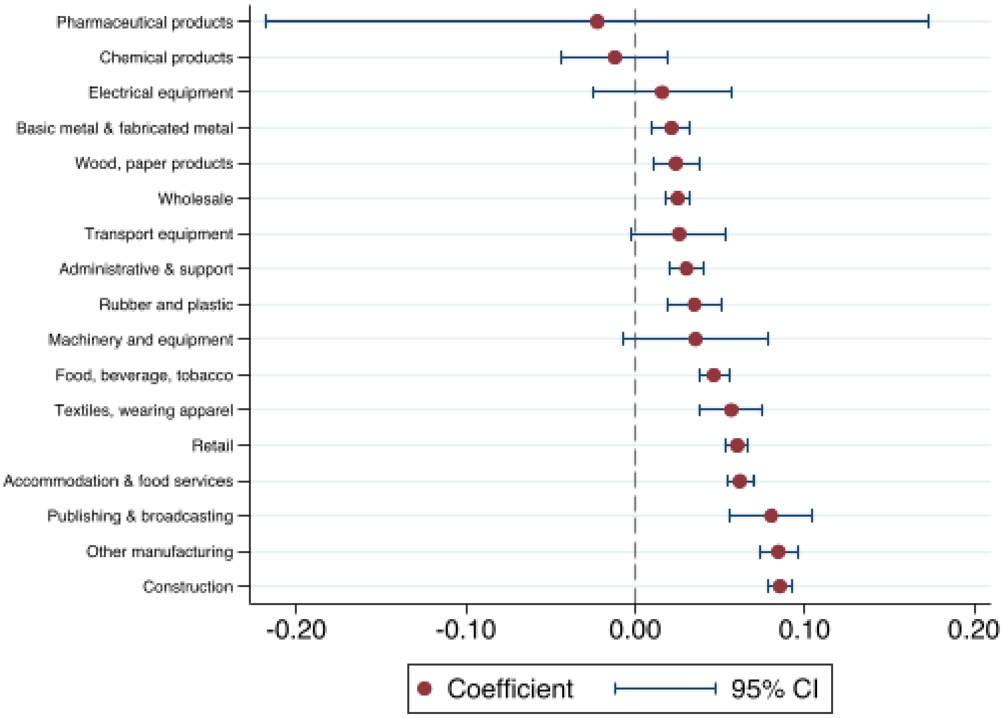


Figure D1: Extensive-margin-only specification, by sector. Coefficient on the extensive-margin indicator $I(T_{f,t-1} > 0)$ when the continuous techie measure $T_{f,t-1}$ is omitted from the right-hand side, estimated separately by A38 sector and ranked by point estimate. Horizontal bars are 95% bootstrap confidence intervals (400 replications).

E General Specification

Table E1 report the estimates of equation 8. It allows us to examine how the impacts of intensive and extensive techie margin increases as productivity rises while Table 6 reports the combined impact effects. Columns 1 and 3 report the baseline estimates (Columns 2 and 8 of Table 4), and columns 2 and 4 report the results when we interact the extensive and intensive techie margins with the lagged productivity.

Interestingly, the extensive techie margin is larger for higher level of productivity and the opposite effect is observed for the intensive techie margin. This result suggests dimininshing return on techie investment.

Table E1: General Specification– GLZ estimates

	Manufacturing		Non-Manufacturing	
	Baseline (1)	Interaction (2)	Baseline (3)	Interaction (4)
$I\left(T_{ft-1}>0\right)$	0.018***	0.014***	0.023***	0.012***
	(0.003)	(0.003)	(0.003)	(0.003)
$I\left(T_{ft-1}>0\right)\times\hat{\omega}_{ft-1}$		0.047***		0.055***
		(0.006)		(0.003)
T_{ft-1}	0.120***	0.122***	0.206***	0.235***
	(0.008)	(0.009)	(0.013)	(0.015)
$T_{ft-1}\times\hat{\omega}_{ft-1}$		-0.165***		-0.263***
		(0.020)		(0.018)
$I\left(x_{ft-1}>0\right)$	0.007***	0.008***	0.007***	0.006**
	(0.002)	(0.002)	(0.002)	(0.002)
$\hat{\omega}_{ft-1}$	0.912***	0.907***	0.873***	0.870***
	(0.003)	(0.005)	(0.002)	(0.002)
Obs.	116,088		484,541	
No. firms	19,571		100,088	

