Techies and Firm Level Productivity *

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

April 21, 2025

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 causal 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 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 technology-related 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). Allocation of more talent to technical occupations is also associated with faster development outcomes (Murphy et al., 1991). Historical studies echo this insight, showing 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. This paper aims to help fill this gap.

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 with information and communication technologies (ICT) and other technology-related functions also contribute significantly. Their effects are pervasive, extending beyond manufacturing to non-manufacturing industries.

We identify techie workers by using the French occupational classification. INSEE (2003) 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

[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).

show that techies are strongly associated with innovation efforts and outcomes (e.g., patents), and are involved in adopting and diffusing technology within firms.

Techies are not homogeneous. 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 expenditure on techies. This methodology enables identification of the causal impact of techies on firm-level productivity.

Our identification strategy relies on two assumptions. First, techies affect firm productivity with a lag. Second, they do not contribute directly to current output. We employ different production function estimators that deal with identification challenges in different ways, while maintaining this approach. This is analogous to how investment or R&D expenditure is modeled in the productivity literature. We implement this approach across a range of specifications, including flexible nonlinear productivity processes and alternative classifications of techies. We also control for exporting status throughout, as exporting is known to correlate with firm productivity (De Loecker, 2013; Barrows et al., 2023). We also show in Section 6.3 that our findings are not sensitive to the inclusion of managers as an additional determinant of firm-level productivity (Bloom et al., 2017).

We find that firms that employ techies experience a substantial increase in future productivity: 4–5 percent higher productivity one year later, and a cumulative long-run effect exceeding 45 percent. While R&D techies drive this pattern in manufacturing, as in Doraszelski and Jaumandreu (2013), ICT and other techies also contribute meaningfully. 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 clearly reject the null hypothesis that techies are no different than other workers, 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% 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 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 causal 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 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.

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 (but do not estimate) to greater productivity, and 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 structural 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 understood. 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 both productivity and its causal sources.

Here our contribution is twofold. First, we are the first to jointly estimate the effects of R&D, ICT, and other techies on firm-level productivity. Second, we extend the analysis beyond manufacturing to include a large set of non-manufacturing firms. This broader scope allows us to capture productivity-enhancing mechanisms that are missed in R&D-focused or IT-focused studies.

Two pioneering studies on structural productivity estimation are De Loecker (2013), who examine exporting, and Doraszelski and Jaumandreu (2013), who focus on R&D expenditure. We build on their frameworks. Doraszelski and Jaumandreu (2013) assume flexible labor markets and specific functional forms; we relax these assumptions using alternative estimation methodologies. In an earlier paper Crepon et al. (1998) also study the effect of R&D on productivity, but rely on a cross-sectional setting and different instruments.[2] Our panel structure and identification strategy allow for sharper causal inference under weaker assumptions.

The rest of the paper is organized as follows. In Section 2 we provide a detailed account of the sources and construction of our datasets. In Section 3 we present a comprehensive analysis of the role of techies, highlighting their technical expertise and their crucial role in adopting, mediating, and diffusing technology at the firm level. Section 4 outlines the theoretical basis for the inclusion of techies in our productivity model and how they can impact productivity. In Section 5 we describe our methodology, where we provide a comprehensive discussion of the econometric challenges and the steps taken to address them. In Section 6 we present the main results of our analysis and perform various sensitivity checks to test the robustness of our 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 panel dataset on firms in the French private sector between 2011 and 2019 by merging three confidential, administrative firm-level datasets.[3] We complement this information with survey data to characterize techies and describe their roles in firms. Matching firms across these datasets is straightforward because firms are identified by the same iden-

[2] 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 explain in Section 5. This differs from how Crepon et al. (1998) address the econometric challenges, who had only a cross section of firms at their disposal.

[3] 2011 is the first year for which our data are available and 2019 is the last pre-pandemic year.

tification number (SIREN) in each of the three datasets. We highlight key features of the data here and relegate other details to Appendix A.

2.1 The composition of labor within firms

Our data on employment is from the DADS.[4] All firms with employees are required to report wages, hours paid, occupation, and the 2-digit sector of activity of the firm. Our labor input measure is hours paid. In a robustness check we also consider quality-adjusted labor input, as in Fox and Smeets (2011).[5] The estimation sample includes firms in 17 industries in both manufacturing and non-manufacturing sectors.[6]

The DADS reports detailed 4-digit PCS occupational codes, almost 500 in total. We use the detailed definitions of these codes (INSEE, 2003) to select the 56 4-digit occupations that we classify as techies. These definitions show that their work is closely related to the installation, management, maintenance, and support of ICT, product and process design, longer-term R&D activities, and other tasks related to technology. In short, the employment of techies is a direct measure of firms' investment in technology. The detailed 4-digit PCS codes allow us to classify techie occupations along two dimensions. The first is whether they are *technical managers and engineers* or *technicians*. The second dimension is their technological orientation: *R&D techies*, *ICT techies*, and *Other techies*, see Table A1.

The documentation in INSEE (2003) makes it clear that techies perform different tasks than workers in other occupations. For example, technical managers and engineers (PCS 38) are distinguished from other managers (PCS 37) by the fact that for the former, "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

[4] *Déclaration Annuelle de Données Sociales.*

[5] Specifically, we multiply the hours of lower-paid, less-qualified workers by the ratio of their average wage to the average wage of higher-paid, highly qualified workers, as do Gandhi et al. (2020).

[6] One sector (coke and refined petroleum) is dropped because it has tiny shares of total hours worked, and one sector (Transportation and storage) is dropped because the estimation of the production function using GLZ failed to converge. We also drop the computers and electronics sector because of its very high intensity in techie workers.

are made between technicians and other occupations.[7] Beyond what is suggested by their occupational titles (reported in Table A1), the INSEE 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 the role of techies is to increase productivity rather than to contribute to current output like other types of workers.

The classification of techies into R&D and ICT techies is unambiguous. For example, all the occupations classified as R&D techies have the phrase "research and development" in their job descriptions, while those classified as ICT techies all have the phrases "Information technology", "computer science" and/or "telecommunications" in their job descriptions. 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 not related to R&D or ICT and new production methods within their firms. 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 revenues, expenditures on inputs, and the necessary series to construct each firm's capital stock.[8] Appendix A describes the source data and explains how we construct firm-level capital stocks. We use data from the French Customs to generate an indicator of export status for each firm-year.

[7] pages 191, 221 and 343 of INSEE (2003),

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

2.3 Survey data

We use three surveys to provide additional information on techies that allows us to describe their role in the firm. First, we provide information on education in STEM fields (Science, Technology, Engineering, and Math, including Computer Science) and STEM training of techie workers from the Training and Professional Qualification (TPQ) survey in 2015. The survey collects data on the specialization of the highest degree obtained by the individual and whether and which training after the highest degree s/he received.

Second, we collect data on firms' expenditures on R&D (both internal and external) from the Annual Survey on the Means dedicated to Research and Development (R&D survey). Among other information, the R&D survey provides information on the labor costs included in R&D expenditures and the share of R&D expenses that are outsourced. It also provides information on firms' innovation activities. Third, we use the Information and Communication Technology survey (ICT survey), which describes the relationship between ICT training and technology diffusion within firms.

Appendix A provides detailed descriptions of all datasets. Both ICT and R&D surveys can be linked to the administrative datasets described above since they use the same SIREN firm identifier. We exploit this feature below.

3 Facts about techies

Using the DADS and survey data, this section provides information on techie's education and training, and on their role in adopting, mediating, and diffusing technology at the firms that employ them. Here we report our key descriptive results, with further results and details on the analysis reported in Appendix B.

Fact 1. The incidence of Techies across industries. Table 1 reports techie wage bill shares by category in our sample and the 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)
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

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. Column (4) reports the share of the Techie wage bill (or subcategories thereof) that is in manufacturing.

In column (1) we see that Techies account for 18.3% of the French private sector's wage bill share, with a larger share of 31.5% within manufacturing (column 2) than the 10.8% within non-manufacturing (column 3). 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. This motivates studying the role of techies beyond R&D tasks.[9] In column (4) we see that most (62.6%) expenditures on techies are in manufacturing, while more than a third are in non-manufacturing industries. This is why we do not confine our analysis of productivity to manufacturing, in contrast to almost all of the relevant literature.