Notes. The table reports estimates of equation (8) in the text. The dependent variable is $\hat{\omega}_{ft}$, log estimated productivity. $I(.)$ is the indicator function. T is the techie wage bill share, x is the value of firm exports. Industry-year fixed effects included in all columns. Bootstrap standard errors clustered by firm in parentheses. *** denotes p-value ≤ 0.01, ** p-value ≤ 0.05, * p-value ≤ 0.10.

F GNR production function estimates

Table F1 reports the average "revenue elasticity" (output elasticity $\times\rho$) across firms, by industry. These estimates arise from the GNR estimator where labor is assumed to be "*dynamic*", i.e., predetermined in time t (like capital), and where we include in the control Markov an indicator for employment of techies and their wage bill share.

Table F1: GNR Production function estimates

Industries	γ^m	γ^l	γ^k	No. of Obs.	No. of Firms
Food, beverage, tobacco	0.429	0.465	0.174	27,201	4,438
Textiles, wearing apparel	0.323	0.556	0.081	7,739	1,146
Wood, paper products	0.289	0.663	0.073	15,529	2,291
Chemical products	0.405	0.470	0.143	6,689	855
Pharmaceutical products	0.263	0.643	0.075	1,526	202
Rubber and plastic	0.363	0.500	0.158	14,040	1,890
Basic metal and fabricated metal	0.272	0.652	0.100	25,449	3,505
Electrical equipment	0.374	0.466	0.137	4,376	586
Machinery and equipment	0.361	0.547	0.095	9,758	1,285
Transport equipment	0.391	0.593	0.086	5,762	786
Other manufacturing	0.248	0.675	0.102	21,345	3,211
Construction	0.247	0.694	0.102	104,860	19,535
Wholesale	0.603	0.361	0.056	169,131	25,050
Retail	0.633	0.326	0.026	238,092	37,367
Accommodation and food services	0.232	0.623	0.188	105,799	20,465
Publishing and broadcasting	0.100	0.793	0.080	10,423	1,756
Administrative and support activities	0.114	0.545	0.225	22,947	4,086

Notes. The table displays "revenue elasticities" (output elasticities times the demand curvature ρ) for materials (γ^m), labor (γ^l) and capital (γ^k). These elasticities vary by firm and year, depending on input use, as explained in Appendix D. Numbers for each industry are calculated by averaging over all firm and year-specific elasticities. The elasticities are computed for the baseline specification, where we include both $I\left(T_{ft-1} > 0\right)$ and T_{ft-1} in the controlled Markov.

G Current vs. lagged techies in the controlled Markov equation

Table G1 reports specifications in which we study the effects of current techies on current productivity, with and without controlling for lagged techies. In Table G2 we report the same specifications in a subsample of "techie-switchers", where we drop firms that always have techies and firms that never have any. Note that adding current techies is a violation of the assumptions that are required for identification. With this caveat in mind, we cautiously interpret these results as being consistent with techies affecting output through future productivity, rather than contemporaneously

Table G1: Current vs. lagged techies in the controlled Markov equation

	Manufacturing			Non-Manufacturing		
	(1)	(2)	(3)	(4)	(5)	(6)
$I\left(T_{ft-1}>0\right)$	0.041***		0.120***	0.056***		0.134***
	(0.003)		(0.006)	(0.003)		(0.005)
$I\left(T_{ft}>0\right)$		0.000	-0.097***		-0.001	-0.103***
		(0.003)	(0.007)		(0.003)	(0.005)
$I\left(x_{ft-1}>0\right)$	0.009***	0.022***	0.015***	0.008***	0.024***	0.015***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
$\hat{\omega}_{ft-1}$	0.910***	0.913***	0.912***	0.872***	0.876***	0.874***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Observations		116,088			484,541	
No. Firms		20,354			114,305	

Notes. The table reports estimates of the controlled Markov equation allowing for current techies. The dependent variable is $\hat{\omega}_{ft}$, log estimated productivity. $I\left(T_{ft-1}>0\right)$ is an indicator for lagged techie employment; $I\left(T_{ft}>0\right)$ is an indicator for current techie employment; $I\left(x_{ft-1}>0\right)$ is an indicator for lagged exports. Industry-year fixed effects included. Standard errors clustered by firm in parentheses. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

Table G2: Current vs. lagged techies in the controlled Markov equation—only firms that switch

	Manufacturing			Non-Manufacturing		
	(1)	(2)	(3)	(4)	(5)	(6)
$I\left(T_{ft-1}>0\right)$	0.046***		0.107***	0.065***		0.121***
	(0.003)		(0.006)	(0.003)		(0.005)
$I\left(T_{ft}>0\right)$		-0.045***	-0.107***		-0.055***	-0.114***
		(0.005)	(0.007)		(0.003)	(0.005)
$I\left(x_{ft-1}>0\right)$	0.005	0.02***	0.013***	0.018***	0.032***	0.025***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
$\hat{\omega}_{ft-1}$	0.876***	0.879***	0.879***	0.874***	0.876***	0.877***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
Observations		38,180			119,902	
No. Firms		7,687			23,040	

Notes. The table reports estimates of the controlled Markov equation allowing for current techies. The dependent variable is $\hat{\omega}_{ft}$, log estimated productivity. $I\left(T_{ft-1}>0\right)$ is an indicator for lagged techie employment; $I\left(T_{ft}>0\right)$ is an indicator for current techie employment; $I\left(x_{ft-1}>0\right)$ is an indicator for lagged exports. The sample is restricted to firms that switch their techie status: we drop firms that always have techies and firms that never have any. Industry-year fixed effects included. Standard errors clustered by firm in parentheses. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

H Third-order polynomial in lagged productivity $\hat{\omega}_{f,t-1}$ in the controlled Markov

The following table reports estimates of similar specifications of the controlled Markov in Table 4 in the main text, where we include higher order of lagged productivity $\hat{\omega}_{f,t-1}$ in the controlled Markov.

Table H1: Third-order polynomial in lagged productivity – GLZ estimates

	Manufacturing						Non-Manufacturing					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(T_{ft-1} > 0)$	0.039***	0.016***					0.051***	0.017***				
	(0.003)	(0.003)					(0.003)	(0.003)				
T_{ft-1}		0.117***						0.214***				
		(0.008)						(0.013)				
$I(T^{RD}_{ft-1} > 0)$			0.014***	0.008***					0.013**	0.001		
			(0.002)	(0.003)					(0.006)	(0.007)		
$I(T^{ICT}_{ft-1} > 0)$			0.019***	0.012***					0.027***	0.015***		
			(0.002)	(0.003)					(0.003)	(0.004)		
$I(T^{OTH}_{ft-1} > 0)$			0.029***	0.011***					0.047***	0.012***		
			(0.003)	(0.003)					(0.003)	(0.003)		
T^{RD}_{ft-1}				0.070***						0.145*		
				(0.024)						(0.087)		
T^{ICT}_{ft-1}				0.118***						0.134***		
				(0.045)						(0.024)		
T^{OTH}_{ft-1}				0.110***						0.242***		
				(0.010)						(0.015)		
$I(T^{38}_{ft-1} > 0)$					0.030***	0.011***					0.044***	0.010***
					(0.002)	(0.003)					(0.003)	(0.003)
$I(T^{47}_{ft-1} > 0)$					0.014***	0.005*					0.030***	0.017***
					(0.002)	(0.003)					(0.002)	(0.003)
T^{38}_{ft-1}						0.148***						0.259***
						(0.014)						(0.019)
T^{47}_{ft-1}						0.084***						0.132***
						(0.013)						(0.017)
$I(x_{ft-1} > 0)$	0.006***	0.004*	0.000	0.002	0.002	0.003	0.005*	0.003	0.002	0.002	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
$\hat{\omega}_{ft-1}$	0.938***	0.941***	0.935***	0.940***	0.937***	0.941***	0.930***	0.932***	0.930***	0.932***	0.930***	0.932***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$\hat{\omega}^2_{ft-1}$	0.027***	0.026***	0.025***	0.025***	0.026***	0.025***	0.018***	0.018***	0.017***	0.018***	0.017***	0.018***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
$\hat{\omega}^3_{ft-1}$	-0.018***	-0.019***	-0.018***	-0.019***	-0.019***	-0.019***	-0.034***	-0.034***	-0.034***	-0.035***	-0.034***	-0.035***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Obs.	116,088						484,541					
No. firms	19,571						100,087					