We observe interesting patterns when we break down techie workers into different categories. Most of the expenditure on R&D techies, 87.3%, is in manufacturing (column 4). Consistent with this, manufacturing is much more R&D techie-intensive that non-manufacturing (comparing column 2 to 3). This implies that studying the impact of R&D on productivity can be largely done within manufacturing. In contrast, 62% of the expenditure on ICT techies is in non-manufacturing, while the ICT techie-intensity is almost identical across sectors. This emphasizes the importance of considering the non-manufacturing sector when

[9] Barth et al. (2017) find that 80 percent of U.S. private sector scientists and engineers worked outside R&D occupations in 2013. This is close to the share of R&D techies in all techies' wage bill (18.5%).

studying the impact of techies on firm-level productivity.

Table 1 also reports the wage bill shares of engineers and technicians. Engineers are twice as large a share of the techie wage bill as technicians.

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 TPQ survey has 26,861 individuals with valid observations, among whom 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.

As we report in Table B1, techies have more STEM education and training than other occupations. In particular, around 63 percent of techies have a degree and/or training in STEM, with about a fifth having both a STEM degree and further STEM training. STEM degrees are somewhat more common among engineers (55%) than technicians (41%).

In contrast, STEM education is uncommon in all other PCS codes, with only 11 percent having a STEM degree, less than a fifth having either a degree or training, and only two percent have both a STEM degree and further training. In particular, this is true for administrative and commercial managers. This is consistent with the distinctions made in the PCS code documentation (INSEE 2003), supporting the idea that techies' skills and role in the firm are technological, and that they are distinct from other workers, including managers.

Fact 3. Most R&D spending is on wages and occurs "in-house". In our structural analysis below, we use the techie wage bill share to measure firm-level resources devoted to improving productivity. Here we compare the techie wage bill to total R&D expenditures.

Firm-level R&D expenditure includes spending on materials and capital goods, which can lead to double-counting when it comes to production function estimation for two reasons. First, total materials are included as an input to production, and it is not possible to extract expenditure on R&D materials from total materials. Second, R&D capital expenditure is part of the firm's total investment, which we use to construct firm level capital stocks. Moreover, capital investment tends to occur in "spikes", which leads to over-estimating effort towards productivity growth when this type of investment occurs and under-estimation of effort in other years. By focusing on wages spent on R&D and excluding techies from current labor input we avoid double counting and obtain a potentially 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.[10] Similarly, Schweitzer (2019) finds that in 2014, labor costs accounted for 60 percent of aggregate R&D spending in France.[11]

One potential threat to our approach is that firms can purchase ICT, R&D, and other technology-related consulting services. This cost would show up as a purchased service, not as a 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.[12]

Fact 4. Techies are positively associated with the diffusion of ICT skills within firms. The ICT survey provides information on whether the firm offers training in developing or improving ICT skills to its workers. ICT training is uncommon, with only 18 percent

[10] 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.

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

[12] 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.

of firms offering training (Table B9). After matching the ICT survey with the DADS dataset, we examine the correlation between techies and ICT training. We use a linear probability model to explain the likelihood of offering ICT training. Our regressions (reported in Table B10) control for firm size, and include sector and year-fixed effects.[13]

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, which makes an effect after training takes place. Techies are associated with this investment-like activity.

Fact 5. Techies are positively associated with patenting and innovation. The R&D survey provides information on firms' patent filings and product and process innovation. As is observed elsewhere, patenting is rare. The firm at the 75^{th} percentile of the patenting distribution 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 correlates positively with all types of R&D expenditures in the R&D survey: internal or external, wages or other expenses (Table B12). Interestingly, the strongest correlation between innovation and patenting is with R&D wages and internal R&D. We find a positive correlation between the techie wage bill (from DADS) and firms' patenting (Table B13). This correlation is particularly strong for R&D techies.

We also find that 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

[13]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.

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.

The analysis reported here (and in greater detail in the appendix) reveals a clear pattern: expenditure on techies' wages are related to patenting and innovation. It also suggests different roles for R&D, ICT and Other techies: R&D techies are associated with both types of innovation, while ICT techies are associated only with product innovation, and Other techies are associated only with process innovation. These findings are consistent with Hall et al. (2010), who argue that R&D is related to product and process innovation.[14] Arora et al. (2017) show how corporate research in the U.S. leads to innovation and patenting, and how the effect on productivity is positively related to the quality of researchers employed in such activities. This quality is likely captured by wages.

4 Why only some firms employ techies: A simple selection model

Despite the potential productivity gains, relatively few firms employ techies. This fact mirrors a well-documented puzzle in trade: although exporting raises firm 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 (Melitz and Ottaviano, 2008). 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

[14] Similar to other studies that examine the impact of R&D on productivity, our methodology does not separately identify whether this occurs through process, product or other types of innovation.

employment T_{ft-1} to maximize profits. The relationship between techies and changes in productivity is

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

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 decision can occur if the effectiveness of techies is perceived to be insufficient to generate a large enough increase in productivity. 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 causal impact of techies on firm-level productivity using structural productivity models. Our empirical strategy addresses two challenges. First, we observe revenue rather than physical output.[15] Second, productivity is unobserved to the econometrician but

[15] 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.

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.

Our key identifying assumption is that techies affect output only through their impact on future productivity. This is analogous to the standard treatment of R&D: it enters the production process only through future productivity, not contemporaneous output (Doraszelski and Jaumandreu, 2013). Beaudry et al. (2016) use a similar framework, where cognitive labor contributes to organizational capital with future returns.

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. 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 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 firm decisions such as techie employment. 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, as we explain below.

Techies enter only in (4), not (3), reflecting our 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). We assess in Section 6.3 the robustness of our results to allowing techies to affect current output.

While we present equation (4) in a linear form, our results are robust to richer specifications of the productivity process. In Appendix F1, 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.

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 causal 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 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 causal 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.

We do not apply the OP/LP/ACF control function approach 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. The GLZ and GNR estimators do not suffer from these issues, and are better suited to our data and identification problem.

The GLZ methodology. 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). These assumptions allow real output to be inferred from observed expenditures.

GLZ use first-order conditions under profit maximization to eliminate unobserved materials and productivity from the estimating equation. The demand elasticity is estimated from variation in revenues, which permits the recovery of firm-level prices and output. The resulting production function is estimated by nonlinear least squares without requiring instruments or assumptions on the productivity process.

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 are adjusted 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, firm-level productivity $\{\widehat{\omega}_{ft}^{GLZ}\}$ can be estimated for each firm and year.

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

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

where $\theta_{i(f)t}$ captures industry-year effects. 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. We compute the long-run impact of techies on productivity by dividing β by $1 - \lambda$.

The GNR methodology. The estimator of Gandhi et al. (2020) imposes no functional form on the production function and allows for arbitrary returns to scale. It requires that at least one input—–typically materials or labor—–is fully flexible and adjusts after productivity is observed. While GNR's baseline estimator assumes physical quantities are observed, they propose an extension that relies only on revenues, building on Klette and Griliches (1996). This version is applicable to our data. Unlike GLZ, GNR identifies the full revenue production function jointly with the productivity process via the moment conditions implied by the controlled Markov equation (4).

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

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

where $\overline{\rho}$ is the average elasticity across industries. 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. While GNR does not recover level effects, it is robust to functional form misspecification and does not require labor to be static.

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 institutional context. French firms face rigid employment adjustment due to both permanent and temporary contract regulations. 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 that use GNR. Our focus is on the estimates of the effects of Techies in a controlled Markov process (4). We control for exporting in all of our specifications (De Loecker, 2013) and in Section 6.3 we show that our results are not sensitive to the inclusion of managers in the controlled Markov (Bloom et al., 2017).