Notes. The dependent variable is $\hat{\omega}_{ft}$. $I(.)$ denotes the indicator function. T is the techie wage bill share. Superscripts $\{RD, ICT, OTH, 38, 47\}$ denote R&D, ICT, other techies, engineers, and technicians. x is firm exports. Industry-year fixed effects included. Bootstrap standard errors clustered by firm. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

I Labor Input Quality

To address differences in labor quality we adjust the labor input of less-qualified workers in our data as in Gandhi et al. (2020). We identify highly qualified workers as those with PCS codes starting with 2 or 3 (PCS codes starting with 1 are in the agriculture sector, which we omit from our analysis). This, largely, corresponds to managers. We adjust downwards the labor input of less-qualified workers (those with PCS codes starting with 4, 5 and 6) by the ratio of their wage to that of qualified labor:

$$\tilde{N}_{ft} = H_{ft} + (w_L/w_H)\, L_{ft},$$

where w_L is the average wage of L and w_H is the average wage of H in the sample. This assumes that less-qualified labor supply is a fraction (w_L/w_H) of that of highly qualified labor input, in efficiency units.

The estimation results after correcting for labor quality as described above using the GLZ estimator are reported in Table (I1), while the results using the GNR estimator are presented in Tables (I2) and (I3).

J Allocating techies to production – GLZ estimates

The following table reports estimates of the same specifications of the controlled Markov in Table 4 in the main text, when all techies are also allocated to production. It is a more elaborate version of Table 7 in the main text.

Table I1: Impact of techies on productivity – GLZ estimates (Adjusting for labor input quality)

	Manufacturing						Non-Manufacturing					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(T_{ft-1} > 0)$	0.047***	0.026***					0.061***	0.030***				
	(0.003)	(0.003)					(0.003)	(0.003)				
T_{ft-1}		0.108***						0.196***				
		(0.009)						(0.012)				
$I(T^{RD}_{ft-1} > 0)$			0.017***	0.013***					0.009	-0.003		
			(0.002)	(0.003)					(0.006)	(0.007)		
$I(T^{ICT}_{ft-1} > 0)$			0.021***	0.016***					0.026***	0.017***		
			(0.002)	(0.003)					(0.003)	(0.004)		
$I(T^{OTH}_{ft-1} > 0)$			0.036***	0.020***					0.058***	0.024***		
			(0.003)	(0.003)					(0.003)	(0.003)		
T^{RD}_{ft-1}				0.042*						0.139		
				(0.025)						(0.090)		
T^{ICT}_{ft-1}				0.081*						0.109***		
				(0.046)						(0.023)		
T^{OTH}_{ft-1}				0.098***						0.229***		
				(0.011)						(0.015)		
$I(T^{38}_{ft-1} > 0)$					0.032***	0.017***					0.047***	0.015***
					(0.002)	(0.003)					(0.003)	(0.004)
$I(T^{47}_{ft-1} > 0)$					0.021***	0.011***					0.039***	0.027***
					(0.002)	(0.003)					(0.003)	(0.003)
T^{38}_{ft-1}						0.117***						0.239***
						(0.014)						(0.018)
T^{47}_{ft-1}						0.089***						0.119***
						(0.013)						(0.017)
$I(x_{ft-1} > 0)$	0.008***	0.006**	0.001	0.002	0.003	0.004	0.008***	0.006**	0.005**	0.005*	0.004	0.004*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
$\hat{\omega}_{ft-1}$	0.912***	0.915***	0.910***	0.913***	0.911***	0.914***	0.875***	0.876***	0.875***	0.876***	0.874***	0.876***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Obs.	116,227						484,659					
No. firms	19,566						100,038					

Notes. The table reports estimates of equation (5) in the text. The dependent variable is $\hat{\omega}_{ft}$, log estimated productivity. $I(.)$ denotes the indicator function. T is the techie wage bill share. Superscripts $\{RD, ICT, OTH, 38, 47\}$ denote R&D, ICT, other techies, engineers, and technicians. x is firm exports. Industry-year fixed effects included. Bootstrap standard errors clustered by firm. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

Table I2: Impact of techies on productivity – GNR estimates assuming labor to be static, adjusting for labor input quality

	Manufacturing						Non-Manufacturing					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(T_{ft-1} > 0)$	0.038***	0.030***					0.023***	0.015***				
	(0.001)	(0.002)					(0.001)	(0.001)				
T_{ft-1}		0.041***						0.045***				
		(0.004)						(0.003)				
$I(T^{RD}_{ft-1} > 0)$			0.014***	0.012***					0.006***	0.006***		
			(0.001)	(0.001)					(0.002)	(0.002)		
$I(T^{ICT}_{ft-1} > 0)$			0.014***	0.012***					0.009***	0.005***		
			(0.001)	(0.001)					(0.001)	(0.001)		
$I(T^{OTH}_{ft-1} > 0)$			0.032***	0.026***					0.022***	0.014***		
			(0.001)	(0.002)					(0.001)	(0.001)		
T^{RD}_{ft-1}				0.011						-0.032		
				(0.010)						(0.023)		
T^{ICT}_{ft-1}				0.025						0.036***		
				(0.017)						(0.008)		
T^{OTH}_{ft-1}				0.031***						0.050***		
				(0.004)						(0.004)		
$I(T^{38}_{ft-1} > 0)$					0.028***	0.026***					0.020***	0.014***
					(0.001)	(0.001)					(0.001)	(0.001)
$I(T^{47}_{ft-1} > 0)$					0.023***	0.019***					0.014***	0.010***
					(0.001)	(0.001)					(0.001)	(0.001)
T^{38}_{ft-1}						0.020***						0.040***
						(0.006)						(0.005)
T^{47}_{ft-1}						0.027***						0.037***
						(0.005)						(0.005)
$I(x_{ft-1} > 0)$	0.015***	0.015***	0.013***	0.013***	0.013***	0.013***	0.006***	0.006***	0.005***	0.005***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\hat{\omega}_{ft-1}$	0.920***	0.921***	0.919***	0.920***	0.917***	0.918***	0.933***	0.933***	0.933***	0.934***	0.933***	0.934***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Obs.	118,796						549,881					
No. firms	19,789						107,649					

Notes. The table reports estimates of equation (6) in the text. The dependent variable is $\widehat{\rho\omega}_{ft}$. $I(.)$ denotes the indicator function. T is the techie wage bill share. Superscripts $\{RD, ICT, OTH, 38, 47\}$ denote R&D, ICT, other techies, engineers, and technicians. x is firm exports. Industry-year fixed effects included in all columns. Bootstrap standard errors clustered by firm. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

Table I3: Impact of techies on productivity – GNR estimates assuming labor to be predetermined, adjusting for labor input quality