Quantification of the control 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, as well as the percentiles of the techie wage bill shares for observations that have positive values and 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

			Percentiles of techie wage bill shares on positive support, percent					
	Percent with positive values	Mean conditional on positive values	10	25	50	75	90	IQR
Manufacturing								
Techies	71.8	22.6	6.4	11.3	19.1	30.4	44.0	19.1
R&D techies	35.4	7.4	1.2	2.6	5.1	9.7	16.2	7.2
ICT techies	22.4	3.6	0.6	1.0	1.9	3.6	7.1	2.5
Other techies	69.7	18.3	5.5	9.5	15.7	24.4	34.8	14.9
Engineers	60.4	14.5	4.3	7.2	12.0	19.0	28.2	11.8
Technicians	60.6	12.3	2.6	5.1	9.6	16.3	25.3	11.3
Non-Manufacturing								
Techies	19.9	16.8	2.2	5.5	12.2	23.4	38.1	17.9
R&D techies	1.3	5.2	0.3	0.9	2.5	6.4	13.1	5.5
ICT techies	5.0	10.5	0.6	1.6	4.0	10.9	31.6	9.3
Other techies	18.3	15.1	2.1	5.1	11.3	21.3	33.8	16.2
Engineers	13.8	13.8	2.1	4.8	10.2	18.9	30.3	14.1
Technicians	13.7	10.5	1.1	2.9	6.6	13.8	25.2	10.9

Table 2 shows that techies are much more prevalent in manufacturing firms (71.8% of the observations) than in non-manufacturing firms (19.9% of the observations). Furthermore, Table 2 shows that the wage bill shares of different types of techies vary across industries. 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 average wage bill share in non-manufacturing. This pattern is even more pronounced for firms with the highest wage bill shares.

In our estimation sample, we find a higher percentage of exporters in manufacturing (56.4%) compared to non-manufacturing (11.5%). While this difference is expected, we find a non-negligible incidence of exporting among non-manufacturing firms, notably in wholesale and in publishing and broadcasting.[16]//

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 nulls 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) for details). 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. Beyond this, the 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 may expect, both in manufacturing and in service sectors.[17]

We relegate the estimates of the "revenue production function" using the GNR methodology to Appendix E. 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).

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

[17]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	0.223	0.597	0.180	2.629	5.339	29277	4721
	(0.002)	(0.006)	(0.009)	(0.199)	(0.249)		
Textiles, wearing apparel	0.341	0.573	0.086	1.752	2.741	8936	1312
	(0.006)	(0.010)	(0.017)	(0.279)	(0.074)		
Wood, paper products	0.283	0.417	0.300	1.362	4.142	17384	2543
	(0.006)	(0.009)	(0.014)	(0.067)	(0.229)		
Chemical products	0.157	0.56	0.283	1.581	4.446	7380	941
	(0.003)	(0.012)	(0.015)	(0.078)	(0.281)		
Pharmaceutical products	0.18	0.451	0.37	1.594	3.303	1703	222
	(0.015)	(0.038)	(0.053)	(0.215)	(0.58)		
Rubber and plastic	0.226	0.532	0.242	1.677	3.895	16100	2143
	(0.004)	(0.009)	(0.012)	(0.095)	(0.169)		
Basic metal and fabricated metal	0.303	0.392	0.306	1.466	3.436	30407	4148
	(0.004)	(0.005)	(0.008)	(0.046)	(0.09)		
Electrical equipment	0.196	0.56	0.244	1.687	3.755	5094	675
	(0.006)	(0.019)	(0.025)	(0.17)	(0.308)		
Machinery and equipment	0.189	0.548	0.263	1.525	3.524	11526	1502
	(0.005)	(0.015)	(0.021)	(0.132)	(0.214)		
Transport equipment	0.177	0.546	0.277	1.818	5.445	6465	873
	(0.005)	(0.017)	(0.022)	(0.205)	(0.588)		
Other manufacturing	0.333	0.424	0.243	1.605	2.872	24178	3601
	(0.006)	(0.007)	(0.013)	(0.084)	(0.077)		
Construction	0.393	0.396	0.211	1.448	2.672	119766	22417
	(0.004)	(0.004)	(0.008)	(0.032)	(0.039)		
Wholesale	0.119	0.735	0.146	1.284	8.931	188565	27882
	(0.000)	(0.002)	(0.002)	(0.018)	(0.186)		
Retail	0.131	0.794	0.074	1.793	6.033	258474	40393
	(0.000)	(0.002)	(0.002)	(0.072)	(0.066)		
Accommodation and food services	0.396	0.265	0.339	1.861	5.518	116511	22411
	(0.006)	(0.004)	(0.017)	(0.053)	(0.298)		
Publishing and broadcasting	0.381	0.062	0.557	1.237	2.272	15771	2680
	(0.018)	(0.003)	(0.021)	(0.023)	(0.119)		
Administrative and support activities	0.465	0.069	0.466	1.702	3.339	31177	5707
	(0.014)	(0.002)	(0.017)	(0.044)	(0.184)		

Notes. The CES production function can be written as: $Q_{ft} = e^{\omega_{ft}}(\alpha_N N_{ft}^{\gamma} + \alpha_K K_{ft}^{\gamma} + \alpha_M M_{ft}^{\gamma})^{1/\gamma}$, 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. We reject the null hypothesis of σ being equal to one in all industries at significance levels well below 1%. We also reject the null hypothesis of η being smaller than one in absolute value in all industries at significance levels well below 1%.

6.2 Baseline results

We capture the effect of techies 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, either overall or by category of techies, T_{ft-1}. We always control for the extensive margin when examining the intensive margin, which identifies the impact of techie-intensity over and above the extensive margin, while allowing for separate effects of each margin.[18] Formally, we estimate specifications of this general form:

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

where we sometimes include only the extensive margin and in other specifications we allow for multiple arguments for techies by their category. $\theta_{i(f)t}$ is an industry $\times$ year fixed effect. We estimate (7) by OLS, with productivity computed from industry-by-industry estimates of equation (3) using the GLZ estimator. As discussed above, we report bootstrapped standard errors that are clustered by firm.

We report our baseline controlled Markov estimates of (5) in Table 4. In Table F1, we report estimates of the controlled Markov process 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 reported 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). Our analysis of non-manufacturing firms contrasts with most of the literature, which restricts attention to manufacturing firms.

Columns (1) and (7) show that firms that employ techies have higher future productivity than firms without techies. The effect is sizable at 4.0 log points in manufacturing industries and 5.7 log points in non-manufacturing industries. Using the persistence coefficient for

[18]Quantitatively, 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.

lagged techies from the final row of the Table, we find that the steady state, cross-sectional effect of techies is virtually identical in both sectors, at around 45 log points. Using equation (5), the steady state effect of $\mathbf{z}$ is $\beta/(1-\lambda)$. While the estimated effects of employing techies are the same in both sectors, 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.

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 samples. This shows that the presence of even a small number of techies raises future productivity, and that the effect increases with greater techie employment. Two simple calculations using Tables 4 and 2 illustrate the magnitudes. First, comparing firms with no techies to those with the median level of positive techies, the latter have 3.9 and 4.9 log points higher future productivity in manufacturing and non-manufacturing, respectively. Second, comparing firms at the 75th percentile of the positive techie distribution to those at the 25th percentile (the inter-quartile range, or IQR), the former have 2.3 and 3.7 log points higher future productivity in manufacturing and non-manufacturing, respectively.

The long-term effects are about 11 times larger than the impact effects for manufacturing firms and 8 times larger in non-manufacturing.[19] These can be seen in Table 5, where we see that firms with the median intensity of techies are estimated to have 57.5% greater productivity in manufacturing, compared to 48.3% in non-manufacturing. The long run intensive margin IQR techie effect on productivity is estimated at 31% in manufacturing and 34.5% in non-manufacturing. Overall, these estimates are not very different across broad sectors.

Columns (3), (4), (9), and (10) in Table 4 display the estimates when techie workers are broken down by their detailed job descriptions. We find that both the presence and

[19]The long-term estimated effects are calculated by multiplying the short-run effects by $1/(1-\widehat{\lambda})$, where the $\widehat{\lambda}$ are taken from the last row of Table 4.

Table 4: Impact of techies on productivity – GLZ estimates

	Manufacturing						Non-Manufacturing					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I\left(T_{ft-1}>0\right)$	0.040***	0.016***					0.057***	0.024***				
	(0.002)	(0.003)					(0.003)	(0.003)				
T_{ft-1}		0.123***						0.207***				
		(0.008)						(0.012)				
$I\left(T_{ft-1}^{RD}>0\right)$			0.017***	0.011***					0.010*	-0.002		
			(0.002)	(0.003)					(0.006)	(0.007)		
$I\left(T_{ft-1}^{ICT}>0\right)$			0.021***	0.014***					0.025***	0.015***		
			(0.002)	(0.003)					(0.004)	(0.004)		
$I\left(T_{ft-1}^{OTH}>0\right)$			0.029***	0.011***					0.053***	0.018***		
			(0.002)	(0.003)					(0.003)	(0.003)		
T_{ft-1}^{RD}				0.069***						0.160*		
				(0.023)						(0.088)		
T_{ft-1}^{ICT}				0.101***						0.117***		
				(0.036)						(0.021)		
T_{ft-1}^{OTH}				0.113***						0.243***		
				(0.010)						(0.015)		
$I\left(T_{ft-1}^{38}>0\right)$					0.030***	0.012***					0.048***	0.013***
					(0.002)	(0.003)					(0.003)	(0.003)
$I\left(T_{ft-1}^{47}>0\right)$					0.017***	0.006**					0.033***	0.022***
					(0.002)	(0.002)					(0.003)	(0.003)
T_{ft-1}^{38}						0.144***						0.263***
						(0.013)						(0.018)
T_{ft-1}^{47}						0.093***						0.112***
						(0.011)						(0.017)
$I\left(x_{ft-1}>0\right)$	0.009***	0.007***	0.002	0.003	0.004*	0.005**	0.008***	0.006**	0.006**	0.005**	0.004*	0.004*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$\hat{\omega}_{ft-1}$	0.911***	0.913***	0.908***	0.911***	0.910***	0.913***	0.874***	0.875***	0.874***	0.876***	0.874***	0.875***
	(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.	131,697						523,877					
No. firms	21,854						106,430					

Notes. The table reports estimates of equation (5) 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 technician 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

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 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. Interestingly, the presence of R&D techie workers at the extensive margin has a smaller impact on productivity than ICT and Other techies in both sectors, especially in non-manufacturing.

Other techie workers have the largest impact on productivity in both manufacturing and non-manufacturing sectors, with a 1.7 times larger impact in manufacturing and 5.3 times larger impact in non-manufacturing than the impact of R&D techie workers. Using the estimates reported in Table 4, we find that in manufacturing, a one IQR difference in R&D and ICT techies leads to 0.49 and 0.26 percent higher productivity, respectively, while the IQR effect of other techies is 1.7 percent. For non-manufacturing firms, the IQR effect of R&D and ICT techies is comparable, at 0.88 and 1.09 percent, respectively, but the IQR effect of Other techies is quite large, at 4 percent. These results convey an important message: firm-level productivity is driven more by non-R&D techies than by R&D techies, especially outside manufacturing.

Columns (5), (6), (11), and (12) in Table 4 display the estimates when we distinguish between engineers (PCS 38) and technicians (PCS 47). Engineers and technicians positively affect productivity, although the engineers exhibit a greater effect than the technicians, both at the extensive and intensive margins. This makes sense, as engineers are more knowledgeable and skilled, and thus matter more in the technology-enhancing and diffusion process. However, technicians' impact is not negligible.

Turning to the effect of exporting, we find a positive impact on productivity, in line with what De Loecker (2013) finds in manufacturing firms. We estimate similar effects in manufacturing and in non-manufacturing firms. We note that only 11.5% of non-manufacturing firms in our sample are exporters (primarily in wholesale, publishing, and broadcasting). This suggests that exporting is not a significant factor accounting for the variability of productivity in non-manufacturing. We estimate smaller impacts of exporting on productivity when we employ more flexible specifications for techies, distinguishing them by their tasks or occupation types, such as engineers versus technicians. This enables us to gauge better the influences of different types of techies on productivity. This finding is in line with De Loecker (2013), who argues that investments in technology partly drive the impact of exports on productivity.

We summarize the main results of the overall impacts of techies on productivity in Table 5, which reports estimates of the magnitudes of the short run impacts and steady-state level effects in percent points. The table illustrates that while the short-run impacts of techies are larger in non-manufacturing, the higher persistence of productivity in manufacturing mitigates these differences in the long run, and in some cases overturns the relative magnitudes.

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	4.03	2.38	5.05	3.77
R&D techies	1.46	0.49	0.20	0.88
ICT techies	1.60	0.26	1.99	1.09
Other techies	2.92	1.70	4.65	4.02
Engineers	2.97	1.71	4.06	3.53
Technicians	1.50	1.05	2.98	1.23
B. Steady state effects				
Techies	57.45	31.00	48.29	34.50
R&D techies	17.72	5.66	1.63	7.35
ICT techies	19.59	2.99	17.20	9.17
Other techies	38.12	20.83	44.28	37.36
Engineers	40.01	21.57	37.52	32.01
Technicians	18.72	12.72	26.51	10.26

Notes. Units are percent points. 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 steady-state 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 steady-state effects are computed by multiplying the impact effects by $1/(1-\widehat{\lambda})$, where $\widehat{\lambda}$ is the estimated coefficient on lagged productivity, reported in the final row of Table 4. These magnitudes are then translated from log points to percent points by taking the exponent, subtracting 1 and multiplying by 100.

The results reported in Table 5 are calculated from estimates of equation (3), which is a simple linear AR(1) version of the general controlled Markov process given by equation (2). We next consider a more general specification of (2) which allows the effect of techies to differ across the distribution of lagged productivity,

$$\widehat{\omega}_{ft}^{GLZ} = \theta_{i(f)t} + \lambda \widehat{\omega}_{ft-1}^{GLZ} + \beta_1 T_{ft-1} + \beta_2 \left(\widehat{\omega}_{ft-1}^{GLZ} \times T_{ft-1} \right) + \xi_{ft}, \tag{8}$$

where T_{ft-1} is firm f's lagged techie wage bill share.

Compared to a firm with no lagged techies, the productivity effect of lagged techies at the p^{th} percentile for a firm with lagged productivity at the q^{th} percentile is then $\beta_1 T_p + \beta_2 \left(\widehat{\omega}_q^{GLZ} \times T_p \right)$. We report estimates of this quantity for $p, q \in \{25, 50, 75\}$ in Table 6. We find that the marginal effect of techies declines somewhat with the levels of both the techie wage bill and lagged productivity, but the effects are not substantially different from the baseline results reported in Panel A of Table 5.[20]

Table 6: Impact of techies on productivity – General specification

	Percentile of lagged ω		
	25	50	75
Percentile of lagged Techies			
Manufacturing			
25	1.68	2.91	4.10
50	3.13	3.78	4.42
75	5.26	5.07	4.89
Non-manufacturing			
25	1.05	2.70	4.63
50	3.39	4.30	5.35
75	7.40	7.01	6.56

Notes. Units are percent points.

[20] Table G1 in the appendix presents the estimates of equation 8, with a short analysis of their implications.

6.3 Sensitivity analysis

Our baseline results reported in Section 6.2 are computed using the GLZ estimator, and include the full range of techies in the estimation of equation (4). In this section, we report sensitivity analysis in two dimensions. We begin by exploring how our results change when we modify the way techies enter the analysis. We next report results using the GNR estimator. In appendix H, we report the results that consider the quality of labor inputs and show that our baseline results are qualitatively unchanged.

Alternative assumption: techies belong 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 methodology makes sense is to compare it to a simple 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, as given by equation (5).

Table 7 reports the results of this exercise. The full results are reported in the Appendix in Table I1. The estimated effects of techies on productivity are somewhat smaller than in our baseline estimates in Table 4, but 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 results in Table 7 contradict this.

Table 7: Allocating techies to production – GLZ estimates

	Manufacturing		Non-Manufacturing	
	(1)	(2)	(3)	(4)
$I\left(T_{ft-1}>0\right)$	0.022***	0.006*	0.028***	0.008***
	(0.003)	(0.003)	(0.003)	(0.003)
T_{ft-1}		0.086***		0.124***
		(0.010)		(0.012)
$I\left(x_{ft-1}>0\right)$	0.009***	0.007***	0.024***	0.023***
	(0.002)	(0.002)	(0.003)	(0.003)
$\hat{\omega}_{ft-1}$	0.917***	0.915***	0.880***	0.880***
	(0.003)	(0.003)	(0.002)	(0.002)
Other controls	Yes		Yes	
Obs.	130,605		525,725	
No. firms	21,744		106,450	

Notes. The table reports estimates of equation (5) 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

Alternative assumption: Other techies belong 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 modified 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.