	Manufacturing						Non-Manufacturing					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(T_{ft-1} > 0)$	0.029***	0.019***					0.013***	0.010***				
	(0.003)	(0.003)					(0.001)	(0.001)				
T_{ft-1}		0.047***						0.017***				
		(0.007)						(0.004)				
$I(T^{RD}_{ft-1} > 0)$			0.002	-0.000					0.003*	0.004**		
			(0.002)	(0.002)					(0.001)	(0.002)		
$I(T^{ICT}_{ft-1} > 0)$			0.008***	0.004*					-0.000	-0.002*		
			(0.002)	(0.002)					(0.001)	(0.001)		
$I(T^{OTH}_{ft-1} > 0)$			0.026***	0.017***					0.010***	0.009***		
			(0.003)	(0.003)					(0.001)	(0.001)		
T^{RD}_{ft-1}				0.023						-0.023		
				(0.015)						(0.016)		
T^{ICT}_{ft-1}				0.080***						0.018**		
				(0.026)						(0.008)		
T^{OTH}_{ft-1}				0.051***						0.012***		
				(0.008)						(0.003)		
$I(T^{38}_{ft-1} > 0)$					0.017***	0.015***					0.009***	0.008***
					(0.002)	(0.002)					(0.001)	(0.001)
$I(T^{47}_{ft-1} > 0)$					0.018***	0.012***					0.011***	0.009***
					(0.002)	(0.002)					(0.001)	(0.001)
T^{38}_{ft-1}						0.026***						0.007
						(0.009)						(0.006)
T^{47}_{ft-1}						0.052***						0.018***
						(0.010)						(0.006)
$I(x_{ft-1} > 0)$	0.027***	0.026***	0.026***	0.026***	0.025***	0.026***	0.008***	0.008***	0.005***	0.005***	0.007***	0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\hat{\omega}_{ft-1}$	0.732***	0.729***	0.734***	0.730***	0.734***	0.731***	0.830***	0.830***	0.838***	0.837***	0.830***	0.830***
	(0.023)	(0.023)	(0.023)	(0.022)	(0.023)	(0.024)	(0.008)	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)
Obs.	118,782						540,153					
No. firms	19,787						105,495					

Notes. The table reports estimates of equation (6) in the text. The dependent variable is $\hat{\rho\omega}_{ft}$. $I(.)$ denotes the indicator function. T is the techie wage bill share. Superscripts $\{RD, ICT, OTH, 38, 47\}$ denote R&D, ICT, other techies, engineers, and technicians. x is firm exports. Industry-year fixed effects included. Bootstrap standard errors clustered by firm. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.

Table J1: Allocating techies to production – GLZ estimates

	Manufacturing						Non-Manufacturing					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I(T_{ft-1} > 0)$	0.023***	0.007**					0.028***	0.010***				
	(0.003)	(0.004)					(0.003)	(0.003)				
T_{ft-1}		0.089***						0.117***				
		(0.010)						(0.011)				
$I(T^{RD}_{ft-1} > 0)$			0.017***	0.007**					0.016**	0.020***		
			(0.003)	(0.003)					(0.006)	(0.008)		
$I(T^{ICT}_{ft-1} > 0)$			0.024***	0.021***					0.038***	0.020***		
			(0.003)	(0.003)					(0.004)	(0.004)		
$I(T^{OTH}_{ft-1} > 0)$			0.014***	0.007*					0.020***	0.010***		
			(0.003)	(0.004)					(0.003)	(0.003)		
T^{RD}_{ft-1}				0.129***						-0.049		
				(0.027)						(0.107)		
T^{ICT}_{ft-1}				0.044						0.200***		
				(0.046)						(0.022)		
T^{OTH}_{ft-1}				0.051***						0.070***		
				(0.012)						(0.012)		
$I(T^{38}_{ft-1} > 0)$					0.016***	0.005					0.017***	0.005
					(0.003)	(0.003)					(0.003)	(0.004)
$I(T^{47}_{ft-1} > 0)$					0.015***	0.007**					0.031***	0.023***
					(0.003)	(0.003)					(0.003)	(0.003)
T^{38}_{ft-1}						0.091***						0.090***
						(0.016)						(0.017)
T^{47}_{ft-1}						0.070***						0.085***
						(0.014)						(0.016)
$I(x_{ft-1} > 0)$	0.009***	0.008***	0.001	0.001	0.005**	0.006**	0.024***	0.023***	0.020***	0.021***	0.021***	0.021***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
$\hat{\omega}_{ft-1}$	0.917***	0.915***	0.915***	0.914***	0.916***	0.915***	0.879***	0.878***	0.878***	0.878***	0.878***	0.878***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Obs.	115,115						486,240					
No. firms	19,456						100,111					

Notes. The table reports estimates of equation (7), but when all techies are allocated to production, in addition to the cvontrolled Markov . The dependent variable is $\hat{\omega}_{ft}$. $I(.)$ denotes the indicator function. T is the techie wage bill share; superscripts $\{RD, ICT, OTH, 38, 47\}$ denote R&D, ICT, other techies, engineers, and technicians. x is firm exports. Industry-year fixed effects included. Bootstrap standard errors clustered by firm. *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$.