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)
$I\left(T_{ft-1} > 0\right)$	0.037***	0.019***					0.058***	0.040***				
	(0.002)	(0.002)					(0.003)	(0.003)				
T_{ft-1}		0.264***						0.192***				
		(0.026)						(0.021)				
$I\left(T^{RD}_{ft-1} > 0\right)$			0.027***	0.012***					0.034***	0.023***		
			(0.002)	(0.003)					(0.006)	(0.007)		
$I\left(T^{ICT}_{ft-1} > 0\right)$			0.024***	0.018***					0.055***	0.038***		
			(0.002)	(0.003)					(0.003)	(0.004)		
T^{RD}_{ft-1}				0.274***						0.296***		
				(0.032)						(0.110)		
T^{ICT}_{ft-1}				0.219***						0.188***		
				(0.051)						(0.022)		
$I\left(T^{38}_{ft-1} > 0\right)$					0.028***	0.011***					0.049***	0.031***
					(0.002)	(0.003)					(0.004)	(0.005)
$I\left(T^{47}_{ft-1} > 0\right)$					0.021***	0.016***					0.036***	0.027***
					(0.002)	(0.003)					(0.004)	(0.004)
T^{38}_{ft-1}						0.331***						0.222***
						(0.040)						(0.031)
T^{47}_{ft-1}						0.149***						0.119***
						(0.039)						(0.036)
$I\left(x_{ft-1} > 0\right)$	0.000	0.002	-0.001	0.000	-0.001	0.000	0.021***	0.022***	0.020***	0.022***	0.020***	0.022***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
$\hat{\omega}_{ft-1}$	0.915***	0.915***	0.914***	0.915***	0.914***	0.915***	0.878***	0.878***	0.878***	0.878***	0.878***	0.878***
	(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.	131,697						523,877					
No. firms	21,854						106,430					

Notes. The table reports estimates of equation (5). Other techies are allocated to production. The dependent variable is $\hat{\omega}_{ft}$, log estimated productivity. $I(.)$ is the indicator function. T is the techie wage bill share, superscripts $\{RD, ICT, 38, 47\}$ denote R&D, ICT, other techies, engineers and technician 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

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 (2). 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) and their associations with innovative and productivity-enhancing activities (Sections 3 and 3). In contrast, we treat managers as part of general labor, whose contributions to output are contemporaneous. In Table 9, we test this implication by including lagged managerial workers (PCS code 37) in the second stage. Columns (1) and (3) reproduce our baseline estimates for convenience, while columns (2) and (4) add lagged managerial labor to the controlled Markov equation.

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

	Manufacturing		Non-Manufacturing	
	Baseline (1)	Managers (2)	Baseline (3)	Managers (4)
$I(T_{ft-1} > 0)$	0.016***	0.016***	0.024***	0.018***
	(0.003)	(0.003)	(0.003)	(0.003)
T_{ft-1}	0.123***	0.119***	0.207***	0.204***
	(0.008)	(0.008)	(0.013)	(0.013)
$I(M_{ft-1} > 0)$		0.002		0.027***
		(0.003)		(0.002)
M_{ft-1}		-0.063***		-0.021***
		(0.010)		(0.006)
$I(x_{ft-1} > 0)$	0.007***	0.008***	0.006**	0.004
	(0.002)	(0.002)	(0.002)	(0.003)
$\hat{\omega}_{ft-1}$	0.913***	0.914***	0.875***	0.873***
	(0.003)	(0.003)	(0.002)	(0.002)
Obs.	131,697		523,877	
No. firms	21,854		106,430	

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, M is the managers (PCS37) 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

The results in Table 9 indicate that including lagged managers does not materially affect the estimated effects of lagged techies. We emphasize that the models estimated in columns (2) and (4) are misspecified because we maintain managers' contribution to contemporaneous labor input. Therefore, the estimated effects on lagged managers do not have a coherent interpretation.

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 labor market rigidities in the French labor market, e.g., firing costs. Table 10 reports the results when labor is assumed to be "static" (like materials, and as we assumed when implementing the GLZ estimator), and Table 11 reports the results for when labor is assumed to be "predetermined" (like capital).

Recall that the estimates here are not directly comparable to our GLZ estimates because GNR does not separately identify the coefficients β in equation (4) from the demand parameter ρ in equation (3). This implies that the numbers we report in Tables 10 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 in the appendix).

Despite differences in methodologies, including assumptions on the response of labor to innovations to productivity and on returns to scale, the results in Tables 10 and 11 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 causal effects of techies on productivity that extend beyond their 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

Table 10: 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)
$I\left(T_{ft-1} > 0\right)$	0.037***	0.029***					0.025***	0.015***				
	(0.002)	(0.002)					(0.001)	(0.001)				
T_{ft-1}		0.041***						0.051***				
		(0.004)						(0.003)				
$I\left(T_{ft-1}^{RD} > 0\right)$			0.014***	0.012***					0.008***	0.007***		
			(0.001)	(0.001)					(0.002)	(0.002)		
$I\left(T_{ft-1}^{ICT} > 0\right)$			0.014***	0.012***					0.010***	0.006***		
			(0.001)	(0.002)					(0.001)	(0.001)		
$I\left(T_{ft-1}^{OTH} > 0\right)$			0.031***	0.026***					0.023***	0.014***		
			(0.002)	(0.002)					(0.001)	(0.001)		
T_{ft-1}^{RD}				0.019*						-0.016		
				(0.011)						(0.025)		
T_{ft-1}^{ICT}				0.017						0.036***		
				(0.016)						(0.008)		
T_{ft-1}^{OTH}				0.029***						0.057***		
				(0.005)						(0.004)		
$I\left(T_{ft-1}^{38} > 0\right)$					0.028***	0.024***					0.022***	0.014***
					(0.002)	(0.002)					(0.001)	(0.001)
$I\left(T_{ft-1}^{47} > 0\right)$					0.021***	0.019***					0.014***	0.010***
					(0.001)	(0.002)					(0.001)	(0.001)
T_{ft-1}^{38}						0.026***						0.050***
						(0.006)						(0.005)
T_{ft-1}^{47}						0.021***						0.034***
						(0.006)						(0.005)
$I\left(x_{ft-1} > 0\right)$	0.015***	0.014***	0.013***	0.013***	0.013***	0.013***	0.008***	0.007***	0.007***	0.006***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\hat{\omega}_{ft-1}$	0.916***	0.918***	0.915***	0.916***	0.913***	0.914***	0.932***	0.933***	0.933***	0.933***	0.932***	0.933***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Obs.	157,660						715,861					
No. firms	22,515						117,594					

Notes. The table reports estimates of equation (6) in the text. The dependent variable is $\rho\hat{\omega}_{ft}$, 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 technician 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

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.

Table 11: 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)
$I\left(T_{ft-1}>0\right)$	0.028***	0.017***					0.014***	0.010***				
	(0.002)	(0.002)					(0.001)	(0.001)				
T_{ft-1}		0.052***						0.024***				
		(0.007)						(0.004)				
$I\left(T_{ft-1}^{RD}>0\right)$			0.004**	-0.001					0.003***	0.005***		
			(0.002)	(0.002)					(0.001)	(0.002)		
$I\left(T_{ft-1}^{ICT}>0\right)$			0.010***	0.005***					0.00017	-0.001		
			(0.002)	(0.002)					(0.001)	(0.001)		
$I\left(T_{ft-1}^{OTH}>0\right)$			0.024***	0.015***					0.010***	0.008***		
			(0.002)	(0.002)					(0.001)	(0.001)		
T_{ft-1}^{RD}				0.044***						-0.024		
				(0.014)						(0.016)		
T_{ft-1}^{ICT}				0.073***						0.018***		
				(0.021)						(0.008)		
T_{ft-1}^{OTH}				0.052***						0.013***		
				(0.008)						(0.003)		
$I\left(T_{ft-1}^{38}>0\right)$					0.019***	0.012***					0.010***	0.008***
					(0.002)	(0.002)					(0.001)	(0.001)
$I\left(T_{ft-1}^{47}>0\right)$					0.016***	0.011***					0.012***	0.009***
					(0.002)	(0.002)					(0.001)	(0.001)
T_{ft-1}^{38}						0.046***						0.011**
						(0.008)						(0.005)
T_{ft-1}^{47}						0.042***						0.026***
						(0.008)						(0.006)
$I\left(x_{ft-1}>0\right)$	0.028***	0.027***	0.026***	0.027***	0.026***	0.026***	0.009***	0.009***	0.006***	0.006***	0.008***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\hat{\omega}_{ft-1}$	0.689***	0.687***	0.690***	0.687***	0.690***	0.687***	0.820***	0.820***	0.846***	0.845***	0.821***	0.820***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.007)	(0.007)	(0.008)	(0.008)	(0.007)	(0.007)
Obs.	157,660						715,861					
No. firms	22,515						117,594					

Notes. The table reports estimates of equation (6) in the text. The dependent variable is $\rho\hat{\omega}_{ft}$, 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 technician 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

Some differences with Table 4 are apparent. For example, in Table 10 we do not identify a statistically significant impact of ICT in the intensive margin in manufacturing. And in Table 11, 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 has shown 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 have often conceived of 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 techies can fully capture overall impact of techies across various industries.

We have conceived of employment of techies as analogous to investment spending: employing techies is profitable because they raise the future productivity of other factors of production, just as investment is profitable because it raises the firm's future capital stock. Our methodology has allowed us to study the causal effects of employing techies on future productivity without having to model the difficult question of optimal employment of techies. To do so we have adopted techniques from the productivity estimation literature, which has similarly shown how to estimate the effect of capital and other factors of production on output without estimating the full system of dynamic factor demands.

Our work has implications for policymakers concerned with promoting economic growth.

Capital accumulation and R&D are rightly central to such policy goals. Our findings about the key role of ICT and other techies suggest that educational, training and other policies that enhance the supply of 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," 2023. 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.

_ , 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.

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, “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.

Melitz, Marc J and Gianmarco IP Ottaviano, "Market size, trade, and productivity," *The Review of Economic Studies*, 2008, *75* (1), 295–316.

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.

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.[21] 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.[22] Our unit of analysis is a firm-year observation.

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. We distinguish between three types of techie workers: ICT, R&D, and other techies. Table A1 reports our classification.

Table A1: Classification of ICT, R&D and other techies

PCS-ESE	Description (see, INSEE (2003))
Research and Development	
383a	Engineers and R&D managers in electricity and electronics
384a	Engineers and R&D managers in mechanics and metalworking
385a	Engineers and R&D managers in the transformation industries (food processing, chemistry, metallurgy, heavy materials)
386a	Engineers and R&D managers in other industries (printing, soft materials, furniture and wood, energy, water)
473b	R&D technicians and manufacturing methods technicians in electricity, electromechanics, and electronics
474b	R&D technicians and manufacturing methods technicians in mechanical construction and metalworking
475a	R&D technicians and production methods technicians in the transformation industries
Information and Communication Technologies	
388a	Engineers and R&D managers in computer science
388b	Engineers and managers in administration, maintenance, support, and user services in computer science

[21] 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.

[22] 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.

388c	IT project managers and IT managers
388e	Engineers and specialist managers in telecommunications
478a	Computer design and development technicians
478b	Computer production and operation technicians
478c	Computer installation, maintenance, support, and user services technicians
478d	Telecommunications technicians and network IT technicians
Other	
380a	Technical directors of large companies
381a	Engineers and management staff in agriculture, fishing, water, and forestry studies and operations
382a	Engineers and management staff in building and public works studies
382b	Architects
382c	Engineers, site managers, and construction supervisors (managers) in building and public works
382d	Technical sales engineers and managers in building and public works
383b	Manufacturing engineers and managers in electrical and electronic equipment
383c	Technical sales engineers and managers in professional electrical or electronic equipment
384b	Manufacturing engineers and managers in mechanics and metalworking
384c	Technical sales engineers and managers in professional mechanical equipment
385b	Manufacturing engineers and managers in transformation industries (food processing, chemicals, metallurgy, heavy materials)
385c	Technical sales engineers and managers in intermediate goods transformation industries
386d	Production and distribution engineers and managers in energy and water
386e	Manufacturing engineers and managers in other industries (printing, soft materials, furniture, and wood)
387a	Industrial purchasing and procurement engineers and managers
387b	Logistics, planning, and scheduling engineers and managers
387c	Production method engineers and managers
387d	Quality control engineers and managers
387e	Maintenance, maintenance, and new works technical engineers and managers
387f	Technical engineers and managers in the environment
388d	Technical sales engineers and managers in IT and telecommunications
389a	Technical engineers and managers in transport operations
389b	Technical and commercial navigating officers and managers of civil aviation
389c	Technical navigating officers and managers of merchant navy.
471a	Technical experts and consultants in agriculture, water, and forestry studies
471b	Technical experts in operation and production control in agriculture, water, and forestry
472a	Building and civil engineering draftsmen
472b	Surveyors and topographers
472c	Quantity surveyors and various building and civil engineering technicians
472d	State and local government public works technicians
473a	Electrical, electromechanical, and electronic draftsmen
473c	Electrical, electromechanical, and electronic production and quality control technicians
474a	Mechanical and metal construction draftsmen
474c	Mechanical and metal construction production and quality control technicians
475b	Production and quality control technicians in the transformation industries
476a	Technical assistants, printing and publishing technicians
476b	Soft materials, furniture, and wood industry technicians
477a	Logistics, planning, and scheduling technicians

477b	Installation and maintenance technicians for industrial equipment (electrical, electromechanical, and mechanical, excluding IT)
477c	Installation and maintenance technicians for non-industrial equipment (excluding IT and telecommunications)
477d	Environmental and pollution treatment technicians
479a	Public research or teaching laboratory technicians
479b	Independent expert technicians of various levels

Source: INSEE (2003): https://www.insee.fr/fr/information/2400059. Own classification.
Notes: The PCS (*Professions et Catégories Socioprofessionnelles*) system of occupational codes is used to classify all workers in France.

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.

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 with the book value of capital recorded in FARE. We follow the methodology proposed by Bonleu et al. (2013) and Cette et al. (2015). Since the stocks are recorded at historical cost, i.e. at their value at the time of entry into the firm i's balance sheet, an adjustment has to be made to move from stocks valued at historic cost ($K^{BV}_{i,s,t}$) to stocks valued at current prices ($K_{i,s,t}$). We deflate K^{BV} by a price by assuming that the sectoral price of capital is equal to the sectoral price of investment T years before the date when the first book value was available, where T is the corrected average age of capital, hence $p^K_{s,t+1} = p^I_{s,t-T}$. The average age of capital is computed using the share of depreciated capital, $DK^{BV}_{i,s,t}$ in the capital stock at historical cost.

$$T = \frac{DK^{BV}_{i,s,t}}{K^{BV}_{i,s,t}} \times \widetilde{A}$$

where

$$\widetilde{A} = median_{i \in S} \left(\frac{K^{BV}_{i,s,t}}{\Delta DK^{BV}_{i,s,t}} \right)$$

with S the set of firms in a sector. We use the median value $\widetilde{A}$ to reduce the volatility in the data, as investments within firms are discrete events.

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.

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 A2 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). .

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

Facts 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 A2: 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.

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 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%).

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

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 .

analysis indicates that most techie workers are employed 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.

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

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 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.

Facts 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.

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.

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.

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.

Facts 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 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.

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 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.[23]

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

[23] $0.12 + 0.35 = 0.47$ which is about half.

Table B7: Exposure to ICT training

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

Source: ICT survey.

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).

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

Source: ICT survey.

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 $\times$ year fixed effects, and the log wage bill excluding techies, named "Ex-techies", as a control for firm size.

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.

Table B9: ICT workers and ICT training

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

Source: Matched dataset.

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.

Table B10: Explaining ICT training

	(1)	(2)	(3)	(4)	(5)	(6)
I(techies > 0)	0.061***					
	(0.006)					
I(ICT techies > 0)		0.203***	0.188***			
		(0.009)	(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***	0.035***
					(0.005)	(0.007)
– R&D techies						0.024***
						(0.006)
– Other techies						0.015
						(0.011)
– Ex-techies	0.087***	0.074***	0.065***	0.068***	0.083***	0.083***
	(0.002)	(0.002)	(0.002)	(0.004)	(0.005)	(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 ≤ 0.01, ** p-value ≤ 0.05, * p-value ≤ 0.10.

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.

Facts 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 causal effects of R&D or techies on these measures of innovation, but the reduced form correlations are informative.

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	Process innovation? 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.

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.

There, we find that ICT techies are also associated with patenting and innovation, In contrast, when using the matched sample, our analysis suggests that Other techies do not significantly impact product innovation, while ICT techies do have an effect. We find a positive correlation between R&D and other workers on process innovation.

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 $\leq$ 0.01, ** p-value $\leq$ 0.05, * p-value $\leq$ 0.10.

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.

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 $\leq$ 0.01, ** p-value $\leq$ 0.05, * p-value $\leq$ 0.10.

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.

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.

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. 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 $I\left(\right)$ 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 $\beta\left(\eta - 1\right) < 1$. Plugging the solution (12) back into the expression for Ω_{ft} and defining the constants $\nu = \beta\left(\eta - 1\right) < 1$ 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

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

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)

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

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

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

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 $\eta = 1/\left(1-\rho\right) > 1$, $\rho \in (0,1)$, *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

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

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

$$\sum_{ft}[s_{ft}^m - \ln\left(\begin{array}{c} \gamma_0' + \gamma_m' m_{ft} + \gamma_l' l_{ft} + \gamma_k' k_{ft} + \gamma_{mm}' m_{ft}^2 + \gamma_{ll}' l_{ft}^2 + \gamma_{kk}' k_{ft}^2 \\ +\gamma_{ml}' m_{ft} l_{ft} + \gamma_{mk}' m_{ft} k_{ft} + \gamma_{lk}' l_{ft} k_{ft} + \gamma_{mlk}' m_{ft} l_{ft} k_{ft} \end{array}\right)]^2. \tag{27}$$

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

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

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

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

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:

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

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 :

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

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

$$\begin{aligned}
s_{ft}^m &= \ln\left[E(e^u)\gamma^m(x_{ft})\right] - u_{ft}^m && (31)\\
s_{ft}^l &= \ln\left[E(e^u)\gamma^l(x_{ft})\right] - u_{ft}^l \ . && (32)
\end{aligned}$$

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

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

This sum of integrals equals

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

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

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

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

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

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 $\widehat{\omega}'(\alpha)_{ft}$ as a residual from (34). Now invoke the Markov assumption for ω_{ft} (17), and estimate

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

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 $\widehat{\xi}'(\alpha)_{ft}$ 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:

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

Once we have estimates of α we can compute one last time $\widehat{\omega}'(\alpha)_{ft}$ 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

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

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.

E Production functions estimates

Table E1 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.

F More lags of $\hat{\omega}_{ft}$

Table E1: GNR Production function estimates

Industries	γ^m	γ^l	γ^k	#Obs.	#Firms
Food, beverage, tobacco	0.429	0.464	0.175	29093	4677
Textiles, wearing apparel	0.326	0.526	0.094	8871	1299
Wood, paper products	0.289	0.673	0.069	17272	2521
Chemical products	0.399	0.482	0.134	7357	938
Pharmaceutical products	0.260	0.640	0.089	1699	222
Rubber and plastic	0.362	0.497	0.161	16068	2137
Basic metal and fabricated metal	0.267	0.646	0.108	30333	4133
Electrical equipment	0.375	0.439	0.155	5080	674
Machinery and equipment	0.359	0.534	0.103	11489	1495
Transport equipment	0.396	0.570	0.094	6435	867
Other manufacturing	0.250	0.665	0.106	23963	3552
Construction	0.224	0.693	0.112	116713	21409
Wholesale	0.592	0.367	0.058	186147	27296
Retail	0.631	0.311	0.051	256347	39837
Accommodation and food services	0.210	0.642	0.173	113923	21554
Publishing and broadcasting	0.055	0.774	0.111	14213	2378
Administrative and support activities	0.070	0.571	0.240	28518	5120

Table F1: Adding lags of productivity – GLZ estimates

	Manufacturing						Non-Manufacturing					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I\left(T_{ft-1} > 0\right)$	0.038***	0.014***					0.053***	0.018***				
	(0.002)	(0.003)					(0.003)	(0.003)				
T_{ft-1}		0.119***						0.215***				
		(0.008)						(0.013)				
$I\left(T_{ft-1}^{RD} > 0\right)$			0.015***	0.009***					0.013**	0.000		
			(0.002)	(0.002)					(0.006)	(0.007)		
$I\left(T_{ft-1}^{ICT} > 0\right)$			0.018***	0.011***					0.025***	0.015***		
			(0.002)	(0.002)					(0.003)	(0.004)		
$I\left(T_{ft-1}^{OTH} > 0\right)$			0.028***	0.009***					0.048***	0.012***		
			(0.002)	(0.003)					(0.003)	(0.003)		
T_{ft-1}^{RD}				0.071***						0.151		
				(0.023)						(0.092)		
T_{ft-1}^{ICT}				0.111***						0.118***		
				(0.037)						(0.022)		
T_{ft-1}^{OTH}				0.114***						0.251***		
				(0.010)						(0.015)		
$I\left(T_{ft-1}^{38} > 0\right)$					0.028***	0.010***					0.046***	0.009***
					(0.002)	(0.003)					(0.003)	(0.003)
$I\left(T_{ft-1}^{47} > 0\right)$					0.015***	0.005*					0.030***	0.019***
					(0.002)	(0.002)					(0.002)	(0.003)
T_{ft-1}^{38}						0.141***						0.271***
						(0.013)						(0.018)
T_{ft-1}^{47}						0.094***						0.117***
						(0.011)						(0.017)
$I\left(x_{ft-1} > 0\right)$	0.007***	0.004**	0.001	0.002	0.002	0.003	0.004*	0.003	0.002	0.002	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)
$\hat{\omega}_{ft-1}$	0.936***	0.940***	0.934***	0.939***	0.936***	0.940***	0.931***	0.933***	0.932***	0.933***	0.931***	0.933***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$\hat{\omega}_{ft-2}$	0.025***	0.024***	0.024***	0.023***	0.025***	0.024***	0.017***	0.017***	0.016***	0.017***	0.016***	0.017***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\hat{\omega}_{ft-2}$	-0.02***	-0.02***	-0.02***	-0.02***	-0.02***	-0.02***	-0.03***	-0.03***	-0.03***	-0.03***	-0.03***	-0.03***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Obs.	131,697						523,877					
No. firms	21,854						106,430					

Notes. The table reports estimates of equation (5) 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 technician 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

G General Specification

Table G1 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 G1: General Specification– GLZ estimates

	Manufacturing		Non-Manufacturing	
	Baseline (1)	Interaction (2)	Baseline (3)	Interaction (4)
$I\left(T_{ft-1}>0\right)$	0.016***	0.013***	0.024***	0.013***
	(0.003)	(0.003)	(0.003)	(0.003)
$I\left(T_{ft-1}>0\right)\times\hat{\omega}_{ft-1}$		0.045***		0.053***
		(0.005)		(0.003)
T_{ft-1}	0.123***	0.121***	0.207***	0.233***
	(0.008)	(0.008)	(0.013)	(0.014)
$T_{ft-1}\times\hat{\omega}_{ft-1}$		-0.163***		-0.266***
		(0.016)		(0.016)
$I\left(x_{ft-1}>0\right)$	0.007***	0.008***	0.006***	0.005**
	(0.002)	(0.002)	(0.002)	(0.002)
$\hat{\omega}_{ft-1}$	0.913***	0.909***	0.875***	0.873***
	(0.003)	(0.004)	(0.002)	(0.002)
Obs.	131,697		523,877	
No. firms	21,854		106,430	

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

H 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:

$$\tilde{N}_{ft} = H_{ft} + (w_L/w_H)\, L_{ft}, \tag{38}$$

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 results using the GLZ estimator are reported in Table (H1) while the results using the GNR estimator are presented in Tables (H2) and (H3).

Table H1: 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)
$I\left(T_{ft-1} > 0\right)$	0.046***	0.024***					0.064***	0.031***				
	(0.002)	(0.003)					(0.003)	(0.003)				
T_{ft-1}		0.109***						0.202***				
		(0.008)						(0.012)				
$I\left(T_{ft-1}^{RD} > 0\right)$			0.018***	0.014***					0.008	-0.003		
			(0.002)	(0.003)					(0.006)	(0.007)		
$I\left(T_{ft-1}^{ICT} > 0\right)$			0.019***	0.014***					0.026***	0.017***		
			(0.002)	(0.003)					(0.004)	(0.004)		
$I\left(T_{ft-1}^{OTH} > 0\right)$			0.035***	0.018***					0.060***	0.024***		
			(0.002)	(0.003)					(0.003)	(0.003)		
T_{ft-1}^{RD}				0.040						0.126		
				(0.025)						(0.098)		
T_{ft-1}^{ICT}				0.081**						0.103***		
				(0.038)						(0.022)		
T_{ft-1}^{OTH}				0.102***						0.242***		
				(0.010)						(0.015)		
$I\left(T_{ft-1}^{38} > 0\right)$					0.030***	0.017***					0.050***	0.015***
					(0.002)	(0.003)					(0.003)	(0.004)
$I\left(T_{ft-1}^{47} > 0\right)$					0.022***	0.010***					0.040***	0.029***
					(0.002)	(0.003)					(0.003)	(0.003)
T_{ft-1}^{38}						0.108***						0.254***
						(0.014)						(0.018)
T_{ft-1}^{47}						0.099***						0.110***
						(0.012)						(0.018)
$I\left(x_{ft-1} > 0\right)$	0.008***	0.006**	0.001	0.002	0.003	0.003	0.008***	0.006**	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)
$\hat{\omega}_{ft-1}$	0.913***	0.916***	0.911***	0.914***	0.912***	0.915***	0.876***	0.877***	0.876***	0.878***	0.876***	0.877***
	(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.			131,697						523,877			
No. firms			21,854						106,430			

Notes. The table reports estimates of equation (5) 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 technician 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

Table H2: 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)
$I\left(T_{ft-1}>0\right)$	0.037***	0.029***					0.026***	0.016***				
	(0.001)	(0.001)					(0.001)	(0.001)				
$\mathrm{T}T_{ft-1}$		0.037***						0.056***				
		(0.003)						(0.003)				
$I\left(T_{ft-1}^{RD}>0\right)$			0.013***	0.012***					0.005***	0.004**		
			(0.001)	(0.001)					(0.002)	(0.002)		
$I\left(T_{ft-1}^{ICT}>0\right)$			0.012***	0.011***					0.011***	0.007***		
			(0.001)	(0.001)					(0.001)	(0.001)		
$I\left(T_{ft-1}^{OTH}>0\right)$			0.031***	0.026***					0.025***	0.014***		
			(0.001)	(0.001)					(0.001)	(0.001)		
T_{ft-1}^{RD}				0.014						-0.016		
				(0.009)						(0.025)		
T_{ft-1}^{ICT}				0.016						0.037***		
				(0.016)						(0.008)		
T_{ft-1}^{OTH}				0.027***						0.063***		
				(0.004)						(0.004)		
$I\left(T_{ft-1}^{38}>0\right)$					0.027***	0.025***					0.023***	0.014***
					(0.001)	(0.001)					(0.001)	(0.001)
$I\left(T_{ft-1}^{47}>0\right)$					0.022***	0.019***					0.015***	0.010***
					(0.001)	(0.001)					(0.001)	(0.001)
T_{ft-1}^{38}						0.016***						0.055***
						(0.005)						(0.005)
T_{ft-1}^{47}						0.024***						0.041***
						(0.005)						(0.005)
$I\left(x_{ft-1}>0\right)$	0.014***	0.013***	0.012***	0.012***	0.012***	0.012***	0.005***	0.004***	0.004***	0.004***	0.004***	0.003***
	(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***	0.922***	0.919***	0.920***	0.917***	0.918***	0.934***	0.935***	0.934***	0.935***	0.934***	0.935***
	(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.	157,660						715,861					
No. firms	22,515						117,594					

Notes. The table reports estimates of equation (6) in the text. The dependent variable is $\rho\hat{\omega}_{ft}$, 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 technician 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

Table H3: 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)
$I\left(T_{ft-1}>0\right)$	0.027***	0.019***					0.012***	0.010***				
	(0.002)	(0.002)					(0.001)	(0.001)				
$\mathrm{T}T_{ft-1}$		0.041***						0.015***				
		(0.006)						(0.004)				
$I\left(T_{ft-1}^{RD}>0\right)$			0.003**	-0.000					0.002**	0.004**		
			(0.001)	(0.002)					(0.001)	(0.002)		
$I\left(T_{ft-1}^{ICT}>0\right)$			0.007***	0.003*					-0.000	-0.002*		
			(0.002)	(0.002)					(0.001)	(0.001)		
$I\left(T_{ft-1}^{OTH}>0\right)$			0.024***	0.017***					0.010***	0.008***		
			(0.002)	(0.002)					(0.001)	(0.001)		
T_{ft-1}^{RD}				0.034***						-0.020		
				(0.013)						(0.016)		
T_{ft-1}^{ICT}				0.057***						0.017**		
				(0.021)						(0.007)		
T_{ft-1}^{OTH}				0.043***						0.011***		
				(0.007)						(0.003)		
$I\left(T_{ft-1}^{38}>0\right)$					0.017***	0.013***					0.008***	0.007***
					(0.002)	(0.002)					(0.001)	(0.001)
$I\left(T_{ft-1}^{47}>0\right)$					0.017***	0.012***					0.011***	0.009***
					(0.002)	(0.002)					(0.001)	(0.001)
T_{ft-1}^{38}						0.029***						0.005
						(0.008)						(0.005)
T_{ft-1}^{47}						0.038***						0.017***
						(0.008)						(0.006)
$I\left(x_{ft-1}>0\right)$	0.026***	0.026***	0.025***	0.026***	0.025***	0.025***	0.007***	0.007***	0.004***	0.005***	0.006***	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)
$\hat{\omega}_{ft-1}$	0.680***	0.678***	0.681***	0.678***	0.681***	0.679***	0.839***	0.839***	0.843***	0.842***	0.839***	0.839***
	(0.021)	(0.021)	(0.020)	(0.021)	(0.021)	(0.021)	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)	(0.009)
Obs.	157,660						715,861					
No. firms	22,515						117,594					

Notes. The table reports estimates of equation (6) in the text. The dependent variable is $\hat{\rho\omega}_{ft}$, 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 technician 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

I Sensitivity

Table I1: Allocating techies to production – GLZ estimates

	Manufacturing						Non-Manufacturing					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$I\left(T_{ft-1}>0\right)$	0.022***	0.006*					0.028***	0.008***				
	(0.003)	(0.003)					(0.003)	(0.003)				
T_{ft-1}		0.086***						0.124***				
		(0.010)						(0.012)				
$I\left(T_{ft-1}^{RD}>0\right)$			0.016***	0.007**					0.017***	0.019**		
			(0.003)	(0.003)					(0.006)	(0.008)		
$I\left(T_{ft-1}^{ICT}>0\right)$			0.022***	0.020***					0.038***	0.019***		
			(0.003)	(0.003)					(0.004)	(0.004)		
$I\left(T_{ft-1}^{OTH}>0\right)$			0.013***	0.005*					0.020***	0.009***		
			(0.003)	(0.003)					(0.003)	(0.003)		
T_{ft-1}^{RD}				0.115***						-0.022		
				(0.027)						(0.112)		
T_{ft-1}^{ICT}				0.038						0.205***		
				(0.040)						(0.020)		
T_{ft-1}^{OTH}				0.054***						0.079***		
				(0.011)						(0.012)		
$I\left(T_{ft-1}^{38}>0\right)$					0.014***	0.004					0.017***	0.004
					(0.003)	(0.003)					(0.003)	(0.003)
$I\left(T_{ft-1}^{47}>0\right)$					0.015***	0.008***					0.031***	0.022***
					(0.003)	(0.003)					(0.003)	(0.003)
T_{ft-1}^{38}						0.086***						0.099***
						(0.015)						(0.017)
T_{ft-1}^{47}						0.070***						0.091***
						(0.013)						(0.015)
$I\left(x_{ft-1}>0\right)$	0.009***	0.007***	0.001	0.001	0.005**	0.005**	0.024***	0.023***	0.020***	0.021***	0.021***	0.021***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
$\hat{\omega}_{ft-1}$	0.917***	0.915***	0.915***	0.914***	0.916***	0.915***	0.880***	0.880***	0.880***	0.879***	0.880***	0.879***
	(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.	130,605						525,725					
No. firms	21,744						106,450					

Notes. The table reports estimates of equation (5) 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 technician 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